METHODS IN MOLECULAR BIOLOGY™ 323

Arabidopsis Protocols

SECOND EDITION

Edited by

Julio Salinas

Jose J. Sanchez-Serrano

Arabidopsis Protocols

METHODS IN MOLECULAR BIOLOGY™

John M. Walker, SERIES EDITOR

352. Protein Engineering Protocols, edited by *Kristian Müller and Katja Arndt, 2006*

351. *C. elegans:* *Methods and Applications,* edited by *Kevin Strange, 2006*

350. Protein Folding Protocols, edited by *Yawen Bai and Ruth Nussinov 2006*

349. YAC Protocols, *Second Edition,* edited by *Alasdair MacKenzie, 2006*

348. Nuclear Transfer Protocols: *Cell Reprogramming and Transgenesis,* edited by *Paul J. Verma and Alan Trounson, 2006*

347. Glycobiology Protocols, edited by *Inka Brockhausen-Schutzbach, 2006*

346. *Dictyostelium discoideum* Protocols, edited by *Ludwig Eichinger and Francisco Rivero-Crespo, 2006*

345. Diagnostic Bacteriology Protocols, *Second Edition,* edited by *Louise O'Connor, 2006*

344. Agrobacterium Protocols, Second Edition: *Volume 2,* edited by *Kan Wang, 2006*

343. Agrobacterium Protocols, Second Edition: *Volume 1,* edited by *Kan Wang, 2006*

342. MicroRNA Protocols, edited by *Shao-Yao Ying, 2006*

341. Cell–Cell Interactions: *Methods and Protocols,* edited by *Sean P. Colgan, 2006*

340. Protein Design: *Methods and Applications,* edited by *Raphael Guerois and Manuela López de la Paz, 2006*

339. Microchip Capillary Electrophoresis: *Methods and Protocols,* edited by *Charles Henry, 2006*

338. Gene Mapping, Discovery, and Expression: *Methods and Protocols,* edited by *M. Bina, 2006*

337. Ion Channels: *Methods and Protocols,* edited by *James D. Stockand and Mark S. Shapiro, 2006*

336. Clinical Applications of PCR: *Second Edition,* edited by *Y. M. Dennis Lo, Rossa W. K. Chiu, and K. C. Allen Chan, 2006*

335. Fluorescent Energy Transfer Nucleic Acid Probes: *Designs and Protocols,* edited by *Vladimir V. Didenko, 2006*

334. PRINS and *In Situ* PCR Protocols: *Second Edition,* edited by *Franck Pellestor, 2006*

333. Transplantation Immunology: *Methods and Protocols,* edited by *Philip Hornick and Marlene Rose, 2006*

332. Transmembrane Signaling Protocols: *Second Edition,* edited by *Hydar Ali and Bodduluri Haribabu, 2006*

331. Human Embryonic Stem Cell Protocols, edited by *Kursad Turksen, 2006*

330. Embryonic Stem Cell Protocols, Second Edition, *Vol. II: Differentiation Models,* edited by *Kursad Turksen, 2006*

329. Embryonic Stem Cell Protocols, Second Edition, *Vol. I***:** *Isolation and Characterization,* edited by *Kursad Turksen, 2006*

328. New and Emerging Proteomic Techniques, edited by *Dobrin Nedelkov and Randall W. Nelson, 2006*

327. Epidermal Growth Factor: *Methods and Protocols,* edited by *Tarun B. Patel and Paul J. Bertics, 2006*

326. *In Situ* Hybridization Protocols, *Third Edition,* edited by *Ian A. Darby and Tim D. Hewitson, 2006*

325. Nuclear Reprogramming: *Methods and Protocols,* edited by *Steve Pells, 2006*

324. Hormone Assays in Biological Fluids, edited by *Michael J. Wheeler and J. S. Morley Hutchinson, 2006*

323. *Arabidopsis* Protocols, *Second Edition,* edited by *Julio Salinas and Jose J. Sanchez-Serrano, 2006*

322. *Xenopus* Protocols: *Cell Biology and Signal Transduction,* edited by *X. Johné Liu, 2006*

321. Microfluidic Techniques: *Reviews and Protocols,* edited by *Shelley D. Minteer, 2006*

320. Cytochrome P450 Protocols, *Second Edition,* edited by *Ian R. Phillips and Elizabeth A. Shephard, 2006*

319. Cell Imaging Techniques, *Methods and Protocols,* edited by *Douglas J. Taatjes and Brooke T. Mossman, 2006*

318. Plant Cell Culture Protocols, *Second Edition,* edited by *Victor M. Loyola-Vargas and Felipe Vázquez-Flota, 2005*

317. Differential Display Methods and Protocols, *Second Edition,* edited by *Peng Liang, Jonathan Meade, and Arthur B. Pardee, 2005*

316. Bioinformatics and Drug Discovery, edited by *Richard S. Larson, 2005*

315. Mast Cells: *Methods and Protocols,* edited by *Guha Krishnaswamy and David S. Chi, 2005*

314. DNA Repair Protocols: *Mammalian Systems, Second Edition,* edited by *Daryl S. Henderson, 2006*

313. Yeast Protocols: *Second Edition,* edited by *Wei Xiao, 2005*

312. Calcium Signaling Protocols: *Second Edition,* edited by *David G. Lambert, 2005*

311. Pharmacogenomics: *Methods and Protocols,* edited by *Federico Innocenti, 2005*

310. Chemical Genomics: *Reviews and Protocols,* edited by *Edward D. Zanders, 2005*

309. RNA Silencing: *Methods and Protocols,* edited by *Gordon Carmichael, 2005*

308. Therapeutic Proteins: *Methods and Protocols,* edited by *C. Mark Smales and David C. James, 2005*

307. Phosphodiesterase Methods and Protocols, edited by *Claire Lugnier, 2005*

306. Receptor Binding Techniques: *Second Edition,* edited by *Anthony P. Davenport, 2005*

305. Protein–Ligand Interactions: *Methods and Applications,* edited by *G. Ulrich Nienhaus, 2005*

304. Human Retrovirus Protocols: *Virology and Molecular Biology,* edited by *Tuofu Zhu, 2005*

303. NanoBiotechnology Protocols, edited by *Sandra J. Rosenthal and David W. Wright, 2005*

302. Handbook of ELISPOT: *Methods and Protocols,* edited by *Alexander E. Kalyuzhny, 2005*

301. Ubiquitin–Proteasome Protocols, edited by *Cam Patterson and Douglas M. Cyr, 2005*

300. Protein Nanotechnology: *Protocols, Instrumentation, and Applications,* edited by *Tuan Vo-Dinh, 2005*

299. Amyloid Proteins: *Methods and Protocols,* edited by *Einar M. Sigurdsson, 2005*

METHODS IN MOLECULAR BIOLOGY™

Arabidopsis Protocols

Second Edition

Edited by

Julio Salinas, PhD

Dept. of Biotechnology,
Instituto Nacional de Investigacion y Tecnología Agraria y Alimentaria (INIA),
Madrid, Spain

and

Jose J. Sanchez-Serrano, PhD

Dept. de Genetica Molecular de Plantas,
Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco,
Madrid, Spain

© 2006 Humana Press Inc.
999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

www.humanapress.com

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise without written permission from the Publisher. Methods in Molecular Biology™ is a trademark of The Humana Press Inc.

All papers, comments, opinions, conclusions, or recommendations are those of the author(s), and do not necessarily reflect the views of the publisher.

This publication is printed on acid-free paper. ∞
ANSI Z39.48-1984 (American Standards Institute)
Permanence of Paper for Printed Library Materials.

Production Editor: Tracy Catanese

Cover illustration: Fig. 2 from Chapter 22, "β-Glucuronidase as Reporter Gene: Advantages and Limitations," by Kye-Won Kim, Vincent R. Franceschi, Laurence B. Davin, and Norman G. Lewis.

Cover design by Patricia F. Cleary

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341; E-mail: orders@humanapr.com; or visit our Website: humanapress.com

Photocopy Authorization Policy:
Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Humana Press, provided that the base fee of US $30.00 per copy is paid directly to the Copyright Clearance Center at 222 Rosewood Drive, Danvers, MA 01923. For those organizations that have been granted a photocopy license from the CCC, a separate system of payment has been arranged and is acceptable to Humana Press Inc. The fee code for users of the Transactional Reporting Service is: [1-58829-395-5/06 $30.00].

Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

eISBN: 1-59745-003-0

Library of Congress Cataloging in Publication Data

Arabidopsis protocols / edited by Julio Salinas and Jose J. Sanchez-Serrano.-- 2nd ed.
p. ; cm. -- (Methods in molecular biology ; 323)
Includes bibliographical references and index.
ISBN 1-58829-395-5 (alk. paper)
1. Arabidopsis--Laboratory manuals. 2. Arabidopsis--Molecular aspects--Laboratory manuals.
[DNLM: 1. Arabidopsis--genetics--Laboratory Manuals. 2. Plant
Preparations--Laboratory Manuals. 3. Plant Components--Laboratory Manuals.
SB 608.A7 A658 2005] I. Salinas, Julio. II. Sanchez-Serrano, Jose J. III. Series: Methods in molecular biology (Clifton, N.J.) ; v. 323.
QK495.C9A735 2005
583'.64--dc22

2005016343

Preface

For several decades, *Arabidopsis thaliana* has been the organism of choice in the laboratories of many plant geneticists, physiologists, developmental biologists, and biochemists around the world. During this time, a huge amount of knowledge has been acquired on the biology of this plant species, which has resulted in the development of molecular tools that account for much more efficient research. The significance that *Arabidopsis* would attain in biological research may have been difficult to foresee in the 1980s, when its use in the laboratory started. In the meantime, it has become the model plant organism, much the same way as *Drosophila*, *Caenorhabditis*, or mouse have for animal systems. Today, it is difficult to envision research at the cutting edge of plant biology without the use of *Arabidopsis*.

Since the first edition of Arabidopsis *Protocols* appeared, new developments have fostered an impressive advance in plant biology that prompted us to prepare Arabidopsis *Protocols, Second Edition*. Completion of the *Arabidopsis* genome sequence offered for the first time the opportunity to have in hand all of the genetic information required for studying plant function. In addition, the development of whole systems approaches that allow global analysis of gene expression and protein and metabolite dynamics has encouraged scientists to explore new scenarios that are extending the limits of our knowledge. These advances will eventually lead to an understanding of how this complex, multicellular organism works, how it copes with the particularities of a sessile life style, and how these strategies compare with those developed in other organisms.

While conceiving this second edition of Arabidopsis *Protocols*, our aim remained as it was in the first: the book should provide both experienced researchers and beginners in the field of plant biology with a comprehensive set of up-to-date protocols covering the many methods developed for work with this species. Readers should have little problem understanding the general design or the specific details of the experimental protocols. Nevertheless, we recommend reading the protocols in advance, before bench work is started. Although a few chapters in this edition have been maintained and updated, most are new and describe technologies that have only very recently been developed in *Arabidopsis*. We hope that Arabidopsis *Protocols, Second Edition* will continue as a standard reference for laboratories working with *Arabidopsis* and other plant species.

Julio Salinas
Jose J. Sanchez-Serrano

Contents

Preface v
Contributors xiii
Color Plates xv

Part I. Growing *Arabidopsis*

1 Growth of Plants and Preservation of Seeds
***Luz Rivero-Lepinckas, Deborah Crist, and Randy Scholl* 3**

2 Prevention and Control of Pests and Diseases
***Jenifer Bush, Georg Jander, and Frederick M. Ausubel* 13**

3 Growth of Other Species Related to *Arabidopsis thaliana*
***Krithika Yogeeswaran and Mikhail Elia Nasrallah* 27**

4 Grafting
***Katherine Bainbridge, Tom Bennett, Colin Turnbull, and Ottoline Leyser* 39**

5 Synchronization, Transformation, and Cryopreservation of Suspension-Cultured Cells
***Margit Menges and James A. H. Murray* 45**

Part II. Genetic Analyses

6 Genetic Analysis
***Maarten Koornneef, Carlos Alonso-Blanco, and Piet Stam* 65**

7 QTL Analysis
***Carlos Alonso-Blanco, Maarten Koornneef, and Johan W. van Ooijen* 79**

8 EMS Mutagenesis of *Arabidopsis*
***YongSig Kim, Karen S. Schumaker, and Jian-Kang Zhu* 101**

9 Low-Resolution Mapping of Untagged Mutations
***M. Rosa Ponce, Pedro Robles, Francisca M. Lozano, M. Asunción Brotóns, and José L. Micol* 105**

10 Gene Identification and Cloning by Molecular Marker Mapping
***Georg Jander* 115**

11 High-Throughput TILLING for *Arabidopsis*
***Bradley J. Till, Trenton Colbert, Christine Codomo, Linda Enns, Jessica Johnson, Steven H. Reynolds, Jorja G. Henikoff, Elizabeth A. Greene, Michael N. Steine, Luca Comai, and Steven Henikoff* 127**

12 Genotyping and Mapping With High-Density Oligonucleotide Arrays
***Justin Borevitz* 137**

13 Forward Genetic Screening of Insertional Mutants
***Anja Schneider and Dario Leister* 147**

14 PCR-Based Screening for Insertional Mutants
Anna N. Stepanova and Jose M. Alonso ***163***

15 Cytogenetic Analysis of *Arabidopsis*
Martin Lysak, Paul Fransz, and Ingo Schubert ***173***

16 Using Information From Public *Arabidopsis* Databases to Aid Research
Margarita Garcia-Hernandez and Leonore Reiser ***187***

Part III. Transient and Stable Transformation

17 *In Planta Agrobacterium*-Mediated Transformation by Vacuum Infiltration
Brian W. Tague and Joanna Mantis ***215***

18 Transient Expression Assay by Agroinfiltration of Leaves
Min Woo Lee and Yinong Yang ***225***

19 Functional Analysis of Transcription Factors by Microparticle Bombardment
Ida Ruberti, Giovanna Sessa, and Giorgio Morelli ***231***

20 Transfection Assays With Protoplasts Containing Integrated Reporter Genes
Shiv Tiwari, Shucai Wang, Gretchen Hagen, and Tom J. Guilfoyle ***237***

21 Chloroplast Genetic Engineering Via Organogenesis or Somatic Embryogenesis
Amit Dhingra and Henry Daniell ***245***

22 β-Glucuronidase as Reporter Gene: *Advantages and Limitations*
Kye-Won Kim, Vincent R. Franceschi, Laurence B. Davin, and Norman G. Lewis ***263***

23 Use of Fluorescent Proteins as Reporters
Guido Jach ***275***

24 Luciferases as Reporter Genes
Megan M. Southern, Paul E. Brown, and Anthony Hall ***293***

25 Aequorin as a Reporter Gene
Christoph Plieth ***307***

26 Applications of Chemical-Inducible Expression Systems in Functional Genomics and Biotechnology
Jianru Zuo, Peter D. Hare, and Nam-Hai Chua ***329***

Part IV. Transcriptomics

27 RNA Extraction
Huazhong Shi and Ray Bressan ***345***

28 Target Preparation for DNA Microarray Hybridization
Tong Zhu, Sherman H. Chang, and Pedro Gil ***349***

29 Statistical Issues in Microarray Data Analysis
Willem A. Rensink and Samuel P. Hazen ***359***

30 Preparation and Quality Assesment of RNA From Cell-Specific Samples Obtained by Laser Microdissection
Regine Kleber and Julia Kehr ***367***

Part V. Proteomics

31 Protein Isolation and Second-Dimension Electrophoretic Separation
Bernhard Schlesier and Hans-Peter Mock* *381

32 Isolation of Nuclear Proteins
Tomasz T. Calikowski and Iris Meier* *393

33 Purification and Fractionation of Membranes for Proteomic Analysis
Anne Marmagne, Daniel Salvi, Norbert Rolland, Geneviève Ephritikhine, Jacques Joyard, and Hélène Barbier-Brygoo* *403

34 Protein Microsequencing
Alberto Paradela and Juan Pablo Albar* *421

35 Peptide-Based Phosphoproteomics With Immobilized Metal Ion Chromatography
Thomas S. Nühse and Scott C. Peck* *431

Part VI. Metabolomics

36 Metabolite Profiling in *Arabidopsis*
Oliver Fiehn* *439

37 Hormone Profiling in *Arabidopsis*
Axel Müller, Petra Düchting, and Elmar W. Weiler* *449

Index *459*

Contributors

JUAN PABLO ALBAR • *Departamento de Proteómica, Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco s/n, Universidad Autónoma de Madrid, Madrid, Spain*
JOSE M. ALONSO • *Department of Genetics, North Carolina State University, Raleigh, NC*
CARLOS ALONSO-BLANCO • *Departamento de Biotecnología, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) and Centro Nacional de Biotecnología (CNB), CSIC, Cantoblanco, Madrid, Spain*
FREDERICK M. AUSUBEL • *Department of Molecular Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, MA*
KATHERINE BAINBRIDGE • *Department of Biology, University of York, York, UK*
HÉLÈNE BARBIER-BRYGOO • *Institut des Sciences du Végétal, CNRS, Gif-Sur-Yvette Cedex, France*
TOM BENNETT • *Department of Biology, University of York, UK*
JUSTIN BOREVITZ • *Ecology and Evolution, University of Chicago, Chicago, IL*
RAY BRESSAN • *Center for Plant Environmental Stress Physiology, Purdue University, West Lafayette, IN*
M. ASUNCIÓN BROTÓNS • *División de Genética e Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, Alicante, Spain*
PAUL E. BROWN • *Department of Biological Sciences, University of Warwick, UK*
JENIFER BUSH • *Department of Molecular Biology, Massachusetts General Hospital, Boston, MA*
TOMASZ T. CALIKOWSKI • *Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland*
SHERMAN H. CHANG • *Diversa Corporation, San Diego, CA*
NAM-HAI CHUA • *Laboratory of Plant Molecular Biology, The Rockefeller University, New York, NY*
CHRISTINE CODOMO • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
TRENTON COLBERT • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
LUCA COMAI • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
DEBORAH CRIST • *Arabidopsis Biological Resource Center, Department of Plant Cellular and Molecular Biology, The Ohio State University, Columbus, OH*
HENRY DANIELL • *Department of Molecular Biology & Microbiology, University of Central Florida, Orlando FL*
LAURENCE B. DAVIN • *Institute of Biological Chemistry, Washington State University, Pullman, WA*
AMIT DHINGRA • *Department of Molecular Biology and Microbiology, University of Central Florida, Orlando FL*

Petra Düchting • *Lehrstuhl für Pflanzenphysiologie, Fakultät für Biologie, Ruhr-Universität Bochum, Bochum, Germany*
Linda Enns • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
Geneviève Ephritikhine • *Institut des Sciences du Végétal, Gif-Sur-Yvette Cedex, France*
Oliver Fiehn • *Max-Planck-Institute of Molecular Plant Physiology, Golm, Germany.*
Vincent R. Franceschi (Deceased, April 30, 2005) • *Swammerdam Institute of Life Science, University of Amsterdam, Amsterdam, The Netherlands*
Paul Fransz • *Institut fur Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany.*
Margarita García-Hernández • *Carnegie Institution Department of Plant Biology, Stanford, CA*
Pedro Gil • *Torrey Mesa Research Institute, San Diego, CA*
Elizabeth A. Greene • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
Tom J. Guilfoyle • *Department of Biochemistry, University of Missouri. Columbia, MO*
Anthony Hall • *School of Biological Sciences, University of Liverpool, Liverpool, UK*
Gretchen Hagen • *Department of Biochemistry, University of Missouri. Columbia, MO*
Peter D. Hare • *Laboratory of Plant Molecular Biology, The Rockefeller University, New York, NY*
Samuel P. Hazen • *The Scripps Research Institute, La Jolla, CA*
Jorja G. Henikoff • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
Steven Henikoff • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
Guido Jach • *Max Planck Institut fuer Zuechtungsforschung, Koeln, Germany*
Georg Jander • *Boyce Thompson Institute for Plant Research, Ithaca, NY*
Jessica Johnson • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
Jacques Joyard • *Laboratoire de Physiologie Cellulaire Végétale, UMR 5168, CNRS/CEA/INRA/UJF, CEA-Grenoble, France*
Julia Kehr • *Max-Planck-Institute of Molecular Plant Physiology, Golm, Germany.*
Kye-Won Kim • *Institute of Biological Chemistry, Washington State University, Pullman, WA*
Yong Sig Kim • *Department of Plant Sciences, University of Arizona, Tucson, AZ*
Regine Kleber • *Max-Planck-Institute of Molecular Plant Physiology,Golm, Germany*
Maarten Koornneef • *Laboratory of Genetics, Wageningen University, Wageningen, The Netherlands and Max Planck Institute for Plant Breeding Research, Cologne, Germany.*

Min Woo Lee • *Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL*
Dario Leister • *Abteilung für Pflanzenzüchtung und Ertragsphysiologie, Max-Planck-Institut für Züchtungsforschung, Köln, Germany.*
Norman G. Lewis • *Institute of Biological Chemistry, Washington State University, Pullman, WA*
Ottoline Leyser • *Department of Biology, University of York, UK*
Francisca M. Lozano • *División de Genética e Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, Alicante, Spain*
Martin Lysak • *Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, UK*
Joanna Mantis • *Department of Biology, Wake Forest University, Winston-Salem, NC*
Anne Marmagne • *Institut des Sciences du Végétal, CNRS, Gif-Sur-Yvette Cedex, France*
Iris Meier • *Department of Plant Cellular and Molecular Biology and Plant Biotechnology Center, Ohio State University, Columbus, OH*
Margit Menges • *Institute of Biotechnology, University of Cambridge, Cambridge, UK*
José L. Micol • *División de Genética e Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, Alicante, Spain.*
Hans-Peter Mock • *Institute of Plant Genetics and Crop Plant Research. Gatersleben, Germany*
Giorgio Morelli • *Istituto Nazionale di Ricerca per gli Alimenti e la Nutrizione, Rome, Italy*
Axel Müller • *CarboGen Laboratories AG, Aarau, Switzerland*
James A.H. Murray • *Institute of Biotechnology, University of Cambridge. Cambridge, UK*
Mikhail Elia Nasrallah • *Department of Plant Biology, Cornell University, Ithaca, NY*
Thomas S. Nühse • *The Sainsbury Laboratory, John Innes Centre, Norwich, UK*
Alberto Paradela • *Departamento de Proteómica, Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco s/n, Universidad Autónoma de Madrid, Madrid, Spain.*
Scott C. Peck • *The Sainsbury Laboratory, John Innes Centre. Norwich, UK*
Christoph Plieth • *Zentrum für Biochemie und Molekularbiologie, Christian-Albrechts-Universität, Kiel, Germany*
M. Rosa Ponce • *División de Genética e Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, Alicante, Spain*
Leonore Reiser • *Carnegie Institution Department of Plant Biology, Stanford, CA*
Willem A. Rensink • *The Institute for Genomic Research, Rockville, MD*
Steven H. Reynolds • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
Luz Rivero-Lepinckas • *Arabidopsis Biological Resource Center, Department of Plant Cellular and Molecular Biology, The Ohio State University, Columbus, OH*
Pedro Robles • *División de Genética e Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, Alicante, Spain.*

NORBERT ROLLAND • *Laboratoire de Physiologie Cellulaire Végétale, UMR 5168, CNRS/CEA/INRA/UJF, CEA-Grenoble, France*
IDA RUBERTI • *Dept. of Genetics and Molecular Biology, Istituto di Biologia e Patologia Molecolari CNR, University of Rome "La Sapienza," Rome, Italy*
JULIO SALINAS • *Dept. of Biotechnology, Instituto Nacional de Investigacion y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain*
DANIEL SALVI • *Laboratoire de Physiologie Cellulaire Végétale, UMR 5168, CNRS/CEA/INRA/UJF, CEA-Grenoble, France*
JOSE J. SANCHEZ-SERRANO • *Dept. de Genetica Molecular de Plantas, Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco, Madrid, Spain*
BERNHARD SCHLESIER • *Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany*
ANJA SCHNEIDER • *Botanisches Institut der Universität zu Köln, Köln, Germany*
RANDY SCHOLL • *Arabidopsis Biological Resource Center, Department of Plant Cellular and Molecular Biology, The Ohio State University, Columbus, OH*
INGO SCHUBERT • *Institut fur Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany.*
KAREN SCHUMAKER • *Department of Plant Sciences, University of Arizona, Tucson, AZ*
GIOVANNA SESSA • *Istituto di Biologia e Patologia Molecolari CNR, Rome, Italy*
HUAZHONG SHI • *Texas Tech University, Deptartment of Chemistry and Biochemistry, Lubbock, TX*
MEGAN M. SOUTHERN • *Department of Biological Sciences, University of Warwick, UK*
PIET STAM • *Laboratory of Plant Breeding, Wageningen University, Wageningen, The Netherlands.*
MICHAEL N. STEINE • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
ANNA N. STEPANOVA • *Department of Genetics, North Carolina State University, Raleigh, NC*
BRIAN W. TAGUE • *Department of Biology, Wake Forest University, Winston-Salem, NC*
BRADLEY J. TILL • *Basic Sciences Division, Fred Hutchinson Cancer Research Center, and Department of Biology, University of Washington, Seattle, WA*
SHIV TIWARI • *Department of Biochemistry, University of Missouri, Columbia, MO*
COLIN TURNBULL • *Division of Biology, Imperial College London, UK*
JOHAN W. VAN OOIJEN • *Kyazma B.V., Wageningen, The Netherlands.*
SHUCAI WANG • *Department of Biochemistry, University of Missouri, Columbia, MO*
ELMAR W. WEILER • *Lehrstuhl für Pflanzenphysiologie, Fakultät für Biologie, Ruhr-Universität Bochum, Bochum, Germany*
YINONG YANG • *Department of Plant Pathology, University of Arkansas, Fayetteville, AR*
KRITHIKA YOGEESWARAN • *Department of Plant Biology, Cornell University, Ithaca, NY*
JIAN-KANG ZHU • *Department of Botany and Plant Sciences, University of California, Riverside, CA*
TONG ZHU • *Syngenta Biotechnology Inc., Research Triangle Park, NC*
JIANRU ZUO • *Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Datun Road, Beijing, China*

Color Plates

Color plates 1–6 appear in an insert following p. 239.

Color Plate 1 Fig. 2 from Chapter 15; for full caption, *see* pp. 176–177.

Color Plate 2 Fig. 1 from Chapter 17; for full caption, *see* p. 218.

Color Plate 3 Fig. 1 from Chapter 18; for full caption, *see* p. 228.

Color Plate 4 Fig. 2 from Chapter 22; for full caption, *see* p. 268.

Color Plate 5 Fig. 3 from Chapter 22; for full caption, *see* p. 269.

Color Plate 6 Fig. 6 from Chapter 23; for full caption, *see* p. 285.

1

Growth of Plants and Preservation of Seeds

Luz Rivero-Lepinckas, Deborah Crist, and Randy Scholl

Summary

This chapter focuses on growth of plants on agar and soil in various environmental settings and especially in growth chambers and greenhouses. Harvesting, seed quality, and seed preservation are also considered. In addition, this chapter elaborates the conditions that are critical to the growth and development of healthy plants that produce high quality and quantity of seeds. The plant and seed management methods are discussed in the chronological order in which they would normally be utilized.

Key Words: *Arabidopsis* growth; seed storage; seed sterilization; seed harvest; seed preservation.

1. Introduction

Arabidopsis can be grown in a variety of environmental settings including growth rooms, growth chambers, greenhouses, window ledges, and outdoors. Peat moss-based mixes ***(1)***, commercial greenhouse mixes, relatively inert media watered with nutrient solutions, and agar can all be employed as plant substrates. This chapter focuses on the growth of plants on agar and soil in various environmental settings and especially in growth chambers and greenhouses. Harvesting, seed quality, and seed preservation are also considered. Additions in this chapter relative to the previous edition are (1) methods for growth of plants in sterile conditions, (2) high-density/high-seed productivity growth protocols, and (3) various modifications to the basic growth protocol, including fertilization, isolation of plants, and use of preventive application of pesticides.

This chapter elaborates the conditions that are critical to growth and development of healthy plants that produce high quality and quantity of seeds. The plant and seed management methods are discussed in the chronological order in which they would normally be utilized.

2. Materials

2.1. Growth and Harvest

1. Petri dishes (10-cm or 15-cm diameter, 10-cm square).
2. Growth media: 0.5X or 1X Murashige and Skoog (MS) mineral salts (*see* **Note 1**), 0.8–1% Bacto™Agar (Becton Dickinson), 0–3% sucrose (*see* **Note 2**), and optional vitamins (included with MS Complete Medium) ***(2)***. Preparation of 0.5X MS agar medium is as follows:
   a. Add 4.31 g of MS salts to 1.8 L of distilled water and stir to dissolve.
   b. Check and adjust pH to 5.7. Adjustments can be made with 1 *M* KOH.
   c. Dilute to final volume of 2 L and add agar (10 g/L).
   d. Autoclave for 15 min at 15 psi, 121°C.
   e. Optional sucrose and vitamins should be added after agar medium cools, before pouring solution into container (e.g., Petri dishes, magenta boxes, culture tubes) (*see* **Note 3**).

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

3. Distilled water.
4. Magnetic stirring device.
5. Beakers (1 L).
6. pH meter.
7. Pasteur pipet and latex bulb.
8. Bleach solution: household bleach (5.25% [w:v] sodium hypochlorite), 0.05% Tween-20.
9. Labeling tape or printable labels.
10. Permanent marker.
11. 8-cm round watch glass.
12. Plastic pots (e.g., 10-cm square, 5.5-cm square) or plastic flats (e.g., approx 26-cm × 53-cm) with holes in bottom.
13. Soil mixture (e.g., PRO-MIX "BX" or other peat moss-based potting mix).
14. Commercial fertilizer in slow-release pellets (e.g., Osmocote 14-14-14).
15. Large spoon or trowel.
16. Plastic transparent floral sleeves (Zwapak, Allsmeer, The Netherlands) for 10-cm square pots, or other devices such as Aracons™ (Lehle Seeds) or plastic bags.
17. Insecticides: Enstar, Tempo, Conserve, Marathon (granular), or similar.

2.2. Postharvest Seed Management

1. Lightweight transparent plastic food storage bags (approx 4-L).
2. Hand sieves (mesh size = 0.425 mm).
3. Small glass jars (125 mL) or other storage containers, such as small manila envelopes (e.g., 6 cm × 9 cm).
4. Cryovials (1 mL) or other sealed containers for permanent seed storage.
5. Forced draft oven.
6. Analytical balance.
7. Desiccator with silica gel.
8. Lightweight heat-resistant dishes with cover (preferably aluminum).
9. Tongs or forceps.
10. Petri dishes (10-cm diameter) or other similar containers.
11. Absorbent paper (e.g., filter paper).
12. Distilled water.
13. Parafilm or tape.

3. Methods

3.1. Plant Growth

3.1.1. Planting and Germination in Sterile Conditions

It is generally necessary to use sterile conditions to grow *Arabidopsis* for specific applications such as selection of drug-resistant plants, transformed plants, early root and shoot phenotypes, lethal mutants, and so on. Contaminants can essentially take over plant cultures. Various shapes and sizes of containers such as Petri dishes, "magenta" boxes, or culture tubes could be used, depending on the required length of the growing time (2–3 wk or to maturation) and characterization of phenotypes (shoots or roots). This section emphasizes the use of Petri dishes. All procedures should be accomplished in a sterile hood or environment.

Seeds can be surface-sterilized by soaking for 8 min in bleach solution and rinsing the seeds 3 to 5 times with sterile distilled water. Be sure that all bleach residue is removed. Maintain seeds in a small amount of water in a watch glass and plant immediately.

There are several methods for placing the seeds on medium, depending on the preferred plant density and type of container used:

1. For planting of individual seeds in low density, a small Pasteur pipet with a latex bulb on the upper end can be used. Exhaust air from the pipet, submerge its tip, and use slow-release pressure on bulb to draw a single seed into the end of the pipet. The seed can be dropped at the

desired location by carefully exhausting the pipet. Do not draw seeds beyond 1 to 2 cm into the pipet. Repeated pipettings are used for the remainder of the seeds.

2. For planting at high densities with uniform distribution on agar, mix seeds in sterile distilled water (or 0.1% cooled top agar), pour onto the dish, and swirl to distribute seeds evenly. A sterile Pasteur pipet tip can be used to move seeds around to adjust the distribution, and to remove excess water. Allow the water or top agar to dry slightly before replacing the lid.

After planting seeds on Petri dishes with agar, replace the cover and seal with Parafilm to prevent desiccation. Place dishes at 4°C (refrigerator temperature) for at least 2 to 4 d (*see* **Notes 4** and **5**) if needed. Dishes can be placed directly into the growth environment. A temperature of 23 to 25°C, 130 to 150 μE/m^2/s illumination are suitable.

3.1.2. Planting and Germination on Soil

Different mixtures and media can be utilized for growing *Arabidopsis*. Growth of plants on "soil" includes all media that can be successfully utilized for nonsterile growth of plants in pots or other similar containers. Mixtures of soil that have substantial peat moss with some perlite and vermiculite for aeration can be used successfully. Peat-based commercial mixes represent a convenient and reliable base for growing plants. Mixes such as Pro-Mix BX support healthy *Arabidopsis* growth and have fertilizer added so that fertilization is not necessary in the very early growth phases. Seeds can be planted in various ways (*see* **Note 6**). However, strict control of numbers of seeds planted can be maintained, and separate rows of different lines can be planted in the same pot for critical comparisons with the technique described here. Preparation of pots and planting can be accomplished as follows:

1. Thoroughly wet soil with tap water and apply Osmocote 14-14-14 (*see* **Note 7**). Alternatively, nutrient solution can be used to wet the soil ***(3)***. Mix well with a trowel or large spoon. Soil can be autoclaved to eliminate pests, but this is not usually necessary.
2. Place soil loosely in pots; level without compressing to give a uniform and soft bed. Pots are ready for planting (*see* **Note 8**).
3. When planting many seeds in a pot, scatter them carefully from a folded piece of filter or other paper, distributing seeds evenly onto the surface of the soil. When planting individual seeds, use a Pasteur pipet (*see* **Note 9** and **Subheading 3.1.1.**).
4. Cover pots with clear plastic wrap, and cut several small slits in the plastic with a knife (*see* **Note 10**). Place pots in 4°C (refrigerator temperature) for at least 2 d (*see* **Notes 4** and **5**).
5. After cold treatment, place pots in the growth area (growth chamber, growth room, greenhouse, and so on). Remove plastic wrap and maintain approx 2 cm of water around the base of the pots during the germination phase. Leave plastic wrap on for plants grown in a growth chamber and do not add additional water.

3.1.3. Care of Plants During Growth

The growth and development of *Arabidopsis*, including time of flowering, is affected by a number of growth conditions in addition to the genetic background. Seeds of the commonly used lines germinate 3 to 5 d after planting under continuous light, 25°C, adequate watering, and good nutrition (*see* **Note 7**). They form rosettes, bolt, and flower within 3 to 4 wk and can be harvested within 8 to 10 wk.

Management of light, temperature, and water during the growth of plants will ensure that healthy plants develop and produce high quality and quantity of seeds. The effect of each factor is discussed separately.

3.1.3.1. LIGHT

Optimum light is approx 130 to 150 μE/m^2/s. Very high output or cool white (VHO or SHO) fluorescent lamps, supplemented by incandescent lighting, are used for growth chambers. Older plants tolerate higher light intensity, up to full sun, although the use of 60% shade cloth in

summer greenhouses helps with light intensity control and temperature regulation. Supplemental evening and morning light is provided in the greenhouse during winter, as the plants generally require a long photoperiod (at least 12 h) for flowering. Continuous light (long photoperiod) can be used to accelerate the reproductive cycle. Photoperiods of 16 h work well for greenhouse growth. Short days (less than 12 h) favor growth of vegetative tissue and prevent or delay flowering.

3.1.3.2. Temperature

The optimum growth temperature range for *Arabidopsis* is 23 to 25°C. Lower temperatures are permissible, but higher temperatures are not recommended, especially for germination through early rosette development. Older plants tolerate higher temperatures, at least up to 30°C. It is advisable to set the greenhouse temperature at 21 to 23°C to avoid fluctuations to higher temperatures. It is recommended that night temperatures be maintained 2 to 4°C lower than the day temperature. In general, high temperatures favor a reduced number of leaves and flowers, and fertility is reduced. At lower temperatures, growth is slow, favoring the vegetative phase, and flowering is delayed. Some late flowering natural accessions (ecotypes) require an additional 4°C incubation (vernalization) of young rosettes for 3 to 4 wk to accelerate flowering.

3.1.3.3. Watering and Humidity

After germination, plants are watered as needed to avoid water stress. Water is best applied by subirrigation when the soil begins to dry. Subirrigation can be achieved by placing pots into flats or trays, allowing proper drainage of the soil. Over-watering should be avoided due to the potential for algal or fungal growth on the soil surface (*see* **Note 11**). Overwatering of greenhouse plants also provides favorable soil conditions for fungus gnat larvae. More frequent watering may be necessary during the first few days, as it is necessary to avoid any drying before the first two true leaves begin expanding. After plants have developed true leaves, watering frequency may be reduced to as low as once or twice per week until the plants flower. The water requirement of plants increases dramatically during silique filling. Daily watering at this stage is necessary for good seed production.

Water requirement is strongly influenced by relative humidity. *Arabidopsis* plants, including seedlings, tolerate low humidity (e.g., 20–30%) although increased humidity (e.g., 50–60%) greatly reduces the risk of accidental drying of the soil surface and subsequent desiccation of the fragile germinating seedlings. Very high humidity (more than 90%) can cause plant sterility. Low humidity (less than 50%) is desirable when siliques begin to mature.

3.1.3.4. Preventive Use of Insecticides

Where local governmental regulations permit and infestation is highly probable, application of insecticide as a preventive measure can be very effective in ensuring plant health throughout growth and avoiding heavy use of chemicals after infestations have occurred (*see also* Chapter 2 in this book). The following procedure is recommended:

1. Add 1.2 mL of Enstar, 1.2 mL of Tempo, and 1.2 mL of Conserve to 12 L of water. Mix well.
2. Spray lightly on rosettes prior to bolting stage—before placement of any isolation devices (*see* **Subheading 3.1.4.**).
3. Marathon (granular) can be applied at 1/8 teaspoon/pot (10-cm size), as directed by the label (*see* **Note 12**).

This treatment will prevent infestation of thrips, aphids, fungus gnats, and white flies.

3.1.4. Plant Isolation and Harvesting

It is necessary to avoid seed mixing among adjacently growing lines and to prevent loss of seeds due to shattering while ensuring quality of the harvested seeds. It is essential to keep plants of one line isolated from neighboring plants to ensure that absolutely no cross-contami-

nation can occur. It is useful to keep inflorescences from sprawling for maximum use of growth area. Various means and devices, such as Aracons™, plastic floral sleeves, and plastic bags can be employed to achieve these goals.

Watering of pots should be discontinued several days prior to harvest so that pots are dry when harvest is conducted. It should be noted that delays in harvesting following physiological maturation of the plant result in seed deterioration, especially under nonoptimal environmental conditions. Several seed collection strategies are compared below.

3.1.4.1. Early Harvest of Individual Siliques

Seeds from individual siliques can be harvested after the siliques have turned completely yellow, if rapid turnover is required. However, such seeds have high levels of germination inhibitors. For normal seed production, seeds are harvested only after the siliques have completely browned and when pressed with fingers do not compress (if the silique has dried even further, it may shatter at this point). At this stage, seeds are completely formed.

Because formation and maturation of siliques occur over time, early siliques can be harvested before later ones mature to avoid seed loss. However, it is usually recommended to wait until the entire inflorescence has browned before harvest.

3.1.4.2. Bulk Production on the Open Bench

For bulk seed production, the best method is to simply grow the plants on the open bench, keep all lines separated by adequate space, avoid disturbance of maturing inflorescences, and harvest when all siliques are dry. The entire inflorescence is cut off at its base and carefully placed into an approx 4-L or larger transparent plastic bag, depending on the size of the bulk of plants. This is compatible with the goals of high seed quality, maximum seed yield, and good pest protection. Some seeds may be lost, but the remainder are almost always healthy and result in vigorously germinating seedlings. After harvest, the entire contents of the bag are allowed to dry in preparation for threshing.

3.1.4.3. Commercial Seed Collectors

Aracons placed soon after bolting are effective for single plant harvesting. Harvesting after pots have been allowed to dry is accomplished easily by carefully cutting off the inflorescence under the device, placing the Aracon plus contents carefully into a large plastic bag (approx 4-L), removing the plant material from the plastic cylinder, and then shaking the seeds into the bag. Alternatively, if plants are totally dry, the plant material can be placed directly onto threshing sieve (*see* **Subheading 3.2.1.**).

3.1.4.4. Use of Plastic Floral Sleeves

For bulk seed production using individual pots, the pots can be placed into transparent plastic floral sleeves (*see* **Note 13**) near the time of bolting, so that all plant inflorescences are maintained within the sleeve, forming a propagator for each pot. At harvest, the sleeves can be cut or torn, the inflorescences cut off at the base, and the plant material placed into plastic bags or, if plants are totally dry, directly onto the threshing sieve (see **Subheading 3.2.1.**). This method is very effective for achieving high densities of lines while maintaining the productivity and purity of each pot.

3.1.4.5. Bagging Inflorescences by Pot

Inflorescences of non-*erecta* lines can simply be trained into an approx 4-L transparent plastic bag before any siliques begin to brown. The bag, however, may potentially collect moisture from transpiration or careless watering and provides a haven for insects when greenhouses are sprayed. To reduce these possibilities, the tops of these bags should be kept widely open at all times. Wait until the inflorescence has browned before harvesting. This method is conducive to strict isolation of the lines, and the bag serves to collect shattered seeds. Harvesting is accom-

plished by carefully cutting the entire inflorescence off at its base after all seeds have matured and shaking the seeds into the plastic bag.

3.2. Post-Harvest Handling and Preservation of Seeds

The longevity of seeds can be affected by (1) genotype, (2) prestorage environment, such as conditions during seed maturation, harvesting and seed handling, and (3) seed storage conditions. The genotype and prestorage conditions are important because they determine the maximum potential for seed longevity. Our experience regarding the effect of genotype is limited, although rapid deterioration of seeds has not been observed for the diverse collections currently maintained at the Arabidopsis Biological Resource Center (ABRC). The abscisic acid mutants are an example of a possible exception.

A slow process of deterioration begins as soon as seeds mature on a plant. Therefore the sooner seeds are placed into storage, the better. Harvested seeds should be processed promptly (including threshing, cleaning, drying and packaging) and then placed into storage. The following procedures form a sequence that ensures that the seeds will be conserved in the best possible condition.

3.2.1. Threshing and Cleaning

If seeds are collected in a plastic bag, the harvested plant material should be allowed to dry for a few days in the opened bag before threshing, as threshing is easier when the inflorescences are dry. Seeds should be threshed when the moisture content is approx 10%, to minimize seed damage during threshing. This moisture content will be reached when all material in the bag appears to be dry. The plastic bags containing dried inflorescences can be gently hand-pressed from the outside, and the seeds will fall to the bottom of the bag (*see* **Note 14**). Most of the dry inflorescence can be removed from the bag by hand before the seeds are sieved to separate them from the chaff.

Hand sieves with graded mesh sizes are recommended to remove debris, with seeds passing through the mesh and collected on clean paper. Totally dry plants from Aracons and sleeves can be placed directly onto the sieve. After sieving, the seeds are still likely to be mixed with soil and residue. A combination of additional sieving, blowing, and visual inspection can be employed to clean the seeds completely. Small samples can be cleaned by hand with the aid of a pointed tool on an opaque glass plate illuminated from below. Cleaned seed samples are placed in open, carefully labeled glass jars (do not use plastic due to static effects), or in small manila envelopes to allow seeds to dry.

3.2.2. Seed Drying

The moisture content of *Arabidopsis* seeds after threshing is usually around 10%. The seeds should be dried to 5 to 6% moisture prior to storage. Higher moisture content can cause seed deterioration. There are many methods available for drying seeds. The safest method is to air-dry the seeds at room temperature for 1 to 3 wk (*see* **Note 15**). Low relative humidity (20 to 30%) is necessary for seeds to reach the desired moisture content. The lower the humidity, the faster the seeds will dry and the lower their final moisture content. If, after testing, the moisture content is found to be not low enough, continue to dry further and check again.

3.2.3. Seed Moisture Content Determination

Moisture testing is necessary to verify that seeds are dry enough for storage. Seed moisture content can be determined by several methods. The following method is a destructive method, and the seeds employed for testing will no longer be viable. This is a "modified low constant temperature oven" method.

1. Preheat oven to 100 to 105°C (*see* **Note 16**).
2. Accurately weigh one clean numbered dish and its cover to four decimal places using an analytical balance. Record the weight (W1) (*see* **Note 17**).

3. Add approx 100 or 200 mg of seeds distributed evenly over the base of the dish, replace the cover, and accurately weigh the dish and cover. Record this weight (W2).
4. Place the dish in a safe place, and continue to prepare the second and/or third replicates in the same way.
5. When all samples have been weighed into numbered dishes, place each dish on top of its numbered lid in the oven. Heat the samples for 15 to 17 h.
6. Remove the dishes from the oven, replace their covers, and immediately place in a desiccator at room temperature to cool for 30 to 45 min. After heating, the dishes must be placed directly into the desiccator so that the dry seeds do not absorb any moisture.
7. Remove the dishes one by one from the desiccator, immediately weigh each dish and lid, and record the weight (W3). Do not leave the desiccator open while weighing the samples.
8. Moisture content is calculated as the percentage loss in weight of the original weight of seeds. This is known as the wet or fresh-weight basis, and is expressed to one decimal place. Algebraically, if W1 is the weight of the dish, W2 the weight of dish and seed before drying, and W3 the weight of dish and seed after drying, then:

$$\% \text{ Moisture Content} = (W2 - W3) / (W2 - W1) \times 100$$

3.2.4. Seed Packaging for Storage

When seed moisture content is within the safe storage limits, dried seeds should be placed in tightly sealed, impermeable containers to prevent rehydration. Cryovials (with threaded lids and gaskets) are convenient for storage. They hold large numbers of seeds, seal tightly, and can be resealed many times.

When packaging seeds, each container should be labeled with relevant information, including date of storage, using a waterproof permanent marker or a suitable printed label. In determining seed quantities, approx 1250 seeds = 25 mg = 50 μL. Seal the container immediately after filling, and visually check. During storage, check the containers at regular intervals to ensure that they remain in good condition.

3.2.5. Seed Storage and Preservation

The major factors influencing seed longevity are storage temperature and seed moisture content. The higher the value of either, the shorter the lifespan of the seeds. Seeds left at ambient temperature and relative humidity lose viability relatively quickly, although they may be viable for about 2 yr if stored in a dry atmosphere at room temperature. The conditions that prolong viability during storage have been well defined for plant seeds ***(4–6)***. Seed storage principles for *Arabidopsis* are similar to those for other plants.

For sealed cryovials or any moistureproof container, where seeds already have 5 to 6% moisture content, there are two storage options:

1. For active collections that are stored for short to medium terms and are accessed often, a convenient temperature is approx 4°C (regular refrigerator temperature).
2. For base collections where seeds are placed in long-term storage without disturbance, a temperature of –20°C is appropriate.

For open containers such as envelopes, the seeds can be stored at 15 to 16°C, with a relative humidity of 15%. Under this controlled environment, the seeds will maintain suitable low moisture content ***(7)***.

Removal of vials from storage represents a potentially very dangerous step. Vials must be warmed to room temperature before opening. Rapid rewarming (placing vial in a 37°C water bath for approx 10 min) serves to minimize freeze/frost damage that can occur during this process. Working in a low relative humidity environment, if possible, also aids in prevention of hydration. If it is suspected that condensation has occurred in a vial during storage or opening, the vial should be left open in a dry location until seeds have desiccated before returning them to cold storage.

The arrangements of vials in storage can vary, but it is important to record the exact location of each line. Codes can be used to indicate boxes, racks, trays, and refrigerators/freezers.

3.2.6. Seed Viability and Testing

Seed viability is the condition in which seeds are alive and have the potential to germinate and develop into normal reproductively mature plants, given the appropriate conditions. Factors that affect viability include the initial viability of the seeds at the start of the storage, seed moisture content, and storage environment. Viability should be monitored at regular intervals (*8,9*). It is anticipated that viability of *Arabidopsis* seeds should remain high for long storage periods, assuming proper conditions. The International Board for Plant Genetic Resources (IBPGR) recommends that seeds in long-term storage under optimal preservation standards should be monitored at least every 10 yr. Seeds in long-term storage with either poor storage life or poor initial viability and seeds in short- to medium-term storage should be monitored at least every 5 yr.

A viability test for *Arabidopsis* seeds can be conducted in 3 to 6 d. Tests should be carried out before seeds are packaged and stored, so that poor-quality seeds can be recognized. A germination test is the best method of estimating seed viability. *Arabidopsis* seeds may fail to germinate because they are dormant or because they are dead. Dormant seeds typically remain firm and in good condition during the germination test, whereas dead seeds soften and are attacked by fungi. Imbibing seeds with water at low temperatures (*see* **Note 4**) can usually break dormancy. The following method is suitable for *Arabidopsis*:

1. Place two layers of filter paper (free from chemical residues that could interfere with the germination of the seeds) firmly in the bottom of a 10-cm diameter Petri dish, labeled with line number and date.
2. Moisten the paper with distilled water. The paper should be totally saturated, but no excess water should be left in the dish.
3. Distribute 100 seeds uniformly on the surface of the paper. Replace the lid and seal the dish with Parafilm or clear tape, to preclude desiccation.
4. Cold-treat seeds by placing Petri dishes in the refrigerator for 2 to 4 d.
5. Place the dishes on an illuminated shelf (or in a growth chamber) under standard light and temperature conditions (*see* **Note 18**).
6. After 3 to 6 d count germinated versus ungerminated seeds, and record germination percentage.

4. Notes

1. "Ready to use" MS complete medium, including vitamins, can be purchased from Sigma (St. Louis, MO).
2. Plants grow more vigorously and more quickly in media containing sucrose, but bacterial and fungal contaminants must be rigorously excluded.
3. Plates with solidified agar can be stored in a container that prevents desiccation for several weeks at 4°C.
4. The use of a cold treatment (4°C for 2–4 d) to break dormancy of seeds, also called *stratification*, is very important for plantings utilizing freshly harvested seeds, which have more pronounced dormancy. Most widely used lines have moderate dormancy, and cold treatment may not be required when planting older seeds of these lines. However, a cold treatment may improve germination rate and its synchrony. For certain lines as many as 7 d of cold treatment are necessary. Cold treatment of dry seeds is usually not effective in breaking dormancy.
5. Seeds suspended in sterile water or top agar can be cold treated (stratified) prior to planting on agar or soil surface.
6. Pots approx 5.5 cm square can be used to grow one plant, 10-cm square pots are suitable for growing up to 60 plants, and rectangular flats that are approx 26 cm × 53 cm can accommodate as many as 200 to 600 plants to maturity. Another option especially suitable for genomic studies is 96-well insets. Higher densities, approx 3000 plants per 30-cm square, can be used if plants are harvested at the seedling stage.

7. Osmocote 14-14-14 (14% nitrogen, 14% phosphate, 14% potassium) is a commercially available extended time-release fertilizer, feeding up to 3 mo from planting. Apply in amounts according to the label.
8. Prepared pots can be stored in covered pans or trays for several days before planting, although pot preparation and planting should be conducted on the same day if possible.
9. Seeds can be planted by various methods depending on the purpose of the plants and availability of seeds. The density of planting depends on the genetic material and the purpose of the planting. For seed production, high yields are achieved utilizing densities of 10 to 20 plants per 10-cm-square pot. Larger populations do not necessarily reduce yield, but production per plant is reduced inversely. Larger populations are necessary for maintenance of representative proportions in a segregating population; this can be achieved with more dense plantings in one or two 10-cm pots or in flats. Planted seeds should not be covered with additional soil; *Arabidopsis* seeds need light for germination.
10. If several pots are planted, they can be placed in a tray or other similar container and covered with clear plastic wrap. In all cases the plastic wrap should not be allowed to contact the soil surface. The wrap is perforated in order to provide some aeration. Clear plastic domes are available for covering flats, but should not be tightly sealed. However, in the greenhouse, this method can overheat the soil surface and kill the germinating seedlings. In this case, the pots are left uncovered and placed in pans filled with 1 to 3 cm of water, which is maintained continually until all plants germinate and have expanded cotyledons.
11. If algae cover the soil as a result of overwatering of young plants, the algae can be carefully scraped from the soil surface if necessary. The pots should then be allowed to dry, which will kill algae before the plants become stressed.
12. Marathon can be added to the soil surface per the product label. However, applications for small plants can be made in the tray with water in cases in which the pots are being subirrigated. This practice reduces the risk of damage to the plants by the insecticide.
13. Floral sleeves fit snugly around a 10-cm pot, extend 60 cm upward, and are wider at the top, allowing for expansion of the developing inflorescences. Sleeves made of biaxially-oriented polypropylene (BOPP) are very clear, maintain upright stiffness, and tear easily for harvesting.
14. Hand rather than machine threshing and cleaning of the small *Arabidopsis* seeds is recommended mainly because the threshing machines need rigorous cleaning between lines to prevent sample cross-contamination, require very careful adjustment and do not accommodate the variable size of *Arabidopsis* seeds well.
15. Because *Arabidopsis* seeds equilibrate to room conditions, it is possible to reliably predict the approximate moisture content of seeds stored in open containers by simply measuring the room humidity. The relationship for *Arabidopsis* is similar to others published for crop seeds having similar chemical composition ***(5)***.
16. High temperatures should not be used because the oil in the seeds may also vaporize and give false results. Temperatures of just over 100°C evaporate water and minimize vaporization of oils.
17. Lightweight dishes should be used, so that the ratio of the weight of the seeds and the dish is not too disproportionate. It is suggested that a minimum of three replicates of 100 mg of seeds or two replicates of 200 mg of seeds per sample be used. Always work with care and finish one sample at a time. Do not leave the dishes open in the laboratory between weighings because the seeds will absorb moisture rapidly from the air, and small changes in weights can result in large differences in the calculations.
18. Environmental conditions for seed germination tests are the same as for growing plants. Two replicates of 100 seeds each provide reliable germination estimates. Cases in which observed germination is <80% may warrant follow-up testing to verify the low rate before it is accepted.

References

1. Kranz, A. and Kirchheim, B. (1987) Genetic Resources in *Arabidopsis*. *Arabidopsis* Information Service, Frankfurt.
2. Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15,** 473–497.
3. Estelle, M. and Somerville, C. (1987) Auxin-resistant mutants of *Arabidopsis* thaliana with altered morphology. *Mol. Gen. Genet.* **206,** 200–206.

4. International Board for Plant Genetic Resources. (1976) Report of IBPGR Working Group on engineering, design and cost aspects of long-term seed storage facilities. IBPGR, Rome.
5. Justice, O. L. and Bass, L. N. (1978) Principles and practices of seed storage. *Agriculture Handbook*. United States Department of Agriculture, Washington, DC. **506,** 34–80.
6. Roberts, E. H. and Ellis, R. H. (1984) The implications of the deterioration of orthodox seeds during storage for genetic resources conservation, in *Crop Genetic Resources: Conservation and Evaluation* (Holden, J. H. W. and Williams, J. T., eds.). George Allen & Unwin, London, pp. 19–37.
7. Hay, F., Mead, A., Manger, K., and Wilson, F. (2003) One-step Analysis of seed storage data and the longevity of *Arabidopsis thaliana* seeds. *J. Exper. Bot.* **54,** 993–1011.
8. Ellis, R. H. and Roberts, E. H. (1984) Procedures for monitoring the viability of accessions during storage, in *Crop Genetic Resources: Conservation and Evaluation* (Holden, J. H. W. and Williams, J. T., eds.). George Allen & Unwin, London, pp. 63–76.
9. Ellis, R. H., Hong, T. D., and Roberts, E. H. (1985) *Handbook of Seed Technology for Genebanks. Vol. I. Principles and Methodology*. International Board for Plant Genetic Resources (IBPGR), Rome.

2

Prevention and Control of Pests and Diseases

Jenifer Bush, Georg Jander, and Frederick M. Ausubel

Summary

A well-controlled growth environment with plants that are not unduly stressed is essential for *Arabidopsis* molecular biology research. Even if they do not kill the plants outright, insect pests and microbial pathogens can cause subtle changes in gene expression or plant metabolism that affect experimental results. Therefore, regular scouting for infestations, frequent cleaning of plant growth areas, proper disposal of dead or diseased plant material, and controlled access to the greenhouses or growth chambers will help to make experiments more reproducible. Powdery mildew, a fungal pathogen, and arthropod pests, including aphids, thrips, fungus gnats, and spider mites, are the most common greenhouse problems. Biological control methods such as parasitoid wasps and *Bacillus thuringiensis* crystal toxin can be used to contain some insect infestations. However, if an infestation gets out of hand despite reasonable precautions, insecticide or fungicide spraying by a licensed applicator may be necessary. Bacterial and viral infections of *Arabidopsis*, though they do occur, tend to be less common and can usually be controlled by maintaining optimal growth conditions and promptly disposing of dead or diseased plant material.

Key Words: Fungus gnats; thrips; aphids; spider mites; powdery mildew; *Botrytis*.

1. Introduction

Most *Arabidopsis* experiments require healthy plants that are free from pests and diseases. As *Arabidopsis* plants are grown from seed, have a short life cycle, and are often raised in growth chambers rather than greenhouses, it is possible to maintain an almost entirely pest-free growth environment. However, pests including insects, mites, and pathogens do sometimes find their way into *Arabidopsis* growth facilities. Early recognition, treatment, and continued monitoring are key to preventing a pest outbreak from getting out of control.

Many greenhouse pests, including thrips, aphids, spider mites, and powdery mildew, can infest *Arabidopsis* (*1*). Therefore, common greenhouse practices used in the care of other plant species can be applied equally well to preventing pest and pathogen infestations of *Arabidopsis*. This chapter focuses on the monitoring and management of pests and pathogens that are frequently found on Arabidopsis. It also includes a general description of plant symptoms, life cycles of the most common pests, and some of the available treatment options. Some specific chemical treatments and biological control organisms are discussed, but this does not represent an exhaustive list, nor is this chapter an endorsement for any particular chemical manufacturer. The specific treatment strategy adopted to control a particular pest depends on many factors and it is not possible to give a general approach that will apply to all *Arabidopsis* growth situations.

2. Pest Prevention Using Good Horticultural Practices

Healthy, vigorously growing plants will generally be less prone to pests and diseases than weak or stressed plants. Attention to the plants and their growth environment minimizes the

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

risk of a pest outbreak (2). It is much easier to prevent a pest problem than it is to contain one after it gets started, and early detection will make any outbreak that does happen much easier to manage. Here, we outline good cultural practices that will help to prevent pest outbreaks.

2.1. Growth Area

1. Keep the growth area clean. Start with a clean shelf or greenhouse bench. Used soil and old plant debris can harbor insects or fungal spores from previous plantings. Therefore, before planting, clean the bench area thoroughly.
2. Periodically schedule a time when the growth area can be completely emptied out. Wash all surfaces with a greenhouse disinfectant such as Zerotol™ (Biosafe Systems), Physan 20™ (Maril Systems), or 10% bleach. Bleach will leave a residue, so if using bleach, it is important that it be rinsed off afterwards. After cleaning, turn up the temperature in the growth facility as high as possible for 2 to 4 d. Because you have removed the host material, sanitized, and "cooked" the area, it is very unlikely that any pests or spores will survive.
3. Clean the floors of the greenhouse or growth room weekly. Sweep up debris and mop the floors.

2.2. Growth Supplies

1. Always use clean or new pots and trays. For potting soil, use bagged, commercial, peat-based soilless mixes. If pots or flats must be reused, they should be cleaned thoroughly and sterilized by soaking overnight in 10% bleach solution and rinsing with water afterward.
2. Some *Arabidopsis* growers choose to autoclave their soil before planting in order to kill any pests or pathogens that may be present. This is not advisable, however, because it may adversely affect the soil properties. Commercial mixes normally contain a wetting agent and a nutrient charge that may be altered after autoclaving. Autoclaving soil also involves much extra labor.

2.3. Dead Plant Material

1. Do not leave dead or drying plants in the growth area. When plants are no longer needed, bag them securely and remove them from the growth facility. Also, remove cells that do not have any germinated seeds from plant flats. These cells will stay wetter than ones with plants growing in them and can breed fungus gnats and diseases.
2. Plants that are dry and awaiting seed harvest should be moved out of the growth facility. This reduces the chance that pests from potentially heavily infested senescing plants will migrate to vulnerable seedlings. The ideal situation is to have a separate seed harvest room, where senescing or dry plants can be stored at room temperature, away from the growth facility, allowing valuable growth space to be used for actively growing plants.

2.4. Controlling Pest Outbreaks

Avoid introducing and spreading pests.

1. If there are already pests in the growth facility, establish an entry protocol whereby anyone who has entered an infested area refrains from entering uninfested areas later on that same day. Make sure that all the users of the growth facility adhere to this protocol, even if it seems inconvenient.
2. Wear a lab coat when working in the growth facility. Pests are easily spread on clothing, both by people moving within the facility and by people coming in from the outdoors. Dedicate a set of "growth area only" lab coats for all users of the facility. Use sticky mats at the entrance of the growth facility to help prevent pests from being tracked in.
3. Quarantine any new plants to avoid introducing pests from another growth area. Isolate the new material for a few days and carefully inspect it for emerging pests. This waiting period is important because pests that are not visible initially may hatch out.
4. The easiest way to control a pest outbreak is often to dispose of the affected plants. Although this may set back a research project for a few weeks, it is worth considering in order to curtail a more serious outbreak. If the infested plants are essential and must be saved, move them to a separate growth chamber or other isolated area that is away from unaffected plants. When

moving infested plants, prevent pests from spreading by covering the flat with a dome or by bagging the flat. Clean the cart and dome afterwards.

5. Keep similarly aged plants together. Plants of the same age will be easier to monitor for pests. Older, senescing plants seem to be more vulnerable to pests and it is best to keep them separated from young seedlings. The ideal approach is to coordinate growth of plants in a particular area so that they mature at about the same time, allowing the growth facility to be cleaned out before planting again.
6. A licensed pesticide applicator or greenhouse manager will be able to handle pesticides safely and properly. In the United States, the regulations of the Worker Protection Standard (www.epa.gov/pesticides/safety/workers/workers.htm) must be followed to protect workers and pesticide handlers from exposure. Local rules may vary, and different regulations for pesticide use will apply in other countries. Some of the pesticides mentioned in this chapter are required by law to be applied by someone who has had pesticide applicator training. In some cases it is necessary to restrict access to the growth facility while the pesticides are being applied, as well as for a period immediately afterward. Pesticide labels describe the specific "reentry interval," as well as the required personal protective equipment (nitrile gloves, respirator, etc.).
7. For spraying fewer than 10 flats, a small hand sprayer will be sufficient. For larger infestations, use a pump sprayer. Set the nozzle to a fine setting. It is best to create a "mist" of small droplets that will evenly cover the surfaces of the plant rather than a "rain" of larger ones.
8. Some pesticides, such as those used to treat fungus gnat larvae, must be applied as a "sprench." The goal is to spray and drench the soil at the same time. For fewer than 10 flats, a common laboratory squirt bottle works well for this. For larger numbers of flats, use a pump sprayer with the nozzle set at the coarsest setting.
9. Always check for phytotoxicity on a few plants before undertaking a large-scale pesticide application. Read and follow the directions on the pesticide label, use proper personal protective equipment, adhere to reentry interval guidelines, and keep good records of all pesticide applications.
10. Consider the growth environment before spraying. Greenhouses are usually designed so that they can be sealed off before spraying, but that is not necessarily the case for growth rooms, which may vent into work spaces. Make sure that the air from the growth rooms does not recirculate with air in neighboring labs and offices. Growth chamber manufacturers recommend that pesticides should not be applied in reach-in chambers.
11. Biological pest controls such as predatory mites, parasitoid wasps, nematodes, and the bacterium *Bacillus thuringiensis* can be used on *Arabidopsis*. However, such control methods are not absolute and usually require some tolerance to low-level pest populations. Extremely keen scouting abilities are also important to ensure that biological controls are implemented early in an infestation. Biological pest control is not always appropriate for *Arabidopsis* research because even a low-level pest problem can alter plant phenotypes beyond experimental parameters. Biological control agents work best when applied before a pest outbreak, rather than once it has started. Therefore, weekly releases of biological control organisms are recommended.

2.5. Early Signs of Pest Outbreaks

Learn the early signs of pest outbreaks and monitor for these. Often, damaged or sick plants can be recognized before the pests themselves are observed *(3)*. Regularly inspect all plants in the growth area. Keep a record of the locations where you suspect a pest outbreak and check those areas closely.

2.6. Maintaining Healthy Plants

1. Provide optimal plant growth conditions (temperature, light, humidity) to generate healthy plants (*4*; *see also* Chapter 1 in this book). *Arabidopsis* grows well at 20 to 23°C, 50 to 65% relative humidity, and a light intensity of 80 to 150 μE.
2. Do not overwater or underwater plants. Thoroughly wet the soil and then allow it to dry out slightly before watering again. In most situations watering two times per week is sufficient, though the watering schedule will change with the age of the plants. An added advantage to

such a watering protocol is that the incidence of fungus gnats and other underground pests is reduced if the soil is not continuously wet.

3. Ensure that there is good air circulation around plants.

3. Arthropod Pests

Fungus gnats and thrips are the most common pests of *Arabidopsis* in growth chambers, growth rooms, and greenhouses. Aphids and spider mites are less problematic, but can also infest *Arabidopsis*, especially in greenhouses where they may already be established on other plants.

3.1. Fungus Gnats

Fungus gnats (*Bradysia* spp.) are one of the most serious *Arabidopsis* pests. They can appear, seemingly from nowhere, and quickly take hold. Often, fungus gnats are not detected until the adult flies emerge. However, it is the larval stage that does all the damage and it is therefore important to monitor plants for signs of larval feeding. Fungus gnat larvae feed on tender roots, fungi, decomposing organic matter, and decaying plant tissue in the soil. They will also chew on leaves that are touching the surface of the soil. Larvae pupate underground and emerge at the surface as adult gnats. Adult fungus gnats do not harm the plants, except by laying more eggs, which soon hatch into more hungry larvae. The full fungus gnat life cycle takes 3 to 4 wk, which may account for the fact that the heaviest infestations are usually seen on older plants.

Shore flies (*Scatella* spp.) are often confused with fungus gnats. These flies do slightly resemble adult gnats, but are stockier, have white spots on their wings, and tend to fly when disturbed. Fungus gnat adults will generally run along a surface when prodded. Shore flies are associated with algae and breed in standing water such as that found in drainpipes and sink traps. They are not a significant pest to *Arabidopsis* and are not covered here.

3.1.1. Fungus Gnat Prevention

1. Fungus gnat eggs can be found in commercial, bagged soilless mixes, especially those that are high in partially decomposed organic matter. Before planting, drench the soil with a suspension containing *Bacillus thuringiensis israelensis*, which is sold as Gnatrol™ (Valent Biosciences). Any fungus gnat larvae that hatch in the soil will be killed after they ingest the *B. thuringiensis* toxin.
2. Allow growth media to dry slightly between watering and do not leave pools or puddles of standing water on the floor or bench. Fungus gnats thrive in damp, humid conditions that also promote fungus growth.

3.1.2 Fungus Gnat Detection

1. Plants grow slowly and lack vigor when fungus gnat larvae chew their roots. Seedling and early-rosette-aged plants are the most vulnerable.
2. Small round holes will appear on leaves touching the soil surface where the larvae have sampled leaf tissue (**Fig. 1A**). Seedlings are extremely vulnerable to fungus gnat larvae. With severe infestations, whole seedlings can be devoured overnight. If suspected fungus gnat holes are seen, carefully lift up the leaf and look for whitish translucent larvae with black heads (**Fig. 1B**). The larvae will avoid light, and usually respond by burrowing into the soil.
3. Yellow sticky cards (Whitmire Micro-Gen) hung throughout the growth area will catch adult gnats (**Fig. 1C**). These cards work only as a monitoring device and should not be considered a control method. Count the number of gnats on each card and keep records as a quantitative measure of the severity of the infestation. Continue monitoring after treating for the gnats to determine the efficacy of the treatment.
4. Cut a small (half-inch) plug of raw potato and place this on the soil surface. After one day, lift the plug and look for larvae feeding on the potato. This is also a good technique to use for monitoring the success of soil treatment, or as part of a regular scouting of the growth facility. Mark the positions where the potato plugs are and check them every few days. They will need to be replaced as they dry out or become moldy.

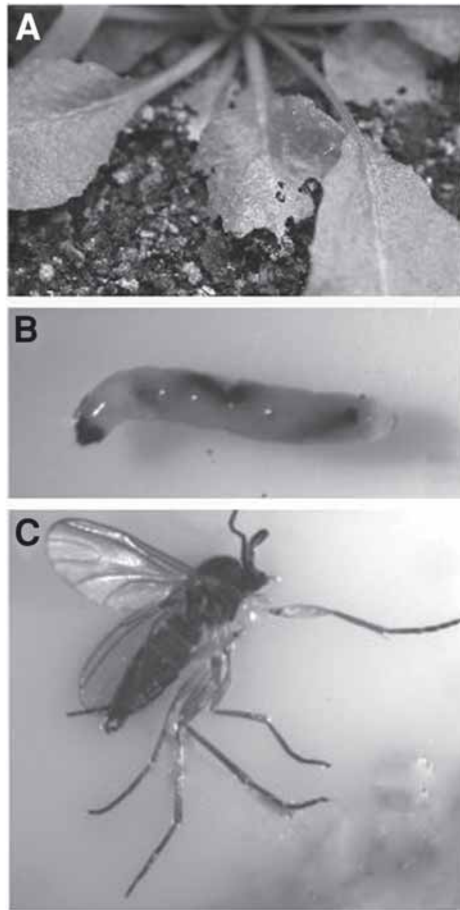


Fig. 1. Fungus gnats (*Braydisia* sp.). **(A)** *Arabidopsis* leaves chewed by fungus gnat larvae. **(B)** Fungus gnat larvae, approx 2 mm. **(C)** Fungus gnat adult trapped on sticky card, approx 2 mm.

3.1.3. Fungus Gnat Treatment

1. Treatment should be targeted at the larval stage. It difficult to control adults because they hide in places where they are not easily reached by pesticides. Fungus gnat larvae live near the soil surface, so the pesticide should be applied to the upper inch of the soil. This can be done using a spray nozzle on a coarse setting and spraying the material onto the soil, or by carefully watering the solution onto the flats. Eggs and pupae are generally not susceptible to pesticides, so it will be necessary to repeat applications, following the directions on the label, every 5 to 7 d until the infestation is eliminated.
2. Gnatrol is not only a good preventive treatment for fresh soil, but also a good pesticide to use on existing fungus gnat infestations. After ingestion of the *B. thuringiensis* bacterial toxin, the gut becomes paralyzed, feeding stops, and the larvae die. The strain of *B. thuringiensis* in Gnatrol specifically targets fly larvae and is not harmful to Arabidopsis or biological control insects that are used in the greenhouse. Gnatrol is effective for only 48 h, so it must be applied repeatedly.
3. Insect growth regulators (IGRs) are effective against fungus gnat larvae. These are pesticides that interfere with the normal development of the insect pest. S-kinoprene, sold as the product Enstar II® (Wellmark International) can be applied as a drench to the soil. Another IGR, azadirachtin, sold as Azatin® XL (Olympic Horticultural Products), is also applied as a drench and is extremely effective against fungus gnat larvae.

4. In addition to *B. t. israelensis* mentioned above, the predatory mite *Hypoapsis miles* and the predatory nematode *Steinernema feltiae* can be used on *Arabidopsis* for biological control of fungus gnats with good results. Both of these organisms need to be incorporated into the soil at planting time. The nematodes are sold under many brand names and, to ensure obtaining viable nematodes, it is best to purchase them through a reputable biological control supplier. A list of biological control suppliers in North America can be found at the Web site: www.cdpr.ca.gov/docs/ipminov/ben_supp/contents.htm

3.2. Thrips

Thrips, in most cases the species *Frankliniella occidentalis,* are another serious pest of Arabidopsis. In their 3-wk life cycle, these insects pass from eggs through two larval and two pupal stages to become adults. Pupae are generally in the soil, but all other growth stages are found on the plant. Thrips are good at hiding in axils, rosette centers, and even flower buds. As a result, they are difficult to control and early detection is particularly important.

3.2.1. Thrip Prevention

1. Regular, methodical monitoring is important for detecting thrips in the early stage of infestation.
2. Old soil and dead dry plant debris can harbor thrips and thrip pupae. When disposing of plant materials, completely remove them from the growth facility. A trash bag full of plants left by the door can serve as a reservoir for thrips waiting to colonize new plants.
3. Thrip-proof screens made from 0.16-mm mesh can be installed on greenhouse vents. Keep in mind, however, that these screens will severely restrict the air flow in the greenhouse. Additional ventilation may be required if such mesh is installed.
4. Keep greenhouse floors clean and clear of weeds that can harbor thrips.
5. Make preventive releases of the predatory mites *Neoseiulus cucumeris* or *Hypoapsis miles* (discussed below).

3.2.2. Thrip Detection and Monitoring

1. Typically thrip damage will be seen before the insects themselves are detected. Thrips scrape the leaf tissue and then consume the plant juices, which results in silvery-white patches that are usually seen at leaf edges, or near the center of the rosette (**Fig. 2A**). Sometimes the white patches will be covered with black dots of thrip frass.
2. Distorted expanding leaves are often caused by thrip damage.
3. Adult thrips are tiny, slender, and difficult to see. They are 1 to 2 mm long and resemble a small splinter (and are just as aggravating) (**Fig. 2B**). Larvae of thrips are slightly smaller, but can also be seen (**Fig. 2C**). If thrip damage is discovered, blow on the plant gently and then look at the damaged area with a hand lens. In response to the CO_2 in your breath, the adult and larval thrips are likely to come crawling out from the center of the rosette where they are hiding.
4. Tap plants, especially flowering or senescing plants, over a piece of white paper, and thrips will be seen crawling on the paper. In a similar manner, thrips are often found when harvesting seeds.

3.2.3. Thrip Treatment

1. Thrips can be suppressed with insecticidal soaps such as Safer®Brand (Woodstream Corp.), or M-Pede® (DowAgroSciences), but complete control is difficult. If plants are stressed, due to severe infestation or some other factor, soaps can sometimes damage leaves.
2. Thrips are known to readily develop resistance to several pesticides. For this reason it is important to rotate pesticides with different modes of action. Conserve™ (Dow AgroSciences), Mavrik Aquaflow® (Wellmark International), and Avid® (Novartis) all work against thrips on *Arabidopsis*. Plan to make two or three applications 7 d apart.
3. Treat all plants in the infested growing area. Adult thrips fly, and can also be transported on clothing.
4. Biological control is particularly challenging with established populations of thrips. However, the predatory mite *Neoseiulus cucumeris* can be used, and it does have the advantage of being

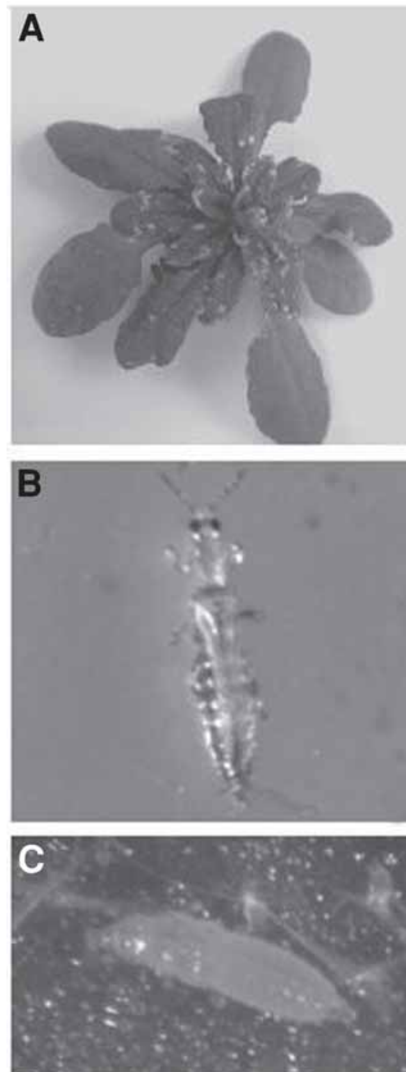


Fig. 2. Thrips (*F. occidentalis*). **(A)** *Arabidopsis* leaves damaged by thrip feeding. **(B)** Adult thrip, approx 1.5 mm. **(C)** Thrip larva, approx 0.5 mm.

able to reach into the thrips' hiding spaces (axils, flowers, etc.) where pesticides may not reach. This mite is especially good at attacking first instar larvae. Use of this mite does require a lot of attention to scouting and monitoring of both the predator population and the thrip population. Several releases of *N. cucumeris* will be required, and, if chemical pesticides are used, they must be compatible with this mite. Ask the biological control supplier about pesticide compatibility when purchasing mites. The soil-dwelling mite *Hypoapsis miles* will attack thrip pupae as well as fungus gnat larvae.

3.3. Aphids

Aphids are common greenhouse pests that also infest *Arabidopsis*. Aphid outbreaks in growth chambers or growth rooms occur less frequently, but there can be transfers of aphids from nearby greenhouses. The green peach aphid, *Myzus persicae,* is the species that is most frequently found on *Arabidopsis*, though other species (e.g., melon aphids and root aphids)

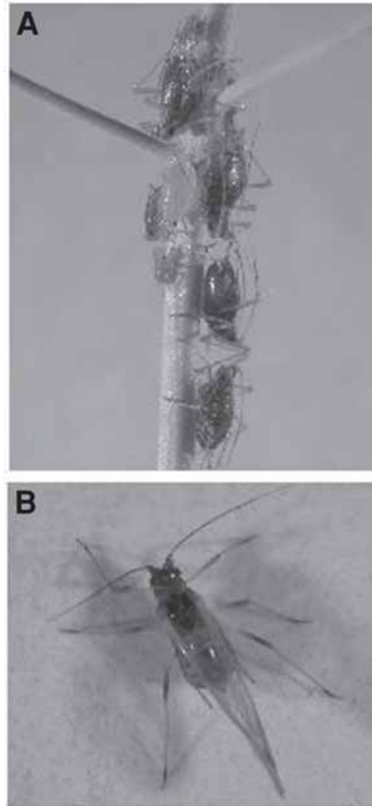


Fig. 3. Aphids (*M. persicae*). **(A)** Nymphs and adults feeding on an *Arabidopsis* stem, 0.5 to 2 mm. **(B)** Winged adult aphid, 1.5 to 3 mm.

have been reported. *M. persicae* are soft-bodied, yellow, green, red, or tan pear-shaped insects, 1 to 3 mm long, that are usually seen in clusters on the undersides of leaves or along the stems. Aphids reproduce by parthenogenesis, with females giving birth to genetically identical first-instar offspring. Because of this, even a single aphid can quickly become a large infestation. Winged asexual female aphids are produced when food becomes limiting—for instance, when plants start to go to seed or are drying down. These winged parthenogens can spread quickly and invariably produce wingless female offspring. Aphids can also reproduce sexually, but this is initiated under cool temperature (15°C) and short daylight conditions that are not commonly found in *Arabidopsis* growth facilities.

3.3.1. Aphid Prevention

Screens on greenhouse vents, diligent inspection of nearby host material, and good housekeeping in the growth facility will help deter aphids.

3.3.2. Aphid Detection and Monitoring

1. Aphids can be found on flower stalks, on young leaves in the rosette center, and on the undersides of leaves (**Fig. 3A**). Winged aphids (**Fig. 3B**) are highly mobile and typically can be found on the rosette leaves. They are often trapped on the yellow sticky cards used to monitor fungus gnats.
2. A sticky, moist "honey dew" will coat the plant surfaces when aphid infestations are severe. Sometimes this honey dew will turn black with sooty mold.

3.3.3. Aphid Treatment Suggestions

1. Insecticidal soaps such as SaferBrand or M-Pede are effective for treating aphids on *Arabidopsis*, as long as the plants are otherwise reasonably healthy. Early detection of aphid infestations is important because soap treatment can burn the leaves of stressed plants. As with any chemical treatment of *Arabidopsis*, it is always a good idea to do a test spray on a few plants first.
2. Other suggested aphid pesticides are Mavrik Aquaflow® and Enstar II® (both from Wellmark International).
3. *Aphidius colemanii* is a small parasitoid wasp that sometimes enters greenhouses without assistance, or can be purchased from a biological control supplier. Female wasps lay eggs in adult aphids, which hatch and feed on the dying aphid. Adult wasps emerge and go on to parasitize more aphids. Evidence of wasp predation can be seen as tan-colored "shells" within clusters of aphids, sometimes referred to as "aphid mummies."

3.4. Spider Mites

The two-spotted spider mite, *Tetranychus urticae,* is a common greenhouse pest that can be found on *Arabidopsis*. Like aphids, spider mites are sap-sucking and can cause leaves and flowers to wilt and die. The entire mite life cycle takes about 3 wk: Eggs hatch into hungry nymphs, the nymphs go through several cycles of feeding followed by dormancy, and adults hatch after the third dormancy period.

3.4.1. Spider Mite Prevention

1. Spider mites thrive in hot, dry conditions. Maintaining a high relative humidity (greater than 50%) will deter spider mites.
2. Sanitation is important for preventing spider mite infestations, in particular if there has been an outbreak recently or nearby. Spider mites are good at hiding in cracks in the floor and in plant debris. Diapause, a type of hibernation, can occur when temperatures drop, days get shorter, or when the food supply gets low, such as when plants are senescing. Fertilized adult females will stop feeding and laying eggs, and will hide in cracks until conditions improve.

3.4.2. Spider Mite Detection

1. All life stages of spider mites are usually found on the undersides of leaves, which makes early detection difficult. Once established, however, they can sometimes be seen crawling across the leaf surface or near the center of the rosette (**Fig. 4A**).
2. Mite feeding causes a stippling effect, in which the leaves become covered with white or yellow chlorotic spots. In early infestations, the stippling tends to be more concentrated near the leaf margins (**Fig. 4b**).
3. When the infestation becomes severe, spider mites will spin threadlike webbing that covers the inflorescences. Mature flowering plants and senescing plants are especially vulnerable.

3.4.3. Spider Mite Treatment

1. Only adult spider mites and feeding nymphs are susceptible to pesticides. The eggs and dormant nymphs will not be harmed by chemical treatment. Therefore, it is important to plan on multiple pesticide applications 5 to 7 d apart. Insecticidal soaps such as Safer Brand or the pesticides Avid and Mavrik Aquaflow are effective chemical controls.
2. Several predatory mites that consume spider mites are available from biological control suppliers. Some will provide more immediate control, whereas other types will establish themselves and provide control over the longer term. One caution is that some species of predatory mites will feed on pollen when they run out of prey species. This could potentially cause crosspollination, which most *Arabidopsis* researchers want to avoid. Ask the supplier for a mite that does not feed on pollen.

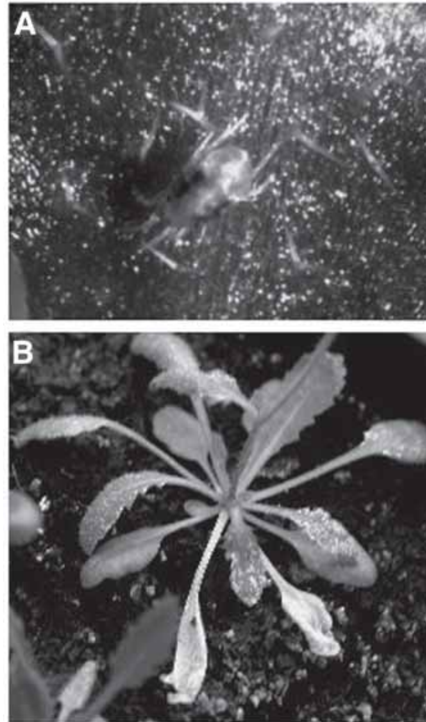


Fig. 4. Spider mites (*T. urticae*). **(A)** Spider mite adult, approx 0.5 mm. **(B)** *Arabidopsis* plant with leaf stippling caused by spider mite feeding.

4. Diseases

The incidence of severe diagnosed microbial diseases in laboratory-grown *Arabidopsis* is relatively uncommon, at least in part due to the short life cycle of *Arabidopsis*, the controlled conditions that most researchers use to grow their plants, and the fact that *Arabidopsis* plants are always grown from seed. In greenhouses, fungal and bacterial pathogens typically occur in situations where there is high humidity, poor air circulation, or continuously wet soil. Viruses are usually spread either by insects or when plants are propagated from cuttings. However, *Erysiphe* sp. (powdery mildew) and *Botrytis cinerea* (gray mold), two fungal pathogens that are familiar in horticultural settings, do occasionally cause spontaneous *Arabidopsis* infections, and are described below.

On the other hand, it is likely that many cases of pathogens infecting *Arabidopsis* remain unidentified, because it is often easier to dispose of a set of sick-looking plants and start over, rather than trying to identify the potential microbial agent causing the problem. We know from studies of plant–pathogen interactions that *Arabidopsis* is susceptible to many pathogens, including a variety of viruses, bacteria, and both necrotrophic and biotrophic fungi. Almost everyone working with *Arabidopsis* has encountered unexplained symptoms such as wilting, yellowing of the entire plant, chlorosis of leaves, anthocyanin accumulation, or general lack of vigor. Often, this is attributed to a "bad batch of soil," but it is possible that these sickly plants are infected with an unidentified pathogen.

*4.1. Powdery Mildew (*Erysiphe *sp.)*

Powdery mildew is an obligate biotrophic pathogen that infects many agricultural crops. Species in the genus *Erysiphe* commonly infect *Arabidopsis* in greenhouses and growth cham-

Fig. 5. Powdery mildew (*Erisyphe* sp.) on *Arabidopsis* leaves.

bers. Airborne fungal spores settle and germinate on living plant tissue. After germination, the mycelium grows over the surface of the leaf, projecting food-absorbing haustoria into the epidermal cells. The pathogen develops in this manner, invisible to the naked eye, until spore-producing conidiophores are formed. This reproductive phase is visible on the leaf as a patch of white "powder" that consists of the conidiophores and the multitudes of spores ready to become airborne, continuing the cycle.

4.1.1. Powdery Mildew Prevention

1. Reduce the relative humidity immediately surrounding the plants by allowing good air circulation around plants. Keep plants spaced apart and use fans to keep air circulating throughout the growth area. Although powdery mildew spores can germinate at a lower humidity than most fungi, the progression of the disease on the plant is increased in high humidity environments. Do not leave water standing in the trays beneath the pots, as this increases the humidity immediately surrounding the plants.
2. Control the temperature and humidity in the growth facility. Warm days with low relative humidity combined with cool nights with high relative humidity favor development of powdery mildew. These conditions are common in winter in greenhouses that do not have good temperature control.
3. Clean the growth area between plantings. After removing all plants, wash all surfaces (walls, floors, benches, etc.), and then turn up the temperature as high as possible for a few days. Cleaning will remove all remnants of plant material on which spores might survive. Heating the area will dry out and kill any remaining spores.

4.1.2. Powdery Mildew Detection

1. White patches of mildew will first become visible at the tips of the upper surfaces of older rosette leaves (**Fig. 5**). When the powdery mildew infection is advanced, cauline leaves, stems, and siliques will also show these patches.
2. Powdery mildew is most easily detected on plants that have reached the large-rosette stage. Plants that are much younger and otherwise appear to be healthy can still have a powdery mildew infection that has not yet reached the reproductive phase and is thus invisible to the naked eye.

4.1.3. Powdery Mildew Treatment

1. Early treatment is important because the airborne spores are easily spread. Avoid spreading powdery mildew by not entering other plant growth areas after exposure to infected plants.

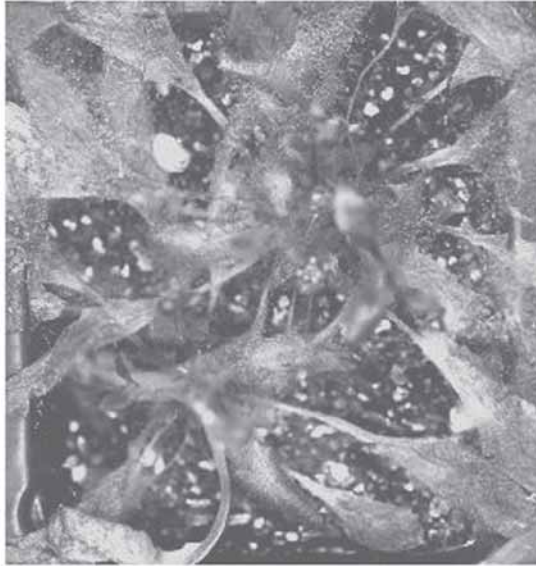


Fig. 6. *Botrytis* infection on an *Arabidopsis ein2* mutant plant. Under normal growth conditions, *Botrytis* infections are uncommon and do not result in extensive lesions. However, some commonly used *Arabidopsis* mutants, such as *ein2*, are hypersusceptible to *Botrytis* infection. *Botrytis* can also cause extensive damage to plants that are grown under high-humidity conditions.

2. When powdery mildew is first discovered, cover the affected plant with a damp paper towel, which reduces the spread of the spores. Spraying plants or moving infected plants drives spores onto neighboring plants.
3. Sulfur-based fungicides sold for use in house and garden are effective against powdery mildew. Use according to the directions on the label. In some cases, however, sulfur sprays may cause older leaves to turn yellow or senesce prematurely. Do a test spray on a few plants.
4. A fungicide that is effective against powdery mildew is 3336F (Cleary Chemical Corp.).

*4.2. Gray Mold (*Botrytis *sp.)*

Botrytis is not very common on *Arabidopsis*, but in high-humidity conditions, plants that are already weak, stressed, or compromised in some way can become infected. *Arabidopsis* mutants in the ethylene signaling pathway (for example, the commonly used *ein2* mutant) are particularly susceptible to *Botrytis* infections.

4.2.1. Botrytis *Prevention*

1. Keep the growth area clean and remove plant debris that can harbor *Botrytis* spores.
2. Healthy, vigorously growing plants are unlikely to become infected with *Botrytis.* Avoid wounding or damaging plants when handling flats. Wound sites are frequent areas of infection.
3. Keep the humidity low around plants by providing good ventilation and not overcrowding plants. In a greenhouse, when night temperatures drop, a thin film of moisture can condense on plant surfaces, creating an ideal condition for *Botrytis* spores to germinate. Airflow around the plants helps to keep leaves dry. Water plants early in the day, and do not leave water standing in the tray.

4.2.2. Botrytis *Detection and Treatment*

1. *Botryis* appears as a gray mold on older leaves (**Fig. 6**).
2. The fungicide 3336F is reported to be effective against *Botrytis.*

North American Biological Control Suppliers can be found at the Web site: www.cdpr.ca.gov/docs/ipminov/ben_supp/contents.htm.

References

1. Anderson, M. (1998) Control of pests and diseases of *Arabidopsis*, in *Methods in Molecular Biology, vol. 82: Arabidopsis Protocols* (Martinez-Zapeter, J. M. and Salinas, J., eds.), Humana Press, Totowa, NJ, pp. 19–26.
2. Nelson, P. V. (1998) *Greenhouse Operation and Management*. Prentice-Hall, Upper Saddle River, NJ.
3. Malais, M. and Ravensberg, W. J. (1992) *Knowing and Recognizing: The Biology of Glasshouse Pests and Their Natural Enemies.* Koppert Biological Systems, Berkel en Rodenrijs, The Netherlands.
4. Preuss, D. (2002) How to grow *Arabidopsis*, in *Arabidopsis: A Laboratory Manual* (Weigel, D. and Glazebrook, J., eds.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1–18.

3

Growth of Other Species Related to *Arabidopsis thaliana*

Krithika Yogeeswaran and Mikhail Elia Nasrallah

Summary

Recent interest in the study of biological phenomena in Brassicaceae members (crucifers) related to the model plant, *Arabidopsis thaliana*, fuels the need for standard protocols for growth and maintenance of these plant species. Like *A. thaliana*, many wild crucifers are amenable to growth in greenhouses, growth rooms, and growth chambers for laboratory study. Some aspects of their care are quite similar to *A. thaliana*. However, several unique traits, such as perennial growth habit, long life cycle, large plant size, need to overwinter for floral initiation, and self-incompatibility, warrant special procedures to ensure healthy growth and good seed set. This chapter attempts to provide sources for acccessions and standard protocols for plant growth and seed harvest of wild crucifers while providing special tips for species with which we have experience.

Key Words: Brassicaceae; crucifers; accessions; *Arabidopsis lyrata*; *Capsella* spp.; perennial; growth; care of plants; environmental conditions; vernalization; self-incompatiblity; self-pollination.

1. Introduction

There has been a growing interest in a number of disciplines of plant molecular biology to extend the study traits of interest across species of the Brassicaceae family (crucifers). The model plant *Arabidopsis thaliana* belongs to this family, along with the economically important genus, *Brassica*, which has been of long-standing interest to plant breeders. Less well known crucifer relatives from the genera *Arabidopsis*, *Capsella, Boechera, Sisymbrium*, and others have been of interest to Brassicaceae systematists for several decades. Only recently have these wild species recieved some attention in basic biology (for examples, *see* ***1–3***). This new trend fuels the need for standard protocols for the growth and maintenance of healthy plants of relatives of *A. thaliana* for laboratory study.

Like *A. thaliana*, these plants are amenable to cultivation in growth chambers, growth rooms, and greenhouses. While some aspects of growth of these crucifers are favored by conditions that mirror those suitable for *Arabidopsis thaliana*, others differ due to contrasting habit, life history, and reproductive behavior (*see* **Table 1** for a list of contrasting traits). *Arabidopsis thaliana* is an annual with a short life cycle, whereas many of its relatives are perennials and have long life cycles. Special measures are often required to speed up the transition from vegetative to reproductive growth. Self-fertility in *A. thaliana* promotes copious seed production, while many other crucifers are self-incompatible and set little or no seed.

General guidelines for indoor cultivation and control of environmental factors will be outlined in this chapter. We will pay special attention to unique aspects that one should consider while growing relatives of *A. thaliana.* These details are based on experience gathered over the past decade, from species we have worked with in our laboratory. These include accessions of *Arabidopsis lyrata* ssp. *lyrata, Arabidopsis lyrata* ssp. *petraea, Capsella rubella, Capsella grandiflora,* and a few varieties of *B. oleracea* and *B. campestris.* As *Brassica* cultivation is

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

Table 1
General Characteristics of Cruciferous Species

	Arabidopsis thaliana	*Arabidopsis lyrata*	*Capsella* spp.	*Brassica* spp.
Common name	Thale-cress	Lyre-leaf rock-cress	Shepherd's purse	Oilseed rape, broccoli, cauliflower, brussels sprouts, kale, cabbage, kohlrabi, radish, etc.
Other names	—	*Arabis lyrata, Arabis petraea, Cardiminopsis petraea* (*8*)	—	—
Geographical distribution	Eurasian origin, Ecotypes found in N. America, Europe, Asia, and parts of Africa (*9*)	*A. lyrata* ssp. *lyrata*: N. America *A. lyrata* ssp. *petraea*: Europe (*8*)	Eurasian origin, worldwide distribution except tropics (*10*)	Eurasian origin, Cultivated *Brassica* spp. grown worldwide in subtropical to subarctic climates (*8*)
Habit	annual	perennial	annual and perennial spp.	annual and biennial spp.
Base chromosome number	5	8	8	8–10
Ploidy	diploid	diploid and polyploid	diploid and polyploid	diploid and polyploid
Self-fertilization barriers	SC	SI	SC & SI	SC & SI
Duration of life cycle	6–8 wk[a]	6–12 mo	6–12 mo	2–12 mo
Unique features	model plant system—genome fully sequenced	aerial rosettes, fragrant flowers	heart-shaped siliques	*B. oleracea* ssp. have been selected for edible traits or seed oil
Economic importance	weed	weed; wild species	weed; wild species	*B. campestris* and *B. napus* —oilseed rape *B. oleracea* ssp.—vegetables
Average plant size:				
Vegetative: rosette diameter	1–3 in.	3–5 in.	8–10 in.	10–14 in. diameter, 6–12 in. tall
Flowering: height and diameter	6–10 in. tall, 2–4 in. in diameter	12–16 in. tall, up to 18 in. diameter when vernalized	14–18 in. tall, up to 24 in. diameter when vernalized	3–4 ft tall, 12–18 in. diameter
Average number of seeds per silique	40–50	20–25	10–15	15–20

[a]Some ecotypes are late-flowering and have much longer life cycle.

fairly well documented (*4*), most of our notes will focus on tips to deal with the wild weedy cruciferous species.

2. Materials

2.1. Plant Material Resources

Currently the Arabidopsis Biological Resource Center (ABRC, Ohio State University) houses no seed stocks for wild relatives of *A. thaliana* but has seed stock from one accession each of *B. oleracea* and *B. rapa* (these may be ordered through The Arabidopsis Information Resource (TAIR) website www.arabidopsis.org). The US Department of Agriculture (USDA) seed banks (www.ars-grin.gov/npgs/acc/acc_queries.html) and Consultative Group on International Agricultural Research (CGIAR) centers (www.singer.cgiar.org/SINGER/singer_test_version.htm) are rich in germplasm of cultivated *Brassica* spp. but show poor representation of wild crucifers. The USDA has one accession of *C. grandiflora* and no *Arabidopsis* spp. accessions. A Brassica Seed Bank maintained at Tohoku University in Japan houses some 177 species from 58 genera of the Brassicaceae, including several wild relatives that are available to the public for research purposes (www.agri.tohoku.ac.jp/pbreed/Seed_Stock_DB/Stock_English_top.html). For most of our laboratory studies we have obtained seed derived from wild accessions from fellow research groups. These include seed derived from accessions of:

- *Arabidopsis lyrata* ssp. *lyrata:* North American subspecies collected from northern Michigan by C. Langley (University of California, Davis).
- *Arabidopsis lyrata* ssp. *petraea:* Northern and Central European subspecies collected from Iceland, and provided to us by M. Schierup (University of Aarhus, Denmark).
- *Capsella rubella* and *C. grandiflora:* Collected originally by H. Hurka (Universität Osnabrück, Germany) and provided to us by R. Schmidt (Max Planck Institute, Germany).
- *B. oleracea* var. *alboglabra* from the Gene Bank Facility courtesy of D. J. Ockendon (Wellesbourne, UK).
- *B. campestris* var. *Yellow* Sarson from India (USDA strain C634) provided by K. Hinata (Tohoku University, Sendai).

2.2. Plant Growth and Cultivation

1. Plastic Petri dishes.
2. Whatman filter paper.
3. Distilled water.
4. Standard plastic 1020 trays and flats of $3 \times 6 = 18$ thin-walled 4-in. maxi square cells to insert in the trays (Griffin Greenhouse and Nursery Supplies, Tewksbury, MA).
5. 4½–6½-in. diameter thin-walled plastic pots and 8–12-in. clay pots (Griffin Greenhouse and Nursery Supplies).
6. Large heavy plastic trays of at least 3-in. depth for use as secondary containers for large plastic and clay pots.
7. Metromix 200 (Scotts Sierra Horticultural Products, Marysville, OH).
8. Large autoclave bags.
9. Water (not distilled).
10. Watering can.
11. Paper.
12. Plastic labeling stakes.
13. Permanent marker.
14. Plastic domes and/or Saran wrap.
15. Commercial fertilizer—Peters® Professional® soluble NPK 20:20:20 with micronutrients (Scotts Sierra Horticultural Products).
16. Pesticides:
    - Gnatrol® Bioinsecticide (Abbott Laboratories, Chicago, IL).
    - Marathon® II (systemic, foliar spray, toxic; Olympic Horticultural Products, Mainland, PA).

- Conserve* SC (nonsystemic, foliar spray, toxic; Dow Agrosciences Canada, Inc., Calgary, Alberta).
- Marathon® 60 WP (systemic wettable powder, toxic; Olympic Horticultural Products).
- Sevin® (organophosphate, highly toxic; Southern Agricultural Insecticides, Palmetto, FL).

17. Yellow and blue insect sensor cards (Whitmore Micro-gen Research Lab).
18. Gardening plastic wire.
19. Preemergent herbicides (Toxic): Surflan AS (Southern Agricultural Insecticides) or XL2G (Setre Chemical Co., Memphis TN) (optional).
20. Magnifying lens.
21. Fine tweezers (Dumont #3 Tweezers, Electron Microscopy Sciences, PA).
22. Jewel tags for labeling (optional).

2.3. Seed Harvest and Preservation

1. Transparent acetate sheets and tape for collars (optional).
2. Cutting shears.
3. Clear plastic storage bags.
4. Sieves of appropriate mesh size (optional).
5. Glass Petri dishes.
6. Small brown coin envelopes: size #1 (2.5 × 3.5 in.) and #3 (2.5 × 4.5 in.).
7. Cryovials.
8. Forceps.
9. Silica gel.
10. Dessicator.
11. Kimwipes.
12. Tape.
13. Parafilm.

3. Methods

3.1. Plant Growth and Cultivation

3.1.1. Growth Environments

Wild relatives of the tiny weed *A. thaliana* tend to be comparatively larger in size, both in aerial and subterranean tissue (*see* **Fig. 1**). While young plants may be easily maintained in growth chambers and growth rooms, large mature plants may be more easily housed in greenhouses. In greenhouses, plants may be shaded from full sun and artificial lights may be used to extend photoperiods when necessary (*see* **Note 1**). Some plants in the flowering stage may require much more spacing than in the vegetative stage (*see* **Note 2**). Many cultivated crucifers form very long erect inflorescences and may be housed only in large growth chambers or greenhouses (*see* **Note 3**). These size and space constraints must be taken into account while planning an experiment (refer to **Table 1** for details of average plant size for various species, at different stages of the life cycle).

3.1.2. Seed Stratification

Dry seeds are often dormant and will need to be imbibed by stratification before planting. We follow the simple protocol outlined below.

1. Cut two pieces of Whatman filter paper to fit a plastic Petri dish.
2. Place the filters in the Petri dish's base and lid and dampen completely so that they are moist but not wet.
3. Gently spread seeds evenly on the surface of the base plate filter.
4. Cover with the other filter and replace the lid.
5. Treat at 4 to 5°C in a cooler or refrigerator for 2 to 5 d (depending on subspecies) to allow imbibition. This treatment does not require light and may be performed under dark conditions.

Alternatively, seeds in starter pots may directly be stratified by placing in a refrigerator or cooler for the same treatment described in **step 5**.

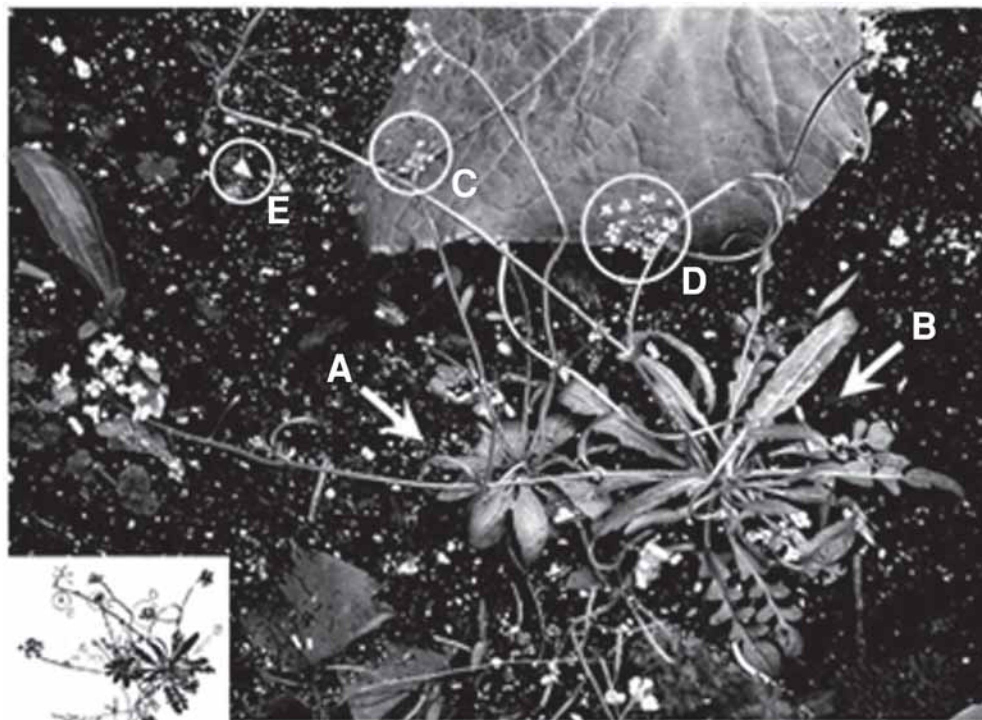


Fig. 1. Crucifer species growing in the wild. Weedy **(A)** *Arabidopsis thaliana* seen growing alongside **(B)** *Capsella rubella* in the author's vegetable garden in Ithaca. Note the differences in flower size between **(C)** *A. thaliana* and **(D)** *C. rubella*. **(E)** *Capsella* spp. have heart-shaped siliques that shatter to disperse seeds over long distances (Insert provided for clarity).

3.1.3. Conditions Favoring Seed Germination

Light is essential for proper seed germination. Continuous light and long daylight regimes (>16 h) favor seed germination. Although overwatering is not recommended, germinating seeds must never be allowed to dry out. This may be achieved by daily watering. High humidity (>60%) favors good seed germination and may be accomplished by keeping pots and flats covered (with lids or Saran wrap) until the first true leaves emerge. A 23°C day temperature prevents any greenhouse-effect type of overheating of the soil surface when the plants are covered.

Germination time in wild crucifers we have worked with tend to be comparable or longer than that for *A. thaliana*, subspecies-specific, and typically in the range of 7 ± 1 days (some subspecies may take longer; *see* **Note 4**).

3.1.4. Planting

The most effective planting strategy to follow depends on the availability of seed. When plenty of seed is available, we suggest denser plantings on 4.5 to 6.5-in. starter pots that can be imbibed directly. These occupy limited space, can be stored in lab coolers or growth chambers, and help one eliminate inviable seed (in the absence of seed viability estimates). A few weeks after germination, healthy seedlings 1 to 2-in. in height can be easily transplanted to individual pots (4.5 to 8-in. diameter). We find this to be both a space- and a time-efficient strategy. When seeds are in limited supply however, it may be best to directly plant in individual pots after stratification on plates—plastic flats with 3 × 6 = 18 individual cells are preferred in this case. For studies involving many individual lines that must be grown and compared under common

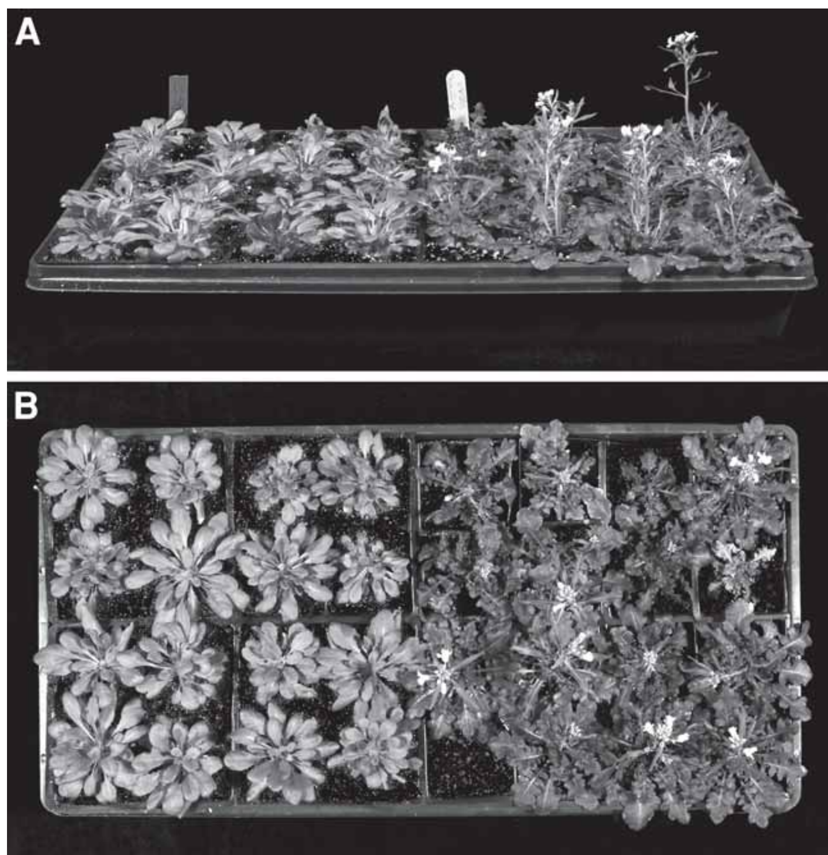


Fig. 2. Flat of young *A. thaliana* and *A. lyrata* plants grown in the greenhouse. Panel **A** shows a side view and Panel **B** an aerial view of a 4 × 8 = 32 flat of young *Arabidopsis* plants. From the left, the first four rows are late-flowering *A. thaliana* ecotype C24 plants at the peak of their vegetative growth. Rows 5–8 show an early-flowering line of *A. lyrata* ssp. *lyrata* plants. Both plant species are about 15 wk old and were grown under continuous incandescent lighting with 25°C day temperatures (16 h) and 21°C night temperatures (8 h). Vernalization was not required to initiate flowering in this particular *A. lyrata* line, although most accessions do require the treatment. Observe the differences in rosette diameter (refer to **Table 1** for average sizes) and leaf morphology (lyrate leaves in *A. lyrata*). Flats are useful to separate different lines and species for comparative studies in which environmental conditions must be uniform.

environmental regimes, one can plant seeds from different accessions in rows on single flats (*see* **Fig. 2**).

We recommend the use of Metromix 200, a moderate-density soil mixture containing peat moss, perlite, vermiculite, and washed sand. Any similar commercial soil mix may be used as a substitute. Peat moss improves the soil's ability to aerate growing roots, which is essential especially under potted conditions. Soil presterilization is a topic of some debate: whereas the advantages include killing contaminating seed material, insect eggs, and fungal spores, disadvantages include killing of native microflora, which makes the plants more susceptible to pathogenic attack. We favor autoclaving of soil as a robust strategy to deter common pests and guard against contamination from stray seed. However, this is a step that may not be necessary or preferred by others. Our planting strategy is outlined below.

1. Autoclave sufficient soil for multiple plantings in autoclave bags (wet cycle, 30 minutes; one 24" × 36" bag full of soil is sufficient for about three flats or about eight 4.5–6.5-in. pots).
2. When using flats, cut out one cell from the bottom right-hand corner to facilitate subirrigation.
3. Fill small pots with soil to about 0.5 in. from the top and large pots to 1 in. from the top.
4. Place pots in a secondary container (flat cells in a tray). Water the soil from above and subirrigate in about 2 in. of water. Leave standing for 30 min to allow the soil to become uniformly moistened.
5. Fold a piece of paper to create a groove (paper boat). Place about 50 seeds in the center.
6. Gently tilt and tap so that only one seed falls at a time. Move carefully over the soil surface to evenly distribute the seeds (if more than a few seeds fall and germinate per spot they can be thinned later). Aim at the center of cells where necessary (for an alternative strategy, *see* **Note 5**).
7. Label pots with plastic stakes including the species, and dates of imbibition and planting.
8. Do *not* cover seeds with soil. Light is required for seed germination. Cover with plastic domes or Saran wrap with a few holes for aeration.
9. Stratify pots with unimbibed seeds as described in **step 5** of **Subheading 3.1.2.**
10. After stratification, place starter pots under lights in a growth chamber (or room) and allow seeds to germinate and seedlings to grow.

3.1.5. Environmental Factors Affecting Plant Growth

Healthy growth of cruciferous plants is favored by a number of environmental conditions that are similar to recommendations for *A. thaliana* (***5***, ***6***; *see also* Chapter 1 in this book).

3.1.5.1. Light

Long days (16–20 h) favor robust plant growth. Typically about 150 μE/m^2/s is favorable and may be obtained from incandescent (or fluorescent) light bulbs. "Cool-white" and "grow" lights are preferred but not necessary. In all cases, overheating from artificial lighting must be prevented. Full sunlight in greenhouses may be too harsh and screens can be used to shade plants appropriately (*see* **Note 1**). Expected plant heights at various stages of development must be considered while setting the bench to lighting source distances (refer to **Table 1** for genus-specific details). Allow at least 1 ft of space between the top of the plant and the light source.

3.1.5.2. Temperature

As with *A. thaliana*, day temperatures of 23 to 25°C and night temperatures of 20 to 22°C are most suitable for healthy growth throughout the plant's life cycle, though a wider range of temperatures are tolerated. Colder temperatures and shorter day lengths may be more suitable for some *Brassica* spp. (*see* **Note 3** for details).

3.1.5.3. Water and Humidity

Control of watering and humidity are most essential at the beginning and end of the plant's life cycle. Daily watering by subirrigation is recommended to maintain soil in a wet (but not saturated) state during these times. For the interim period, watering 2 to 3 times per week is sufficient. The range of humidity tolerated during most stages of plant growth is quite broad, and ranges from about 30 to 70%. Within this range, proper seed germination is favored by high humidity, while low humidity favors rapid seed maturation.

3.1.5.4. Fertilizers

When using Metromix 200 soil mix, fertilizers are not essential during the early stages of growth when the plants have limited biomass. Biweekly fertilization is recommended after the first month to replenish major nutrients being depleted from the soil. We recommend any commercially formulated soluble NPK 20:20:20 fertilizer with micronutrients. We use Peters Professional fertilizer, as it is compatible with Metromix 200. Apply fertilizer of choice as per manufacturer's suggestions.

3.1.5.5. Pest Management

Thrips and fungus gnats are pests encountered regularly in greenhouses. Thrips are quite devastating to the vegetative and reproductive tissue of plants they feed on and can cause stress responses, transmit disease, and even cause sterility in extreme cases. Soil sterilization, disposal of infected dead plant material, and care during handling help avoid crosscontamination. We regularly have our greenhouse material sprayed with a systemic commercial pesticide, Marathon® II, once a month to deter foliar thrip activity. Marathon II is a good prophylactic measure against thrips. In case of a thrip outbreak, our greenhouse staff prefer to use Conserve* SC, which is a nonsystemic foliar spray that effectively destroys pests attacking all external parts of the plant, including the floral tissue. It is important to apply this insecticide exactly according to the manufacturer's instructions for the best outcome.

Fungus gnats are less damaging as they feed on soil fungi, but their soil-inhabiting larvae can do significant damage to roots and any leaves that come in contact with the soil. Use of autoclaved soil and avoidance of overwatering the plants are good preventive strategies. We use a commercially available *B. thuringiensis* biocontrol agent called Gnatrol, which is highly specific to fungus gnat larvae and effective in controlling this pest. Treatment may be repeated once a month per manufacturer's instructions.

In our growth rooms and greenhouses we use sensor cards to help monitor the presence of these pests as well (blue cards for thrips and yellow cards for fungus gnats).

Aphids and spider mites are other less frequent pests. These insects reside on the aerial surfaces of the plants and feed on phloem; hence they can detrimentally affect overall plant health. We use commercially formulated organophosphate pesticides such as Sevin or systemics such as Marathon 60 WP to control these pests. The latter is a wettable powder that allows rapid control of systemic aphid infections as well as suppression of thrip activity.

(For more information, *see also* Chapter 2 in this book).

3.1.6. Care of Plants During Vegetative Growth and Favorable Conditions

Within 3 to 4 wk of germination, a healthy growth of rosette leaves can be expected. At this stage, young plants are ready to be transplanted to individual pots.

1. Follow **steps 1**, **3**, and **4** from **Subheading 3.1.3.** using 8- to 12-in. pots.
2. Using two or three fingers, create 1- to 3-in. deep depressions in the center of the soil surface of about 1" diameter.
3. Transplanting within a half hour of watering allows easy removal and separation of seedlings.
4. Gently separate individual seedlings (or small clumps of seedlings for smaller plants; they can be thinned later if necessary) with as little root damage as possible (transplant with adhering soil clumps to minimize transplantation shock) and gently insert into soil depressions in new pots.
5. Pat down surrounding soil to fill the hole and water to remove air pockets and settle soil. Shading of plants against direct harsh lighting that might cause wilting allows the roots time to settle well in new pots and effectively transpire water. After a few days in the growth environment previously used (again, to minimize transplantation shock) the pots can be moved to the greenhouse if warranted.

Light and temperature regimes are as recommended in **Subheading 3.1.5.** Crucifers tend to be long-day plants, so shortening of day length can be used to prolong the vegetative stage. Low day temperatures (16–20°C) also tend to delay flowering. Most wild crucifers overwinter before flowering, so floral initiation may require a vernalization period; avoiding or delaying this treatment can extend the vegetative growth stage. Growing plants need to be watered once every two days from above or by subirrigation to prevent drying out (*see* **Note 6**). Regular fertilizing is recommended after 3 to 4 wk as described in **Subheading 3.1.5.** Repotting in fresh soil and new pots every 8 to 12 mo may also be advisable for some perennial subspecies being maintained for several years. Some species are highly amenable to vegetative propagation and lines may be effectively maintained for long periods of times (*see* **Note 6** for example and strategy).

3.1.7. Care of Plants During Reproductive Growth and Favorable Conditions

While wild crucifers we have worked with are expected to flower within 5 to 6 mo when left to their own devices, floral initiation may be expedited by *vernalization* (cold treatment). This procedure also encourages much more prolific flower production than would otherwise be expected.

1. Plants that have formed large rosettes, 2 to 3-in. in diameter, are selected (rosettes of this size can be expected within about 2 mo from the time of planting).
2. Plants are placed in a growth chamber or cold room at 4 to 5°C for 4 to 8 wk (or more, depending on subspecies). The day-night light regime is maintained during the vernalization period.
3. After cold treatment, plants can be returned to the greenhouse.

Most crucifers we have worked with are long-day plants, so 16- to 20-h photoperiods to continuous light favor rapid flowering. Incandescent lights are preferred for floral induction (*see* **Fig. 2** for a rapid flowering *A. lyrata* line). Water becomes more critical at higher temperatures (up to 30°C nondetrimental) because transpiration is increased, but the most favorable day temperature is still 25°C. The flowering period typically lasts a few weeks to several months (*see* **Note 7**).

In self-compatible (SC) species, regular manual removal of siliques can extend the flowering period because seed pods are strong sinks for nutrients and shift the nutrient supply away from young flowers (this can be observed by the reduction of flower size as seed setting progresses). For self-incompatible (SI) species this process is not necessary, as self-fertilization is inhibited and consequently silique production is significantly reduced. We note that under greenhouse conditions, in the absence of insect vectors, crosspollination is fairly negligible when plants of different genotypes are adequately spaced. Vernalized *C. rubella* and derivatives of the North American accession of *A. lyrata* ssp. *lyrata* both produce dense "canopies" of inflorescences over time. Regular pruning of such prolifically flowering plants to remove old inflorescence stalks is critical in promoting better aeration to vegetative tissue below and encouraging new inflorescence stalk formation (*see* **Note 8**).

3.1.8. Conditions Favoring Good Seed Set

For good seed set, regular watering and fertilizing are essential. A continuous-light regime can speed up the reproductive cycle to seed set. Temperatures up to 30°C are tolerated but 25°C is still optimal. Low humidity also encourages silique maturation and drying.

For studies involving strict maintenance of genetic stocks, greenhouse plants must be monitored for "volunteers" or "escapes," i.e., plants derived from seed dispersed from neighboring plants. These escapes are easy to catch when young but may be harder to distinguish from the potted plant as they grow older and mature, and they may result in seed stock contamination (*see* **Note 9** for strategies to combat this problem).

In striking contrast to self-compatible crucifer subspecies, which produce plenty of seed in many large siliques (*see* **Table 1** for the number of seeds to expect per pod for different species), self-incompatible plants will produce predominantly small empty siliques with hardly any naturally-selfed seed pods. Bulking these from several plants may provide a reasonable amount of seed for some purposes but this may not be a suitable strategy when a lot of seed is required. SI plants often require intervention by forced self-pollination in order to set sufficient seed and maintain the line. Molecular self-incompatibility barriers are functional from the –1 stage (1 day to anthesis) onward but not before this stage *(7)*. One can therefore collect pollen from open flowers post-anthesis and manually pollinate –2 and younger buds (2 or more days to anthesis) with the help of a magnifying lens as follows:

1. Gently tease open –2 flower buds of a plant to expose the stigma surface.
2. Using tweezers, remove dehisced anthers manually from an open flower of the same plant.
3. Dust the stigmatic surface of the pistil with the adaxial surface of the anther, which has exposed pollen grains on it.
4. To distinguish self-seeded siliques from open-pollination-derived siliques, jewel tags are used to label selfed branches by tying directly beneath the pollinated buds (*see* **Note 10**).

In the case of SI species, the number of seeds representing a line is strictly limited by the amount of manual pollination that can be performed.

3.2. Seed Harvest and Preservation

3.2.1. Seed Harvest

Seed harvesting strategies for sister crucifer species are quite similar to those for *A. thaliana*. However, for species with larger plants and seeds, there are unique challenges and advantages. For bulked seed collections, inflorescences may be bagged by pot or flat before the siliques dry and shatter on the plants, to contain the seeds. Alternatively, mature inflorescences may be cut off the plants and collected in large bags to dry. For SI plants and plants for which seed from individual lines is desired, one can enclose the inflorescence in transparent plastic collars to contain the seeds and prevent crosscontamination. We make our own collars by rolling and taping transparent acetate sheets of appropriate dimensions into tubes of about 2 to 4-in. in diameter. Collars must be inserted before the inflorescence tissue dries, when still green and pliant. If seed is a limiting factor, one can remove and bag inflorescences as they mature on the plants over several weeks and bulk them to ensure minimal loss. In all cases, the seeds should be allowed to dry in the siliques on the inflorescences.

When dry, the siliques will shatter naturally. The dried inflorescence stalks can be threshed to release all seed. Sieves of suitable mesh size may be used to separate the chaffs from the seed, but we find that gentle blowing is sufficient to remove most of the chaffs. Harvested seed can be stored in open glass Petri dishes in very low-humidity environments to encourage further drying (*see* **Note 11**). The Petri dishes may be further dried in silica gel-containing dessicators overnight. Approximately 5% moisture content is desired for seed storage.

3.2.2. Seed Preservation

Seeds must be preserved in environments of very low humidity (~5–10%) to prevent rehydration before use. For short- and intermediate-term storage, we store seeds in coin envelopes, labeled and organized in cardboard boxes at room temperature or at 4°C. Homemade silica gel packets using Kimwipe bags sealed with tape may be added to each box to keep the boxes dry. For long-term storage of seed stocks, dried seeds may be placed in cryovials and stored at –80°C. In cases of cold storage, seed containers must be brought back to room temperature before opening under low-humidity conditions to prevent rehydration of the seed and freezing damage.

3.3. Conclusions

Our growth protocols have been developed over the years by tweaking time-tested strategies to grow *A. thaliana* and *Brassica* spp. until they work for each new crucifer we study. We hope that the general guidelines outlined here will be suitable for other wild crucifers, either directly or with a few modifications. Some wild crucifers may be recalcitrant to laboratory growth. To date, we have not had success in inducing some species (e.g., *Cardamine* spp.) to flower even though they thrive in the vegetative stage. Knowledge of native habitats, environmental regimes, and the plants' general characters and traits will help one effectively customize standard protocols to work effectively for at least some crucifer species of interest.

4. Notes

1. We find that for plant populations grown in our greenhouses, plants close to glass walls and plants in pots on the centers of the benches tend to be more prone to poor growth, stress, and premature death. These locations receive intense natural sunlight or direct artificial lighting, and such sites must be protected by shading screens or left unoccupied if space permits.
2. The flowering plants of *Capsella* spp. have long inflorescences with heart-shaped siliques that tend to tightly intertwine with those from adjacent pots in the absence of adequate spacing; this can be quite a nuisance as untangling plants can damage the inflorescence tissue and cause siliques to shatter, dispersing seeds prematurely. Care should be taken to prevent this when inflorescence tissue and individual seed are important for the researcher's purpose.

3. Subspecies of *B. oleracea* may be several feet tall with fairly thick stems at maturity and therefore must be grown in individual pots and preferably in the greenhouse. *B. oleracea* var. *alboglabra* may grow up to 4 ft tall, whereas *B. oleracea* var. *gemmifera* grows to only about 2 ft in height. Some *B. campestris* varieties such as the Indian "Yellow Sarson" are much smaller and can be grown in flats in growth chambers (1 to 2 ft tall, more delicate and thin inflorescences than *B. oleracea* ssp.). Greenhouse conditions with 14-h days at 20°C and 10-h nights at 18°C favor *B. oleraceae* growth and flowering.
4. We have observed germination time in *A. lyrata* ssp. *lyrata* to be significantly longer than this average and of the order of 9 to 10 d. Seed stratification in this case should be carried out for 5 to 7 d.
5. For very small seed or seed of limited quantity, one can use a moist-tipped toothpick to pick up the seeds one by one and place them on the soil surface, one per cell on flats (instead of **steps 5–6** in **Subheading 3.1.4.**).
6. *A. lyrata* plants form aerial rosettes to varying degrees depending on original growth habitat. For example, we have observed that subarctic *A. lyrata* ssp. *petraea* plants form dense soil covers of profuse aerial rosettes. In the vegetative stage, care must be taken to ensure adequate watering of such plants, as very little water can penetrate through the dense leaf cover; we recommend subirrigation for such plants to prevent water stress. Aerial rosettes are excellent vegetative propagules and can be repotted to maintain lines indefinitely. Pots with aerial rosettes can be placed on greenhouse misting benches for 2 to 3 wk to encourage root formation and growth before returning to normal growth conditions.
7. Prolific flowering in SI perennial *A. lyrata* ssp. *lyrata* plants has been observed for up to 18 mo following 8 wk of vernalization and regular pruning, making these plants an ideal source of floral tissue for studies of reproductive processes.
8. Dense canopies of *C. rubella* inflorescences coupled with careless overwatering may cause rosette leaves and the tissue at the root–shoot interface to rot due to poor aeration. Monitored subirrigation and regular pruning of old inflorescence stalks can improve aeration to covered tissue and prevent such damage. Because the rosette leaves are completely covered, the small cauline leaves of the inflorescence stalks take over most of the photosynthetic function; adequate nutrient supply for developing flowers and seeds must be maintained by regular fertilizing and watering.
9. In our experience, SC *Capsella rubella* plants easily crosscontaminate other plants in the greenhouse and each other if mature siliques are allowed to dry and shatter while on the plants. Seed dispersal distances can be 3 ft or more, making this plant a potential greenhouse weed if not carefully monitored. For experimental purposes that require lines to be clearly distinguished, adequate spacing of plants and precautions to prevent or contain seed dispersal from plants (e.g., regular removal of mature inflorescence stalks) are necessary. Alternatively, preemergent herbicides such as Surflan AS or XL2G (granular) (active ingredients: oryzalin, benefin) may be used to treat the soil when plants are in the vegetative stage prior to vernalization. These herbicides act by suppressing the emergence of seedlings of contaminating seed and are effective against broadleaf and grass weeds including *Capsella* and *Brassica* spp. They are activated by watering within 21 d of application and need be reapplied only every 90+ days (*see* www.utextension.etk.edu/publications/pbfiles/PB1728.pdf for further details).
10. We like to include the cross made (female parent × male parent), date of the cross, and the number of buds crossed on our jewel tag labels to keep track of our crosses.
11. Glass Petri dishes are preferred to plastic ones, as static effects in the latter cause seeds to adhere to walls and make them very hard to work with, especially in the case of small seeds.

Acknowledgments

The authors acknowledge useful input and helpful suggestions to improve the manuscript from Prof. June Nasrallah, Dr. Amanda Garris, our former technician, Stephen Snyder, and lab members Suzy Strickler and Sue Sherman-Broyles. We thank Kent Loeffler for taking and improving photographs of our flats of plants.

References

1. Fiebig, A., Kimport, R., and Pruess, D. (2004) Comparisons of pollen coat genes across Brassicaceae species reveal rapid evolution by repeat expansion and diversification. *Proc. Natl. Acad. Sci. USA* **101,** 3286–3291.
2. Kusaba, M., Dwyer, K., Hendershot, J., Verbalov, J., Nasrallah, J. B., and Nasrallah, M. E. (2001) Self-incompatibility in the genus *Arabidopsis*: Characterization of the S-locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell* **13,** 627–643.
3. Schein, M., Yang, Z. H., Mitchell-Olds, T. and Schmid, K. J. (2004) Rapid evolution of a pollen-specific oleosin-like gene family from *Arabidopsis thaliana* and closely related species. *Mol. Biol. Evol.* **21,** 659–669.
4. Williams, P. H. (1985) Maintenance of CrGR Stocks, in *CrGC Resource Book*, University of Wisconsin, Madison, WI.
5. Scholl, R., Rivero-Lepinckas, L., and Crist, D. (1998) Growth of plants and preservation of seeds, in *Methods in Molecular Biology, vol. 82: Arabidopsis Protocols* (Martinez-Zapater, J. and Salinas, J., eds.). Humana Press, Totowa, NJ.
6. Preuss, D. (2002) How to grow Arabidopsis, in *Arabidopsis—A Laboratory Manual* (Weigel, D. and Glazebrook, J., eds.) Cold Spring Harbor Press, Cold Spring Harbor, NY.
7. Nasrallah, J. B., Nishio, T., and Nasrallah, M. E. (1991) The self-incompatibility genes of Brassica: Expression and use in genetic ablation of floral tissues. *Annu. Rev. Plant Mol. Biol*. **42,** 393–422.
8. Mitchell-Olds, T. (2001) *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol. Evol.* **16,** 693–700.
9. www.naturalvariation.org.
10. Neuffer, B. and Hurka, H. (1999) Colonization history and introduction dynamics of *Capsella bursa-pastoris* (Brassicaceae) in North America: isozymes and quantitative traits. *Mol. Ecol.* **8,** 1667–1681.

4

Grafting

Katherine Bainbridge, Tom Bennett, Colin Turnbull, and Ottoline Leyser

Summary

Grafting provides a simple way to generate chimeric plants with regions of different genotypes, and thus to assess the cell autonomy of gene action. The technique of grafting has been widely used in other species, but in *Arabidopsis*, its small size makes the process rather more complicated. However, there are now several well-established grafting procedures available, which we described here, and their use has already contributed greatly to understanding of such processes as shoot branching control, flowering, and disease resistance.

Key Words: *Arabidopsis thaliana*; grafting; graft-transmissible signal.

1. Introduction

The assessment of the cell autonomy of a mutant phenotype can provide highly informative information about gene function. This kind of analysis requires the construction of chimeric plants with cells of different genotypes. There are several ways to achieve this, including the tissue-specific expression of a wild-type gene in a mutant background (e.g., ***1***) and the generation of sectors of different genotypes following somatic recombination or chromosome breakage (e.g., ***2***), or transposition (e.g., ***3***) or site-specific homologous recombination (e.g., ***4***) to remove an insertional mutagen. These methods are versatile in allowing different amounts and positions of the tissues of each genotype to be generated. However they are all very time-consuming, requiring transgenesis and/or construction of lines of particular genotypes, and a system to mark the different sectors and thus identify their genotypes.

In contrast, grafting is an extremely simple method for making a chimeric plant. In some ways, it is more restricted in its applications than those mentioned above, because only a limited number of options are available for connecting tissues of different genotypes. However, the methods are straightforward, do not require construction of complex transgenics or other genotypes, and enable an almost infinite number of genotype combinations to be tested.

The best-described graft combination in *Arabidopsis* is a simple root-shoot graft, performed on young seedlings, to generate a plant where the genotype of the root differs from that of the shoot ***(5)***. This method, with variations, is described below. In addition it is possible to graft a seedling shoot into the hypocotyl of a second seedling, a so-called Y-graft, to generate a plant with two genetically different shoot systems ***(5)***. There are also previous reports of success with bolting stem grafts ***(6)*** and more recently with mature rosette grafts ***(7)***, and there is no reason why other versions should not be equally successful. To date, *Arabidopsis* grafting has been reported in relation to several diverse biological processes including shoot branching ***(1,5,8,9)***, flowering time ***(7,10)***, leaf development ***(11)***, nutrient transport ***(12)***, wound responses ***(13)***, and disease resistance ***(14)***, indicating that it is an approach with wide applicability in this species.

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

2. Materials

1. Sterilized, cold-treated, good-quality *Arabidopsis* seed of appropriate genotypes.
2. 0.3-mm diameter silicon tubing (SF Medical, cat. no. SMF3-1050, available through VWR International), cut into 2- to 3-cm sections and autoclaved (*see* **Note 1**).
3. Razor blades or No. 15 scalpel blades (*see* **Note 2**).
4. Microsurgery knife: 15° disposable stab knife (Fine Science Tools, cat. no. 10315-12).
5. Fine forceps.
6. 10-cm square Petri dishes.
7. ATS (*Arabidopsis thaliana* salts *[15]*) or equivalent, agar (0.8%), sucrose (1%).
8. Dissecting microscope.
9. 22/18°C growth cabinet.
10. 27°C growth cabinet.

3. Methods

3.1. Root-Shoot Grafts

1. Under sterile conditions sow the seed onto square Petri dishes containing ATS-agar (or equivalent), with a spacing of 1 cm between seeds. Place the sealed plates vertically in a growth cabinet under standard axenic growth conditions (*see* **Note 3**). Leave the seedlings to germinate and grow for 3 d.
2. After 3 d move the seedlings to a growth cabinet set at 27°C (*see* **Note 3**) for a further 2 d.
3. Cut the sterile silicone tubing into lengths of roughly 2 mm (*see* **Note 4**).
4. Under sterile conditions, grafting can now be performed. Cut selected seedlings transversely across the hypocotyl (*see* **Note 5**) while on the agar plates. The root should not be disturbed and should essentially remain in place. Remove the apical part of the seedling and place a collar (*see* **Note 4**) over the cut hypocotyl of the rootstock. The top of the rootstock should be about halfway along the length of the collar. Feed the hypocotyl of a suitably excised scion (*see* **Note 5**) into the collar such that the base of the scion meets the rootstock. Thus a whole seedling is reconstituted. Note that as well as the reciprocal genotype combinations, it is necessary to include appropriate controls in which self-grafted plants are used to reconstruct the original genotypes, to ensure that the grafting process itself does not affect the phenotype of interest.
5. Using a dissecting microscope, inspect the graft junctions. The two graft parts should be in contact across the whole of the grafting surface with no gaps. If this is not the case, the scion should be pushed further into the collar until it does meet the rootstock. As the success rate of the protocol is 50 to 70%, it is recommended that twice the number of seedlings needed be grafted.
6. When grafting is complete, if suitably moist (*see* **Note 6**), return the plates to the 27°C growth cabinet for 3 to 4 d (**Fig. 1**).
7. After this time, grafts can be assessed for healing using a dissecting microscope (*see* **Note 7**). Transfer successful grafts to soil (*see* **Note 8**) and use a propagator lid to keep humid for about a week.
8. At an appropriate time thereafter, plants can be phenotypically assessed. When appropriate phenotypic data have been recorded, the plants can be assessed for graft integrity, thus allowing confirmation of the validity of the results (*see* **Note 9**).

3.2. Wedge Grafts and Y-Grafts

Instead of cutting the hypocotyl transversely, grafting can also be achieved with V-shaped “wedge-slit” connections. These are similar to many horticultural graft types. Precise cuts are essential and are best made under well-lit dissecting microscope conditions; magnification of ×25–×40 is ideal.

3.2.1. Single Wedge Graft

1. Make the rootstock by cutting the hypocotyl transversely with a razor or scalpel blade about one-fourth of the distance from the top, then slit down middle of the hypocotyl with a microsurgery knife (*see* **Note 2**).

Fig. 1. Detail of a root-shoot graft, showing the collar of sterile silicone tubing that holds the graft junction.

2. Make the scion by cutting a very shallow angled V-shape with a microsurgery knife. The first cut should extend more than halfway across the hypocotyl, but do not sever the root completely; otherwise the shoot will move around a great deal when making the second cut. This second cut should result in a symmetrical wedge.
3. Push the scion wedge gently into the slit (which should be same length as the wedge) in the rootstock. Tissue elasticity and surface tension will keep these grafts together without the aid of a collar.

Some practice is needed for these cutting procedures—mainly to achieve a very fine sawing action with the knife, rather than pushing down with large strokes.

3.2.2. Two-Shoot Y-Graft

This is a modification of above—a wedge-shaped scion connected into a cut in the *side* of an otherwise intact rootstock plant, to generate a graft with two shoots on a single root system. The rootstock plant keeps its roots. Y-grafts can be easier to cut and assemble if hypocotyls are curved: rotate pairs of vertical plates 60° left and right 1 d before grafting. The two shoots are then aligned with curves facing away from each other.

1. Make a shallow angled slit into side of hypocotyl, starting about one-third of the way from the top and extending no more than halfway across the diameter so that the central vascular tissue is penetrated but not severed.
2. Make a wedge-shaped scion as above.
3. Assemble by aligning the shoots as well as possible for maximum contact area.

4. It is often also necessary to trim off the majority of one cotyledon on each shoot, to allow the two shoots to sit close together.

4. Notes

1. Collars are used to support the graft and hold the rootstock and scion together during graft healing. We have found that they increase the proportion of successful grafts dramatically. However, it is possible to perform hypocotyl grafting without collars. Although this is a less efficient process, it allows greater flexibility. The protocol is essentially the same as grafting with collars, with only slight alteration. For single grafts a normal transverse cut can be used, but a "slit and wedge" graft can give better results, as it holds the scion and rootstock together more effectively. The major advantage of collarless grafting is the ability to perform two-shoot Y-grafts to test shoot-to-shoot signaling, which is not possible when a collar is used.
   Grafting can also be performed on short-day grown seedlings. The seedlings should be grown at a constant 23°C (~100 μmol/m^2/s) for 7 to 9 d, and then grafted. After grafting, they should be returned to this temperature regime for at least 1 wk but up to 6 wk, at which point successfully grafted plants can be transferred to soil.
2. The razor blades should have very fine edges in order to make clean cuts and avoid squashing the hypocotyl. Standard industrial razor blades are not appropriate. Number 15 scalpel blades may be used, but we find the best results are given by Wilkinson Sword "Classic" double-sided razor blades (or equivalent). The razor blades must be sharp at all times, and so should be changed frequently. A different blade should be used for cutting the collars (*see* **Note 4**). For wedge-shaped single and Y-graft connections, a disposable microsurgery knife is ideal because of its thin and ultra-sharp blade, but care is needed to avoid damage to the delicate cutting edge.
3. For the initial 3-d period, a standard regime of 16 h light/8 h dark, 22/18°C, and 100 μmol/m^2/s should be used. For the second two days, and the graft-healing period, a regime of 16/8 h light/dark, constant 27°C temperature, and 60 μmol/m^2/s should be used. Growing the seedlings at 27°C increases the levels of endogenous auxin in the plant, which in the first instance increases hypocotyl length ***(16)***, allowing easier grafting; and in the second instance promotes callus formation and healing. The reduction in light intensity reduces twisting of the hypocotyls. Such twisting makes grafting more difficult, and disrupts graft healing.
4. The collars used to hold grafts together are made from sterile 0.3-mm i.d. silicone tubing by slicing the tubing into ~2-mm sections. Difficulties will be experienced in fitting the rootstock and scion together if the collars are too long. The collars can also be slit longitudinally before use, which allows the collar to open up as the plant grows, or for the collar to be removed after the graft has healed fully. A pointed scalpel blade (e.g., No. 11) is best for this, and slitting can be facilitated by first threading the uncut 3 cm lengths of tubing onto suitable diameter of fine wire.
5. There are two key points to assembling successful grafts. The first is to select the most appropriate seedlings on each plate. For most situations, the selected seedlings should have long, straight hypocotyls and strong root growth. There is a small range of hypocotyl thicknesses that can be used. It is often difficult to distinguish which seedlings have the correct dimensions; trial and error is required to some degree. Seedlings with hypocotyls that do not fit into the collars easily should be discarded, as forcing them in will damage the seedling. Similarly, seedlings that are a very loose fit in the collar should also be discarded, as the graft will not be held together effectively. If seedlings of the correct size are used, the graft should fit effortlessly together.
   The second key element is in the cutting of the hypocotyl. Cuts should be as clean and straight as possible. The hypocotyl should not be squashed during cutting, and it should not be necessary to cut into the agar to cut through the seedling. These problems can be avoided by use of a new blade. A sharp razor should slice through the hypocotyl with almost no resistance. In addition, preventing seedlings sinking into the gel can be achieved by using double-strength gel in the media, and/or growing the seedlings on a "raft" of cellulose membrane filter (Millipore type) on the surface of the gel. Initially, it may take some practice to be able to cut the hypocotyls in the correct way, and it is advisable to have a few hours' experience on other seedlings before attempting grafting itself.

It is also important to cut the seedlings in the correct place. Best results appear to be produced if the rootstock donor is cut three-fourths of the way up the hypocotyl and the scion donor is cut halfway up the hypocotyl. In this case both the root of the scion donor and the shoot of the rootstock donor cannot be used for further grafts and should be discarded. It is possible to use all excised parts by cutting all seedlings halfway up the hypocotyl and simply swapping scions between rootstocks, but this may increase the risk of adventitious rooting and make insertion into the tubing more difficult.

6. Plates used for grafting should be as moist as possible at all times, as high humidity aids the graft healing process. It may, however, be necessary to remove excess surface water before grafting. If this is the case, or if the plants appear to be drying out (e.g., indicated by dull, soft, or wilting cotyledons), a small amount of sterile water can be added to the plates as needed during grafting and before they are sealed up at the end of the procedure.
7. Only truly grafted seedlings should be used, otherwise results may be erroneous. This can only be shown definitively when the plants are harvested (*see* **Note 8**), at which stage it is generally obvious if a graft has succeeded. Visual inspection using a dissecting microscope should show if the scion and rootstock have fused. However, if further confirmation is needed, a very light pull of the scion with forceps will determine whether the graft has united. Grafts are often connected by 4 d, but obviously strengthen further with time. Normally transfer of successful grafts can be done 6 to 7 d after grafting, or a little longer for Y-grafts, which need to be stronger.

   Usually a proportion of the scions will have produced adventitious roots from hypocotyl tissue within the collar, which displace the rootstocks. These seedlings should clearly be discarded. Scions that produce adventitious roots above the level of the collar, but which have also joined to the rootstock, can theoretically be used, as long as the adventitious roots are excised. However, adventitious root formation is often a sign of poor graft connection, so rescuing grafts by root excision may be futile.
8. Transfer plants to soil as soon as shoot (and root) growth seems reestablished, usually 6 to 7 d after grafting. To minimize stress, keep everything wet during transfer—add extra water to plates, saturate potting mix, spray plants with fine mister and cover the tray as soon as it has been filled with plants. Pick plants off plates carefully–"hook up" with fine forceps or grab edge of cotyledon. With Y-grafts, be careful not to bend the graft union—it will probably break. Drop roots into a prebored hole, gently push potting mix across to hold roots in place. Do not bury the graft union; otherwise it is hard to inspect and adventitious rooting will be promoted.

   Keep tray vents closed for the first 3 d or so, then open vents for another 3 d. Remove the lid after about a week. Keep growth cabinet humidity high if possible. Often a few casualties are seen soon after the lid is removed—these have poor root systems (poor grafts or adventitious root removal was too much for them).
9. Confirming that the plants have grafted successfully, and can therefore be included in the dataset, is normally a destructive process, and is thus best performed after phenotypic assessment. Plants should be removed from the growth medium intact and the graft union found. Often the silicon collar is split by the broadening of the stem, and may be absent, but the union is usually identifiable by the clear scarring at the site. Depending on the nature of the experiment, either the majority, or all of the root tissue must originate beneath the level of the union. Otherwise, the plants are essentially in an "ungrafted" state.

   Use of a β-glucoronidase (GUS) reporter gene can aid in the verification of graft integrity. If one of the genotypes of plant carries a broadly expressed promoter-GUS transgene (for example, CaMV 35S::GUS), then it is possible to use GUS activity to verify the correctly grafted plants, and also to identify adventitious roots of the "wrong" genotype.

References

1. Booker, J. P., Chatfield, S. P., and Leyser, O. (2003) Auxin acts in xylem-associated or medullary cells to mediate apical dominance. *Plant Cell* **15,** 495–507.
2. Furner, I. J., Ainscough, J. F. X., Pumfrey, J. A., and Petty, L. M. (1996) Clonal analysis of the late flowering *fca* mutant of *Arabidopsis thaliana*: cell fate and cell autonomy. *Development* **122,** 1041–1050.

3. Jenik, P. D. and Irish, V. F. (2000) Regulation of cell proliferation patterns by homeotic genes during Arabidopsis floral development. *Development* **126,** 1267–1276.
4. Woodrick, R., Martin, P. R., Birman, I., and Pickett, F. B. (2000) *Arabidopsis* embryonic shoot fate map. *Development* **127,** 813–820.
5. Turnbull, C. G. N., Booker, J. P., and Leyser, H. M. O. (2002) Micrografting techniques for testing long-distance signalling in *Arabidopsis. Plant J.* **32,** 255–262.
6. Rhee, S. Y. and Somerville, C. R. (1995) Flat-surface grafting in *Arabidopsis thaliana*. *Plant Mol. Bol. Rep.* **13,** 118–123.
7. Ayre, B. G. and Turgeon, R. (2004) Graft transmission of a floral stimulant derived from CONSTANS. *Plant Physiol*. **13,** 2271–2278.
8. Sorefan, K., Booker, J., Haurogné, K., et al. (2003) *MAX4* and *RMS1* are orthologous dioxygenase-like genes that regulate shoot branching in *Arabidopsis* and pea. *Genes Dev.* **17,** 1469–1474.
9. Booker, J., Auldridge, M., Wills, S., McCarty, D., Klee, H., and Leyser, O. (2004) MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule. *Curr. Biol*. **14,** 1232–1238.
10. An, H.L., Roussot, C., Suárez-L—pez, P., et al. (2004) CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis*. *Development* **131,** 3615–3626.
11. Van Norman, J. M., Frederick, R. L., and Sieburth, L. E. (2004) BYPASS1 negatively regulates a root-derived signal that controls plant architecture. *Curr. Biol*. **14,** 1739–1746.
12. Green, L. S. and Rogers, E. E. (2004) FRD3 controls iron localization in *Arabidopsis*. *Plant Physiol*. **136,** 2523–2531.
13. Xia, Y. J., Suzuki, H., Borevitz, J., et al. (2004) An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. *EMBO J.* **23,** 980–988.
14. Hawkes, V. and Turner, J. (2004) How are signalling pathways involving jasmonate and calcium linked to the wound response in *Arabidopsis*? 15th International Conference on Arabidopsis Research, Berlin. Abstract T008-002.
15. Wilson, A. K., Pickett, F. B., Turner, J. C., and Estelle, M. (1990) a dominant mutation in Arabidopsis confers resistance to auxin, ethylene and abscisic acid. *Mol. Gen. Genet.* **222,** 377–383.
16. Gray, W. M., Ostin, A., Sandberg, G., Romano, C. P., and Estelle, M. (1998) High temperature promotes auxin-mediated hypocotyl elongation in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **95,** 7197–7202.

5

Synchronization, Transformation, and Cryopreservation of Suspension-Cultured Cells

Margit Menges and James A. H. Murray

Summary

We have recently described the selection of rapidly dividing *Arabidopsis* cell suspension cultures MM1 and MM2d that provide a powerful platform for plant cell-cycle research. Here we provide detailed protocols and procedures to achieve high levels of synchronization, either by starving the cell cultures of sucrose or by applying the toxin aphidicolin. Cell-cycle activity during cell-cycle reentry (starvation-induced synchrony) or further cell-cycle progression (aphidicolin-induced synchrony) can be conveniently followed by using various validation procedures, such as determination of labeling index and metaphase/anaphase index or flow cytometry. We also describe a procedure that allows clonal transformed cell-suspension lines to be produced using *Agrobacterium*-mediated transformation, and an optimized and straightforward method for the cryopreservation and recovery of both parental and transformed lines which is applicable both to *Arabidopsis* and the tobacco BY2 cell lines. Cell-cycle synchronization capacity of the parental lines is maintained after both transformation and recovery from cryopreservation. The techniques described here require no specialized equipment and are suitable for routine laboratory use, greatly facilitating the handling and maintenance of cell cultures. The ability to store easily large numbers of transformed lines opens the possibility of using *Arabidopsis* cell suspension cultures for future high-throughput cell-cycle analysis.

Key Words: *Arabidopsis* cell suspension; synchronization; starvation; aphidicolin block; cell-cycle activity; *Agrobacterium*-mediated transformation; selection and regeneration; cryopreservation; rapid regrowth ability; viability.

1. Introduction

Dispersed plant cell suspension cultures allow processes such as cell proliferation and growth to be studied in the absence of developmental and differentiation programs. They clearly represent a very useful system to investigate the cell cycle ***(1)*** although the application of functional genomic procedures considerably extends their value for other aspects of cell biology. For cell-cycle research their particular value is that they provide a homogeneous population of nearly identical cells that offer material for synchronization, which may be achieved either by removal and subsequent resupply of a compound required for growth, such as phosphate, nitrate, hormones, or sucrose ***(2–8)***, or by applying reversible blocks at different stages of the cell cycle using specific inhibitors ***(8–16)***. Such synchronized cultures can therefore provide material representative of specific cell-cycle phases that can be used to detect gene expression at the RNA or protein level. However, only a few plant cell suspensions can be synchronized to a high degree; one of these, widely used in cell-cycle studies, is the tobacco Bright Yellow-2 (BY-2) cell line ***(13,17)***. Few *Arabidopsis* cell cultures are available, such as the widely used T87 cell line (***18,19***), various other Col-lines (***20,21***) and a suspension culture of *Arabidopsis*

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

thaliana ecotype Landsberg *erecta* (**22,23**). This last line was the source from which we derived the cell lines MM1 and MM2d (**8**). However, until recently it was reported to be difficult to achieve synchrony in *Arabidopsis* (**19,23–26**). Here we review procedures we have developed to achieve highly synchronous cultures of MM1 and MM2d, suitable for following both the reentry of cells into the cell cycle (sucrose-starvation induced) and progression from S phase following an aphidicolin block/release (**8**). These highly synchronizable *Arabidopsis* suspension cultures have proved to be powerful transcript-profiling platforms for cell-cycle studies using Affymetrix ATH1 microarrays (**27,28**) and massively parallel signature sequencing (MPSS).

To extend further the utility of the *Arabidopsis* cell suspension lines, we established an *Agrobacterium*-mediated transformation protocol (**29**). In previous studies the integration of DNA into plant cells has been achieved by using various methods, including direct gene transfer by biolistic transformation (microprojectile bombardment using DNA-coated gold particles), polyethylene glycol-mediated DNA uptake into isolated protoplasts, electroporation, and the use of *Agrobacterium*, plant viruses and liposomes as DNA carriers (**30–32**). In *Arabidopsis*, *Agrobacterium*-mediated transformation has been used in a wide range of tissues, such as roots, shoots, calli, and cell suspension (**20,29,33,34**). Here we describe the establishment of an *Agrobacterium*-mediated transformation procedure using the MM1 and MM2d cell lines. Such generated transgenic-derivative cell lines retain their synchronization properties, a prerequisite for further investigation of the effects on the cell cycle on over-expression or downregulation of target genes (**29**).

Routine maintenance of cell cultures by repeated weekly subculture is not only labor-intensive but also creates the risk of transgene loss or other acquired variation in growth rate, chromosome cytology or other undesirable genetic changes. Therefore we also sought a reliable method for the long-term storage and regeneration of transgenic and parental cell cultures. Despite the widespread use of techniques for storing animal cells in liquid nitrogen (LN), few laboratories have used this method for the routine storage of plant cells. Controlled freezing and low-temperature storage of cultured plant cells in LN might enable the characteristics of newly initiated cultures to be preserved for indefinite periods. However, the few reports of such methods being used for plant cells typically suggest low viabilities, long lag periods before a rapidly growing cell suspension is recovered, and the requirement for specialized apparatus to provide a controlled slow cooling rate (**35–38**). Here we describe the development of a simple and effective cryopreservation technique in LN of the *Arabidopsis* cell lines MM1, MM2d and transgenic-derivative lines, that offers a high viability rate and rapid post-thaw growth during the reestablishment of cultures. In addition, desirable characteristics such as synchronization potential (and transgene expression) are unaffected by the frozen storage and recovery (**29**). The techniques involved require no specialized equipment and are equally effective with the tobacco BY-2 cell lines, suggesting their probable general utility for the frozen storage of rapidly growing plant cells. They therefore create the potential for high-throughput cell-cycle studies using plant cell suspension cultures with possible applications in functional genomic analysis.

2. Materials

2.1. General

1. MSS medium: 1X MS-salt (without sucrose, without indole-3-acetic acid [IAA], without kinetin, without agar), supplemented with 3% (w/v) sucrose, 0.5 mg/L naphthalene acetic acid (NAA) and 0.05 mg/L kinetin, pH 5.8 (adjusted with 1 *M* KOH). Autoclave and store at room temperature (RT).
2. MS medium: same medium composition as described in *1.* but without the addition of sucrose. Autoclave and store at RT.
3. 0.8% MSS agar: MSS medium, supplemented with 0.8% (w/v) Bacto™ Agar (cat. no. 0140-01, Difco Laboratories, MI). Autoclave and store at 4°C.

4. PBS buffer: phosphate-buffered saline, dissolve 11.5 g of Na_2HPO_4, 2.96 g of $NaH_2PO_4 \cdot 2H_2O$, and 5.84 g of NaCl in 1 L of distilled water and autoclave. Alternative formula and commercially available tablets may be used. Store at RT.
5. Hemocytometer: counting chamber of 0.2-mm depth.

2.2. Synchronization Protocols and Validation Procedures

1. 2 *M* sucrose solution: prepare a 2 *M* stock solution and sterilize by autoclaving. Store at RT.
2. Aphidicolin: prepare a 5 mg/mL stock solution by dissolving the powder in dimethylsulfoxide (DMSO). Store at 4°C.
3. Filtration unit used to wash aphidicolin-blocked cells: Cut the base from a screw-top plastic container of an approximate volume of 300 mL, and cut out the lid. Use the remaining screw ring of the lid to clamp a nylon mesh over the top (mesh size 47 μm, Spectra/Mesh® Nylon Filters, Spectrum Laboratories).
4. Labeling reagent: the labeling reagent is part of the RPN20-cell proliferation kit (cat. no. RPN20/OL/96/06, Amersham Pharmacia Biotech, UK) and is supplied as a concentrated aqueous solution of 5-bromo-2'-deoxyuridine (BrdU) and 5-fluoro-2'deoxyuridine (10:1 ratio) with a BrdU concentration of 3 mg/mL in the stock solution. Store at 4°C.
5. Fixative solution: 3.7% formaldehyde in PMEG-buffer. PMEG is 50 m*M* PIPES, 2 m*M* $MgSO_4$, 5 m*M* EGTA, 2% glycerol (v/v), pH 6.8. Store at RT.
6. Enzyme solution: 1% (w/v) cellulase-R10 (Yakult Honsha Co. Ltd, Japan), 0.1% (w/v) pectolyase Y-23 (Kikkoman, Japan) in 0.4 *M* mannitol. Store at –20°C.
7. Igepal solution: 0.1% (v/v) Igepal CA-630 (Sigma) in PBS buffer. Store at RT.
8. Precoated microscope slide: coat a normal glass slide with Vectabond™ reagent, Vector Laboratories, Cat. No. SP-1800.
9. Humidified incubation chamber: for example, a 10 mL square Petri dish lined with damp paper. Place the microscope slides on top of two Pasteur pipets (glass).
10. Glycine blocking solution: 0.1 mol/L glycine, 1% (w/v) bovine serum albumin (BSA), 0.05% (v/v) Triton X-100, in PBS buffer. Store at 4°C.
11. Primary antibody: anti-5-bromo-2'-deoxyuridine/nuclease; part of the RPN20-cell proliferation kit (Amersham Pharmacia Biotech, UK). Store at 4°C.
12. Secondary antibody: Texas Red-conjugated-donkey-antimouse antibody; Jackson Immuno Research, cat. no. 715-076-150. Dilute the antibody 1:400 in PBS buffer and store at 4°C.
13. Mounting-solution: Vectashield containing 4,6-diamidino-2-phenylindole (DAPI); cat. no. H-1000, Vector Laboratories, Burlingame, CA. Store at 4°C.
14. Release and staining of cell nuclei: Solution A, nuclei-extraction buffer; Solution B: DAPI staining buffer (CyStain UV Precise P, Partec).
15. Flow cytometry analysis: Flow cytometer (PASIII, Partec GmbH, Münster, Germany); CellTrics (yellow 50 μm, Cat. no. 06-4-2317, Partec); sample tubes (cat. no. 55.484, Sarstedt); High Resolution Kit for plant ploidy level analysis (CyStain UV Precice P, Cat. no. 05-5002, Partec); HBO lamp excitation (Partec, mercury lamp HBO 100 long life, 100W, excitation filters KG1, BG38, UG1) and detection of emission using a blue fluorescence emission filter GG435 (long-pass color glass); sharp razor blade to chop cells and release nuclei.
16. Multicycle for Windows software to analyze cell-cycle phases (Phoenix Flow Systems, San Diego, CA).

2.3. Transformation

1. Luria-Bertani (LB) medium: 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl in distilled water. Autoclave and store at RT.
2. Tetracycline: antibiotic to select for *Agrobacterium* (strain-specific). Prepare a 10 mg/mL stock solution in distilled water, filter-sterilize and store at –20°C.
3. Acetosyringone: 3',5'-dimethoxy-4'-hydroxyacetophenone (cat. no. 115540010, Acros Organics). Prepare a 20 m*M* stock solution in distilled water, filter-sterilize and store at –20°C.
4. Timentin: ticarcillin/clavulanic acid, 1500 mg/100 mg; cat. no. T0152, Duchefa Biochemie BV. Prepare a 250 mg/mL stock solution in distilled water, filter-sterilize and store at –20°C.

5. Kanamycin: antibiotic to select for transgenic calli (construct-specific). Prepare a 100 mg/mL stock solution in distilled water, filter-sterilize and store at –20°C.
6. MSS+selection-agar: Autoclaved 0.8% MSS-agar supplemented with 250 μg/mL timentin and 100 μg/mL kanamycin. Store at 4°C.

2.4. Cryopreservation and Regeneration

1. MSS+sorbitol medium: MSS medium supplemented with 0.5 *M* D-sorbitol. Autoclave and store at RT.
2. Dimethylsulfoxide (DMSO): filter-sterilize the DMSO solution prior to addition to the cell suspension. Store at RT.
3. Cryovials to store cells in LN (cat. no. E31100012, Starlab, Helsinki, Finland).
4. Cryo-freezing container: Nalgene™ Cryo 1°C. Freezing container, cat. no. 5100-0001, Fisher Scientific, Loughborough UK; filled with isopropanol according to the manufacturer's instructions.
5. Styropore box to use as additional freezing unit: in order to reduce the cooling rate of –1°C to approx –0.5°C per min during the incubation period at –80°C, the cryofreezing container is placed in a styropore box. A standard box of about 2 L volume (width, 14 cm; length, 14 cm; height, 10 cm) and 2.5 cm wall thickness (as used for the delivery of biological supplies on dry ice) is found to be suitable.
6. Nylon membrane: CellMicroSieves™, 70 micron pore size, BioDesign, New York.
7. 0.4% Trypan blue solution.

3. Methods

3.1. Maintenance of Arabidopsis Cell Cultures MM1 and MM2d

3.1.1. Cell Line MM1 (see **Note 1**)

1. Weekly maintenance of cell culture: Subculture 3.5 mL of an early stationary stage MM1 cell suspension (7 d after previous subculture) into 100 mL of fresh MSS medium (300 mL narrow necked Erlenmeyer flask, covered with a double layer of domestic aluminium foil).
2. Culture conditions: Grow MM1 cells under continuous light conditions (24 h per day at an average 1300 Lux; for example, use Philips lamps with the specification TLD HF 50W/840 to ensure light level). Rotate the cell suspension at 120 rpm with 25 mm orbit at a temperature of 23°C.

3.1.2. Cell Line MM2d (see **Note 1**)

1. Weekly maintenance of cell culture: Subculture 3.5 mL of an early stationary stage MM2d cell suspension (7 d after previous subculture) into 100 mL fresh MSS medium (300 mL Erlenmeyer flask).
2. Culture condition: Grow MM2d cells under continuous darkness. Rotate at 130 rpm in a darkened incubator shaker with 19 mm orbit at a temperature of 27°C.

3.1.3. Growth-Curve Determination

1. Start of cell culture (day 1): Inoculate 5 mL of an early stationary phase culture (7 d after previous subculture) into 100 mL of fresh MSS medium and incubate under the conditions described in **Subheading 3.1.** Take 1-mL samples at approx 24-h intervals.
2. Cell density determination (days 1–8): Vigorously pipet the cell suspension up and down through a 1-mL Gilson pipet tip several times to disrupt possible cell clumps. Count cells using a hemocytometer (counting chamber, 0.2 mm depth) to determine the cell density (cells no/mL, *see* **Note 2**).

3.2. Achieving Synchronization in Arabidopsis Cell Suspension

The methods described in this section outline (**Subheading 3.2.1.**) protocols to achieve synchronous cell-cycle transition during cell-cycle reentry by arresting cells in G0/G1 (sucrose-

starvation-induced) or (**Subheading 3.2.2.**) further synchronous cell-cycle progression from S-phase onward after blocking cells at the G1/S transition (aphidicolin-induced), and (**Subheading 3.3.3.**) procedures to monitor cell-cycle activity.

3.2.1. Sucrose Starvation-Induced Synchronization

1. Start of cell culture before synchronization (day 1): Subculture 5 mL of an early stationary phase MM2d cell suspension (7 d after previous subculture) into 100 mL of fresh MSS medium. Incubate at 27°C, 130 rpm in the dark for a period of 4 d to have an exponentially growing culture (*see* **Note 3**).
2. Wash cells to remove sucrose from the medium (day 4): Transfer 100-mL aliquots of the cell suspension into 50-mL Falcon polypropylene tubes, centrifuge at 387*g* for 2 min without applying brake force, and discard supernatant (*see* **Note 4**). Wash the cell pellets twice with 50 mL of MS medium (lacking sucrose) by gently resuspending and repeating the centrifugation step. Finally resuspend and pool cells in a total of 100 mL of MS medium.
3. Dilution of the cell culture and nutrient starvation (day 4): Transfer 20 mL of the pooled resuspended cell suspension into 100 mL of fresh MS medium (in a 300 mL Erlenmeyer flask) to achieve a dilution factor of approx 1:5, and incubate at 27°C, 130 rpm in the dark for 24 h.
4. Sucrose addition and start of time course (day 5): Add sucrose to a final concentration of 3% (add 4.55 mL of a sterile 2 *M* stock solution) and incubate under the conditions described above. Take a T0 sample directly after starvation and hourly after sucrose is resupplied for validation procedures to follow, in particular, the transition from G1 into S phase and further cell-cycle progression through mitosis. A typical result of marker gene expression during reentry (Northern blot analysis of histone H4 and cyclin B1;1) and results of labeling index (LI) and metaphase/anaphase index (M/AI) determination is shown in **Fig. 1**. Typical flow cytometry analysis results are shown in **Fig. 2**.

3.2.2. Aphidicolin-Induced Synchronization

1. Start of cell culture before synchronization (day 1): Subculture 5 mL of an early stationary phase MM2d cell suspension (7 d after previous subculture) into 100 mL of fresh MSS medium. Incubate at 27°C, 130 rpm in the dark for a period of 7 d to have an early stationary phase cell suspension (*see* **Note 5**).
2. Dilution of the cell suspension (day 7): Transfer 20 mL of the cell suspension into 100 mL of fresh MSS medium to achieve a dilution factor of approx 1:5.
3. Add aphidicolin to arrest cells in late G1/early S phase (day 7): Add 4.16 µg/mL aphidicolin (add 100 µL of a 5 mg/mL stock solution) to the 120 mL of diluted cell suspension and incubate at 27°C, 130 rpm in the dark for a period of 24 h (*see* **Note 6**).
4. Wash the cell culture to remove aphidicolin (day 8): Filter the cell suspension through a homemade sterile filtration unit (*see* **Subheading 2.**). Transfer the filtration unit into a sterile glass beaker and resuspend the cell suspension in an aliquot of fresh MSS medium. Carefully swirl the beaker to wash the cells. Change the medium several times (use a total of 1 L of MSS medium) by simply lifting the filtration unit out of the glass beaker and discarding the washing solution. After approx 10 min resuspend the cell suspension in around 100 mL of MSS medium and vigorously swirl before pouring the cell culture from the filtration unit into a second sterile glass beaker. Transfer the cell culture into 50-mL Falcon polypropylene tubes and centrifuge at 387*g* for 1 min without braking applied. The washing procedure should not take more than 15 min in total (*see* **Note 7**).
5. Resuspend cell pellets and start of time course (day 8): Resuspend cell pellets in 120 mL of fresh MSS medium (in a 300 mL Erlenmeyer flask) and incubate under cultivation conditions as described above. Take a T0 sample directly after washing to release the block and hourly for validation procedures. Typical results after LI and M/AI determination are shown in **Fig. 3**, which clearly demonstrate a separation of S phase and M phase cells after synchronization. A typical flow cytometry analysis result to follow synchronous S phase transition of up to 80% of cells after aphidicolin block and release is shown in **Fig. 4**.

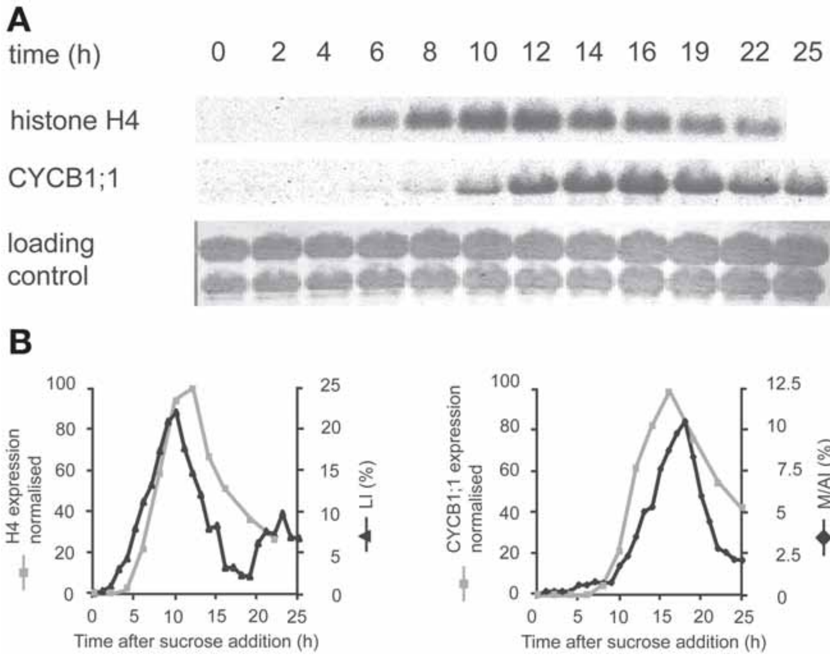


Fig. 1. Validation of cell-cycle activity after sucrose starvation-induced synchrony of MM2d cells. Exponential-phase *Arabidopsis* MM2d cells were subjected to sucrose removal and resupply *(8)*. **(A)** Northern blot of RNA samples prepared at 2–3 h intervals from MM2d cells synchronously reentering the cell cycle from a G1 state (t = 0) probed with histone H4 (marker gene for S phase) and cyclin B1;1 (mitotic marker gene). **(B)** Left: LI determination of S phase (▲) compared to quantified histone H4 expression (■). Signals after Northern blot analysis were quantified using NIH Image 1.62. The level of expression (in arbitrary units) was normalized by correcting against a loading control and expressing as a proportion of the maximum signal. Right: M/AI determination of cells showing mitotic figures in metaphase or anaphase (◆) compared to quantified cyclin B1;1 expression (■).

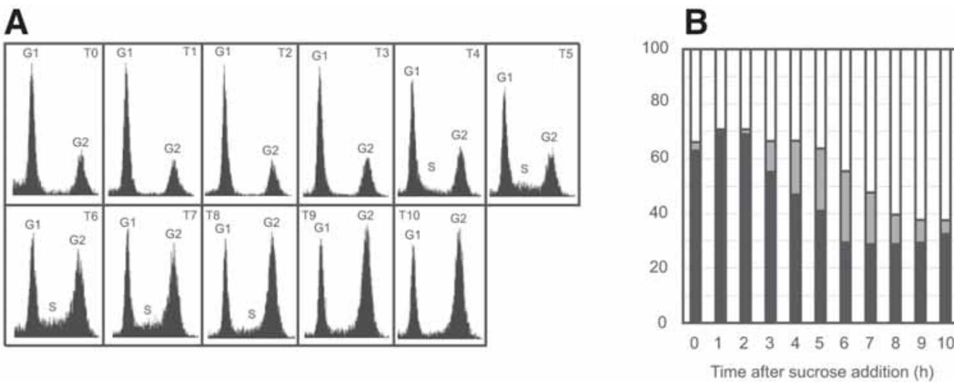


Fig. 2. Validation of cell-cycle activity after sucrose starvation-induced synchrony of MM2d cells. Exponential-phase *Arabidopsis* MM2d cells were subjected to sucrose removal and resupply. **(A)** Flow cytometry analysis of MM2d cells synchronously reentering the cell cycle from a G1 state (t = 0) showing the coherent population of cells progressing through S phase (plateau). Each block represents a sample taken at 1-h intervals from T0 (time of sucrose addition) to T10 (10 h later). **(B)** DNA histogram of flow analysis results in **A**.

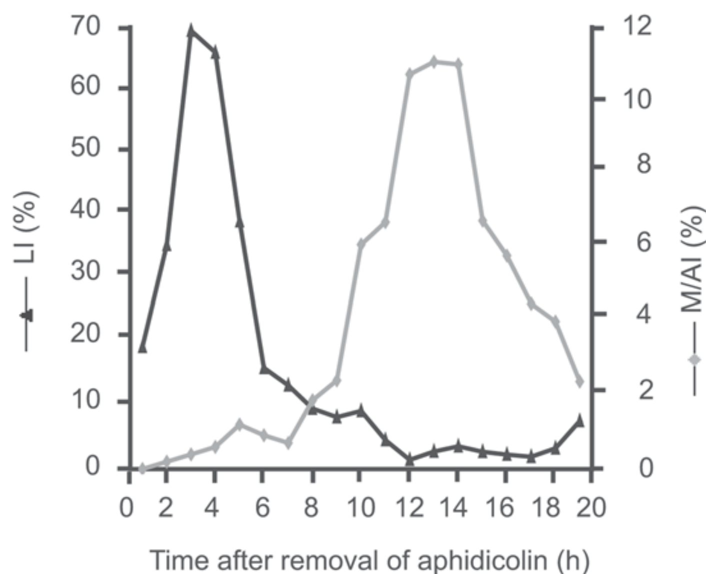


Fig. 3. Aphidicolin block/release of *Arabidopsis* cell line MM2d. LI and M/AI determination to detect S phase (LI, ▲) and metaphase/anaphase cells (M/AI, ◆) after the aphidicolin block is released (t = 0). The M/A index was determined as the proportion of cells with DAPI-stained mitotic figures.

3.3. Procedures to Monitor Cell-Cycle Activity

3.3.1. BrdU Incorporation to Determine LI and M/AI

1. Labeling of cells with bromodeoxyuridine (BrdU incorporation): Each hour of the synchrony, transfer 5 mL of cell suspension into a 100 mL Erlenmeyer flask. Add 10 μL of labeling reagent (dilution 1:500; *see* **Note 8**), to use BrdU at a final concentration of 6 μg/mL, and incubate for 60 min under the conditions described.
2. Harvest cells by centrifugation: Transfer the cell suspension into a 15-mL Falcon polypropylene tube and spin at 387*g* for 1 min. Discard the supernatant and resuspend the cell pellet in 2 mL of PBS buffer to wash the cells. Transfer the cells to a 2-mL microfuge tube, spin at 420*g* for 1 min, and discard the supernatant.
3. Fixation of cells with 3.7% formaldehyde: Add 1 mL of the fixative solution and incubate on a rotary wheel at 4°C (cold room) overnight. Fixed cells can be stored at 4°C prior to further analysis.
4. Enzyme digestion to break open the cell wall: Transfer a 100-μL aliquot of fixed cells into a 1.5-mL microfuge tube and wash three times with 1ml of PBS buffer by centrifugation 420*g*/1 min). Discard supernatant and add 1 mL of the enzyme solution. Invert the tubes several times during the incubation period of 25 min at RT (*see* **Note 9**). After centrifugation 420*g*/1 min) remove the enzyme solution.
5. Permeabilization of the cell wall: Resuspend the cell pellet in 1 mL of the Igepal solution and leave for 25 min at RT for permeabilization. Wash cells twice with 1 mL of PBS buffer by centrifugation 420*g*/1 min) and resuspend the cell pellet in 100 μL of PBS buffer. Transfer cells to a precoated microscope slide and leave to settle overnight in a humidified incubation chamber at RT (*see* **Note 10**).
6. Labeling of cells with primary antibody: Remove the PBS buffer with a pipet. Add 100 μL of the glycine blocking solution to block the cells and incubate for 10 min at RT. Remove the blocking solution and add 100 μL of the primary antibody (anti-BrdU). Incubate for 60 to 90 min at 30°C in a humidified incubation chamber (in the dark). Carefully wash the attached cells three times with 100 μL of PBS buffer for 5 min each (on slide) and remove the buffer solution.

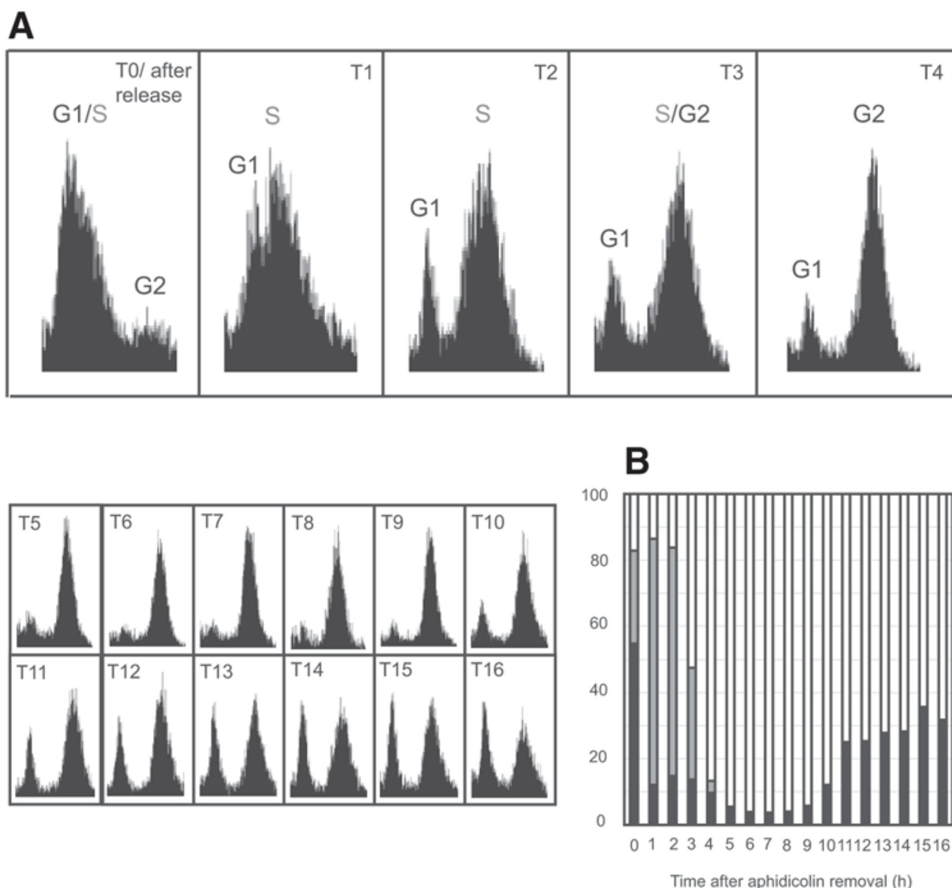


Fig. 4. Aphidicolin block/release of *Arabidopsis* cell line MM2d. **(A)** Flow cytometry analysis of MM2d cells after release of aphidicolin block in late G1/early S phase showing the coherent population of cells progressing synchronously through S phase (peak). Each block represents a sample taken at 1-h intervals from t = 0 (top panel, left) to t = 16 (bottom panel, right). **(B)** DNA histogram of flow analysis results in **A**.

7. Labeling of cells with secondary antibody: Add 100 µL of a Texas red-conjugated donkey-antimouse antibody and allow binding for 60 to 90 min by incubating at 30°C in a humidified incubation chamber (in the dark). Wash the attached cells carefully three times with 100 µL of PBS buffer for 5 min each (on slide) and remove any excess buffer solution. Mount the cells in DAPI-containing mounting solution.
8. Microscopic examination of antibody-labeled cells: Examination with light of wavelength 575 nm results in bright red cells where BrdU incorporation occurred during DNA synthesis. To determine the labeling index (LI), manually count antibody-marked (bright red) cells and calculate the percentage of labeled cells compared with the total number of cells. Score at least 1000 cells (*see* **Figs. 1** and **3**).
9. Microscopic examination of cells showing mitotic figures: To determine the metaphase/anaphase index (M/AI) in the same field, count the proportion of DAPI-stained cells in mitosis (metaphase and anaphase figures) compared to the total number of cells. Score at least 1000 cells (*see* **Figs. 1** and **3**).

3.3.2. Determination of DNA Content by Flow Cytometry

1. Harvest cells by centrifugation: Transfer 5 mL of cell suspension to a 15-mL Falcon polypropylene tube and centrifuge at 387*g* for 1 min. Discard supernatant and snap-freeze the cell pellet in liquid nitrogen.
2. Release of cell nuclei: Transfer a sample of the frozen cell pellet into a small Petri dish (5 cm diameter). Add 150 µL of solution A (nuclei-extraction buffer) and carefully chop the cell pellet with a sharp razor blade to release cell nuclei. Add 150 µL more of solution A and pipet the cell suspension up and down several times before leaving for approx 2 min at RT. Transfer the cell suspension into a CellTric (Partec) placed on a sample tube to filter the solution through a 50-µm nylon mesh.
3. DAPI stain of released nuclei: Add 1 mL of solution B (staining buffer) to the filtrate, mix carefully, and leave for approx 2 min.
4. Flow cytometry analysis of DNA content: Use HBO lamp excitation (filters KG1, BG38, UG1) and detection of emission (filter GG435) to count on average 10,000 stained particles (*see* **Figs. 2** and **4**).
5. Cell-cycle analysis: Analyze the observed peak areas using appropriate software, such as Multicycle for Windows, to determine the DNA content of cell nuclei (or precisely stained particles). The analysis results in the percentage of cells belonging to different cell-cycle phases (G1, S, and G2 phase) that can be represented in a DNA histogram (*see* **Figs. 2** and **4**).

3.4. Generation of Transgenic Arabidopsis *Cell Cultures (*Agrobacterium*-Mediated Transformation)*

3.4.1. Cocultivation With Agrobacterium *to Stably Transform* Arabidopsis *Cell Cultures*

1. Start of cell culture before cocultivation (day 1): Subculture 2.5 mL of an early stationary phase MM2d cell suspension (7 d after previous subculture) into 25 mL of fresh MSS medium (dilution 1:10, in a 100 mL Erlenmeyer flask). Incubate at 27°C, 130 rpm for a period of 2 d.
2. Preparation of *Agrobacterium* culture (day 2): Inoculate an *Agrobacterium* culture (such as *Agrobacterium* strain LB4404 harboring an appropriate construct) in 50 mL of LB medium containing the appropriate *Agrobacterium* strain-specific antibiotic selection (e.g., 10 µg/mL tetracycline) and leave to grow at 30°C, 130 rpm overnight. Take a 1-mL aliquot and measure the optical density (OD) of the overnight culture using a spectrophotometer (the OD_{600nm} should be between 1 and 1.5). Transfer 100 µL of the exponentially growing *Agrobacterium* culture to a 1.5 mL microfuge tube, add 900 µL of MSS medium (dilution 1:10) and wash three times with 1 mL of MSS medium by centrifugation 14,962*g*/2 min). Remove the supernatant and resuspend the pellet in 1 mL MSS medium (*see* **Note 11**).
3. Preparation of cells (day 3): Subculture 10 mL of the exponentially growing MM2d cell suspension (2 d old) into 50 mL of fresh MSS medium (dilution 1:5). Use this culture in part for (a) *Agrobacterium* infection and in part for (b) the production of conditioned medium.
4. Cocultivation of *Arabidopsis* cells with *Agrobacterium* (day 3): Transfer 10 mL of diluted (1:5, *see* **step 3a**) MM2d cell suspension into a 250-mL Erlenmeyer flask. Add acetosyringone to a concentration of 500 µ*M* (*see* **Note 12**). Infect the cell suspension with 100 µL of diluted (1:10, *see* **step 2**) *Agrobacterium* culture. Cocultivate the *Agrobacterium*-infected MM2d culture in darkness at 27°C without agitation for 2 d. As a positive/negative control, add to 10 mL of cell suspension 500 µ*M* acetosyringone without *Agrobacterium* infection and follow the same procedure.

3.4.2. Recovery of Putative Transgenic Calli

1. Cultivation of cells to obtain conditioned medium (day 3): Incubate the remaining diluted cell suspension (*see* **Subheading 3.3.1., step 3b**) at 130 rpm, 27°C in the dark for a further 2 d. Transfer the cell suspension to a 50-mL Falcon polypropylene tube and clear cells from conditioned medium by centrifugation 387*g* for 5 min without braking). Transfer the supernatant to a new 50-mL polypropylene tube (day 5).

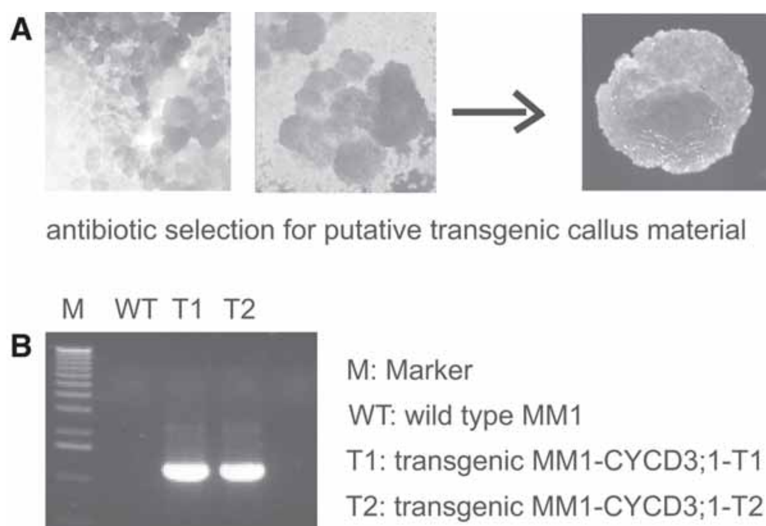


Fig. 5. Kanamycin selection and assaying for the presence of the transformed construct by PCR. *Arabidopsis* cell line MM1 was transformed as described. **(A)** After 2–3 wk incubation on selective medium, putative transformed callus material was transferred onto new MSS-agar plates (+ selection) and left to grow further (right picture). **(B)** After reestablishment of liquid cultures, DNA was prepared and used in a PCR reaction to confirm the presence of the transformed construct by using construct-specific primers.

2. Wash cells to remove *Agrobacterium* (day 5): Transfer the infected cell culture into a 50-mL Falcon polypropylene tube and centrifuge for 5min at 387*g*, without applying a brake force (*see* **Note 4**). Gently wash the cells (*see* **Note 13**) three times with 50 mL of fresh MSS medium 387*g* for 5 min without breaking) and resuspend the cells in 20 mL of conditioned medium (*see* **step 1**). To select against remaining *Agrobacterium* cells, add timentin to a concentration of 500 µg/mL (*see* **Note 14**).
3. First recovery of putative transformed cells in liquid culture (day 5): Transfer the 20 mL of cell suspension to a 250-mL Erlenmeyer flask and incubate the putative transformed MM2d cells for a further 3 d, rotating at 130 rpm, 27°C in the dark (without any further selection; *see* **Note 15**).
4. Recovery and selection of putative transgenic calli on solidified medium (day 8): Transfer the cell suspension to a 50-mL Falcon polypropylene tube and centrifuge for 5 min at 387*g*, without applying the brake force. Resuspend the cell pellet in 10 mL of fresh MSS medium, supplemented with 50 µg of NAA and 5 µg of kinetin (*see* **Note 16**). Spread 2.5 mL of cell suspension each on plates containing 0.8% MSS agar (containing no NAA or kinetin), 250 µg/mL timentin (to select against remaining *Agrobacterium* cells), and 100 µg/mL kanamycin (construct-specific antibiotic to select for transgenic calli). As a control for growth ability, spread 2.5 mL on 0.8% MSS agar (without antibiotics). Also spread 2.5 mL of the control-treated culture (*see* **Subheading 3.3.1.**, **step 4**) on MSS agar (without/with antibiotics: positive/negative control). Seal the plates with tape (3M, Micropore™) and incubate at 27°C in darkness (*see* **Note 17**).
5. Selection of putative transgenic calli: After 2 to 3 wk transfer putative transgenic calli carefully with a sterile yellow pipet tip onto new MSS agar plates (250 µg/mL timentin, 100 µg/mL kanamycin, 0.5 mg/L NAA, 0.05 mg/L kinetin) and leave to grow for a further 1 to 2 wk (*see* **Fig. 5**).

3.4.3. Establishment of Transgenic Liquid Cultures

1. Establishment of liquid cultures: Transfer the putative transgenic callus material to a 5-cm Petri dish. Carefully squeeze the soft callus material with a sterile yellow pipet tip and resuspend the cells in 10 mL of fresh MSS medium (supplemented with 100 µg/mL kanamycin) by

pipetting the solution up and down several times. Transfer the callus-derived cell suspensions into 100-mL Erlenmeyer flasks and incubate at 130 rpm, 27°C in the dark for 7 d (week 1).

2. Adjustment of cell growth to produce an inoculum of approx the same cell number as WT: Add an additional 10 mL of MSS medium to the suspension cultures and leave for a further 7 d under the conditions described above (week 2). For the next 2 wk, subculture (at 7-d intervals) 5 mL of each putative transgenic cell suspension into 50 mL of fresh MSS medium (supplemented with 100 µg/mL kanamycin) and leave for a further 7 d under the conditions described above (week 3–4).
3. Maintenance of transgenic-derivative cell lines: Subsequently, subculture weekly 3.5 mL into 100 mL of fresh MSS medium (supplemented with 100 µg/mL kanamycin) in 300-mL Erlenmeyer flasks. Growth should be similar to untransformed control cells.

3.5. Cryopreservation of Transgenic and Wild-Type Arabidopsis Cell Suspension

3.5.1. Cryopreservation and Storage in Liquid Nitrogen

1. Start of cell culture before cryopreservation (day 1): Subculture 5 mL of an early stationary-phase MM1 cell suspension (7 d after previous subculture) into 100 mL of fresh MSS medium. Incubate at 23°C, 120 rpm for a period of 3 d to obtain an exponentially growing culture (*see* **Note 18**).
2. Harvest cells by centrifugation (day 3): Transfer the cell suspension into a 50-mL Falcon polypropylene tube and spin at 387*g* for 1 min without applying brake force (*see* **Note 4**). Discard the supernatant and determine the weight of the cell pellet.
3. Sorbitol preincubation (day 3): Add an appropriate volume of MSS+sorbitol-medium (supplemented with 0.5 *M* D-sorbitol; *see* **Note 19**) to the weight cell pellet to reach a packed-cell volume (PCV) of approx 40% (*see* **Note 20**). Transfer the resuspended cells into a 100-mL Erlenmeyer flask and incubate for a further 2 d (*see* **Note 21**).
4. Cryoprotection with DMSO (day 5): Transfer the flasks to the cold room (4°C; *see* **Note 22**). Pre-cool the cell suspension for 25 min at 4°C (cold room) with rotation at 140 rpm in ambient light (*see* **Note 23**). Add DMSO to a final concentration of 5% (*see* **Note 24**). Incubate the cell culture for 1 h at 4°C (cold room) with rotation at 140 rpm.
5. Two-step freezing procedure (day 5): Transfer 1.8 mL each of the cell suspension into 2 mL of cryo-vials. Place the vials into the controlled rate (1°C/min) cryofreezing container and place the container in a styropore box (*see* **Note 25**). Transfer the cryoprotected cells to a low-temperature freezer and incubate at –80°C (freezing step 1; *see* **Note 26**). After 4 h of incubation at –80°C, remove the vials very quickly from the cryofreezing container to prevent thawing and immediately plunge the vials in the liquid phase of liquid nitrogen (LN, freezing step 2; *see* **Note 27**). Store the cryoprotected cells under LN (–196°C) until further use.

3.5.2. Recovery of Viable Cells and Regeneration of Cell Suspension

1. Rapidly thaw cryoprotected MM1 cells: Place cryovials immediately from LN into a water bath at 40°C for 3 min (*see* **Note 28**).
2. Spread cell suspension onto a nylon membrane to separate the toxic cryoprotectants: Place a sterile nylon membrane on a plate containing 0.8% MSS agar. Spread the thawed culture on the nylon membrane and seal the plate with tape (3M, Micropore). Incubate for 3 h under continuous light conditions (1300 Lux) at 23°C. Carefully transfer the membrane carrying the spread cell suspension onto a fresh 0.8% MSS agar plate (*see* **Note 29**) and incubate for a further 7 d under the conditions described above (*see* **Fig. 6**).
3. Regeneration of cell suspension: Resuspend the recovered cell-callus material in a total of 50 mL fresh MSS medium. Add 25 mL of medium each directly on top of the nylon membrane. Resuspend the calli by carefully pipetting the cells up and down several times to rinse off the cell lawn and transfer the cell suspension into a 100-mL Erlenmeyer flask. Incubate the cell cultures for 7 d under the conditions described above (week 1).
4. Adjustment of cell growth to produce an inoculum of approximately the same cell number as unfrozen control line: Subculture 10 mL of the regenerated cell suspension into 100 mL of fresh MSS medium (in a 300-mL flask) and leave for a further 7 d under the conditions described above (week 2). For the next week, subculture 5 mL into 100 mL fresh MSS medium (week 3).

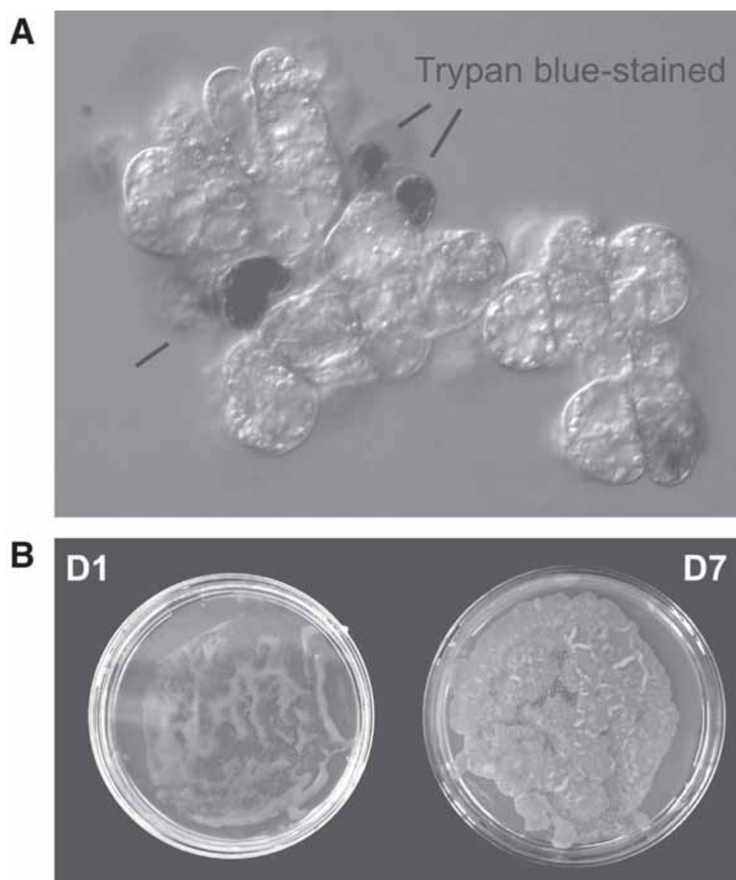


Fig. 6. Post-thaw viability and growth of MM1 cell suspension after cryopreservation. **(A)** Trypan blue viability test to determine the percentage of viable cells by microscopic examination in cryopreseved samples directly post-thaw. Dark blue stained cells indicate dead or dying cells. **(B)** Post-thaw growth of the cryopreserved MM1 cell suspension after transfer to solid medium; the photographs of post-thaw growth at day 1 (D1) and day 7 (D7) clearly demonstrate rapid regrowth ability.

5. Maintenance of regenerated cell lines: Subsequently subculture weekly 3.5 mL into 100 mL fresh MSS medium.

3.5.3. Test of Viability in Regenerated Cell Cultures After Cryopreservation

1. Rapidly thaw cryoprotected cells: Place cryovials immediately from liquid nitrogen into a water bath at 40°C for 3 min.
2. Stain cells with the chemical Trypan blue: Transfer an aliquot of 100 µL into a 1.5-mL microfuge tube and add 100 µL of a 0.4% Trypan blue solution (*see* **Note 30**).
3. Microscopic examination: Count the percentage of blue-stained (dead) cells as a proportion of the total number of cells (*see* **Fig. 6**). Score at least 1000 cells.

4. Notes

1. A suspension culture of the fast-growing *Arabidopsis* cell lines MM1 and MM2d ***(8)*** were previously selected from a cell suspension originally produced from Landsberg *erecta* stem explants by May and Leaver ***(22)***.

2. If necessary, dilute the cell suspension with PBS buffer prior to cell counting to determine the cell density (cell no/mL). Remember to take into account the possible dilution factor.
3. Stage of culture prior to start of starvation period: The same procedure could be followed by using an early stationary-phase cell suspension (7 d after previous subculture) or the alternative cell line MM1.
4. General handling advice while working with cell suspension: Throughout the synchronization procedure try to treat the cell suspension as gently as possible and avoid applying too much physical disturbance. It is therefore very important to ensure that the brake is switched off during the centrifugation steps to minimize the physical stress.
5. Aphidicolin-induced synchrony is equally effective using the alternative cell line MM1 or transgenic derivatives ***(8,29)*** in which case all steps are carried out under the conditions described for culturing MM1 cells.
6. The level of achieved synchrony is similar after applying aphidicolin for a time period ranging between 20 and 24 h.
7. Importance of limiting the overall time taken to 15 min during washing procedure: After washing to release the aphidicolin-induced cell-cycle arrest in late G1/early S phase, cells start immediately with progression through S phase.
8. General advice while working with BrdU: The labeling reagent contains 5-bromo-2'-deoxyuridine and 5-fluoro-2'deoxyuridine in dilute solution. 5-bromo-2'-deoxyuridine and 5-fluoro-2'deoxyuridine are classified as toxic, teratogenic, and mutagenic when undiluted and should be considered as potentially hazardous.
9. Use of 25 min enzyme treatment to effectively break open cell walls: For the BrdU-labeling procedure using *Arabidopsis* cells, the enzyme digestion is the most critical step. Preliminary time-course experiments have shown that the enzymatic digestion of the cell wall for a period of 25 min at RT is necessary to ensure the best antibody probing. The cell wall must be partially digested to allow the antibody to enter the cells. On the other hand, to ensure that cells attach to the Vectabond-treated glass slides, it is also important to avoid the complete removal of the cell wall.
10. Importance of cell attachment to a Vectabond-treated glass slide while settling down overnight: It is important to ensure that cells are effectively attached to the precoated microscope slides and that such attached cells are handled with care during the following washing procedure. To avoid high levels of cell loss while washing with PBS buffer, wash cells carefully directly on the microscope slides without too much physical disturbance.
11. It is very important to wash the *Agrobacterium* culture with MSS medium prior to cocultivation to ensure that the antibiotic (used as an *Agrobacterium* strain-specific selection) is sufficiently removed and does not affect the growth of the *Arabidopsis* cell suspension.
12. Acetosyringone is a phenolic compound produced by wounded plant tissues, and stimulates the transfer of T-DNA from *Agrobacterium* to plants. The addition of acetosyringone during the cocultivation step enhances the transformation efficiency and results in an increased number of transgenic calli after selection. During the 2-d cocultivation period, the transition to a coagulated cell suspension should be observed, an indication of likely higher transformation activity.
13. Importance of effective washing procedures: One of the major problems during the procedures described is the effective removal of the *Agrobacterium* by washing without applying too much physical stress to the cells that would prevent or inhibit further growth. In some experiments, overgrowth of *Agrobacterium* cells may prevent the further regeneration of transgenic calli, indicating that the washing procedure was not sufficient to remove the *Agrobacterium* cells after the cocultivation period. In other experiments, no growth may be observed at all, indicating that the procedure used was either too stressful or toxic, preventing further growth. A compromise condition is needed between vigorous washing and the recovery of viable cells. Washing the cells after cocultivation using gentle and stress-free conditions, such as repeated centrifugation (three times, 387*g*, without applying the brake force) and gentle resuspension of cells between centrifugation steps by carefully inverting the tubes has been found to be most effective.

14. The thorough removal of *Agrobacterium* cells is one key problem in the regeneration procedure after transformation. The clumpiness of *Arabidopsis* cell cultures compared with BY-2 cells increases the difficulty of thorough removal of *Agrobacterium* cells by washing alone. The addition of timentin after gentle washing resolves this problem. Timentin is a 15:1 mix of tricarcillin (a penicillin derivative) and the β-lactamase inhibitor clavulanic acid, and is inhibitory to the growth of *Agrobacterium* without having a toxic effect on plant cells. To prevent entirely the growth of *Agrobacterium*, 500 μ*M* timentin is added directly after washing and during the first incubation steps on solid medium.
15. Initial cell growth, and therefore increased cell density, is achieved by incubating the washed cell suspension after cocultivation for a further 3 d in liquid culture (containing 500 μ*M* timentin) before the start of the selection procedure for transgenic calli with kanamycin. This is mainly done to increase the transgenic cell population after transformation and therefore the chance of recovering viable calli. The disadvantage of introducing this additional incubation period without applying kanamycin for selection is the possibility that multiple putative transgenic calli for further analysis might have been derived from a single original transformed cell. However, incubation for a further 3 d in MSS medium or conditioned medium, which is harvested from a cell suspension in the same growth stage, together with the addition of 500 μ*M* timentin, results in optimized growth on selective medium and seems to be a necessary and critical step for the improvement of the regeneration procedure. In any case it should be noted that only one clone should be chosen from each transformation to ensure independent origins of transformation events.
16. Addition of hormones directly to cell suspension to boost initial growth: Direct addition of NAA and kinetin to the resuspended and washed cell suspension before spreading on selective medium enhances growth. The hormone treatment remarkably increases callus growth and results in a lawn of cell growth after a short cultivation period on solid medium. Initially an apparent lawn of growing cells is seen, which subsequently resolves into distinct calli. A dense layer of fast-growing cells clearly increases the chance of regenerating viable putative transgenic callus material after the untransformed cells have died due to kanamycin selection. In contrast, poor growth and eventual death of putative transformed calli is observed on plates where the spread cell density is low. The direct addition of hormones to the washed, resuspended cell suspension, compared with addition of the similar concentration of hormones to the solid medium, represents one key step in the improvement and optimization of the regeneration procedure to select for viable transgenic calli.
17. Use of untransformed control-treated cultures and further controls for growth ability: To ensure that the cell suspension in general survives the applied procedures, it is advisable to always include control-treated cultures that have not been infected with *Agrobacterium* but have undergone every other treatment and handling step. The positive control (control-treated culture) should grow on MSS agar (without antibiotic selection). If, however, no growth is observed, this indicates that the procedure was too stressful. The negative control (control-treated culture) should not grow on MSS agar (supplemented with timentin and kanamycin). If growth is observed, the MSS agar most likely contains no antibiotic selection. As an additional control, spread the putative transformed cell culture on MSS agar (without antibiotic selection) to observe cells' general growth ability. If no growth is observed on such control plates, again the procedure was most likely too stressful, therefore preventing further growth.
18. Exponentially growing cells that have undergone active division are small and display a dense cytoplasm (small vacuole, relatively small water content). These cells are reported to survive the freezing-thawing procedure better than large, more highly vacuolated cells ***(39)***.
19. Sorbitol is reported to be the best cryoprotectant among other sugars tested ***(40)***. Osmotica such as sorbitol act to protect cells from freezing injury by reducing the cellular water content (shrinkage). In general, the use of a combination of two or more cryoprotectants is advisable, as the concentration of each cryoprotectant, and therefore the toxic effect on the cells, can be reduced.
20. *Packed-cell volume (PCV):* A minimum cell density is required during the post-thaw procedure for survival and regrowth ***(35,38,41–43)***. There are various reports about the concentration of the cell suspension prior to cryoprotection in a range of 10 to 60% PCV. To reach a PCV of 40% (w/v, 40 g of cell pellet in a total volume of 100 mL), determine the weight of the

cell pellet and calculate the volume of MSS-medium (supplemented with D-sorbitol) that must be added accordingly.

21. Prior to further cryoprotection and freezing, it is important that cells are grown under conditions where they develop a maximum content of cytoplasm and minimum cell size (shrinkage). This can be achieved by preincubation of cells in media of enhanced osmotic potential (such as MSS medium, supplemented with sorbitol) for 1 to 2 d ***(40)***.
22. The compounds used as a second cryoprotectant, such as DMSO, are often added at ice temperature (0°C, or ice water). The rationale is to preserve the stability of vital cellular systems by reducing the metabolic activity and the toxic effect of the cryoprotectant ***(44)***. Based on practical considerations, a simple shaker can be used in the cold room (4°C). Precooling the cells for 25 min at 4°C is needed to achieve an equilibrium temperature.
23. For the cultivation of dark-grown MM2d cell lines, Erlenmeyer flasks can be covered with aluminium foil.
24. DMSO is a commonly used compound for cryopreservation utilized at a wide range of concentrations ***(40)***. DMSO is reported to enter the cells (penetrating cryoprotectant) to reduce the freeze-induced cellular dehydration. In comparison, sorbitol, a nonpenetrating cryoprotectant, reduces the cellular water and therefore may reduce the rate of initial ice crystallization during freezing.
25. There is no need to purchase any expensive equipment, such as an electronic controlled-freezing apparatus. The cryopreservation procedure can be performed in any lab, provided that it has a low-temperature freezer (–70 or –80°C) and a simple passive controlled cooling jacket.
26. During the initial stages of freezing, a slower freezing rate of the extracellular solution reduces the stress on the plasma membrane. Additionally, dehydration by extracellular freezing, thereby minimizing intracellular ice formation, can be induced by slow freezing which results in active protection against cryoinjury. Such partially dehydrated cells are reported to be able to survive the liquid nitrogen (LN) temperature ***(38,40,42,45)***.
27. Transfer temperatures are reported in literature in a range of –40°C to –100°C (***36,39,40,46,47***). For convenience, a transfer temperature of –80°C is used in this protocol to prevent cells in the vials from reaching the critical thawing temperature during the unavoidable handling time (such as removing the container from the freezer, and transferring to LN container).
28. Cell survival is reported to be best at the highest thawing rate (120°C/min) ***(36,39,43,48)***.
29. Post-thaw washing of cells decreases the viability of frozen-thawed cells due to a rapid change of the osmotic pressure that can result in deplasmolysis injury. Cells are therefore spread on a nylon membrane to separate the toxic cryoprotectants in solution after the transfer onto a fresh agar plate with minimum physical disturbance (after the 3-h incubation period). The use of a nylon membrane first ensures a good contact between the frozen-thawed cells and the solidified medium and, second aids the diffusion of cryoprotectants from the recovery cell suspension without the need to remove any supernatant beforehand by pipetting or centrifugation. Furthermore, it is reported that a minimum effective initial cell density is required for growth (*see also* **Note 20**), as freshly frozen cells do not survive at an extremely low density (a problem if the aim is to directly start a liquid culture). Another problem with liquid cultures is that cells might be diluted too rapidly and die, due to rehydration or desplasmolysis injury. Therefore a compromise is necessary that minimizes the time cells are exposed to the toxic level of cryoprotectants. This is achieved efficiently by the diffusion of cryoprotectants through a nylon net into the agar, followed by a transfer onto a new agar plate.
30. The chemical Trypan blue will stain dead or dying cells; viable cells repel the dye and therefore do not stain. The use of this stain, however, is time-sensitive. Viable cells do absorb Trypan blue over time, and this can affect counting and hence viability results; therefore samples should be stained and counted one at a time.

Acknowledgments

The authors are very grateful to Dr. Graham Armstrong for critical reading of the manuscript. We thank Prof. Keith Lindsey for the original gift of the cell suspension from which the MM1 and MM2d cell lines were derived, Dr. Schumacher of the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) for information on cryopreservation of

plant cells, and Dr. Kathleen D'Halluin and Dr. Bart den Boer for advice on the transformation protocol. This work was financially supported by grant 8/C15792 from the UK Biotechnology and Biological Sciences Research Council (BBSRC).

References

1. Gould, A. R. (1984) Control of the cell cycle in cultured plant cells. *CRC Crit. Rev. Plant Sci.* **1,** 315–344.
2. Amino, S., Fujimura, T., and Komamine, A. (1983) Synchrony induced by double phosphate starvation in a suspension-culture of *Catharanthus roseus*. *Physiol. Plant.* **59,** 393–396.
3. King, P. J., Mansfield, K. J., and Street, H. E. (1973) Control of growth and cell division in plant cell suspension cultures. *Can. J. Bot.* **51,** 1807–1823.
4. Kodama, H., Ito, M., Hattori, T., Nakamura, K., and Komamine, A. (1991) Isolation of genes that are preferentially expressed at the G1/S boundary during the cell cycle in synchronized cultures of *Catharanthus roseus* cells. *Plant Physiol.* **95,** 406–411.
5. Nishida, T., Ohnishi, N., Kodama, H., and Komamine, A. (1992) Establishment of synchrony by starvation and readdition of auxin in suspension-cultures of *Catharanthus roseus* cells. *Plant Cell Tissue Organ Cult.* **28,** 37–43.
6. Riou-Khamlichi, C., Huntley, R., Jacquard, A., and Murray, J. A. H. (1999) Cytokinin activation of *Arabidopsis* cell division through a D-type cyclin. *Science* **283,** 1541–1544.
7. Riou-Khamlichi, C., Menges, M., Healy, J. M., and Murray, J. A. H. (2000) Sugar control of the plant cell cycle: differential regulation of *Arabidopsis* D-type cyclin gene expression. *Mol. Cell. Biol.* **20,** 4513–4521.
8. Menges, M. and Murray, J. A. H. (2002) Synchronous *Arabidopsis* suspension cultures for analysis of cell-cycle gene activity. *Plant J.* **30,** 203–212.
9. Fukuda, H., Ito, M., Sugiyama, M., and Komamine, A. (1994) Mechanisms of the proliferation and differentiation of plant cells in cell culture systems. *Int. J. Dev Biol.* **38,** 287–299.
10. Glab, N., Labidi, B., Qin, L. X., Trehin, C., Bergounioux, C., and Meijer, L. (1994) Olomoucine, an inhibitor of the cdc2/cdk2 kinases activity, blocks plant cells at the G1 to S and G2 to M cell cycle transitions. *FEBS Lett.* **353,** 207–211.
11. Ito, M., Marie-Claire, C., Sakabe, M., et al. (1997) Cell-cycle-regulated transcription of A- and B-type plant cyclin genes in synchronous cultures. *Plant J.* **11,** 983–992.
12. Magyar, Z., Bako, L., Bogre, L., Dedeoglu, D., Kapros, T., and Dudits, D. (1993) Active cdc2 genes and cell-cycle phase-specific cdc2-related kinase complexes in hormone-stimulated alfalfa cells. *Plant J.* **4,** 151–161.
13. Nagata, T., Nemoto, Y., and Hasezawa, S. (1992) Tobacco BY-2 cell line as the "HeLa" cell in the cell biology of higher plants. *Int. Rev. Cytol.* **132**, 1–30.
14. Perennes, C., Qin, L. X., Glab, N., and Bergounioux, C. (1993) Petunia p34 (cdc2) protein kinase activity in G2/M cells obtained with a reversible cell cycle inhibitor, mimosine. *FEBS Lett.* **333,** 141–145.
15. Planchais, S., Glab, N., Inze, D., and Bergounioux, C. (2000) Chemical inhibitors: a tool for plant cell cycle studies. *FEBS Lett.* **476,** 78–83.
16. Roudier, F., Fedorova, E., Gyorgyey, J., Feher, A., Brown, S., Kondorosi, A., and Kondorosi, E. (2000) Cell cycle function of a *Medicago sativa* A2-type cyclin interacting with a PSTAIRE-type cyclin-dependent kinase and a retinoblastoma protein. *Plant J.* **23,** 73–83.
17. Breyne, P., Dreesen, R., Vandepoele, K., et al. (2002) Transcriptome analysis during cell division in plants. *Proc. Natl. Acad. Sci. USA* **99,** 14,825–14,830.
18. Axelos, M., Curie, C., Mazzolini, L., Bardet, C., and Lescure, B. (1992) A protocol for transient gene-expression in *Arabidopsis thaliana* protoplasts isolated from cell-suspension cultures. *Plant Physiol. Biochem.* **30,** 123–128.
19. Callard, D. and Mazzolini, L. (1997) Identification of proliferation-induced genes in *Arabidopsis thaliana*—characterization of a new member of the highly evolutionarily conserved histone H2A.F/Z variant subfamily. *Plant Physiol.* **115,** 1385–1395.
20. Mathur, J., Szabados, L., Schaefer, S., et al. (1998) Gene identification with sequenced T-DNA tags generated by transformation of *Arabidopsis* cell suspension. *Plant J.* **13,** 707–716.
21. Mathur, J. and Koncz, C. (1998) Establishment and maintenance of cell suspension cultures. *Methods Mol. Biol.* **82,** 27–30.
22. May, M. J. and Leaver, C. J. (1993) Oxidative stimulation of glutathione synthesis in *Arabidopsis thaliana* suspension cultures. *Plant Physiol.* **103,** 621–627.

23. Fuerst, R. A. U. A., Soni, R., Murray, J. A. H., and Lindsey, K. (1996) Modulation of cyclin transcript levels in cultured cells of *Arabidopsis thaliana*. *Plant Physiol.* **112,** 1023–1033.
24. Breyne, P. and Zabeau, M. (2001) Genome-wide expression analysis of plant cell cycle modulated genes. *Curr. Opin. Plant Biol.* **4,** 136–142.
25. Richmond, T. and Somerville, S. (2000) Chasing the dream: plant EST microarrays. *Curr. Opin. Plant Biol.* **3,** 108–116.
26. Stals, H., Casteels, P., Van Montagu, M., and Inze, D. (2000) Regulation of cyclin-dependent kinases in *Arabidopsis thaliana*. *Plant Mol. Biol.* **43,** 583–593.
27. Menges, M., Hennig, L., Gruissem, W., and Murray, J. A. H. (2002) Cell cycle-regulated gene expression in *Arabidopsis*. *J. Biol. Chem.* **277,** 41,987–42,002.
28. Menges, M., Hennig, L., Gruissem, W., and Murray, J. A. H. (2003) Genome-wide gene expression in an *Arabidopsis* cell suspension. *Plant Mol. Biol.* **53,** 423–442.
29. Menges, M. and Murray, J. A. H. (2004) Cryopreservation of transformed and wild-type *Arabidopsis* and tobacco cell suspension cultures. *Plant J.* **37,** 635–644.
30. Klein, T. M., Harper, E. C., Svab, Z., Sanford, J. C., Fromm, M. E., and Maliga, P. (1988) Stable genetic-transformation of intact *Nicotiana* cells by the particle bombardment process. *Proc. Natl. Acad. Sci. USA* **85,** 8502–8505.
31. Potrykus, I. (1991) Gene-transfer to plants—assessment of published approaches and results. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42,** 205–225.
32. Seki, M., Iida, A., and Morikawa, H. (1998) Transient expression of foreign genes in tissues of *Arabidopsis thaliana* by bombardment-mediated transformation. *Methods Mol. Biol.* **82,** 219–225.
33. Forreiter, C., Kirschner, M., and Nover, L. (1997) Stable transformation of an *Arabidopsis* cell suspension culture with firefly luciferase providing a cellular system for analysis of chaperone activity in vivo. *Plant Cell* **9,** 2171-2181.
34. Gallego, M. E., Sirand-Pugnet, P., and White, C. I. (1999) Positive-negative selection and T-DNA stability in *Arabidopsis* transformation. *Plant Mol. Biol.* **39,** 83–93.
35. Jain, S., Jain, R. K., and Wu, R. (1996) A simple and efficient procedure for cryopreservation of embryogenic cells of aromatic Indica rice varieties. *Plant Cell Rep.* **15,** 712–717.
36. Jekkel, Z., Heszky, L. E., and Ali, A. H. (1989) Effect of different cryoprotectants and transfer temperatures on the survival rate of hemp (*Cannabis sativa* L.) cell suspension in deep freezing. *Acta Biol. Hung.* **40,** 127–136.
37. Kim, S. I., Choi, H. K., Son, J. S., et al. (2001) Cryopreservation of *Taxus chinensis* suspension cell cultures. *Cryo-Letters* **22,** 43–50.
38. Maddox, A. D., Gonsalves, F., and Shields, R. (1983) Successful preservation of suspension-cultures of 3 *Nicotiana* species at the temperature of liquid-nitrogen. *Plant Sci. Lett.* **28,** 157–162.
39. Withers, L. A. and Street, H. E. (1977) Freeze preservation of cultured plant cells. III. The pregrowth phase. *Physiol. Plant.* **39,** 171–178.
40. Chen, T. H. H., Kartha, K. K., Constabel, F., and Gusta, L. V. (1984) Freezing characteristics of cultured *Catharanthus roseus* (L) Don, G. cells treated with dimethylsulfoxide and sorbitol in relation to cryopreservation. *Plant Physiol.* **75,** 720–725.
41. Benson, E. E., Lynch, P. T., and Jones, J. (1992) The detection of lipid-peroxidation products in cryoprotected and frozen rice cells—consequences for postthaw survival. *Plant Sci.* **85,** 107–114.
42. Chen, T. H. H., Kartha, K. K., Leung, N. L., Kurz, W. G. W., Chatson, K. B., and Constabel, F. (1984) Cryopreservation of alkaloid-producing cell-cultures of periwinkle (*Catharanthus roseus*). *Plant Physiol.* **75,** 726–731.
43. Lynch, P. T., Benson, E. E., Jones, J., Cocking, E. C., Power, J. B., and Davey, M. R. (1994) Rice cell cryopreservation—the influence of culture methods and the embryogenic potential of cell-suspensions on postthaw recovery. *Plant Sci.* **98,** 185–192.
44. Finkle, B. J. and Ulrich, J. M. (1982) Cryoprotectant removal temperature as a factor in the survival of frozen rice and sugarcane cells. *Cryobiology* **19,** 329–335.
45. Withers, L. A. and King, P. J. (1979) Proline: a novel cryoprotectant for the freeze preservation of cultured cells of *Zea mays* L. *Plant Physiol.* **64,** 675–678.
46. Nag, K. K. and Street, H. E. (1973) Carrot embryogenesis from frozen cultured cells. *Nature* **245,** 270–272.
47. Nag, K. K. and Street, H. E. (1975) Freeze preservation of cultured plant cells. I. The pretreatment phase. *Physiol. Plant.* **34,** 254–260.
48. Nag, K. K. and Street, H. E. (1975) Freeze preservation of cultured plant cells. II. The freezing and thawing phase. *Physiol. Plant.* **34,** 261–265.

6

Genetic Analysis

Maarten Koornneef, Carlos Alonso-Blanco, and Piet Stam

Summary

The Mendelian analysis of genetic variation, available as induced mutants or as natural variation, requires a number of steps that are described in this chapter. These include the determination of the number of genes involved in the observed trait's variation, the determination of dominance relationships between alleles of the same locus, and epistatic interactions with related genetic variants. A new variant should be compared with previously identified genetic variants, which is most efficiently done by allelism tests in case of recessive mutants. In addition the locus should be localized on the genetic and physical map by linkage analysis. The mapping of mutant loci in *Arabidopsis* is facilitated by the genomic resources available in *Arabidopsis* and often consists of two steps, a crude and a fine mapping, the latter enabling the approximate location of a variant on the sequence map of *Arabidopsis*.

Key Words: *Arabidopsis*; genetic analysis; mutants; gene mapping.

1. Introduction

The genetic behavior of *Arabidopsis* is not different from that of other diploid organisms; therefore, one can find the topics of this chapter in almost any handbook of genetics. It is only because of the specific (reproductive) biology of a species that sometimes a certain analytical procedure is favored over another. For *Arabidopsis* the facts that it is a self-fertilizing species and that it has relatively small flowers are important in this respect.

A classical genetic analysis of a certain trait implies that observed genetic variation for such a trait is described in terms of certain genetic parameters. Genetic variation, the starting material for such an analysis, can be induced by mutagenic treatments or can be found as "natural" variation present among wild accessions or ecotypes.

The analysis of genetic variation should answer the following questions:

- How many genes control the observed variation?
- Which are the dominance relationships of the various alleles?
- Is the gene (as a genetic entity) allelic to previously described genes?
- Which are the epistatic relationships with other genes in the same pathway?
- Where is the gene located on the linkage map?

At the moment genetic variation has been identified (e.g., a mutant with a specific phenotype), the variant should be crossed with its wild type. The resulting F1 hybrid is used to determine if either the wild type or the variant is dominant. In addition, it can be used to obtain the F2 generation by harvesting (selfed) seeds from the F1 plant. The detailed scoring of individual F2 plants for the trait of interest and the analysis of the obtained numbers of each phenotypic class allow the determination of the mono-, di- or multigenic nature of the variant phenotype that one identified. In a case in which the mutant has a complicated phenotype—that is, several

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

characters seemed to be changed simultaneously—the segregation analysis furthermore indicates which of these changed properties are due to mutations in the same gene (pleiotropic) and which are due to mutations in other genes. The F1 plants can also be crossed (back) to the original variant to give the backcross (BC) generation, but in *Arabidopsis* often only an F2 analysis is performed. From the F2 generation a number of plants with the variant phenotype (and without other visible mutations) are selected and selfed and the individual F3 progenies are grown to be retested for being "true type" and for the absence of segregation of other unlinked mutations. Preferentially this backcross-and-reselection procedure is repeated several times so that a "clean" mutant line showing monogenic inheritance is obtained. If the mutant resembles a mutant phenotype isolated or described before, it should be crossed with such mutants to test for allelism with these "old" mutants. In many situations several loci are identified that are involved in the same process or pathway. The phenotype of the double mutants is often indicative of the interaction between the two respective genes. To identify these double mutants in segregating generations several precautions should be taken, as the double mutant phenotype is not known *a priori*.

A next step in a genetic analysis is the location of the new gene (as identified by its mutant phenotype) on the genetic map of *Arabidopsis*. Placing a gene on the map based on its mutan phenotype has changed with the availability of the complete physical and sequence maps of *Arabidopsis* (*1*). Gene mapping now consists of locating a gene on the physical map as accurately as possible in order to identify the DNA sequence. Several recent reviews (*2–4*) describe this process of efficient map-based cloning. Compared with classical linkage analysis based on a genome-wide screen for linkage, where often crude map positions resulted, much more accurate positions are obtained using larger populations genotyped with many markers in the vicinity of the gene. Mapping is mostly divided into two steps. Phase 1 is the determination of a crude map position using a small-sized population and a limited number of markers, making use of efficient multimarker systems such as amplified fragment length polymorphism (AFLP) or multiplex single-nucleotide polymorphism (SNP) or microsatellite techniques (*see* Chapter 9 in this book). This step provides a set of markers between which the gene is located. Hereafter, fine or high-resolution mapping is performed on larger populations (up to several thousands of F2 plants) and recombinants between the markers surrounding the locus are preselected. The obtained recombinants are then analyzed with new markers between the flanking markers until the closest recombination event on both sides of the gene will delimit the position of the locus between those markers (*see* Chapter 10 in this book). Therefore, mapping a mutant on the physical map (of Columbia [-Col-] strain) does not require building a genome-wide genetic linkage map of *Arabidopsis*, making use of multiple recombination estimates. However, this construction of genetic maps is done in *Arabidopsis* when new segregating populations are generated for the purpose of mapping polygenic traits. This especially applies to the analysis of quantitative trait in immortal mapping populations such as recombinant inbred lines (RILs) (*see* Chapter 7 in this book).

For linkage analyses the mutant (backcrossed and reselected) must be crossed with a genotype that differs from the original mutant for several previously mapped loci. Molecular marker loci (such as RFLPs, RAPDs, SSCPs, ARMSs, CAPSs, AFLPs, etc.; *see* Chapters 9, 10, and 12), are preferred nowadays because they are abundantly available among wild accessions or ecotypes of *Arabidopsis*. Such a cross often involves one or both of the accessions Col and L*er*, because many mutants are obtained in either of these genetic backgrounds and because many polymorphisms between these strains are known. A complication may arise when the genetic background affects the expression of the traits. This is rather common, especially in traits in which a quantitative difference is analyzed as a qualitative trait—for instance, flowering time—and natural variation for several genes exist, which will segregate in almost any cross among wild accessions. A solution is either to cross with near isogenic lines (NILs; which need to be developed) or the selection of heterogeneous inbred families (HIFs), e.g., selected as F3 lines, that show a clean monogenic segregation.

Genes are said to be genetically linked when they are located on the same chromosome arm region. For genes located far apart on the same chromosome their chromosomal assignment and/or position is inferred from their linkage to one or several other genes on the same chromosome. Linkage of two genes will be detected by a deviation from independent segregation in any type of segregating population such as an F2, a backcross, a set of RILs, and so on. The detailed map position of a locus to be mapped relies on the identification of a locus (or loci) with which it shows a close (preferentially almost absolute) linkage.

The quantification of genetic linkage between two loci is based on the estimation of the fraction of recombinant (with respect to the parental) gametes originating from a meiosis of a double heterozygote for the loci under study. Recombination between these two loci is a direct consequence of crossing over, the recombinant frequency (r) being the basis of the genetic distance. Genetic distances between loci are given as map distances (D), usually expressed in centiMorgans (cM). Genetic distance is defined as the average number of crossover events between two loci during a single meiosis and per chromatid. Recombinant frequency increases with map distance but a doubling of map distance does not result in a doubling of the recombinant frequency. This is because of the fact that a second crossover between two loci may or may not involve the same chromatids of a given chromosome pair. A second crossover involving the same chromatid "cancels" the effect of the first one and therefore gives rise to a parental-type gamete. The mathematical relationship between map distance and recombinant frequency is given by the genetic mapping function. Its form depends on the degree of interference in crossing over. In the absence of interference the recombination events in adjacent regions are independent, i.e., a crossover at a given position does not influence the probability of another crossover in its neighborhood.

When estimates of the map distances between at least three genes are available, these loci can be ordered in a linear linkage group. In most cases the order of genes that segregate in the same population can be determined unambiguously both from the estimates of the recombination percentages and from the presence or absence of joint recombination of the different markers. However, the order of genes that are closely linked often cannot be determined with certainty. When recombination data originate from different populations, there may be several reasons for this unresolvability, i.e., (1) the criterion of joint recombination cannot be applied; (2) standard deviations of the estimates of r are often relatively high; and (3) variations in the recombinant frequency and the distribution of recombination events along chromosomes can occur among different mapping populations.

2. Materials

1. Pair of sharp-pointed tweezers.
2. Stereo microscope.
3. Labels for plants and crosses (e.g., colored threads).
4. The growing conditions are standard (*see* Chapter 1 in this book) but should allow the handling of individual plants (for making crosses). Specific mutants may need selective growing conditions to be able to score the various phenotypes.
5. Systems for the analysis of molecular markers.
6. Computers and specific software to estimate recombinant frequencies (e.g., LINKAGE ***(5)***, etc.) and to construct linkage maps, such as MAPMAKER ***(6)***, Joinmap ***(7)***, or GMENDEL ***(8)***.

3. Methods

3.1. Crossing Arabidopsis

1. Select the parents for the cross. These can be plants that just started flowering (*see* **Notes 1** and **2**).
2. Remove from the female parent all young flower buds and flowers with petals from the main inflorescence until 3 or 4 large buds (white petals not yet visible) are left (*see* **Note 3**).
3. Open the buds carefully and remove the six anthers without damaging the pistil. This emasculation should be performed under low magnification on a stereo microscope (*see* **Note 3**).

4. When all buds have been emasculated, carefully check that no anthers are left.
5. Take a fresh fully open flower from the male parent and "brush" the stigma with this flower so that the pollen is visible on the stigma as a yellow powder (*see* **Note 3**).
6. Mark the group of flower buds with a label attached to the stem just under the buds, which indicates the cross and the parent that was used as female parent, and take note of the crossing date (*see* **Note 3**).
7. Three to 4-d later the pistil should have developed as a young silique indicating that the cross was successful.
8. Seeds should be harvested when the siliques are brown and have not dehisced. This is 2 to 3 wk after pollination, depending on the growing conditions.

3.2. Segregation Analysis

The analysis of the genetic segregation of the trait of interest in the studied population (usually an F2) will allow the determination of the number of genes that segregates, and in addition, will give information on the dominance relationship between alleles.

1. To obtain F2 generations, the F1 seeds should be planted, checked whether they are the result of unintentional selfing, and harvested individually (*see* **Note 4**).
2. The phenotype of the F1 already gives an indication about dominance, which will be confirmed by the type of segregation in the subsequent generations (*see* **Note 5**).
3. The segregating population should be large enough to be able to distinguish between the different segregation ratios that one theoretically may expect (*see* **Note 5**).
4. Score each individual plant of the population taking specific precautions to allow the identification of all possible genotypes.
5. Determine the number of plants with each phenotype and test whether the data are not in conflict with the expected ratio for the simplest genetic explanation (e.g., 3:1 in case of a single gene controlling the trait) using a chi-square test for goodness of fit (*see* **Note 5**).
6. If the chi-square test indicates that the observed data are not in agreement with this simple genetic explanation, test if the obtained data agree with a digenic genetic model, and so on (*see* **Note 5**).
7. In order to identify the F2 individuals that are either homozygous or heterozygous, progeny testing of individual F2 plants with a wild-type phenotype can be performed (*see* **Note 6**).

3.3. Allelism Tests

Allelism tests will tell whether the new mutant allele corresponds to (is allelic with) an already known gene or whether it represents a new locus.

1. Cross the new mutant with all available independent mutants with a similar phenotype.
2. Compare the phenotype of the F1 hybrids with that of the wild-type and the parental mutants. If two recessive allelic mutants are crossed their F1 hybrid has a mutant phenotype: *a-1/a-1* × *a-2/a-2* → *a-1/a-2* (mutant phenotype).
   When two non-allelic mutants are crossed, the F1 hybrid has a nonmutant phenotype; that is, the two mutants complement each other: *a/a B/B* × *A/A b/b* → *A/a B/b* (wild type phenotype) (*see* **Note 7**).
3. In the case of dominant alleles, the F1 hybrid of two homozygous parents is not informative, and one must self this hybrid or backcross it with a recessive genotype to find, in case of nonallelism, recessive (wild-type) phenotypes in the progeny (*see* **Note 8**).

3.4. Isolation of Double Mutants to Analyze Epistatic Relationships

The isolation of double mutants is usually started from an F2 population segregating for two genes controlling similar or related traits. The difficulty of identifying the double mutant depends on its phenotype, which is not known *a priori*. The double mutant phenotype can be similar to one of the parents, it may be a novel phenotype, or it may even resemble the wild type. Additional work will be required to isolate it (larger number of plants to analyze) when both mutations are genetically linked.

1. When, after crossing two recessive mutants, a novel (often more extreme) phenotype appears in the F2 generation with a frequency of approx $\frac{1}{16}$, this most likely is the double recessive. When the mutants are not recessive, other frequencies such as $\frac{9}{16}$ (in case of two dominant mutants) and $\frac{3}{16}$ (in case of one recessive and one dominant mutant) can be expected for the novel phenotype.
2. Cross the putative double mutants with both parental mutants and check for the absence of complementation in the hybrids. In case of dominant mutants, see the previous section.
3. In case of complete epistasis the phenotype of one mutant is not visible in the double mutant and one must take specific precautions to detect this genotype.
   a. When the single recessive mutants have a distinguishable phenotype, plants with either one or the other mutant phenotype can be identified in the F2 and the corresponding F3 lines can be checked for segregation of the second phenotype (*see* **Note 9**). Allelism tests with the parent mutants should confirm the genotype of the double mutants identified in such F3 lines.
   b. If both recessive mutants cannot be distinguished phenotypically and no novel phenotypes are observed in an F2, this usually results in a 9:7 segregation in an F2. Test crossing of F2 plants with mutant phenotype to both parental mutants will be necessary to identify the double mutant (*see* **Note 10**).

3.5. Gene Mapping

3.5.1. Genetic Mapping Up to Chromosome Arm Assignment

3.5.1.1. Linkage Analysis

Linkage analysis is performed to detect which loci are linked. By identifying markers linked to the mutant locus, the latter can be assigned to a particular chromosome arm region.

1. Make an F1, which is a multiple heterozygote for the markers to be analyzed, and derive from this F1 a segregating F2 or back-cross (BC) generation. Parents of this F1 are the mutant and an accession, which differs from the accession that is the background of the mutant. Generating different F2s that can function as a backup mapping population when complications due to lack of polymorphism, local suppression of recombination, or modifiers of the mutant that are present in the opposite parent might be useful (*see* **Note 11**).
2. Score all individual plants for the mutant/wild-type phenotype (*see* **Note 12**).
3. Genotype 25 to 100 plants or bulks (*see* **Note 13**) of these with a set of 20 to 30 markers evenly spaced in such a way that all 10 *Arabidopsis* chromosome arms are well covered or use a multilocus marker test system such as AFLP™ (*see* **Note 14**)
4. Determine which markers show linkage to the mutant locus. These markers will show in the mutant plants of the segregating population, predominantly the same allele of the genetic background of the mutant (*see* **Note 14**).
5. Test the independent segregation of the mutant locus and the markers by performing the chi-square test for the contingency tables derived for all pair combinations. Alternatively, the likelihood of observed data (LOD) score associated to the recombination fraction can be calculated for each pair combination (*see* **Notes 15–17**).

3.5.1.2. Estimation of Recombination Frequencies and Map Distances

Subsequently, the frequency of recombinant gametes is estimated from the observed segregation ratios; these estimates serve as basis to establish the linear order of markers and the construction of genetic linkage maps.

1. Estimation is usually done by maximum likelihood (ML). Apart from the simple backcross population, where recombinants can be counted directly, computer programs are used for this purpose (LINKAGE). In addition to the estimate itself, the standard error (SE) of the estimate and/or the LOD score is produced. Both SE and LOD score can be seen as measures of “linkage information” (*see* **Notes 17–20**).

2. Recombination estimates and their standard errors can be converted into genetic map distances. Usually either Haldane's (assuming absence of interference) or Kosambi's (assuming interference) mapping functions are used (*see* **Note 21**).

3.5.1.3. Construction of Linkage Maps

The pairwise map distances for three or more linked genes allow the ordering of these genes into a linear linkage map. When more than five loci must be ordered, this is better done by using the software packages developed for this purpose. These packages implicitly perform the estimation and calculation of LOD scores.

1. Prepare a crude data file to be used as input for the mapping program. The format of these data files depends on the software package that is being used; however, packages require a similar layout. Crude data files can be prepared with any text editor or spreadsheet program that enables export of plain-text files (*see* **Notes 22** and **23**).
2. Determine the linear order of the markers in relation to the mutant (*see* **Note 24**). Depending on the software that is being used, intermediate results are produced before a final map is calculated. Joinmap, for example, produces a suggested assignment of genes to linkage groups, as well as a list of pairwise estimates and LOD scores. It also allows testing for segregation distortion for each gene (*see* **Notes 22** and **23**).
3. Graphical representations of the maps can be produced with specific graphic software such as Drawmap ***(9)*** or MapChart ***(10)*** or with any other design computer method (*see* **Note 23**).

3.5.2. Fine Genetic Mapping

1. Grow a large F2 mapping population. This can be the same population used for crude mapping or a new (available as backup [*see* **Note 11**]), especially when one expects that modifiers of the mutant phenotype will segregate.
2. Isolate DNA from all plants or from only the mutants segregating in the population.
3. Genotype the plants for molecular markers flanking the locus under study.
4. Identify all recombinants between both flanking markers (*see* **Notes 24** and **25**).
5. Harvest seeds from all recombinants because it might be necessary to grow the F3 progeny of plants that are recombinant between both markers (*see* **Note 25**).
6. Identify molecular markers located between both flanking markers, making use of Web sites where sequence comparisons between accessions are accessible (*see* **Note 26** and Chapter 16 in this book).
7. Genotype the recombinant plants for the new linked markers.
8. Locate the mutant gene in relation to the molecular markers according to the recombination events in the genotyped recombinant plants. Markers that still identify one recombinant on each side of the mutant locus delimit the region in which the gene is located (*see* **Note 27**).

4. Notes

1. Whenever possible, use the genotype with a recessive trait as the female parent because this will allow checking for unwanted selfing of the parent in the F1 generation. This recessive parent is often the mutant but it can also be the parent that carries another recessive allele, such as the *erecta* mutation (in L*er* and its mutants), when the other parent carries the dominant allele of this gene (in the other wild accessions or ecotypes).
2. When a mutation is lethal or leads to complete sterility, crosses must be made with heterozygous individuals that can be identified by progeny testing. It may happen that at the moment the cross is made this information is not available (e.g., embryo lethals can often be identified in older siliques of the same plant). In this situation several putative heterozygous plants (wild-type sister plants in a line that segregates for the mutant) must be crossed and seeds of these parents have to be harvested for progeny testing.
3. Although all flower buds can be used for crosses, the flowers at the lower part of the main stem are often preferred because these flower buds seem "stronger." The use of these early flowers allows one to remake the cross when it has failed. Removing petals (accidentally) is not a problem; however, the pistil should have no damage. To check for pollen availability, touch-

ing a nail with the open flower should show some yellow pollen. *Arabidopsis* buds that are emasculated just before the petals become visible (1 d in practice) can be pollinated on the same day, although the stigma is also receptive 1 d later. The transfer of an overdose of pollen on a stigma prevents crosspollination by open flowers from the same or neighboring plants; therefore, "bagging" of the crosspollinated buds is not necessary.

When several crosses must be made, use different plants or inflorescences. Making different pollinations in a group of buds easily leads to mistakes due to mislabeling or problems with harvest.

4. Although the different F1 plants from a cross of two homozygous parents should all be genetically identical, it is advised not to bulk seeds from different F1 plants but to test the individual progenies. When one F1 is wrong due to selfing, unwanted crosspollination, seed contamination and the like, this can be recognized by its unexpected segregation and the progeny of such a plant can be eliminated from the analysis. In cases of recently isolated mutants they can be heterozygous for other mutations; when such a plant has been crossed half the F1 will be heterozygous for this second mutation and half will not. The data of progenies that are identical can always be added but data from a contaminated bulk of F1 plants are useless.
5. A 1:1 segregation in the BC with the recessive parent (variant) and a 3:1 (wild-type:variant) in the F2 indicates monogenic inheritance and recessiveness of the variant. Dominance of the variant yields 1:3 (wild-type:variant) ratios for the F2 and a 1:1 segregation only in the BC with wild-type. Other monogenic ratios are due to intermediate expression of the heterozygotes (1:2:1 in the F2) and/or reduced viability or even lethality of certain genotypes (often the homozygous mutant). Another factor that may lead to a deficit of a class of genotypes is a reduced transmission of certain (mostly mutant) alleles through the gametophyte. This phenomenon, which is called *certation*, is assumed to occur only in pollen and is the main cause for recessive deficits in induced mutants ***(11)***. Digenic inheritance of a certain character gives ratios in the F2 that are variations of the classical 9:3:3:1 (*A/_B/_:a/aB/_:A/_b/b:a/a b/b*) ratio for two genetically independent genes. These ratios are, depending on the epistatic relations between the two genes, 9:7; 13:3; 15:1.

   When traits (e.g., seed traits) are controlled by the genotype of mother plant (maternal inheritance), one must be aware that such a trait needs to be analyzed in the progeny of the mother plant.
6. Use at least 16 F3 plants per selfed progeny (i.e., line or family) from an F2 plant to distinguish between a progeny that originates from either a heterozygous genotype or from a progeny of a homozygous plant with 99% certainty ($0.75^n < 0.01$ at $n = 16$). Using fewer than 16 plants may result in some progenies that contain, because of chance, only wild-type plants, although they should segregate 25% recessive ones.
7. The outcome of the complementation test can be complicated when allelic complementation occurs. A well-documented case of this in *Arabidopsis* is some alleles at the *py* locus ***(12)***. However, in contrast to nonallelic complementation of true recessive mutations, allelic complementation is often only partial and does not occur in all allele combinations; therefore, incorrect conclusions are rare.
8. Very close linkage of two loci with a dominant allele for gene 1 in one parent and a dominant allele of gene 2 in the second parent (resulting in similar phenotype) can not be distinguished from allelism until a recombinant is found.
9. The genotype of the epistatic mutant will be inferred from the segregation in F3 lines obtained from F2 plants having either one mutant phenotype or the other. Most likely, one type of mutant phenotype F2 plants will not segregate in F3, indicating that in this phenotypic class is included the double-mutant phenotype. Two out of each three F2 plants of the other mutant phenotype will produce F3 lines that segregate for plants with either one or the other mutant phenotype (in the progeny of *A/a b/b,* the *a/a b/b* segregants will resemble the *a/a B/B* parent and not the *A/A b/b* parent when the *a* mutant is epistatic over the *b* mutant). When the double-mutant phenotype resembles wild-type, both types of single-mutant phenotype F2 plants will segregate for wild-type and the corresponding mutant phenotypes.
10. The double mutant will comprise 1/7 of the mutants when no novel phenotypes are observed and the two parents can not be distinguished. In this case at least 20 plants [$(6/7)^n < 0.05$ at $n = 20$] must be tested to be sure that the double mutant will be identified.

11. When mutants are either in Col or L*er* background, a cross with either L*er* or Col, respectively, is preferred because most publicly available polymorphism data are derived from them ***(2)***. Nevertheless, polymorphism data for many other combinations are becoming available (www.mpiz-koeln.mpg.de/masc/). The use of marker lines carrying mutations of known genetic position can also be considered. Most of these lines are in L*er* or mixed L*er*/Col background and they allow a rapid rough location of the new mutation on the linkage map. When the mutant is in a background different from these marker lines this type of mapping can be combined with the analysis of molecular markers, which allow a more precise localization because of the high marker density ***(3)***. A problem with this approach is that sometimes the marker phenotypes interfere with the scoring of the new mutant. However, an advantage of doing the mapping in one genetic background is that modifiers present in another genetic background do not interfere with the scoring. Examples of traits for which modifying genes are segregating in almost any cross between wild accessions are flowering time, plant length, and seed dormancy. To reduce the effects of modifiers two solutions are possible: (1) Make use of introgression lines of segments of one accession in the background of the accession in which the mutant is present. (2) In the F2 select several plants heterozygous for the mutant (tested by progeny testing) and use as mapping population, a segregating F3 line in which scoring of the mutant phenotype can be done unambiguously (such a heterogenous inbred family is selected for nonsegregation of modifiers). Introgression lines are now available for Cvi into the L*er* background and L*er* into Col background ***(13)***, and several other combinations will become available in the near future.
12. The analysis of linkage might be performed with mutant plants only. The wild-type phenotypes are also informative but require knowledge about their heterozygosity for the wild-type phenotype. Heterozygosity for the mutation requires progeny testing of F2 plants, which, in cases of embryo lethal mutants, can be done by inspecting immature siliques. In case of lethal mutations, instead of analyzing mutant F2 plants one will need to use heterozygous mutant plants and the homozygous wild-type F2 plants. In principle other mapping populations, such as backcrosses to the recessive mutant, may also be used. However, in *Arabidopsis* the use of F2s is common practice.
13. The advantage of pooling plants in small bulks is that one needs to do fewer marker analyses. However, analysis of pools with linked markers does not allow the determination of the order of markers, which can be obtained when individual plants are analyzed. It is advised that DNA be isolated from individual (mutant) plants so that in a later phase, when linked markers are found, their linkage can be confirmed. Assuming that one cannot fully rely on the quantification of marker bands in pooled DNA, only the absence of the allele from the nonmutant parent will be really informative. With five F2 plants in a bulk, one recombinant allele leads to an estimate of $R = 0.10$, which implies that only closely linked markers can be detected even in case of small pool size.
14. For the determination of a crude map position a large population is not required. Only the detection of significant linkage is necessary. The latter can be achieved by increasing the marker density and/or by increasing the population size. With 22 to 25 evenly spaced markers, Lukowitz et al. ***(3)*** and Jander et al. ***(2)*** recommend 100 to 150 mutant F2 plants, whereas Peters et al. ***(4)*** recommend 20 to 30 plants when 85 well-dispersed markers are used. Population sizes around 100 (25 mutants) are in general sufficient to detect linkage, but with $r > 35\%$ this population size often results in observed segregation ratios that are not significantly different from independent segregation.
15. In case of dominance, four phenotypic classes are expected in an F2, whereas one expects six classes when one of the markers is codominant and nine when both are codominant. In the case of dominance of linked markers (morphological markers and also AFLP markers), the deviation from the "classical" independent segregation ratio in an F2 (i.e., 9:3:3:1 for *A/_B/_:a/a B/_:A/_b/b:a/a b/b*) depends on the so-called "phase" of the recessive and dominant alleles in the double-heterozygous F1 parent. We distinguish "coupling phase" (F1 derived from a cross such as *a/a b/b* × *A/A B/B*) and "repulsion phase" (F1 derived from the cross *a/a B/B* × *A/A b/b* or reciprocal). With absolute linkage the F2 ratio for coupling phase is 12:0:0:4, whereas for repulsion phase this is 8:4:4:0. As it is easier to tell the difference between 9:3:3:1 and 12:0:0:4

than between 9:3:3:1 and 8:4:4:0, it is intuitively clear that it is harder to detect linkage from a repulsion-phase F2 than from a coupling-phase F2. When only the mutant class of the F2 is analyzed, linkage in coupling shows that most mutants have the allele of the background parent of the mutant, whereas repulsion is seen as the absence of the allele of the opposite parent.

16. Independence of segregation of two genes can be tested with the chi-square test. It is better to use a contingency table analysis instead of a test of deviation of the 9:3:3:1 for F2 or 1:1:1:1 for BC generations (or any other) as the expected ratios can be distorted when the monogenic ratios are not Mendelian (e.g., due to certation or reduced viability of certain genotypes). If one class (only the mutants) is analyzed for markers, one can only test deviations from 3:1 and 1:2:1 ratios and when disturbed segregation occurs (not uncommon in ecotype crosses), one may interpret the reduced overall transmission of a marker as linkage of that marker to the mutant to be mapped. In the latter case, also testing the wild-type class for recombinants avoids this pitfall.
17. Information on linkage can be expressed in a LOD value, the logarithm of the likelihood odd ratio. Mathematically the LOD reads:

$$LOD = \log \frac{(\text{likelihood of observed data with } p = \hat{p})}{(\text{likelihood of observed data with } p = 0.5)}$$

A LOD score of, e.g., 5.2 means that for the given data, the true recombination frequency is $10^{5.2}$= 158,489 times more likely to be equal to the estimated *r* than to be equal to 0.5.
The use of LOD scores is especially useful when the recombination estimate equals zero. If this estimate is obtained from a backcross, i.e., from a binomial sample, its variance formally equals zero, irrespective of the sample size. The LOD score, however, also indicates that finding no recombinants in a sample of 100 is more informative than finding no recombinants in a sample of 10 (LOD = *n* log2, in this case; *n* is sample size). Usually a LOD score > 3 is taken as evidence for linkage. A separate test for independent segregation is often omitted when a large number of markers are tested and LOD scores, recombinant frequencies and linkage maps are estimated directly with the appropriate software.
18. The fraction of recombinant gametes (which is used as the estimate of recombination frequency) can be derived from the frequencies of the various phenotypes in segregating generations. The analysis of the segregation in a test cross, that is, the progeny of a cross of a double recessive with the diheterozygote (*a/a b/b* × *A/a B/b*), is an effective way to determine the recombination percentage between the two loci. The number of nonparental (recombinant) progeny gives a direct estimate of the recombination fraction *r*, of which standard deviation can be derived from the binomial distribution. A drawback of test crosses is that they require double-recessive genotypes and accurate crossings, which is not easy in some genotypes of *Arabidopsis*. For these practical reasons and the fact that selfed seeds can be easily obtained in large quantities in *Arabidopsis*, mapping has been performed mainly with F2 and F3 populations, which can also be very efficient and which averages the often significant differences of male and female meiosis ***(14)***.
19. From the observed F2 segregation frequencies, the recombination fraction, *p*, can be estimated by the product ratio (PR) method. The product ratio is the product of the two non parental phenotypic classes divided by the product of the two parental classes. This PR relates to recombination fractions, which are listed in the tables ***(15,16)***. These tables also give values from which the standard deviations can be estimated, taking into account the total number of F2 plants. The PR method can be used only for the "classical" F2 (dominance, all phenotypes equally viable, resulting in a 9:3:3:1 ratio in case of independent inheritance). In more complex situations, such as F2s with dominance or recombinant inbred lines, the recombination rate is usually estimated by maximum likelihood. Allard ***(17)***, Ritter et al. ***(18)***, and Maliepaard et al. ***(19)*** have compiled a large number of ML esztimation equations for many different situations.
Computer programs such as LINKAGE ***(5)***, and also the mapping programs (discussed in the following) can be used to estimate recombination frequencies with standard deviations.

20. The analysis of the selfed (F2) progeny of a diheterozygote, where both markers are fully recessive, is much more efficient when the genes are linked in coupling phase as compared with linkage in repulsion phase.
    The problem with F2 repulsion data, especially in the case of close linkage, is the statistical inaccuracy of the estimate of recombination frequencies. The analysis of F3 lines derived by selfing from specific F2 plants allows an accurate estimate of the recombinant fraction (r) in this case. Progeny testing is more efficient than increasing the size of the F2 population at $r < 0.11$. Because the aim of this progeny test is to determine whether an F2 plant was either homozygous or heterozygous for the dominant allele, 16 plants per line are sufficient to be sure about the classification with 99% certainty.
    Progeny tests are more efficient when the F3 line can be screened for an early and easy-to-score marker. For example, when one is interested in linkage between a seed coat gene and a trichome gene, the seed coat recessive plants are harvested in the F2 and their progeny is screened for the trichome mutation at the seedling stage. Recombinant fractions (r) can be estimated by combining the F3 data with F2 data using the maximum likelihood procedure or by the following formulas:

$$\text{if } x = \frac{a/a\,B/b}{n} = \frac{\textit{number of segregating lines}}{\textit{number of lines tested}}$$

$$\hat{r} = \frac{x}{2-x} \quad \textit{with} \quad S_r = \frac{2\sqrt{x(1-x)}}{(2-x)^2\sqrt{n}}$$

    With low recombination values the number of segregating F3 lines is almost the same as the number of recombinants found among the number of gametes that are tested, which is twice the number of progenies tested (each F2 plant comes from two gametes).
21. Recombination data can be converted into map distances using the Haldane ***(20)*** or Kosambi ***(21)*** mapping functions.
    The Haldane mapping function reads:

$$r = \frac{1}{2}(1 - e^{-2x})$$

$$x = -\frac{1}{2}\ln(1-2r)$$

    with x in Morgans; or expressed with centiMorgans (D) and recombination frequency (r):

$$r = 50\left(1 - e^{(-D/50)}\right)$$

$$D = -50\ln[(50 - r/50)]$$

    A commonly used mapping function that accounts for a certain degree of interference is the Kosambi function ***(21)***.

$$r = \frac{1}{2}\tanh(2x)$$

$$x = \frac{1}{4}\ln\left(\frac{1+2r}{1-2r}\right)$$

or

$$r = 50 \tanh (D / 50)$$

$$D = 25 \ln\left(\frac{100+2r}{100-2r}\right)$$

Transforming estimates of recombination percentages into map distances requires transforming of standard errors as well. Writing s_r for the standard error of an estimate of recombinant percentage, the corresponding standard error of the map distance approximately equals:

$$S_D = \frac{50}{50 - r} S_r \text{ (Haldane)}$$

$$S_D = \frac{2500}{2500 - r^2} S_r \text{ (Kosambi)}$$

The best mapping function is one that gives the best additivity of the calculated map distances.

22. The mapping programs differ in the type of data they can handle. Both MAPMAKER and GMENDEL can use only crude data, i.e., phenotypic scores for the markers to be mapped. For these programs the phenotypes must have been scored in a single mapping population, e.g., a backcross, an F2, or a set of recombinant inbred lines. Joinmap uses crude data to produce a list of pairwise recombination estimates. This list is used to construct the linkage map. Various pairwise recombination lists, obtained from different mapping populations (not necessarily of the same type, e.g., from literature) can be merged into a single input list. So with Joinmap, linkage maps obtained from various data sets can be combined to calculate an integrated map *(7)*. The various software packages for map construction differ, furthermore, in the way in which they are being used and the methods of finding the “best fitting” gene order. MAPMAKER can be used in an interactive way, i.e., the user can specify any gene order for which a map and the corresponding joint likelihood are calculated. This allows “manual” insertion by trial and error of additional genes to an established map. The map construction part of Joinmap is completely noninteractive. However, Joinmap consists of a number of distinct modules, each one performing a specific task (calculating pairwise estimates and LOD scores, assignment of genes to linkage groups, testing for segregation distortion, testing for heterogeneity among estimates, ordering the loci, etc.), allowing flexibility and data checking. Furthermore, Joinmap, in contrast to MAPMAKER, can deal with all types of mapping populations that are being used in plants.
23. The MAPMAKER program can be obtained from MAPMAKER, c/o Dr. Eric Lander, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142 (mapmaker@genome.wi.mit.edu) or through the Web by anonymous FTP from host “genome.edu.wi.mit.edu.”
The Joinmap and Drawmap programs can be requested from Dr. J. W. van Ooijen, Kyazma B.V., P.O. Box 182, 6700 AD Wageningen, The Netherlands (www.kyazma.nl/index.p). E-mail: Sales@Kyazma.nl.
The MapChart program can be obtained upon request from Dr R. E. Voorrips (Roeland.Voorrips@wur.nl), Plant Research International, P.O. Box 16, 6700-AA Wageningen, The Netherlands.
The updated genetic, physical, and sequence map, as well as information concerning all kinds of markers, can be found at The *Arabidopsis* Information Resource (TAIR) (www.arabidopsis.org).
24. Analyze the rare recombinants between the mutant allele and markers. These will mostly be homozygous for the mutant locus and heterozygous for some of the other linked markers. When finding in the same population genotypes such as *M1A/M1A*, *m/m*, *M2A/M2B*, *M3A/M3B*, and *M1A/M1B m/m M2AM2A M3AM3A* (with the *M3* marker showing more recombinants with m than *M2*), the mostly likely locus order is *M1*, *m*, *M2*, *M3*. Finding which markers unambiguously surround the mutant locus is an important outcome of this crude mapping exercise.

25. The recombinants between the flanking linked markers *M1* and *M2* will have the genotypes *M1A/M1A*, *M2A/M2B* or *M1A/M1B*, *M2B/M2B*. In these recombinant plants it is necessary to know the genotype of the mutant, *m/m*, *M/m*, or *M/M*. The latter two genotypes are among the wild-type F2 plants and can be distinguished by progeny testing. Harvesting the seeds of these recombinants will be important to test their progeny and to have plant material available when one runs out of DNA or when desiring to recheck the analysis.
26. Additional markers located between the flanking markers need to be identified. Such polymorphic markers can be found easily when sequence information from both accessions is available (www.arabidopsis.org). When the sequence is available for only one or none of the parents, one can test if L*er*/Col polymorphisms (or those between other accessions) are also present between the two parents. This may require testing several markers at nearby positions. At a later stage, the development of new markers might be required to find new polymorphisms in particular regions. This can be done by sequencing short DNA fragments from those regions in both parental lines. A higher polymorphism rate is observed in noncoding (intergenic or introns) than in coding regions.
27. A marker absolutely linked to the mutation might be the gene disrupted in the mutant. However, the last marker identifying recombinants is usually located several genes away from the mutant gene. This will be the case when several closely linked markers are absolutely linked to the mutation because no recombination events occurred between these genes in the population investigated.

References

1. The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana. Nature* **408,** 796–815.
2. Jander, G., Norris, R. R., Rounsley, S. D., Bush, D. F., Levin, I. M., and Last, R. L. (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol*. **129,** 440–450.
3. Lukowitz, W., Gillmor, C. S., and Scheible, W. R. (2000) Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. *Plant Physiol.* **123,** 795–805.
4. Peters, J. L., Cnudde, F., and Gerats, T. (2003) Forward genetics and map-based cloning approaches. *Trends Plant Sci.* **8,** 484–491.
5. Suiter, K. A., Wendel, J. F., and Case, J. S. (1983) LINKAGE-1: a PASCAL computer program for detection and analysis of genetic linkage. *J. Heredity* **74,** 203–204.
6. Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., and Newberg, L. (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1,** 174–181.
7. Stam, P. (1993) Construction of integrated linkage maps by means of a new computer package: Joinmap. *Plant J.* **3,** 739–744.
8. Liu, B. H. and Knapp, S. J. (1990) GMENDEL: a program for Mendelian segregation and linkage analysis of individual or multiple progeny populations using log-likelihood ratios. *J. Heredity* **81,** 407.
9. van Ooijen, J. W. (1994) Drawmap: A computer program for drawing genetic linkage maps. *J. Heredity* **85,** 66.
10. Voorrips, R. E. (2001) MapChart version 2.0: Windows software for the graphical presentation of linkage maps and QTLs. Plant Research International, Wageningen, The Netherlands.
11. Dellaert, L. M. W. (1980) Segregation frequencies of radiation-induced viable mutants in *Arabidopsis thaliana* (L.) Heynh. *Theor. Appl. Genet.* **57,** 137–143.
12. Li, L. and Rédei, G. P. (1969) Direct evidence for models of heterosis provided by mutants of *Arabidopsis* blocked in the thiamine pathway. *Theor. Appl. Genet.* **39,** 68–72.
13. Koumproglou, R., Wilkes, T. M., Townson, P., et al. (2002) STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. *Plant J.* **31,** 355–364.
14. Vizir, I. Y. and Korol, A. B. (1990) Sex differences in recombination frequency in *Arabidopsis*. *Heredity* **65,** 379–383.
15. Stevens, W. L. (1939) Tables of the recombination fraction estimated from the product ratio. *J. Genet.* **39,** 171–180.
16. Rédei, G. P. (1982) *Genetics*. Macmillan, New York, p. 736.
17. Allard, R. W. (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* **24,** 235–278.

18. Ritter, E., Gebhardt, C., and Salamini, F. (1990) Estimation of recombination frequencies and contruction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics* **125,** 645–654.
19. Maliepaard, C., Jansen, J., and van Ooijen, J. W. (1997) Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Gen. Research* **70,** 237–250.
20. Haldane, J. B. S. (1919) The combination of linkage values, and the calculation of distance between linked factors. *J. Genet.* **8,** 299–309.
21. Kosambi, D. D. (1944) The estimation of map distances from recombination values. *Ann. Eugen.* **12,** 172–175.

7

QTL Analysis

Carlos Alonso-Blanco, Maarten Koornneef, and Johan W. van Ooijen

Summary

The study of the intraspecific natural variation existing in *Arabidopsis* provides a useful resource for the dissection of its genome at the functional, ecological, and evolutionary levels. A major step in these studies is the identification and location of quantitative trait loci (QTLs) determining the variation for trait(s) of interest, so-called QTL analysis. Here we provide general methods to achieve this goal, including (1) the selection of wild accessions that can serve as parental lines for QTL mapping; (2) the development of mapping populations; (3) the construction of the required genome-wide linkage maps of the populations; and (4) the mapping and characterization of QTLs affecting the trait(s).

Key Words: *Arabidopsis*; genetics; quantitative trait locus (QTL); natural variation; mapping; recombinant inbred lines (RILs); marker; polymorphism; epistasis; genetic interaction; pleiotropy; linkage.

1. Introduction

Arabidopsis is a wild species with a wide geographical distribution covering a diversity of environments. This distribution embraces a large intraspecific natural genetic variation for many traits, presumably reflecting adaptations ***(1)***. In a few cases, the variation among wild accessions is due to single gene differences with large discrete effects, such as those determining the glabrous and the erecta phenotypes of some accessions and certain disease resistances ***(2)*** that can be studied by Mendelian genetic analysis (*see* Chapter 6 in this book). However, most of the natural phenotypic variation existing among *Arabidopsis* wild accessions is of a quantitative nature, which has historically hampered its genetic analysis. Traits displaying quantitative variation are called quantitative traits, and they are generally under multigenic control. Currently, the genomic resources available in *Arabidopsis* allow the analysis of this source of genetic variation up to the molecular identification of the underlying genes and polymorphisms ***(1)***.

The first step in the forward genetic analysis of natural variation corresponds to the identification, location, and genetic characterization of genomic regions containing the loci that affect a quantitative trait, so-called quantitative trait loci (QTLs). This procedure is commonly referred to as QTL analysis and QTL mapping, and is the main scope of this chapter. For that, we need to generate an experimental population (mapping population) that segregates for the trait(s) of interest. Subsequently, this population is thoroughly genotyped for segregating markers evenly distributed over a genome-wide genetic linkage map, and characterized phenotypically for the quantitative trait(s). Because the phenotype of each individual is determined by the environment and an unknown number of genes, the genotype at the segregating QTLs cannot be inferred directly from the phenotype. Consequently, the location of QTLs in the linkage maps cannot be based on the direct estimation of recombination frequencies, as for Mendelian loci. In contrast, QTL positions are indirectly inferred from linked markers, by using specific statistical methods aiming to find and compare associations between marker genotypic classes and the phenotypic trait.

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

QTL mapping populations are based upon the development of F1 hybrids derived from crosses between different *Arabidopsis* wild accessions, usually distinct for the trait of interest. From this, a segregating population with large linkage disequilibrium (LD) between loci is established, which provides the foundation of the genetic linkage analysis. This mapping population may be crossed or selfed further, thereby decreasing the LD through recombination. As a result only the markers more closely linked to a QTL will detect the association, and therefore the mapping resolution is increased. Many kinds of mapping populations can be used for QTL analysis, such as F2 or backcross (BC) populations. However, families of recombinant inbred lines (RILs) have been the main choice in *Arabidopsis*, owing to the advantages derived from their homozygosity and their increased effective recombination. RILs are produced by single-seed descent from an F2 population: from each F2 plant a single seed is harvested and grown into an F3 plant, from which again a single seed is harvested and grown into an F4 plant, and so on. During this process, the level of heterozygosity per locus is halved in each generation, so plants become practically homozygous at the F8 generation and onward. For this reason, RILs are considered immortalized or permanent populations that can be propagated indefinitely and need to be genotyped for segregating markers only once. In addition, during this process of inbreeding, plants undergo several rounds of meiosis, increasing the recombination frequency between closely linked markers. RIL populations can be used to map QTLs very efficiently because the influence of the environment on the quantitative trait can be much reduced by assessing several plants of the same genotype instead of just a single plant. Furthermore, once they have been obtained and genotyped, RIL populations can be used in an unlimited number of studies of different traits in different environments, allowing efficient comparative QTL mapping.

There are other types of immortalized populations, such as families of doubled haploid lines (DHLs), which may be obtained faster than RILs. These can be developed in some plant species, such as rice, barley, or *Brassica*, where haploid plants can be generated from gametic cells, and whose chromosomes are subsequently duplicated. DHLs are mostly generated from an F1 and therefore they undergo a single round of recombination, resulting in about half the recombination and mapping resolution of RILs. However, this type of population is not feasible in *Arabidopsis* because no procedure for efficient haploid induction has been described for this species. Another type of mapping population with an increased effective recombination is the family of advanced intercross lines (AILs), which are derived by randomly intercrossing the individuals of the F2, F3, and F4 generations ***(3,4)***. Yet another population type is the so-called advanced backcross population or family of backcross inbred lines (BILs), consisting of introgression lines (ILs) whose introgressed regions together cover the whole genome. They are developed by recurrent backcrossing the individuals of a BC population to a reference genotype, usually followed by several generations of single-seed descent inbreeding. The development of BILs might be combined with marker-assisted selection of the BC individuals to improve the coverage of the genomic regions introgressed. BILs are useful because they allow a more precise estimation of QTL effects due to the reduced genetic background variation, but especially because they provide an important material for the subsequent QTL characterization. Despite their advantages, these population types and others with more complex experimental designs involving multiple crosses between more than two parental accessions ***(5)*** require intensive labor and have not yet been exploited in *Arabidopsis*. However, another type of immortalized population consisting of a collection of wild accessions is recently being evaluated for the dissection and mapping of QTLs in *Arabidopsis* ***(6)***. Mapping using such complex heterogeneous populations of wild accessions is (confusingly) referred to as LD mapping, and aims to exploit the residual LD naturally persisting among accessions as a consequence of their common ancestry and evolutionary history. It has been shown that such "natural" LD extends physically up to 250 kb when comparing wild accessions of *Arabidopsis* ***(6)***. This relatively large natural LD combined with the small genome of *Arabidopsis* and its high polymorphism frequency (necessary to obtain the dense coverage by marker data needed) might provide a new mapping tool for the near future.

Table 1
Some Internet Resources for Analysis of Natural Variation and QTLs in *Arabidopsis*

Content	Title	Internet address
Genotypic and phenotypic information on *Arabidopsis* collections of wild accession and RIL and NIL populations		
	NATURAL	www.natural-eu.org
	Naturalvariation	naturalvariation.org/
	INRA-Versailles	www.inra.fr/qtlat/
Arabidopsis seed stock centers		
	ABRC	www.arabidopsis.org/abrc/
	NASC	nasc.nott.ac.uk/
	SASSC	www.brc.riken.jp/lab/epd/SASSC/
Arabidopsis markers and polymorphisms		
	TAIR	www.arabidopsis.org/
	Cereon polymorphisms	www.arabidopsis.org/Cereon/index.jsp
	MIPS polymorphisms	www.mpiz-koeln.mpg.de/masc/
Genetic mapping software		
	Alphabetic list of genetic analysis software	www.nslij-genetics.org/soft
	Joinmap	www.joinmap.nl
	MAPMAKER, MAPMAKER/QTL	www.broad.mit.edu/genome_software
	Map Manager QTX	www.mapmanager.org
	MapQTL	www.mapqtl.nl
	MultiQTL	www.multiqtl.com
	PlabQTL	www.uni-hohenheim.de/~ipspwww/soft.html
	QTL Cartographer	statgen.ncsu.edu/qtlcart
	QTLexpress	qtl.cap.ed.ac.uk

Mapping populations are genotyped with any marker technique, such as polymerase chain reaction (PCR)-based markers, including microsatellites and codominant cleaved amplified polymorphic sequences (CAPSs), or any of the high-throughput technologies currently available to detect single-nucleotide polymorphisms (SNPs) or insertion/deletions (INDELs). Genotyped markers are arranged in a genetic linkage map based on their pairwise recombination frequencies, which provides the basis for the subsequent QTL analysis. For this, several statistical tools are used, aiming to estimate the genetic locations and effects of QTLs, their possible interactions with other QTLs (epistasis) or with the environment, and their pleiotropic effects on other traits. The power of QTL detection and the accuracy of QTL location will depend on the intrinsic characteristics that determine the genetic basis of the trait, such as the number of QTLs, their heritability and additive effects, the interactions among QTLs, and the linkage between QTLs. In addition, the quality of QTL mapping will also depend on several factors that can be partly manipulated in the experimental design, such as the type of mapping population, the population size, the number of observations per genotype, the marker density of the linkage map, and the statistical method used to map QTLs. Due to the current interest in the analysis of *Arabidopsis* natural variation, several biological and genomic resources have been generated that are specifically devoted to it. These include collections of wild accessions, many RIL mapping populations, and several polymorphism databases involving various wild accessions (*see* **Table 1**). These resources make QTL analysis in *Arabidopsis* an efficient routine methodology that can be used to identify the major effect loci determining natural variation for any trait of interest. Here we describe the particular characteristics of Arabidopsis QTL analysis derived from its genome and resources. The formal theory behind QTL analysis is beyond the scope of this protocol and can be found in textbooks ***(7–10)***.

To fully understand the natural genetic variation, after the QTL analysis of a trait one needs to pursue the molecular characterization of the identified QTLs. This includes the molecular isolation of the genes underlying individual QTLs (referred to as quantitative trait genes [QTGs]) and the identification of the DNA polymorphisms altering the function of the gene and causing the phenotypic variation (so-called functional polymorphisms or quantitative trait nucleotides [QTNs]). The characterization of individual QTLs requires the development of genotypes differing only in the alleles at a single genomic fragment, called near isogenic lines (NILs), which differ only at a single locus (or several closely linked loci) affecting the trait of interest. These plants can be derived by recurrent backcrossing as introgression lines carrying a single genomic fragment from a donor parent introgressed in a reference genetic background. Alternatively, NILs can be obtained in a heterogeneous genetic background using heterogeneous inbred families (HIFs) derived from plants heterozygous for a particular region of interest in an otherwise homozygous background ***(11)***. When NILs are intercrossed, their offspring will segregate only for the heterozygous region, hopefully allowing the Mendelian analysis of a targeted QTL. This might include the QTL fine mapping and the study of interaction effects between QTLs. In addition, NILs can be used for the further genetic, physiological, and molecular characterization of the QTL, in a similar way as is done with *Arabidopsis* mutants. Nevertheless, the molecular isolation of QTLs still requires a major effort ***(12)***, mainly due to the large amount of nucleotide diversity existing among wild accessions; two *Arabidopsis* accessions reveal, on average, six polymorphisms per kilobase of DNA ***(13,14)***, which hampers the identification of the causal QTGs and QTNs. Several genomic resources available in *Arabidopsis* are making QTG and QTN identifications feasible ***(1)***, such as efficient fine- or high-resolution mapping technologies, genome-wide gene expression tools, collections of insertional or ethyl methanesulfonate (EMS) mutants for almost any *Arabidopsis* gene, possibility of full-sequence comparison and association studies in small genomic regions, and complementation tests by plant transformation. After the QTGs are identified, their genetic variation among other *Arabidopsis* accessions may be evaluated, either by direct sequencing or with other techniques like the so-called ecotilling (***15***; *see* Chapter 11 in this book). The existing molecular variation at QTGs can then be analyzed with the tools of population genetics in combination with the functional resources. Thus, the analysis of natural variation of *Arabidopsis* is becoming a useful resource for the functional dissection of its genome and for the understanding of the ecological and evolutionary bases of plant adaptation to different environments.

2. Materials

1. Seeds of *Arabidopsis* accessions and mapping populations.
2. Equipment to cross plants (tweezers, stereo microscope, and labels).
3. Equipment to grow many plants simultaneously, under the assay conditions, which is necessary in order to perform the quantitative analysis of whole accession collections or mapping populations. Specific requirements will depend on the particular test conditions.
4. Equipment to genotype molecular genetic markers. This might be as simple as oligonucleotides and PCR ingredients, a thermocycler and an agarose gel system, for standard microsatellites and CAPS markers; or apparatus and reagents for high-throughput genotyping of SNPs or INDELs.
5. Equipment to measure the quantitative trait(s) of interest. Depending on the biological parameter to be measured (*see* **Notes 1** and **2**) this might be, for instance, from a simple ruler, up to a luciferase luminometer or a microarray scanner.
6. Software for general statistical analysis (e.g., SAS, SPSS, or STATISTICA packages), for linkage mapping analysis (e.g., MAPMAKER or Joinmap) and for QTL analysis (e.g., MAPMAKER/QTL, Map Manager QTX, MapQTL, MultiQTL, PlabQTL, QTL Cartographer or QTL express).

Table 1 shows some Internet addresses containing relevant information on these materials.

3. Methods

3.1. Statistical Overview and Concepts in QTL Analysis

3.1.1. Basics of QTL Analysis

QTL analysis basically consists in testing the effect of each marker of the linkage map on the trait of interest by an advanced analysis of variance (ANOVA). The analysis of variance tests whether any additional variance observed after some treatment (here, a change in QTL alleles) might have arisen just by chance or should be attributed to the treatment. In standard ANOVA the mean values of the (QTL genotype) classes are compared using the residual variance within the classes, and statistically tested with a *t*-test or *F*-test. A major problem in QTL analysis is that the QTL genotype, i.e., the treatment, cannot be assessed, but one uses the genotype of linked markers instead. One can perform an ANOVA for each segregating marker and study the results in relation to the marker map positions; this type of analysis is sometimes referred to as single-marker ANOVA. An important improvement called interval mapping (IM) uses a statistical technique called maximum likelihood (ML) and recombination probabilities to estimate QTL genotype classes. Thus, in contrast to single-marker analysis, IM allows one to test QTL effects at every map position, and not just at the marker positions. Haley and Knott ***(16)*** introduced the regression IM technique as a good approximation to the ML IM technique, because of the slow ML computations. A more recent development is the use of so-called Bayesian methods in QTL analysis, which are based on computerized sampling schemes to obtain posterior distributions of all parameters of interest ***(17)***. Their advantage derives from their powerful capacity to evaluate complex genetic models with many unknown parameters, such as complex family structures, complex experimental setup, residual multigenic effects, and interactions. Their disadvantages are the need for a thorough understanding of the problem in order to set up proper sampling schemes, the high computational demands, and the lack of available software that can be easily applied.

In IM, the ANOVA *F*-test statistic is replaced by the logarithm of the odds (LOD) score or deviance (D). The LOD score is a likelihood ratio test statistic, which is a generally applicable type of statistic: the likelihood (L) of the acquired data under the assumption of one model (H_1), $L(data|H_1)$, here a QTL segregate, is compared to the likelihood of the acquired data under the assumption of a null hypothesis model (H_0), $L(data|H_0)$, here no QTL segregates. The LOD score is the 10-base logarithm of the likelihood ratio: $^{10}\log[L(data|H_1) / L(data|H_0)]$. Alternatively the deviance (D) is used, which is twice the natural logarithm of the ratio of the likelihoods, $2*\ln[L(data|H_1) / L(data|H_0)]$. Under the null hypothesis the deviance has a chi-square distribution that can be used for the test. When H_1 is more likely than H_0, the LOD and D will have large (positive) values (LOD = 0.217D and D = 4.605LOD).

In IM, the whole linkage map is scanned for QTL effects and the LOD score is calculated at every marker position and between markers (e.g., every centiMorgan), resulting in the so-called LOD profile. This profile will show a peak at the position where a QTL is expected to be located, and will have low values outside QTL regions. When complete marker data are available, interval mapping yields on the marker positions the same results as the analysis of variance, i.e., the same class means and residual variance are estimated, but the test statistic is expressed as an LOD instead of an *F*-statistic. Although interval mapping is named after its ability to calculate LODs between marker positions, the technique's real power lies much more in the ability to accommodate for missing marker observation using the genetic linkage map and the neighboring markers, which is something that single-marker ANOVA cannot do because ANOVA just ignores those individuals.

3.1.2. Experimental Factors Affecting QTL Detection

If there is a QTL segregating, one needs the experiment to statistically prove its effect, i.e., to detect the QTL. For this, the variance generated by the segregating QTL must be sufficiently large with respect to the residual variance. Therefore, one needs to know the genotypic vari-

ance of the QTL (generated by the QTL effect) and the value of the residual variance. In a segregating population the residual variance is determined by the environment as well as by all other segregating QTLs, because the IM test statistic evaluates the presence of only a single QTL at the same time. A large genotypic effect or a small residual variance will increase the probability of detecting the QTL. Knowledge of these parameters allows the calculation of the number of plants needed to achieve (with reasonable certainty) a positive statistical conclusion in the test.

The environmental variance (σ_e^2) can be estimated straightforwardly by growing and evaluating several plants per homozygous genotype of a set of wild accessions. From the mean accession values obtained, it is usually best to choose the two opposing extremes of the set of accessions as the future parents for the QTL analysis experiment. These extremes are likely to have the largest differences in QTL genotype means, the largest number of differing QTLs, and their QTL effects in the same direction within the genotype; needless to say, this is a very wild guess, but one cannot do much better without further extensive experimentation.

However, to get the remaining parameters one needs to apply some further simplifying assumptions: the difference between the two parental accessions is determined by a certain number of QTLs for which they carry different alleles, for instance, five (in some cases one might have indications for more or less); the QTLs have equal genotypic effects; the effects are in the same direction; there is no dominance nor epistasis (QTL × QTL interaction) and the QTLs are unlinked. From the mean accession values of the future parents of the cross to be used for QTL analysis, m_1 and m_2, and under the above assumptions, we might estimate the expected additive genotypic effect per QTL: $a_q = [(m_1 - m_2) / 2] / 5$ ***(9)***. The expected additive genotypic variance per QTL depends on the type of segregating population: $\sigma_q^2 = fa_q^2$ with f as ¼, ½, 1, and 1 for a backcross, F2, doubled haploid, and RIL family, respectively ***(10)***. In addition, when there is a dominance effect, this will increase the genotypic variance of the F2, and it may increase or decrease the genotypic variance in a backcross depending on the sign of the effect; doubled haploid and RIL families are unaffected by dominance because they are homozygous. Furthermore, in advanced backcross generations the reduction in the variance is important due to the unbalance in QTL genotype frequencies: the factor f is $3/16$ and $7/64$ for the second- and third-generation backcross populations, respectively, and nearly halves every next generation. The total phenotypic variance becomes $\sigma_t^2 = 5\sigma_q^2 + \sigma_e^2$, where $5\sigma_q^2$ would be the total genotypic variance (σ_g^2).

3.1.2.1. Population Type, Population Size, and Number of Observations per Genotype

The power to detect a QTL depends on the variance explained by the single QTL in relation to the total variance: $\sigma_{exp}^2 = \sigma_q^2 / \sigma_t^2$. As described above, the population type determines the expected additive genotypic variance of a given additive effect and, consequently, this affects the detection power. In addition, the precision with which the QTL genotype means are estimated will depend on the population size, with large populations increasing the probability to detect the QTL. From various studies it is clear that QTLs with an explained variance of 10% stand roughly an 80% chance of being detected in a population of 200 plants, the probability of detection decreasing more or less linearly with smaller population sizes ***(18–22)***. Thus, if one is only interested in very large-effect QTLs one might use populations of fewer than 200 individuals; if the expected explained variance per QTL is less than 10% one should resort to larger-sized populations. Another option to increase detection power is to reduce the environmental variance by increasing the number of observations per genotype, *i.e.*, to grow several plants per genotype, ***(18,23)***; the environmental variance of the genotype mean is $\sigma_{em}^2 = \sigma_e^2 / n$, with n being the number of plants per genotype. Keeping the same final number of individual plants analyzed for the trait of interest, the power of an experiment may be enlarged either through population size, *i.e.*, by increasing the number of genotypes (mapping lines or families), or by taking more plants per genotype. As far as the power of QTL detection is concerned, it is better to double the number of genotypes than phenotyping two plants per genotype. How-

ever, the mapping populations used in QTL analyses are often previously developed elsewhere, and increasing their size implies an additional extra cost involved in the genotyping of more genotypes.

3.1.2.2. Marker Spacing

The power to detect a QTL depends little on the density of markers employed because reducing marker spacing below 10 cM does not increase the probability of detecting a segregating QTL ***(20,21)***. When preparing a QTL analysis on a new mapping population, all marker genotypes are still to be established. If possible, markers should be chosen as an evenly distributed sample, based on known linkage map positions. Because map distances vary across populations and experiments, these should be calculated and verified for the population under investigation, and the resulting map should be employed in the subsequent analysis.

3.1.3. Multiple QTL Models

Nowadays, most QTL analysis software packages have the option to fit approximate multiple QTL models, usually called composite interval mapping ***(24)*** or multiple QTL model (MQM) mapping ***(25)***. These methods increase the QTL detection power compared with the standard IM approach. In a multiple QTL model, the search for a single QTL is enhanced by taking into account in the statistical model all the detected QTLs, which is done by using as cofactors markers closely linked to the QTLs. In this way, the total variance determined by other QTLs is reduced with the cofactors and, consequently, the relative variance explained by the searched QTL is increased. Applying a multiple QTL model has effect only when the detected QTLs determine a large part of the total variance, such as 10% or more. Using marker cofactors that do not relate to a detected QTL, as is sometimes done in composite interval mapping, can negatively affect the power to detect a QTL ***(22)***. Therefore, it is advised to use in the final model only markers that are linked to positions where QTLs are concluded to be present.

When mapping QTLs, there is the possibility of having two linked QTLs. If their effects are in the same direction, the QTLs are said to be in *coupling phase*. In this case, it is very likely that, at first instance, a single QTL is concluded to be segregating between the two actual QTLs; this is called a ghost QTL ***(26)***. If the QTLs are counteracting each other, the QTLs are said to be in *repulsion phase*, and the QTLs may go undetected. Detection of linked QTLs can be improved when using some automatic cofactor selection procedure, in which the statistical process of backward elimination should play an important part.

3.1.4. LOD Thresholds

After the LOD profile has been obtained with (composite) interval mapping, one needs to establish whether the maximal values of the profile are statistically significant, i.e., to establish a significance LOD threshold. Although an individual LOD score follows an approximate chi-square distribution under the null hypothesis, the chi-square distribution cannot be used for interval mapping because many interdependent tests are being performed (the LOD is calculated at many positions on all linkage groups). The LOD threshold will depend on the number of linkage groups, the type of mapping population, the total genetic length, the number of marker intervals, and the phenotypic distribution of the trait. To solve this, basically three solutions have been derived: (1) to use some analytical formula ***(20,27)***; (2) to use the permutation test ***(28)***; or (3) to use the results of simulations ***(29)***. The analytical formulae are not very accessible, and therefore they are not employed very often. The permutation test is a time-consuming method that repeatedly resamples the phenotypic data (and thereby destroys any possible linkage between markers and QTL) and subsequently performs interval mapping to determine the maximum LOD for that random sample. After at least 1000 such iterations (but preferably 10,000 or more), the distribution of the maximum LOD score is obtained and from this, the significance threshold can be derived. The use of the permutation test is often advocated because it is based on the actual experimental data, which do not need to behave accord-

ing to the normal distribution (after fitting the QTL), an assumption of the analytical formula and of the LOD score. However, the robustness of the LOD against deviations from normality has been shown by a simulation study using a nonnormal (skewed) distribution ***(30)***. Therefore, use of the third method might offer the best compromise because despite the fact that its application might be biased due to violation of the normality assumption, it relies on millions of simulations. Thus, the large number of simulations will compensate for accuracy, especially when compared with applying the results of just 1000 permutations.

3.1.5. QTL Localization: Obtaining Confidence Intervals

So far we have only discussed QTL detection; another important aspect is the actual mapping of the QTL, i.e., the QTL localization. Although detection and localization are more or less confounded, one could say that QTL detection only proves the existence of a QTL on a certain linkage group (it is a limited form of localization), whereas the actual position within that linkage group must be derived from the linkage group's LOD profile. The point estimate of the position of a QTL is the map position at which the highest LOD was reached. For all practical purposes a simple point estimate is insufficient because one wants to have a reasonable certainty that the map region used in future work (e.g., fine mapping, introgression, positional cloning) does indeed contain the QTL. Therefore, a confidence interval is needed, i.e., a genomic interval that will include such a QTL with a certain probability, usually 95%.

There are four main techniques to obtain a confidence region, although none of them provides a confidence interval in the strict statistical sense: (1) to take plus and minus two times the estimated standard error of the QTL position; (2) to apply a simulation, based on the current estimates of QTL position and effects (the so-called parametric bootstrap); (3) to apply a (nonparametric) bootstrap technique; and (4) to obtain an LOD support interval. All four methods suffer from the problem that the so-called LOD profile on a linkage group actually is a contiguous set of LOD profiles corresponding to each interval flanked by two markers; this is illustrated by the cusps in the LOD profile charts at the marker positions. The first technique produces intervals that are too optimistic (i.e., too narrow), their actual confidence being much lower than the intended 95%, and therefore it is not advised. As a replacement, an analytical formula with a proper realized confidence has been proposed for a backcross population ***(31)***; similar formulae may be derived for other population types. Simulations also produce too-narrow confidence intervals, especially when the genetic effect and the population size are large. The nonparametric bootstrap performs well, although there are some unsatisfactory differences in behavior between QTLs detected on markers and QTLs detected between markers ***(32)***. The LOD support interval methodology is the easiest to establish; for example, a 1-LOD support interval is constructed by taking the two positions left and right of the maximum value in the LOD profile that have an LOD value of 1 less than this maximum. The confidence of the LOD support intervals varies with size of the QTL effect, population type, and population size. Theoretically, a 1-LOD support interval corresponds to an approx 97% confidence level, although in practice this is about 90%; increasing the support to 1.5-LOD should provide about 95% confidence for population sizes between 200 and 400 ***(19,20)***, but drops to 90% confidence for smaller population sizes of 100 plants.

3.1.6. Experimental Factors Affecting QTL Localization

The factors that increase the power to detect a QTL also increase mapping accuracy because the larger the QTL genotypic variance in relation to the residual variance, the smaller the size of the confidence interval. Increasing the population size will increase the precision with which the QTL genotypic class means are estimated, and, thus, the mapping accuracy. In addition, the population type will affect the resolution by its effect on the genetic variance. Choosing an F2 instead of a backcross enhances the resolution, and taking RILs or doubled haploids increases the genetic variance even more. However, the population type also affects mapping resolution through the level of recombination; in RILs at advanced generations, the level is nearly twice that

of that in F2 or BC populations *(9)*. This will result in reduced association (i.e., lower LODs) of markers located at certain distance away from the QTL, whereas the most closely linked markers to the QTL will retain high LODs; therefore, the LOD profile will contain a sharper peak and the support interval becomes smaller in length. Dupuis and Siegmund *(20)* argue that the size of the confidence interval is inversely proportional to the level of recombination; this means that using RILs in the experiment will lead to confidence intervals of half the length of those obtained when using a DH under an identical level of genotypic effect. Advanced intercross lines have even higher levels of recombination, and thus a higher mapping resolution. For populations smaller than 200 individuals, marker spacing below 10 cM doesn't have much effect on the size of the confidence interval, except for situations with high levels of recombination. In those cases, such as with advanced intercross lines, a 5-cM spacing is recommended to prevent loss of power *(20–22)*. Nevertheless, as a matter of QTL verification, a marker density of 5 cM is also advisable in other population types, within the regions of the support intervals.

3.2. Selection of Parental Accessions for QTL Mapping: General Quantitative Genetic Analysis

When initiating the analysis of natural variation for a trait of interest, it is useful to first explore the overall amount of variation existing for the trait(s) and its heritability. In setting up the subsequent QTL analysis experiment one will lack important information necessary for making decisions about the size of the mapping population because it is not known how many QTLs will segregate, whether they are genetically linked (and if so, how), and how large their genotypic effects are. Performing a preliminary general quantitative genetic study with available accessions will determine whether a QTL analysis could be successful and it will improve the choice of the parents and the design of the resulting QTL mapping experiment. To do this:

1. Set up the appropriate quantitative assay to measure precisely the trait(s) of interest (*see* **Notes 1** and **2**).
2. Grow, under the appropriate assay conditions and in a random design, 4 to 20 individuals per accession of a collection of wild accessions, and measure them for the trait(s) of interest (*see* **Notes 3–6**).
3. Enter the quantitative trait data in a spreadsheet, estimate for each trait the mean and standard deviation per accession, and analyze the phenotypic frequency distribution of the trait(s) (*see* **Note 7**).
4. Perform a quantitative analysis of the data and estimate the overall genetic variance, the environmental variance, and the heritability of the trait (*see* **Note 8)**. If several traits have been measured in the same assay, correlations between the traits can be estimated to determine whether they have a common genetic basis (*see* **Note 9**).
5. The study can be completed by repeating the analysis under different environmental conditions, which might identify more suitable environments and/or accessions that respond differently to the environment (*see* **Note 10**).
6. Choose the two parental accessions (*see* **Notes 11** and **12**) and design the experimental setup (quantitative assay, environmental conditions, population type, population size, and the number of replications per genotype) for the QTL analysis (*see* **Notes 4–6**, **8**, **10**, **13**).

3.3. Construction of QTL Mapping Populations

When the wild accessions chosen for QTL analysis are not included in any of the existing mapping populations (*see* **Note 3** and **Table 1**), a new mapping population must be developed, generally as an F2, a backcross or a family of recombinant inbred lines. To do this:

1. Grow together the two parental accessions so that they flower at the same time, make a (reciprocal) cross between them by hand-pollination, and harvest the F1 generation seeds (*see* **Notes 14** and **15**).
2. Grow the F1 plants, check with morphological or molecular markers to determine whether they are the expected result of the intended cross (*see* **Note 16**), and self them to obtain the F2

population seeds. To develop a BC population, cross the F1 plants to one (or both) of the parents (*see* **Notes 13–17**).

3. To obtain an RIL population, grow the F2 population and harvest the F3 self-progeny seed of each F2 individual plant (*see* **Notes 18** and **19**).
4. Grow one (or several) randomly chosen F3 seed derived from each F2 individual and harvest the F4 self-progeny seed (*see* **Note 18**).
5. Repeat the previous single-seed descent procedure up to the F8/F9 generation. Thereafter, plants are practically homozygous and become lines that might be bulk-harvested (*see* **Notes 18**, **20**, **21**).

3.4. Development of Genome-Wide Linkage Maps

After a new mapping population has been constructed, it needs to be characterized genetically with molecular markers to obtain its molecular linkage map. The genotypic information of markers distributed throughout the genome, and ordered according to the linear DNA structure of the five *Arabidopsis* chromosomes, will provide the frame for QTL mapping. General information on the estimation of recombination and linkage mapping is provided in Chapter 6 of this book. To develop a genome-wide genetic map:

1. Grow the BC, F2, or RIL population and the corresponding parental wild accessions, and isolate the DNA of the individual plants (*see* **Note 22**). To propagate the mapping population, obtain and harvest the self-progeny seed of all the individuals (*see* **Note 23**).
2. Determine the marker density desired for the linkage map, search for potential markers, and test the parental accessions for the DNA polymorphisms that will be used as molecular markers (*see* **Notes 24–26**).
3. Analyze the DNA of all the individuals of the mapping population for the molecular markers.
4. Score the genotype of all the individuals of the mapping population for each analyzed marker, taking into account the type of marker, the parental origin of the segregating alleles (the linkage phase), and doubtful or missing data (*see* **Notes 27–29**). Name the marker genotypes according to the instructions of the linkage mapping software to be used (*see* **Table 1**).
5. Enter the individual genotype scores of the molecular markers in a spreadsheet and build the appropriate locus genotype data file for the corresponding mapping software.
6. Estimate the recombination frequency and its statistical significance for all pair combinations of markers (*see* **Notes 30–32**).
7. Perform a segregation analysis for each marker and each individual of the mapping population, to test if the observed genotypic frequencies fit the expected segregation ratios. Use these analyses and the recombination frequencies to inspect the locus data file for possible errors (*see* **Note 33**).
8. Assign the markers to different linkage groups on the basis of the recombination frequencies and statistical significances. A marker is assigned to a particular linkage group when it shows significant linkage to any marker belonging to that group (*see* **Notes 31** and **34**).
9. Analyze each linkage group separately to calculate its genetic map, following the instructions of the linkage software used. The linear order of markers and the map distances between adjacent markers are determined by comparing the recombination frequencies with their statistical significances (*see* **Notes 35** and **36**).
10. Inspect the genetic map of each linkage group for inconsistencies, and recalculate the maps after correction of detected errors (*see* **Note 37**).
11. The linkage map might be completed by including new markers within the genetic intervals flanked by adjacent markers distanced more than 10 to 15 cM (*see* **Notes 25** and **38**).

3.5. Mapping QTLs

A mapping population for which its genetic map has been developed can be used to identify and locate QTLs accounting for the variation present in the population, for the trait(s) of interest. To do this:

1. Grow in a random design and under the appropriate test conditions, 4 to 10 individuals of the following genotypes: each RIL or the self-progeny of each F2 or BC individual; the parental

accessions; the reciprocal F1 hybrids derived from the parental accessions (*see* **Notes 3–6**, **13**, **19**, **23**, **39**).
2. Measure the individual plants for the trait(s) of interest (*see* **Notes 1**, **2**, **4**).
3. Enter the quantitative trait data in a spreadsheet, estimate the mean and standard deviation for each genotype, and analyze the phenotypic frequency distribution of the trait (*see* **Notes 7** and **40**).
4. Perform a "classical" quantitative analysis of the data to estimate the heritability of the trait(s) in the mapping population (*see* **Note 8**), and the overall dominance and the maternal genetic effects of the trait(s) (*see* **Notes 41** and **42**). Correlations between traits measured in the same mapping population might be estimated (*see* **Note 9**), which in the case of permanent RIL populations may include data publicly available from other experiments (*see* **Table 1**).
5. Combine the quantitative data file with the molecular marker and linkage map files according to the instructions of the QTL mapping software to be used (*see* **Note 43**).
6. Process the data of each trait separately to identify the QTLs affecting each trait of interest by using single-marker mapping, (composite) interval mapping, or any other QTL mapping method implemented in the software used (*see* **Notes 44–46**). Determine the map position (QTL support intervals), the additive allele effect, and the percentage of explained variance of the individual QTLs detected; estimate the total percentage of the phenotypic variance explained by all QTLs (*see* **Notes 25**, **47–51**).
7. Analyze the epistatic interactions among QTLs and perform a comparative QTL analysis to identify QTL with putative pleiotropic effects (*see* **Notes 52** and **53**).
8. The study can be completed by repeating the QTL mapping analysis under different environmental conditions to identify QTLs that interact with environments (*see* **Notes 54** and **55**).

4. Notes

1. In a broad sense, a quantitative trait is any trait measurable on a quantitative scale. This can be a continuous scale such as length, or a discrete ordinal scale such as, for instance the number of leaves or infection spots. The measurement should avoid subjective classifications. Standard QTL analysis software uses the assumption of (statistical) normality, which hampers the correct analysis of data on ordinal scales including binary data. Averaging multiple scores on an ordinal scale usually generates sufficient normality for a correct analysis.
2. When setting the assay, it is advisable to take into account precise environmental, temporal, and spatial conditions to reduce variation that might come from the environment, the age, the developmental stage, or the particular organ of the plants. When analyzing a mapping population, always include the parents of the cross to estimate their trait values under identical conditions as the experimental population.
3. Currently, there are around 750 different wild accessions of *Arabidopsis* (originally coming from more than 300 locations) publicly available in stock centers (*see* **Table 1**). To estimate the amount of variation that exists within the species for the trait(s) of interest, one should grow as many accessions as possible. However, to efficiently reduce the amount of work, one might select a limited number of accessions that can be chosen according to different criteria: (a) the 60 parental accessions of the current permanent mapping populations (www.inra.fr/qtlat/NaturalVar/RILSummary.htm) or the reference collection of 96 accessions used for LD studies (http://naturalvariation.org/); (b) a set of accessions chosen on the basis of nucleotide diversity, such as the 8 to 48 core collections established based on DNA sequence data ***(33)***; or (c) a partial collection selected according to geographical and ecological information, presumed to be relevant for the trait(s) of interest.
4. The standard error of the trait mean value of a genotype is inversely proportional to the square root of the number of observations. Thus, using measurements of several individual plants per genotype will reduce the error of the means. When the measurement of the trait(s) is laborious it should be kept in mind that working with homozygous lines such as wild accessions or RILs, more than four replications will improve the accuracy of the measurement only marginally (*see* **Note 8**). Alternatively, some traits might be measured on bulks of several sister plants, such as, for instance, in biochemical assays. Using a single large bulk per genotype prevents the estimation of the variance within a genotype but it will reduce the environmental effects; two or more bulks per genotype allow the residual variance to be estimated.

5. In order to obtain unbiased estimates of means and variances, plants should be grown in random designs such as a fully randomized design or a randomized block design. Growing the plants of the same genotype close to each other might lead to underestimations of the environmental variance effects and overestimation of the genetic variance and heritability. In case of a very large number of genotypes, one might have to use an incomplete block design.
6. For traits with presumably large maternal effects (such as, for instance, seed germination, seed chemical content and seedling-related traits), the environmental influence on those effects should be reduced by growing all maternal material within the same environment.
7. A careful inspection of raw data and of the frequency distribution may identify unusual values of the trait (outliers) that can affect the results of statistical analyses. These are often due to misscoring or defective sampling (and should be corrected or discarded), but may also be the result of a rare combination of alleles.
   For the analysis, most statistical tests of quantitative genetics and QTL mapping (parametric tests) assume (a) a normal frequency distribution of the trait residuals, i.e., after fitting the QTLs (if there are large-effect QTLs the phenotypic distribution will not be normal, but the analysis will be correct if the residuals are normally distributed); (b) that the residual error is independent from the mean, i.e., large values have the same error (on average) as small values. There are several kinds of observations for which these assumptions do not hold and diagnostic tests should be performed to assess these requirements. For example, proportions (or percentages) close to 1.0 can never be larger than 1.0, whereas those around 0.5 can vary up or down; in time-related events, such as flowering time, early events have low variation, whereas late events have much larger variation. Such data are to be transformed before analysis; usual transformations are (a) square root transformation $\left[\sqrt{(x+1)}\right]$ for counts of rare events or when the variance is proportional to the mean (μ / σ^2 is constant), (b) angular transformation (i.e., $\arcsin\left[\sqrt{(x)}\right]$) for proportions, and (c) logarithmic transformation [log (x + 1)] for when the standard deviation is proportional to the mean (μ / σ is constant). Alternatively, one might use nonparametric statistical tests for their analysis, as it is usually applied to ordinal data such as disease rates on an (arbitrary) scale. Interval mapping with ordinal data may sometimes reveal false-positive QTLs (*see* **Note 1**).
8. Such quantitative analysis might consist in an ANOVA using genotypes (accessions or mapping lines) as a random effect factor, preferably using restricted ML (REML) in a general statistical software package. With this analysis the genetic variance (σ^2_g) and the environmental variance (σ^2_e) can be estimated; from them, the broad sense heritability of the trait (on a single individual basis) (h^2) can be calculated as $h^2 = \sigma^2_g / (\sigma^2_g + \sigma^2_e)$. The heritability of genotype mean values is estimated as $h^2 = \sigma^2_g / [\sigma^2_g + (\sigma^2_e / n)]$, which increases with the number of biological replicates (n) (individuals measured of the same genotype or line). For easier comparability, it is advisable to report the heritability based on single individuals (i.e., $n = 1$).
9. The individual and the mean values can be used to calculate the Pearson (i.e., the standard linear) correlation coefficients between traits. Significant correlations might be due to pleiotropic effects of some QTLs on several traits, or to linkage disequilibrium between different QTLs affecting the different traits (which might due to close genetic linkage or to other factors operating during the evolutionary history of the wild accessions). Both possibilities will be distinguished by subsequent QTL mapping, although pleiotropy and close linkage can be discriminated only by QTL fine mapping or, in the case of tight linkage, by cloning the QTL(s).
10. The expression of a trait may differ across environments ("environments" is used here in the widest sense possible). Measuring traits in different environments may help to establish the ideal environment for doing the QTL analysis (i.e., the environment with the largest variation). Alternatively, one might be especially interested in the interaction between QTLs and the environment, allowing the detection of the genotypes that exhibit the largest (or most interesting) interaction. In addition, one might find an environment in which the variation for the trait of interest derived from genetic variation in other traits is reduced. For instance, accessions differ considerably in the timing of flowering, which might indirectly affect many other traits. Growing the accessions under short-day photoperiod conditions will increase the duration of the vegetative phase and might reduce variation in traits measured in this developmental phase. Genotype by environment interaction should be tested with two-way ANOVA.

11. The choice of parents for QTL analysis is based mostly on their large phenotypic differences, ideally corresponding to the extremes from the phenotypic distribution of the accessions studied. However, it is also advisable to take one of the common laboratory strains, Columbia (Col), Landsberg *erecta* (L*er*), or Wassilewskija (Ws), as one of the parents, because more experimental resources might be available. In addition, a reference genetic background is provided in this way, which allows comparison of the effects of QTLs and mutant alleles. To accelerate the construction and analysis of QTL mapping populations, late flowering or very dormant accessions should be avoided as parents, unless these are the traits of interest.
12. If accessions carry alleles with opposite effects at several QTLs affecting a trait, their joint effects may result in a similar phenotype, and consequently this genetic variation will go undetected when studying the accessions directly. For this reason, the analysis of segregating populations may reveal the presence of individuals with phenotypes outside the range of the parental accessions (transgressive genotypes). Transgression may determine the choice of parents for QTL analysis, by revealing most of the genetic variation.
13. Decision on the population size and the number of replicates aimed to eliminate environmental effects (and maximize heritability) in an assay for QTL analysis might be assisted using the estimates of environmental variance and the total genetic variance derived from the two accessions selected as parents (*see* **Subheading 3.1.2.** and **Note 8**). Roughly, population sizes of 200 RILs without replications will allow detection of large-effect QTLs with an explained variance above 10% in confidence genetic intervals of 10 to 20 cM. Detection of small-effect QTLs or mapping accuracy below 5 cM will require increasing the population size to at least 300 RILs. Similar sizes of F2 or BC populations will be considerably less powerful.
14. Although *Arabidopsis* is practically a strict autogamous species, one cannot be absolutely sure of the sister plants from the same accession being fully homozygous and/or genetically identical. For this reason, always use a single plant of each accession to generate the F1 hybrid and the BC populations, and generate the BC and F2 populations from a single F1 plant. Thus, it is ensured that one allele per locus from each parent will segregate in further generations. To be able to test whether there is a significant maternal cytoplasmic effect on the trait(s) of interest, it is suggested to use both parental accessions in reciprocal crosses, to derive F1 hybrids and BC and F2 populations with both maternal cytoplasms.
15. The number of viable seeds obtained in crosses between wild accessions is usually as high as that of crosses within an accession (and self-pollinations), ranging between 10 and 50 seeds per fruit. Nevertheless, some crosses among wild accessions are almost sterile in one direction but fertile in the reciprocal way, and some accessions can even behave as incompatible (*34*). This should be taken into account when planning the number and direction of crosses needed to obtain sufficient F1 and BC seeds.
16. F1 hybrid plants derived from crosses between accessions can often be distinguished morphologically (color, shape, and size of whole plant and organs, flowering time, etc.) from both parents. However, it is advisable to test using molecular markers that the seed harvested is indeed the F1 and not the product of selfing, unwanted crosspollination, or seed contamination. For that, isolate the DNA of the F1 plants from a small sample of tissue and analysed a limited number of molecular markers. Using just a few (3–5) multiallelic markers (such as micro- satellites) enables a rapid confirmation of the pedigree of the F1 plant (the more alleles a marker shows among accessions, the higher chance to detect a unique combination of alleles).
17. F2 populations allow the estimation of QTL dominance, whereas homozygous RIL populations do not. BC populations enable only the mapping of QTLs at which the recurrent parent does not carry a fully dominant allele. Developing two BC populations using both parental accessions as backcross parents will enable the detection of QTLs independently of their dominance. Furthermore, significant differences between male and female meiosis have been found in *Arabidopsis* (*35*). Using the F1 parent as mother and father in reciprocal crosses to obtain the BC populations would average possible recombination differences between sexes.
18. When growing the F2 population and the subsequent RIL generations, the acquisition of the progeny seed may be affected by an unknown level of "natural" selection, even under very controlled conditions. For instance, this might operate through certain degree of sterility or poor growth of some lines. In addition, one should take all possible precautions to avoid any

artificial selection because this will lead to segregation distortion in related genomic regions. Distortion from the expected Mendelian segregation ratios will reduce the effective population size and, as a consequence, the power for QTL analysis in those affected regions. For instance, nonrandom selection for flowering time and seed dormancy can easily and unintentionally be directed toward early flowering and fast germination, respectively. Therefore, for the production of an RIL population, it is advisable to grow several sister plants of every line in each generation, to be able to compensate for accidental losses of individual plants, and to use a random system (e.g., pot position) to determine which plant(s) will be chosen for the subsequent generation.

19. At this stage, other types of mapping populations might be initiated, such as advanced intercrossed lines or advanced backcross populations, which differ in the amount of recombination and background genetic variation (*see* **Subheading 1.**). Furthermore, when the mapping population is aimed for QTL analysis of only one particular trait, phenotyping the population and selecting the 25% individuals from both extremes might reduce the cost of the QTL analysis by reducing the number of individuals to genotype (so-called selective genotyping; ***36***). Selective genotyping will lead to an overestimation of QTL effects, but it will increase the power to detect QTLs in relation to the number of genotyped plants ***(36)***. Selective genotyping is not practised very often because mapping population are developed mostly for the analysis of, at least, several related traits; then, the selection of plants for the various traits will not coincide, reducing its advantage. In addition, multiple QTL analysis (composite interval mapping) (*see* **Note 46**) is hampered by selective genotyping.
20. After the F8 generation, single-plant propagation and bulk propagation (i.e., harvesting together several sister plants) of the RIL population will lead to "fixation" of the very low residual heterozygosity (on average, 0.8% per locus: $[1/2]^{g-1}$ with g being the generation number). However, bulk propagation during successive generations will be slower in fixing the genotypes and therefore will lead to populations with slightly higher heterozygosity compared with single-plant propagation.
21. The single-seed-descent procedure of RIL propagation might be finished at the F5 or F6 instead of the F8 generation to maintain an increased heterozygosity per locus in the population. This residual heterozygosity of RILs will facilitate the future development of near isogenic lines (NILs) from heterogenous inbred families (HIFs) ***(11)***. At the F5 generation most *Arabidopsis* lines will be heterozygous for one genomic region and small population sizes such as 100 will provide parents to obtain HIF and NILs for almost any locus (1/16 = 6.2% of the RILs will be heterozygous per locus).
22. Using standard protocols for DNA isolation from single plants (*see* Chapters number 11 and number 14) will preserve the plants alive and will render 1 to 5 µg of DNA from a small amount of tissue consisting of 1 or 2 inflorescences of flower buds and 1 or 2 leaves. This will be enough to genotype hundreds of markers because most current molecular marker technologies require very small amount of DNA per marker (*see* Chapter number 12). A larger amount of DNA is needed from the parents to perform the polymorphism screening; in this case, DNA might be isolated from whole plants or bulks.
23. The genotyping and phenotyping of the mapping population often cannot be performed on the same individual plants and progeny tests might be required. Homozygous RIL populations are usually genotyped at the F6–F8 generation (*see* **Note 21**), whereas the quantitative analyses of traits are performed in the following generations. In F2 and BC populations where heterozygous individuals are segregating, one might choose to either genotype or phenotype the self-progeny of the population. Genotyping in the self-progeny will require pooling tissue or DNA from at least 16 or 7 sister plants per F2 or BC plant, respectively, to distinguish with 99% certainty between a progeny that derives from a heterozygous genotype and a progeny from a homozygous plant ($0.75^n > 0.01$ at $n = 16$ for a dominant marker in an F2 population; $0.5^n > 0.01$ at $n = 7$ for any marker in a BC population). Phenotyping the progeny will increase significantly the family mean-based heritability (*see* **Note 8**) ***(18,23)***.
24. For standard *Arabidopsis* populations of fewer than 200 individuals, 50 to 100 markers evenly distributed are sufficient to detect and locate QTLs, providing a marker density of 1 marker per 10–5, cM respectively (the average *Arabidopsis* genetic map length is 500 cM). Databases including more than 30,000 *Arabidopsis* polymorphisms with known physical and genetic

positions among various accessions (*see* **Table 1**) provide potential PCR-based markers such as microsatellites, CAPSs, or SNPs. From these databases a set of markers can be developed with an even distribution throughout the genome, including markers at the telomeres and centromeres of the five linkage groups, which will define the upper and lower ends of every chromosome arm. Alternatively, other marker techniques can be used without *a priori* knowledge of the map position, such as amplified fragment length polymorphism (AFLP) or some high-throughput protocols; in this case, the number of markers to be analyzed to ensure the same marker density will be larger. For instance, at least 113 randomly distributed markers should be scored to obtain a marker every 10 cM with a 99% chance [$n = \ln(1-p) / \ln(1-(2c/L))$ provides a rough underestimation of the number of markers, n, needed to obtain markers within an interval size c in cM, with a probability p, in a map of total length L; ***8***).

25. Because map distances may vary across experiments due to random variation and to genetic variation for recombination, a stepwise approach could be worthwhile (depending on the logistics and expenses of the marker technique). A preliminary map might be calculated starting with markers sampled at 10 cM from the known genetic map (which would correspond to average physical distances of 2.4 megabases); thereafter, the marker density might be adjusted to 10 cM or less in regions where gaps in the current map occur. After a subsequent preliminary QTL analysis, the density in the map can be further increased in genomic regions with (near) significant LODs.
26. In F2 populations where the heterozygosity is high, dominant markers, such as those obtained with some AFLP procedures, are less informative than codominant markers, such as microsatellites and CAPSs. In BC populations, dominant markers with the recessive allele derived from the recurrent parent are the only informative dominant markers, although these will provide the same information as codominant markers. In addition, dominant markers are almost equally informative as dominant markers in RIL populations advanced to the F8 or higher selfing generation, where the heterozygosity might be negligible.
27. The number of genotypic classes will depend on the type of mapping population and marker. Codominant markers will have three classes in F2 and RIL populations and only two in BC populations. Informative dominant markers will show only two genotypic classes in all populations.
28. Genotypes should be named and scored according to the parental origin of the alleles, so alleles derived from the same parent should all have the same code to provide the correct linkage phase (coupling or repulsion).
29. Doubtful genotype scores should be genotyped again or recorded as missing observation to reduce the amount of errors, which will overestimate the length of genetic maps. In addition, despite their low frequency, heterozygous RILs should be scored as heterozygous instead of missing, to make use of all the information. Although interval mapping methods have the ability to deal with missing marker observations, this should not be used as an excuse to have many of them; on the contrary, having few missing marker scores enhances the quality of QTL localization.
30. Genetic linkage between two markers is measured by the recombinant frequency (r) among them (r = number of recombinant gametes/total number of gametes). The recombination frequency among pairs of markers and their standard errors can be calculated by hand, using the equations estimated for most genetic situations (*see* Chapter 6 in this book). However, because the number of marker pairs rises with the factorial of the number of markers, this soon becomes unmanageable by hand. Therefore, to estimate the recombination frequencies involved in building a genome-wide linkage map and to calcualte the map, mostly a specific mapping software package is used (*see* **Table 1**).
31. Statistical information on the linkage between two markers is usually expressed with an LOD value, which is the 10-base logarithm of the odds ratio:

    LOD = log[(Likelihood of observed data with $r = \hat{r}$) / (Likelihood of observed data with r = 0.5)]. For instance, an LOD score of 5 means that for the given data, the true recombination fraction is 10^5 = 100,000 times more likely to be equal to the estimated $r = \hat{r}$ than to be equal to 0.5 (unlinked or independent). Originally, LOD scores larger than 3 were taken as evidence for linkage; current experience shows that this threshold should often be considerably larger. Statistical information on linkage can also be provided by the standard error associated with

the $\hat{r}$ estimate. However, LOD values are more useful when the recombination estimate equals zero because they take into account the population size. For instance, if a value of $\hat{r} = 0$ is estimated in a BC population, i.e., from a binomial distribution sample, its variance formally equals zero irrespective of the population size. In contrast, the LOD score is linear with the population size (e.g., LOD = *n*log(2) for *n* plants in a backcross) and therefore, finding no recombinant in a population of 100 individuals will be more informative than finding no recombinant in a population of 10. Nevertheless, a problem of the linkage LODs is that they are sensitive to segregation distortion, i.e., large LOD values can be obtained when there is no linkage due strong segregation distortion. Therefore, it is advised to test for linkage using the chi-square test for independence in a two-way contingency table of pairwise marker frequencies (markers located on nonhomologous chromosomes segregate independently, unless there is a specific selection on both markers simultaneously); this test can also be taken as a likelihood ratio test in the form of a so-called G^2 statistic, which may be translated to the LOD scale and presented as a LOD.

32. In a practically homozygous RIL population (at the F8 generation and onward), the observed fraction of recombinant (nonparental) RILs between two markers (R) will be higher than the recombinant frequency (r) in a single meiosis such as that of F_2 or BC populations. Estimates of r can be obtained manually using the following Haldane and Waddington equation ***(37)***:

$$r = \frac{\hat{R}}{2-2\hat{R}}\text{; the standard error of } r \text{ being } sd(r) = (1/2)\sqrt{\frac{\hat{R}}{N(1-\hat{R})^3}}$$

N is the total number of RILs analyzed. For closely linked markers ($r < 0.05$), $\hat{R}$ is twice the r value.

33. Errors in the scores of individual genotypes will lead to wrong marker orders and distances. Some of these may be detected before the start of the actual map construction, by performing a careful check of the locus data file for possible errors from different sources. Include the following inspections: (a) Check whether there are individuals of the mapping population identical to the parental lines or whether there are identical individuals within the population, which are likely to be contaminant plants or DNAs. (b) Test the segregation ratio of the marker genotypes in each individual of the population to detect unexpected segregations (such as, for instance, a highly heterozygous RIL or a parent-like individual); individuals with distorted ratios might be due to undesired outcrossing or to external plant or DNA contaminations. (c) Test the segregation ratio of each marker in the population to detect unexpected ratios that might be due to misscoring and not to true segregation distortion (*see* **Notes 18** and **31**). (d) Check the recombination frequencies among all pairs of markers searching for suspicious high values that might be due to an error in the coding of one marker leading to the wrong linkage phase.

34. Unassigned markers, i.e., markers genetically independent from the rest, are separated from further analysis. These might be added manually to particular linkage groups if one knows their presumed genetic position. However, caution should be taken when adding markers manually to specific linkage groups, because that might limit the detection of possible chromosomal rearrangements, such as translocations, that have been reported among *Arabidopsis* accessions ***(38)***. In addition, the expected location of a marker may vary between crosses because some techniques may pick up different segregating DNA fragments in different crosses.

35. Missing and erroneous marker scores and double recombination events complicate the procedure to calculate a map (*see* **Notes 33** and **37**). Various, quite distinct algorithms have been developed to calculate a map. Given a map order, its probability is usually calculated based on maximum likelihood or on a regression model, and various orders can be compared with likelihood or some goodness-of-fit measure. Because the enormous number of possible orders determined by the usual large numbers of markers (½ factorial of the number of markers [*m*], 0.5*m**!,; e.g., with 50 markers there are 1.5×10^{64} possible orders) it is virtually impossible to test all map orders. Therefore, there are various different approaches to limit the search among all possible orders to those that are more likely; thus, the calculations are reduced to reasonable limits. With MAPMAKER this search is guided more or less manually, whereas with Joinmap this search is automated. Graphical representations of linkage maps can be produced by specific graphic software such as MapChart (***39***; www.biometris.nl/uk/software/MapChart).

36. Recombination frequencies among neighboring markers are not additive due to the existence of double crossovers. For this reason, frequencies are converted in additive map distances in Morgans or centiMorgans (cM) using the mapping functions. The best mapping function is that one giving the best additivity among the calculated map distances, although usually the Haldane ***(40)*** (assuming no genetic interference) or Kosambi ***(41)*** (assuming certain interference) mapping functions are used (*see* Chapter 6 in this book).
37. Markers might be sorted in the wrong order due to errors from different sources through the mapping procedure (*see* **Note 33**). In the final map of each linkage group, we may carefully compare the estimated map distances and the corresponding recombination frequencies between each marker and the linked markers. Strong inconsistencies between both parameters might indicate wrong marker orders that should be evaluated. In addition, the genotype of all individuals of the mapping population can be analyzed according to the marker order of the final genetic maps to detect possible individuals carrying unlikely double crossovers between closely linked markers. Such apparent double crossovers might be due to genotyping errors. Some mapping programs, such as Joinmap, include automated tests to verify the calculated maps, which strongly facilitates this task.
38. Permanent RIL populations can be used to locate DNA sequences or markers whose genetic position cannot be found in the existing genome sequence obtained from the Col accession. A polymorphism between both parental accessions should be found to proceed as in **step 2** in **Subheading 3.4.**
39. When the analysis of the trait(s) of interest is laborious and/or expensive one might reduce the population size by selecting a subset of RILs, or BC or F2 families, carrying more recombination events than average, evenly distributed. Reducing the population size through selection of the individuals with more recombination will reduce QTL detection power but will limit the reduction in QTL localization power. This selection may be facilitated using certain software developed for graphical representation and analysis of mapping populations, such as GGT (***42***; www.dpw.wau.nl/pv/pub/ggt/).
40. Occasionally, some traits might show a phenotypic frequency distribution that does not fit a normal distribution but shows several distinct phenotypic classes and modes, such as, for instance, some pathogen resistances or certain flowering time variants. These traits are not quantitative traits in a strict sense because they do not have a continuous phenotypic distribution. Such *qualitative* distributions might fit monogenic or digenic segregations that can be analyzed by Mendelian genetics (*see* Chapter 6 in this book) to map the major qualitative trait locus determining the observed variation.
41. The maternal genetic effects on the trait(s) and the overall dominance of the trait(s) can be estimated only when reciprocal F_1 hybrids have been included in the assay. When significant phenotypic differences between the reciprocal F1 hybrids are found, the distinction of its maternal genetic source, either cytoplasmic or nuclear, will require the analysis of further progenies differing exclusively in their cytoplasm because F1 plants differ in both the genetic components (*see* **Notes 14** and **42**).
42. When the mapping population consists of individuals carrying either of the two parental cytoplasms, its genetic effect on the trait(s) of interest can be tested (*see* **Note 14**). For that we might perform an ANOVA using the cytoplasm as fixed effect factor. When significant statistical differences are found, the percentage of the phenotypic variance explained by the maternal cytoplasm might be estimated as the R^2 value.
43. Marker and phenotypic data of the RIL populations already existing can be obtained from Web pages such as NATURAL or TAIR (*see* **Table 1**). Linkage group assignments and map positions of markers should be calculated and verified for the population under investigation because sometimes administrative errors are made. The map used in the QTL analysis should be the map that applies to the studied population (or subpopulation, when only part of it is used; *see* **Note 39**), because in the QTL analysis, QTL genotype probabilities and missing marker genotype probabilities are calculated based on this map; using a different map may lead to very deviant probability estimates and, thus, to incorrect results.
44. Single-marker mapping can be performed with standard statistical software packages using a parametric method such as one-way ANOVA including the markers as fixed effect factors. When the trait is not normally distributed one can apply a transformation or apply a non-

parametric test such as Kruskal-Wallis (*see* **Note 7**). Applying a nonparametric test will result in a decreased detection power. The statistical significance threshold (p value) for the identification of a QTL in single-marker mapping methods is advised to be stringent to reduce false QTL identifications; p values of 0.005 or smaller should be used.

45. Genome-wide LOD thresholds for IM have been established by simulations to provide an overall α level of 5% ***(29)***, which correspond to LOD values of 2.8, 3.4, and 2.6 for *Arabidopsis* RIL, F2, and BC populations, respectively. Given the dependency of the LOD threshold on the number of chromosomes, it has been proposed to use a chromosome-wide significance of α = 5% to define suggestive linkage of QTLs ***(29,43)***. This corresponds to LOD values of 2.1, 2.7, and 1.9 for RIL, F2, and BC populations, respectively. Alternatively, several mapping software packages include permutation analysis to derive empirical LOD significance thresholds based on the phenotypic data used to map the QTLs.
46. When using approximate multiple QTL mapping methods (composite interval mapping) ***(24,25)***, marker cofactors must be carefully selected. One may follow an approach in which markers close to significant peaks in the interval mapping LOD profile are chosen as cofactors, for the subsequent composite IM. QTL detection power will then increase, and new significant LOD peaks may be observed; from these, additional cofactor markers are chosen, and the mapping procedure repeated. This approach can be described as a statistical forward selection approach. It has the drawbacks of potential detection of ghost QTLs and missing linked QTLs in repulsion. Therefore, when applying this method, these weaknesses should be investigated by modifying the cofactors to neighboring markers, and studying the LOD profile afterward. The number of cofactors (QTLs) that can be included in the model will depend on the population size and the marker and population type. It is recommended that the number of parameters in the model is less than twice the square root of the population size. Thus, for population sizes of 200 individuals, we might theoretically include up to 27, 27, and 13 QTL cofactors for RIL, BC, and F2 populations, respectively. However, these large numbers of cofactors affect the significance threshold, and therefore the number of cofactors should be kept moderated.
47. When using LOD support intervals as confidence intervals for QTL map positions, it is advised to apply a 1.5-LOD interval for populations of 200 plants or more, and a 2-LOD support for smaller populations.
48. The additive allele effect of a QTL (*add*) is given in single-marker mapping by the effect of the closest linked individual marker. For RIL and F2 populations, this is calculated as half of the mean phenotypic difference between the two homozygous groups of the population classified according to the genotype at the corresponding marker $(add = (\bar{A} - \bar{B}) / 2)$. For BC populations, this is the mean phenotypic difference between the heterozygous and the homozygous groups of the population classified according to that marker $(add = \bar{A} - \bar{H} \text{ or } \bar{H} - \bar{B})$. The additive allele effect of a QTL detected by IM will be estimated in a similar way, but using the mean phenotypic values of the corresponding QTL genotypic classes estimated by the IM procedure at the map position with the maximum LOD score, or at the nearest linked marker.
49. F2 populations allow the estimation of dominance effect (*dom*) and mode of action of QTLs as: $dom = \bar{H} - [(\bar{A} + \bar{B}) / 2]$ where $\bar{H}$, $\bar{A}$, and $\bar{B}$ are the mean phenotypic values of the heterozygous and the two homozygous genotypic groups of the population classified according to the genotypes at the corresponding marker or QTL. Practically-homozygous RIL populations (at the F8 generation and onward) do not allow estimations of QTL dominances, but in RIL populations with residual heterozygosity this can be estimated similarly to that in F2 populations. In BC populations, the dominance effects cannot be separated from the additive allele effects (*see* **Notes 17** and **48**) due to the lack of one homozygous class.
50. The percentage of the phenotypic variance of the trait explained by an individual marker or by a QTL detected by ANOVA is usually given as the 100%*R^2 value. In IM this is given as 100%*$(\sigma^2_{H0} - \sigma^2_{H1}) / \sigma^2_{H0}$, with σ^2_{H0} the residual variance under the null hypothesis (no QTL) and σ^2_{H1} the residual variance under the alternative hypothesis (a QTL).
51. The total phenotypic variance explained by all QTL detected affecting a trait is estimated as the total R^2 value attributed to the closest markers to each QTL included simultaneously in a linear regression model.

52. Genetic interactions between pairs of QTLs can be tested by two-way ANOVA, using the closest markers to the QTLs as fixed effect factors. Alternatively, a multiple regression model with (multiplicative) interaction terms can be used for this purpose. With the latter approach, the number of interaction terms should not be large in order to retain degrees of freedom for the test. Two-way QTL interactions might be searched not only among the detected QTLs, but also among all pairs of markers. Such genome-wide search of QTL interactions might identify new QTLs with only epistatic effects that could be undetected with standard QTL mapping procedures, which are based mostly on detection of additive effects. A multiple regression model can also be used to estimate the total genetic variance explained by all the QTLs (R^2; *see* **Note 51**) including not only the additive effects of all QTLs (or closest markers), but also significant additive effects of the maternal cytoplasm (*see* **Notes**) and the significant interaction terms. Some software packages, such as QTL Cartographer (*see* **Table 1**), include one-dimensional searches of QTL interactions; new developments in multiple interacting QTL methods ***(5)*** are expected to be implemented in this and other packages.
53. When several traits have been analyzed in the same population, QTLs detected in the same genomic regions will identify putative pleiotropic effects of the same QTL. In case of RIL populations, traits analyzed in other laboratories and/or experiments can also be compared, which constitute a major advantage of permanent mapping populations (*see* **Note 43**). The simultaneous analysis of several traits is included as multitrait analysis in software packages such as MultiQTL (*see* **Table 1**).
54. When the same mapping population is grown in different environments (years, locations, or treatments) this allows studying genotype by environment (G × E) interaction (*see* **Note 10**). Significant G × E interaction, denoting differential response of the genotypes, is due to the presence of QTLs whose effects are differently expressed in the various environments. The simplest way to identify such QTLs showing QTL by environment interaction (QTL × E) is to perform a separate QTL analysis for each environment. QTLs interacting with environments will show different relative effects in the various environments. This can be statistically tested using a multiple linear regression model including the QTL marker and environment as factors in a design with repeated measurements of the genotypes. Alternatively, QTL × E interactions can be included in the QTL mapping models to improve the mapping accuracy by using different methods ***(44,45)***, and which are implemented in some software packages (*see* **Table 1**). These approaches might detect that some QTLs are expressed in specific environments only, whereas other QTLs display their effect in all environments.
55. Standard *Arabidopsis* QTL mapping studies will locate QTLs in genetic intervals ranging from 5 to 50 cM. These intervals correspond to average physical intervals of 1.2 to 12 Mbp of DNA, which will contain on average 300 to 3000 predicted open reading frames. Therefore, further analysis is needed to characterize the QTL (*see* **Subheading 1.**). The follow-up analysis might include (a) Development of NILs with single QTL differences (“monogenic”) affecting the trait (*see* **Note 21** and **Table 1** for availability of NIL populations). NILs might be used to confirm the QTL effects and QTL × E interactions, to estimate the dominance in F1 hybrids and to further characterise the QTL in a similar way to mutant loci. (b) Fine mapping of the QTL in populations derived from the cross of NILs, or of two RILs or mapping individuals differing in only one of the QTLs identified. Selection of such genotypes from the mapping population is facilitated by softwares developed for that purpose, such as GGT (***42***; www.dpw.wau.nl/pv/pub/ggt/). Fine mapping will enable to determine whether the mapped QTL is a single locus or several closely linked QTLs, to precisely determine its/their location and, eventually, the map-based cloning of an interesting QTL.

Acknowledgments

C.A.-B. was supported by a Ramón y Cajal salary contract. Research in M.K. and C.A.-B.’s laboratories was supported by the European Union program NATURAL (contract QLG2-CT-2001-01097).

References

1. Koornneef, M., Alonso-Blanco, C., and Vreugdenhil, D. (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. *Ann. Rev. Plant Biol.* **55,** 141–172.
2. Holub, E. B. (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* **2,** 516–527.
3. Darvasi, A. and Soller, M. (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141,** 1199–1207.
4. Liu, S.-C., Kowalski, S. P., Lan, T.-H., Feldmann, K. A., and Paterson, A. H. (1996) Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. *Genetics* **142,** 247–258.
5. Jannink, J.-L. and Jansen, R. (2001) Mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* **157,** 445–454.
6. Borevitz, J. O. and Nordborg, M. (2003) The impact of genomics on the study of natural variation in *Arabidopsis. Plant Physiol.* **132,** 718–25.
7. Kearsey, M. J. and Pooni, H. (1996) *The Genetical Analysis of Quantitative Traits*. Chapman and Hall, London.
8. Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA.
9. Bulmer, M. G. (1985) *The Mathematical Theory of Quantitative Genetics.* Clarendon Press, Oxford, UK.
10. Mather, K. and Jinks, J. L. (1977) *Introduction to Biometrical Genetics*. Chapman and Hall, London.
11. Tuinstra, M. R., Ejeta, G., and Goldsbrough, P. B. (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theor. Appl. Genet.* **95,** 1005–1011.
12. Paran I. and Zamir, D. (2003) Quantitative traits in plants: beyond the QTL. *Trends Genet.* **19,** 303–306.
13. Aguadé, M. (2001) Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **18,** 1–9.
14. Schmid, K. J., Rosleff Sörensen, T., Stracke, R., Törjek, O., Altmann, T., Mitchell-Olds, T., and Weisshaar, B. (2003) Large-scale identification and analysis of genome-wide single nucleotide polymorphs for mapping in *Arabidopsis thaliana*. *Genome Res.* **13,** 1250–1257.
15. Comai, L., Young, K., Till, B. J., et al. (2004) Efficient discovery of DNA polymorphism in natural populations by ecotilling. *Plant J.* **37,** 778–786.
16. Haley, C. S. and Knott, S. A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69,** 315–324.
17. Beaumont, M. A. and Rannala, B. (2004) The Bayesian revolution in genetics. *Nat. Rev. Genet.* **5,** 251–261.
18. Knapp, S. J. and Bridges, W. C. (1990) Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. *Genetics* **126,** 769–777.
19. Van Ooijen, J. W. (1992) Accuracy of mapping quantitative trait loci in autogamous species. *Theor. Appl. Genet.* **84,** 803–811.
20. Dupuis, J. and Siegmund, D. (1999) Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151,** 373–386.
21. Charmet, G. (2000) Power and accuracy of QTL detection: simulation studies of one-QTL models. *Agronomie* **20,** 309–323.
22. Cornforth, T.W. and Long, A.D. (2003) Inferences regarding the numbers and locations of QTLs under multiple-QTL models using interval mapping and composite interval mapping. *Genet. Res.* **82,** 139–149.
23. Zhang, Y.-M. and Xu, S. (2004) Mapping quantitative trait loci in F_2 incorporating phenotypes of F_3 progeny. *Genetics* **16**6, 1981–1993.
24. Zeng, Z.-B. (1994) Precision mapping of quantitative trait loci. *Genetics* **136,** 1457–1468.
25. Jansen, R. C. and Stam, P. (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136,** 1147–1455.
26. Martinez, O. and Curnow, R. N. (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* **85,** 480–488.
27. Rebaï, A., Goffinet, B., and Mangin, B. (1994) Approximate thresholds of interval mapping tests for QTL detection. *Genetics* **138,** 235–240.
28. Churchill, G. A. and Doerge, R. W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics* **138,** 963–971.

29. Van Ooijen, J. W. (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* **83,** 613–624.
30. Doerge, R. W. and Rebaï, A. (1996) Significance thresholds for QTL interval mapping tests. *Heredity* **77,** 459–464.
31. Mangin, B. and Goffinet, B. (1997) Comparison of several confidence intervals for QTL location. *Heredity* **78,** 345–353.
32. Walling, G. A., Visscher, P., and Haley, C. (1998) A comparison of bootstrap methods to construct confidence intervals in QTL mapping. *Genet. Res.* **71,** 171–180.
33. McKhann, H. I., Camilleri, C., Bérard, A., et al. (2004) Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J.* **38,** 193–202.
34. Pérez-Pérez, J. M., Serrano-Cartagena, J., and Micol, J. L. (2002) Genetic analysis of natural variation in the architecture of *Arabidopsis thaliana* vegetative leaves. *Genetics* **162,** 893–915.
35. Vizir, I. Y. and Korol, A. B. (1990) Sex differences in recombination frequency in *Arabidopsis*. *Heredity* **65,** 379–383.
36. Darvasi, A. (1997) The effect of selective genotyping on QTL mapping accuracy. *Mammalian Genome* **8,** 67–68.
37. Haldane, J. B. S. and Waddington, C. H. (1931) Inbreeding and linkage. *Genetics* **16,** 357–374.
38. Kowalski S. P., Lan T. H., Feldmann K. A., and Paterson A. H. (1994) QTL mapping of naturally occurring flowering time of Arabidopsis thaliana. *Mol. Gen. Genet.* 245, 548–555.
39. Voorrips, R. E. (2001) MapChart version 2.0: Windows software for the graphical presentation of linkage maps and QTLs. Plant Research International, Wageningen, The Netherlands.
40. Haldane, J. B. S. (1919) The combination of linkage values, and the calculation of distance between linked factors. *J. Genet.* 8, 299–309.
41. Kosambi, D. D. (1944) The estimation of map distances from recombination values. *Ann. Eugen.* **12,** 172–175.
42. Berloo, R. (1999) The development of software for the graphical representation and filtering of molecular data: graphical genotypes (GGT). *J. Heredity* **90,** 328–329.
43. Lander, E. and Kruglyak L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* **11,** 241–247.
44. Van der Schaar W., Alonso-Blanco C., Léon-Kloosterziel K. M., Jansen R. C., van Ooijen, J. W., and Koornneef, M. (1997) QTL analysis of seed dormancy in *Arabidopsis* using recombinant inbred lines and MQM mapping. *Heredity* **79,** 190–200.
45. Ungerer, M. C., Hallsdorsdottir, S. S., Purugganan, M. D., and MacKay, T. F. C. (2003) Genotype-environments interaction in quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. *Genetics* **165,** 353–365.

8

EMS Mutagenesis of *Arabidopsis*

YongSig Kim, Karen S. Schumaker, and Jian-Kang Zhu

Summary

A powerful approach for determining the biological functions of genes in an organism is to produce mutants with altered phenotypes and physiological responses. Various approaches for mutagenesis involving chemical, irradiation, and insertional methods have been developed; each has advantages and disadvantages for the study of gene function. In this post-genomic era, the use of reverse genetic approaches to understanding the role of genes in growth and development has become widespread. With development of new techniques such as targeting induced local lesions in genomes (TILLING), ethyl methanesulfonate (EMS) mutagenesis can be used for both forward and reverse genetic studies. Generation of diverse mutant alleles in the same gene provides critical tools to understand the role of these genes in the function of the organism. Here we describe the general method of EMS mutagenesis for the molecular genetic model plant *Arabidopsis thaliana*.

Key Words: EMS; chemical mutagenesis; *Arabidopsis*; mutants.

1. Introduction

Alkylating agents such as ethyl methanesulfonate (EMS) induce chemical modification of nucleotides, which results in mispairing and base changes. Strong, biased alkylation of guanine (G) residues results, forming O^6-ethylguanine, which can pair with thymine (T) but not with cytosine (C). Through subsequent DNA repair, the original G/C pair can then be replaced with A (adenine)/T ***(1)***. The functions and mechanisms of action of other monofunctional alkylating agents such as diethyl sulfate (DES) and dimethyl sulfate have been extensively reviewed ***(2)***. Comparisons of DES and EMS have shown that they share similar alkylating activities and genetic effects.

The majority (99%) of the time, EMS induces C-to-T changes resulting in C/G to T/A substitutions, whereas methyl methanesulfonate produces T/A to G/C transversion and A/T to G/C transitions ***(1,3,4)***. At a low frequency, EMS generates G/C to C/G or G/C to T/A transversions by 7-ethylguanine hydrolysis or A/T to G/C transition by 3-ethyladenine pairing errors ***(3)***. Based on codon usage in *Arabidopsis*, the frequency of EMS-induced stop codon and missense mutations has been calculated to be ~5% and ~65%, respectively ***(5)***. EMS mutagenesis generates randomly distributed mutations throughout the genome in *Arabidopsis* ***(1)***. As a result, chemical mutagenesis can be used not only to search for loss- or gain-of-function mutants but also to understand the role of specific amino acid residues in protein function. Results of many studies suggest that use of chemically induced mutants can also provide useful information for understanding the functions of essential genes by generating weak nonlethal alleles. In addition, EMS mutagenesis can be used for generating breeding lines ***(6)***.

To achieve saturation of EMS mutagenesis in *Arabidopsis*, at least 125,000 M1 lines should be generated ***(7)***. Evaluating the degree of saturation requires independent calibration through screening of visible recessive traits such as albinism, embryo lethality, and trichome pheno-

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

types ***(8)***. The frequency of mutation depends on the position of the gene in the genome and the treatment conditions during mutagenesis ***(4)***. Effects such as embryonic lethality and chlorophyll mutations are directly related to the length of exposure to a fixed concentration of EMS ***(9)***. High concentrations (over 0.9%) of EMS have been shown to result in a decreased frequency of induced mutations in soybeans ***(10)***.

Use of *Arabidopsis* with amber, opal, and ocher stop codons in an introduced β-glucuronidase (*uidA*) gene determined the spontaneous (natural somatic) mutation rate to be between 10^{-7} and 10^{-8} events per base pair per generation ***(4)***. Using this frequency as a base of comparison, it is clear that significant advantages of chemical mutagenesis include its high efficiency and frequency in achieving saturation mutagenesis, especially as compared to genetic mutagenesis including insertion of transfer DNA (T-DNA) or transposable elements. Isolation of mutants requires the use of relatively small-scale screens with chemically induced mutagenized pools. In addition, chromosomal rearrangements caused by T-DNA or transposon insertion can be avoided with chemical mutagenesis ***(8)***. Compared with irradiation mutagenesis, EMS induces relatively few strand breaks that lead to inversion or deletion mutations ***(9)***. In addition, in Arabidopsis EMS induces four times as many embryonic or chlorophyll mutants as irradiation at the M1 level ***(9)***.

The difficulty in detecting single-nucleotide polymorphisms or substitutions can be overcome by use of **t**argeting **i**nduced **l**ocal **l**esions **in g**enomes (TILLING) complemented with denaturing high-performance liquid chromatography (DHPLC; **5**). These technologies allow chemically induced mutant pools to be used for reverse genetics. With help of automation, robust and rapid detection makes it possible to screen a wide range of mutant pools in a short time and to avoid the often laborious process of forward genetic screening ***(11,12)***.

2. Materials

1. *Arabidopsis thaliana* seeds.
2. Disposable 50-mL plastic tubes.
3. Nutating mixture.
4. 100 m*M* phosphate buffer (pH 7.5).
   a. Prepare 100 mL of 1 *M* K_2HPO_4 and 100 mL of 1 *M* KH_2PO_4.
   b. Pour 70 mL of 1 *M* K_2HPO_4 into a beaker.
   c. Add 20 mL of 1 *M* KH_2PO_4 and mix the solution.
   d. Adjust the pH to 7.5 with the addition of 1 *M* KH_2PO_4.
   e. Dilute the 1 *M* phosphate buffer with water to make 100 m*M* phosphate buffer (*see* **Note 1**).
5. Ethyl methanesulfonate (EMS, F.W. 124.16; *see* **Note 2**).
6. Sterilized water.
7. Squeeze bottle.

3. Methods

3.1. EMS Mutagenesis

1. Weigh 2.5 g of well-dried *Arabidopsis thaliana* seeds (approx 125,000 seeds).
2. Soak the seeds in a 50-mL plastic tube with 40 mL of 100 m*M* phosphate buffer at 4°C overnight.
3. If EMS mutagenized seeds are to be planted in soil, then do not sterilize the seeds. However, if the screening is to be performed with M1 seeds on plates, M0 seeds should be sterilized (*see* **Note 3**).
4. Stand the tube upright to allow the seeds to settle and decant the excess phosphate buffer.
5. Add 40 mL of fresh 100 m*M* phosphate buffer.
6. Add EMS to a final concentration of 0.4% (*see* **Note 4**).
7. Incubate the mixture for 8 h at room temperature with gentle nutation.
8. Wash the seeds thoroughly 20 times with water (40 mL per wash).
9. After EMS mutagenesis, seeds should be planted immediately in soil or on plates. Preserve M1 seeds by drying them on filter paper.

3.2. Planting EMS Mutagenized M1 Seeds

M1 seeds can be planted in soil or on plates for direct screening for dominant mutants. Planting in soil and selfing of the M1 plants will yield M2 seeds for screening for recessive as well as dominant mutants. M1 seeds can be planted in soil uniformly with the use of a squeeze bottle.

1. Prepare lightly wetted soil in flats.
2. Add 1 mL of M1 seeds to 400 mL of water in a squeeze bottle.
3. Squirt the M1 seeds onto the soil.
4. Keep the flat at 4°C for 2 to 4 d for seed stratification.
5. Transfer the flat to a growth chamber at 25°C with a 16/8 light/dark cycle (*see* **Note 5**).
6. Harvest the seeds from approx 500 to 2000 M1 plants and divide the M2 seeds into pools of desired sizes.

4. Notes

1. If EMS mutagenesis requires aseptic conditions, the 100 m*M* phosphate buffer can be sterilized by autoclaving at 121°C for 20 min. Phosphate buffer can be stored at room temperature.
2. Alternate names are methanesulfonic acid ethyl ether, ethyl methanesulfonic acid or ethyl mesylate. Ethyl methanesulfonate liquid can be purchased from Sigma-Aldrich, product number M0880.
3. Phosphate buffer-imbibed M0 seeds can be sterilized by replacing the buffer with 40 mL of 6% sodium hypochlorite. After a 5-min incubation, seeds should be washed with sterilized water at least four times. Add 40 mL of phosphate buffer to sterilized seeds for EMS treatment (**step 6**).
4. **Caution: Steps 6** to **8** should be done in a fume hood because EMS is volatile and highly toxic. The waste containing EMS should be collected separately from that of other chemicals.
5. To prevent contamination by nonmutagenized seeds, use a growth chamber or room dedicated to this purpose.

References

1. Greene, E. A., Codomo, C. A., Taylor, N. E., et al. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* **164,** 731–740.
2. Hoffmann, G. R. (1980) Genetic effects of dimethyl sulfate, diethyl sulfate, and related compounds. *Mut. Res.* **75,** 63–129.
3. Krieg, D. R. (1963) Ethyl methanesulfonate-induced reversion of bacteriophage T4*r*II mutants. *Genetics* **48,** 561–580.
4. Kovalchuk, I., Kovalchuk, O., and Hohn, B. (2000) Genome-wide variation of the somatic mutation frequency in transgenic plants. *EMBO J.* **19,** 4431–4438.
5. McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000) Targeted screening for induced mutations. *Nat. Biotech.* **18,** 455–457.
6. Lee, S. Y., Cheong, J. I., and Kim, T. S. (2003) Production of doubled haploids through anther culture of M_1 rice plants derived from mutagenized fertilized egg cells. *Plant Cell Rep.* **22,** 218–223.
7. Jander, G., Baerson, S. R., Hudak, J. A., Gonzalez, K. A., Gruys, K., and Last, R. L. (2003) Ethyl methanesulfonate saturation mutagenesis in *Arabidopsis* to determine frequency of herbicide resistance. *Plant Physiol.* **131,** 139–146.
8. Grini, P. E., Schnittger, A., Schwarz, H., Zimmermann, I., Schwab, B., Jürgens, G., and Hülskamp, M. (1999) Isolation of ethyl methanesulfonate-induced gametophytic mutants in *Arabidopsis thaliana* by a segregation distortion assay using the multimarker chromosome I. *Genetics* **151,** 849–863.
9. Koornneef, M., Dellaert, L. W. M., and van der Veen, J. H. (1982) EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mut. Res.* **93,** 109–123.
10. Zhu, B., Gu, A., Deng, X., Geng, Y., and Lu, Z. (1995) Effects of caffeine or EDTA post-treatment on EMS mutagenesis in soybean. *Mut. Res.* **334,** 157–159.
11. Colbert, T., Till, B. J., Tompa, R., et al. (2001) High-throughput screening for induced point mutations. *Plant Physiol.* **126,** 480–484.
12. Henikoff, S. and Comai, L. (2003) Single-nucleotide mutations for plant functional genomics. *Annu. Rev. Plant Biol.* **54,** 375–401.

9

Low-Resolution Mapping of Untagged Mutations

M. Rosa Ponce, Pedro Robles, Francisca M. Lozano, M. Asunción Brotóns, and José L. Micol

Summary

The mapping method detailed here is based on the multiplex polymerase chain reaction (PCR) coamplification of 32 molecular markers, using fluorescently labeled oligonucleotides as primers. For the genotyping of a single plant from a mapping population, only two simultaneous amplifications are required, the products of which are finally electrophoresed in an automated DNA sequencer controlled by fragment analysis software. An analysis of the genotypes of 50 plants allows mapping of the mutation of interest within a candidate genomic interval of about 15 cM (3 Mb, corresponding to about 40 BAC clones).

Key Words: Positional cloning; linkage analysis; SSLP; In/Del; multiplex PCR.

1. Introduction

A typical departure point for genetic approaches to identify which genes are involved in a given biological process is to isolate mutants. Once a mutant phenotype has proved to be unequivocally distinguishable from the wild type and stable through several generations, it is relatively easy to approach the gene damaged by mutation.

Mutations tagged by insertional mutagenesis have been, and continue to be, instrumental for the rapid identification of the genomic sequences flanking the insertion. However, there are a number of drawbacks in the use of T-DNA, which in many cases generates multiple unlinked insertions or internally reorganized and large concatameres, and in some instances causes untagged mutations as a result of abortive insertions. Such behavior of T-DNA-based constructs sometimes complicates the cloning of the genes damaged by their insertions. In addition, most T-DNA-induced mutations are null and some cause recessive lethality.

The use of physical or chemical mutagens represents an alternative to insertional mutagenesis. A major objection raised against the use of these mutagens is that positional approaches must be followed to clone the corresponding genes. However, although map-based cloning has until recently been considered time-consuming and expensive, the available techniques and the information provided by the *Arabidopsis* Genome Initiative make it possible to complete positional cloning projects in a short period of time ***(1)***. In addition, physical and chemical mutagens have mutation rates higher than those of insertional mutagens, and they induce null and lethal alleles as well as hypomorphic and viable alleles.

To map an untagged mutation, a mapping population should first be obtained by crossing the mutant of interest to a wild-type line. The genetic background of the latter must be different and as polymorphic as possible compared with that of the mutant under study. With this regard, the most usual crosses in *Arabidopsis thaliana* are those involving Landsberg *erecta* (L*er*) and a mutant with a Columbia-0 (Col-0) genetic background, or vice versa, whose F1 progeny is allowed to self in order to obtain F2 plants. Although a single cross is usually sufficient for the

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

low-resolution mapping of a mutation, minimal genotyping of its F1 progeny is convenient in order to confirm heterozygosis.

Phenotypically recessive F2 plants are collected to individually extract DNA, which is then used to assess the linkage between the mutation under study and molecular markers more or less evenly spaced along the genome of *Arabidopsis thaliana*. Low-resolution mapping of mutations is facilitated by the availability of a plethora of polymorphic microsatellites, also named simple sequence length polymorphisms (SSLPs; ***2***), and small-insertion and deletions (In/Dels; ***3***), all of which are easy to score codominant markers, given that their alleles yield polymerase chain reaction (PCR) amplification products of distinguishable sizes. Differences between the widely used L*er* and Col-0 accessions at the DNA sequence level mean that it is possible to find at least one polymorphism per BAC clone, many of which can be used to develop molecular markers ***(3)***.

The method detailed here is based on multiplex PCR coamplification of a large set of molecular markers using fluorescently labeled oligonucleotides as primers ***(4)***. Only two simultaneous PCR amplifications are required for the genotyping of a single F2 plant for the 32 markers, which include 20 SSLPs and 12 In/Dels. An analysis of the genotypes of 50 plants allows mapping of the mutation of interest within a candidate genomic interval of about 15 cM (3 Mb, corresponding to about 40 BAC clones).

Once a mutation has been low-resolution mapped, there are several procedures that can be used to further refine the candidate genomic interval ***(3,5)***, some of which are covered in other chapters of this book. Needless to say, a prudent practice is to survey the literature and congress proceedings for already-described phenotypes associated to mutations in the candidate interval, as recloning a gene already cloned by others is becoming increasingly common in the postgenomics era.

2. Materials

Use autoclaved distilled water to prepare all solutions. All plastic materials should be sterile.

2.1. Equipment and Supplies

1. 1.5-mL Eppendorf tubes.
2. 0.2-mL thin-walled PCR tubes.
3. 3-mm tungsten carbide beads (Qiagen, cat. no. 69997).
4. Vortex.
5. Microfuge.
6. Heat block.
7. MM 300 Mixer Mill (Qiagen, 85120).
8. Autoclavable plastic pestles (Sigma, Z35,994-7).
9. Cordless motor grinder (Sigma, Z35,997-1).
10. 96-well optical reaction plate (Applied Biosystems, part no. 4306737).
11. 3100 Capillary Array 36 cm (Applied Biosystems, 4315931).
12. ABI PRISM 3100 Genetic Analyzer (Applied Biosystems).
13. GeneAmp PCR System 2700 (Applied Biosystems).
14. Genescan 3.7 software (Applied Biosystems).
15. Mapmaker 3.0b program ***(6)***.

2.2. Reagents and Buffers

1. Hi Di formamide (Applied Biosystems, 4311320).
2. 20% SDS.
3. 5 *M* KOAc.
4. Isopropanol.
5. 3 *M* NaOAc.
6. 70% ethanol.
7. *Taq* DNA polymerase.

8. 10X PCR buffer.
9. Oligonucleotide mixes (*see* **Table 1**).
10. 50 m*M* $MgCl_2$.
11. DNA extraction buffer: 100 m*M* Tris-HCl, pH 8.0, 50 m*M* EDTA, 500 m*M* NaCl.
12. Deoxyribonucleotide mix: 2.5 m*M* dATP, 2.5 m*M* dCTP, 2.5 m*M* dGTP, and 2.5 m*M* dTTP.
13. GeneScan-500 LIZ Size Standard (Applied Biosystems, 4322682).
14. 3100 POP-6 polymer (Applied Biosystems, 4316357).

3. Methods

3.1. DNA Extraction

Select 50 homozygous plants (*see* **Note 1**) from the F2 mapping population derived from an outcross of the mutant of interest to a wild-type plant. This should allow the detection of at least one marker linked to the mutation under study (*see* **Note 2**).

1. Collect 50 to 80 mg of plant material from each individual to be genotyped in 1.5 mL-Eppendorf tubes.
2. Add to each tube one tungsten carbide bead and 500 µL of DNA extraction buffer.
3. Grind the samples twice for 1.5 min at 30 Hz in the Mixer Mill (*see* **Note 3**).
4. Add 35 µL 20% SDS, mix gently, and incubate for 5 min in a heat block at 65°C.
5. Add 130 µL 5 *M* KOAc, mix gently, and incubate on ice for 8 min.
6. Centrifuge at 13,000 rpm for 15 min in a microfuge.
7. Transfer the supernatant to properly labelled 1.5 mL-Eppendorf tubes. Repeat **steps 6** and **7** if the supernatant is not completely free of cellular debris.
8. Add 640 µL isopropanol and 60 µL 3 *M* NaOAc and mix gently.
9. Precipitate DNA at –20°C for 20 min.
10. Centrifuge at 13,000 rpm for 10 min in a microfuge.
11. Remove supernatant.
12. Wash the DNA precipitate with 300 µL 70% ethanol.
13. Centrifuge at 13,000 rpm for 5 min in a microfuge.
14. Remove supernatant and let the pellet air dry (usually 1 h).
15. Resuspend the pellet in 100 µL of water (or 50 µL if the starting plant material was less than 50 mg).
16. Store the DNA samples at 4°C or –20°C.

3.2. Multiplex PCR Amplifications

3.2.1. Preparation of Oligonucleotide Mixes

The method described here allows the simultaneous genotyping of a total of 32 markers (**Fig. 1**) by simply performing two PCR amplifications per plant. Two oligonucleotide mixes are required, which include 17 and 15 primer pairs, respectively (**Table 1**; *see* **Note 4**). Each primer pair includes an oligonucleotide that has been labeled with one of four fluorescent dyes (NED, 6-FAM, VIC, or PET), with nonoverlapping emission wavelengths. Both oligonucleotide mixes are prepared in the same way (*see* **Notes 5** and **6**), as follows:

1. Prepare 10 µ*M* oligonucleotide stock solutions (*see* **Note 7**).
2. Prepare primer mixes 1 and 2, both of which contain some oligonucleotides at a 250 n*M* concentration and some others at 375 n*M*. In this way, their final concentrations in the PCR mix will be 50 or 75 n*M*, respectively (*see* **Note 8**).

   To prepare primer mix 1, add the following to an Eppendorf tube:

   a. 2.5 µL of each 10 µ*M* oligonucleotide stock solution of the primers, which should be at a 50 n*M* final concentration in the PCR mix (11 × 2 = 22 primers, 22 × 2.5 µL = 55 µL) (*see* **Table 1**).
   b. 3.75 µL of each 10 µ*M* oligonucleotide stock solution of the primers that should be at a 75 n*M* final concentration in the PCR mix (6 × 2 = 12 primers, 12 × 3.75 µL = 45 µL) (*see* **Table 1**).

Table 1
Oligonucleotide Sets for Low-Resolution Mapping

Marker	Type of marker	Oligonucleotides: Sequences (5'→3') Unlabelled reverse primer	Labelled forward primer	Label	Final concentration in PCR mix	Mix	Amplification product size (bp) Ler	Col-0
nga59	SSLP	GCCTAACAAATTTAAAGTTAAGACT	ATCTGTGTTCACTCGCCGCC	NED	50 n*M*	1	173	169
JV18/19	SSLP	AATTCAGTATCGAGATACCCCT	TGTCGTATATCAATCGAAAAAGAG	6-FAM	75 n*M*	1	284	290
SGCSNP10026	In/Del	ACTATTGACTTCGTGTTAACAAAG	CAAGCCCTCTTATGCTTTAGAAA	6-FAM	50 n*M*	1	238	231
F1I21	In/Del	ACCTGACTAGCACTTGCATCTT	GTTGGTAGTGGTTGCAAAACAG	VIC	50 n*M*	1	453	468
nga1145	SSLP	GAGCACATACCCACAACCAGA	CTTCACATCCAAAACCCACCT	VIC	50 n*M*	1	220	214
SGCSNP922	In/Del	AACTCGAACCCGAAAATACCCA	GTCAAATACTAGTTACAATCTCAG	VIC	50 n*M*	1	316	302
SGCSNP8895	In/Del	ACTTTGGGAGAGCGATTTCCG	GGAGTTAGTCAGTTAGCTAACAA	6-FAM	50 n*M*	1	339	327
nga6	SSLP	CTTCACCGATTTCTTAGATCTTC	AATCCGAAAATAATGGAGAAGCTT	NED	75 n*M*	1	117	137
nga1111	SSLP	GTTGGTTCGGTTACAATCGTGT	GTTACCAGATTGAGCTTTGAGC	VIC	50 n*M*	1	158	151
AthF28J12	SSLP	GCCTTTATCAATGGATGAGGAA	AAATCTCAGATCCGTCATTTCCA	VIC	50 n*M*	1	257	265
AthCTR1	SSLP	ATCACCAAAATATGAAATACCAAATT	CCACTTGTTTCTCTCTCTAGTTT	6-FAM	50 n*M*	1	359	377
nga76	SSLP	TCCACCGCCATCACCGTCG	TGGATCTTTTGACAAATCAGGAT	6-FAM	75 n*M*	1	248	177
SO191	SSLP	TAATGCTCTCTTAATTGTGGTTAC	CTCCACCAATCATGCAAATGTTT	NED	75 n*M*	1	492	482
MNF13	In/Del	AATCGTTTACGTCGGAGAAAAC	GGATTTAATTCTCAGCCATCGC	PET	75 n*M*	1	434	446
MCL19	In/Del	GGTAATGAGTAAAGTTTCAATATCA	GATGTTGATTTCTCATAAAGTCATA	VIC	50 n*M*	1	422	411
MNB8	In/Del	TCCTGCACATTTCATAGCGGAA	CTCTACACGTGTATATATCAACA	NED	75 n*M*	1	390	402
K8K14	SSLP	CACAAATCCTTGTGGATGGAAC	CCCAGGATTTGTCTCAATCGA	VIC	50 n*M*	1	345	351
AthZFPG	SSLP	AGAAATGGGTCAATTCACATGTA	TTGCGTTTCCACATTTGTTTAAAC	PET	75 n*M*	2	143	149
nga128	SSLP	TAAATCTTGAAACCTTTAGGGAG	GGTCTGTTGATGTCGTAAGTC	PET	75 n*M*	2	193	183
SGCSNP10490	In/Del	GGGAAAGAAAAGACTTAATTAGCT	GAGTCGCAACGGCTATTTCGT	NED	50 n*M*	2	283	271
nga111	SSLP	TGTTTTTTAGGACAAATGGCGAAA	CTCCAGTTGGAAGCTAAAGGG	6-FAM	50 n*M*	2	160	124
nga1126	SSLP	GGTTAATGTTTCTTCAGTGCTTG	ATCTTTCTCCTCTACGCTTTTC	NED	50 n*M*	2	479	459
nga168	SSLP	GCCAACATTGTGTATCATCAATC	CTACAGAGCTGTCATCGCTGA	VIC	50 n*M*	2	381	397
nga126	SSLP	CGTAATGATTGTGACTTTTTTGC	GAAAAAACGCTACTTTCGTGGTA	PET	75 n*M*	2	408	382
MYF24	In/Del	CGCCAGCGGCCTCAAAAACATA	GGTAAAATAGCCTGAAATCACAC	NED	50 n*M*	2	308	340
AthGAPab	SSLP	GCCACCTTAAGCTTGGCCTC	TCTGGTCACCATGGCTTCGG	PET	75 n*M*	2	476	467
T32N15	In/Del	TATCGGAAATAGGATTGACTACG	TACTTTCATGGGAGTAGGCACT	6-FAM	50 n*M*	2	243	250
F28P10	In/Del	GTTATTGTTGACAACTTTTAGCCT	ATCCACCAATTCCAGGAGCCA	VIC	75 n*M*	2	353	361
nga1139	SSLP	GGCTCGGGTGAGTCACCAT	TTTTTCCTTGTGTTGCATTCCG	PET	75 n*M*	2	309	297
nga151	SSLP	TCGTTGTAGTAGCAGTTTACTATA	TGACCCATATGTTCCAATGTGTT	6-FAM	50 n*M*	2	420	448
AthPHYC	SSLP	GAAACTCGAGAGTTTTGTCTAGA	CTCAGAGAATTCCCAGAAAAATC	NED	75 n*M*	2	225	209
MUA2	In/Del	CCGTTATTTCTCAAATGATCTTTC	GCTATGATGAAGAAGAAAGAAGG	NED	75 n*M*	2	319	368

To prepare primer mix 2, add the following to an Eppendorf tube:

a. 2.5 µL of each 10 µ*M* oligonucleotide stock solution of the primers, which should be at a 50 n*M* final concentration in the PCR mix (7 × 2 = 14 primers, 14 × 2.5 µL = 35 µL) (*see* **Table 1**).

b. 3.75 μL of each 10 μ*M* oligonucleotide stock solution of the primers, which should be at a 75 n*M* final concentration in the PCR mix (8 × 2 = 16 primers, 16 × 3.75 μL = 60 μL) (*see* **Table 1**).
c. 5 μL of water.

3.2.2. PCR Amplifications

PCR amplifications are performed in a 5-μL volume, thus minimizing the amount of labeled primers required. In order to avoid pipetting errors, a master mix for at least 10 reactions should be prepared.

1. Prepare PCR master mixes 1 and 2, both of which should include the following components per plant to be genotyped: 1.86 μL of water, 0.5 μL 10X PCR buffer, 0.4 μL dNTP mix (2.5 m*M* each), 1.0 μL primer mix 1 (or primer mix 2), 0.2 μL 50 m*M* $MgCl_2$, and 0.04 μL 5 U/μL *Taq* polymerase (*see* **Note 9**).
2. Dispense 4-μL aliquots of the PCR master mixes into 100 0.2-mL thin-walled tubes (50 tubes per PCR master mix). Add to each tube 1 μL of plant DNA. Two amplifications per plant are required to genotype the whole set of markers.
3. Place the tubes in a thermocycler, and run the following program: 3 min at 94°C, followed by 40 cycles (30 s at 94°C, 15 s at 50°C, and 2 min at 70°C), and a final 7-min incubation at 70°C (*see* **Note 10**).

3.3. Electrophoresis and Sizing of PCR Amplification Products

1. Dilute each PCR mix 1:4 by adding 15 μL of water in order to avoid saturation of the genetic analyzer.
2. Prepare a loading buffer by mixing 1 mL of Hi Di formamide and 25 μL of GeneScan-500 LIZ size standard.
3. Add 1 μL of the diluted PCR mix and 10 μL of loading buffer per well of a 96-well optical reaction plate.
4. Immediately before electrophoresis in an ABI PRISM 3100 Genetic Analyzer, heat the 96-well optical reaction plate containing the samples at 94°C for 2 min in a thermocycler or a water bath, and place it on ice for at least 5 min (*see* **Note 11**).
5. Perform the electrophoresis of the amplification products using a capillary array of 36 cm, the POP-6 polymer and the Data Collection 3.1 program with the GeneScan36_POP6Default Module.
6. After the end of the electrophoresis, use the Genescan 3.7 software to size the electrophoresed molecules (**Fig. 2**).

3.4. Genotyping

1. Genotype all the F2 plants under study for all the 32 markers (*see* **Note 12**). Although the Genotyper 3.6 NT software can be used for this purpose, manual genotyping is usually possible (*see* **Note 13**).
2. Represent the genotypes in a data matrix as shown in **Table 2**.

3.5. Testing Linkage and Estimating Map Distances

1. Calculate recombination frequencies for each marker and all the F2 plants studied as the recombinant chromosomes/total chromosomes ratio.
2. Map distances between a mutation and its neighboring markers, as well as their relative positions (their maximum-likelihood map), can be determined using the Mapmaker 3.0b program ***(6)***.

4. Notes

1. If the mutation to be mapped is recessive, choose phenotypically mutant F2 plants to extract their DNA; if the mutation is completely dominant, select the wild-type plants; if the mutation is semidominant, select those with the most extreme phenotypes, either the unequivocally wild type or the most severely mutant ones, in order to avoid selecting unnoticed heterozygotes.

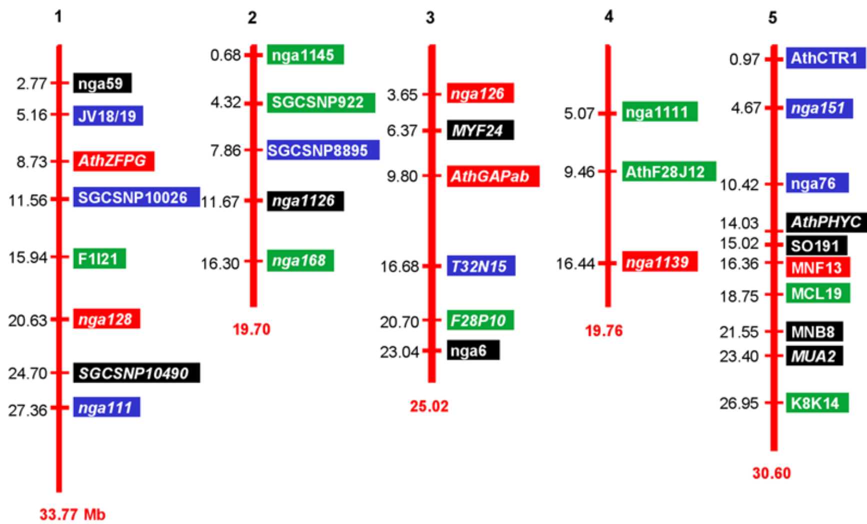


Fig. 1. *Arabidopsis thaliana* physical map with indication of the positions of the markers used in the linkage mapping method described here. All figures indicate Mb, the only exceptions being chromosome numbers. Markers are colored to indicate the fluorophore used to label one of their corresponding oligonucleotide PCR primers (VIC, green; 6-FAM, blue; NED, black; PET, red). Markers indicated in italics are genotyped using primer mix 2 and the remaining ones with primer mix 1 (*see* **Table 1**).

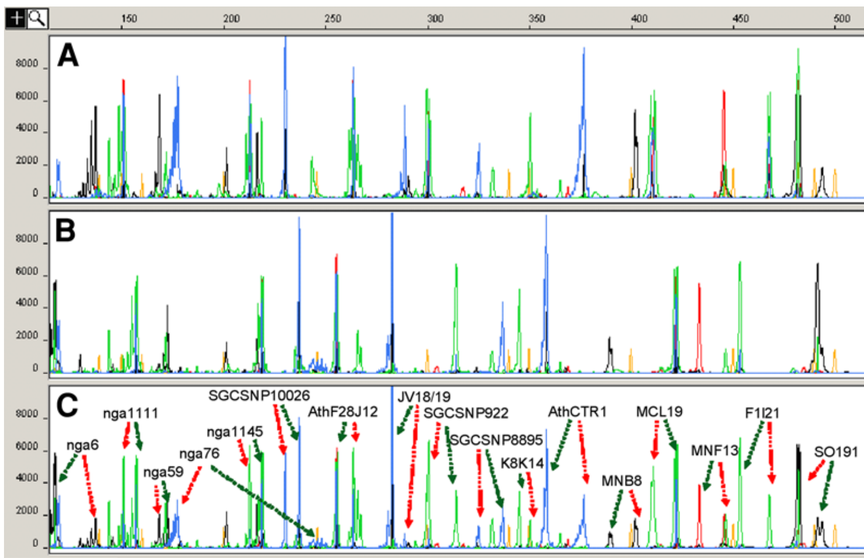


Fig. 2. Electrophoregram illustrating the results of a mapping experiment following the method described in this chapter. The horizontal and vertical axes indicate, respectively, the size of the amplification products (in nucleotides) and the intensity of fluorophore emission (in arbitrary units). The profiles represent the molecules obtained from the multiplex PCR coamplification of 17 markers (primer mix 1), using as a template genomic DNA from single plants of the Col-0 (panel A) and L*er* (panel B) accessions, and their F1 progeny (panel C). Green, blue, black, and red peaks indicate VIC, 6-FAM, NED, and PET fluorochromes, respectively. Peaks corresponding to the internal molecular weight standard (GeneScan-500 LIZ Size Standard) appear in orange. In panel C, peaks are denoted by the name of their corresponding marker and the arrows point to the L*er* (green arrows) or Col-0 (red arrows) alleles.

Table 2
Example of Results of Low-Resolution Mapping Experiment

Marker	Chromosome	Physical map position (bp)	F_2 plants										r (%)	Sr	D (cM)	Sd	*n*
			1	2	3	4	5	6	7	8	9	10					
nga59	1	2.768.000	C	C	L	H	C	L	C	L	H	C	50.00	5.05			98
JV18/19	1	5.160.594	C	C	L	L	C	L	C	L	-	C	67.07	5.19			82
AthZFPG	1	8.727.056	C	C	L	H	C	L	H	L	C	C	55.10	5.02			98
SGCSNP10026	1	11.562.061	C	C	L	H	C	L	H	L	H	C	53.06	5.04			98
F1I21	1	15.937.110	C	C	L	H	C	L	H	L	H	C	53.26	5.20			92
nga128	1	20.633.251	H	C	L	L	C	L	L	L	H	H	59.18	4.96			98
SGCSNP10490	1	24.696.198	H	C	H	H	C	L	L	H	L	H	57.14	5.00			98
nga111	1	27.356.874	L	C	H	H	C	L	L	H	L	L	53.06	5.04			98
nga1145	2	683.625	H	H	L	L	C	C	L	H	L	C	59.18	4.96			98
SGCSNP922	2	4.321.924	H	H	L	L	C	C	L	H	L	H	54.26	5.14			94
SGCSNP8895	2	7.860.251	H	C	L	L	C	L	L	H	L	L	57.14	5.00			98
nga1126	2	11.670.000	H	H	C	H	H	L	L	H	L	L	53.06	5.04			98
nga168	2	16.298.919	H	H	C	H	H	L	L	H	L	H	52.04	5.05			98
nga126	3	3.649.000	C	C	H	H	H	L	H	H	L	C	39.80	4.94	54.37	13.49	98
MYF24	3	6.370.809	C	C	H	H	H	L	H	H	L	C	36.73	4.87	46.94	10.58	98
AthGAPab	3	9.796.456	**C**	**C**	**C**	**C**	**H**	**L**	**H**	**C**	**H**	**C**	**23.47**	4.28	25.46	5.49	98
T32N15	3	16.676.053	**C**	**C**	**C**	**C**	**C**	**C**	**H**	**C**	**H**	**C**	**7.14**	2.60	7.19	2.66	98
F28P10	3	20.702.928	**C**	**H**	**C**	**H**	**H**	**C**	**C**	**H**	**C**	**C**	**15.31**	3.64	15.81	4.01	98
nga6	3	23.042.025	**C**	**H**	**C**	**H**	**H**	**C**	**C**	**H**	**H**	**H**	**27.55**	4.51	30.99	6.48	98
nga1111	4	5.074.681	H	H	H	H	L	H	C	H	H	C	62.24	4.90			98
AthF28J12	4	9.456.646	H	H	H	H	L	H	C	H	H	C	55.10	5.02			98
nga1139	4	16.444.155	H	C	C	H	C	H	H	H	H	C	44.90	5.02	73.08	25.94	98
AthCTR1	5	974.507	C	L	H	C	L	C	H	L	H	C	45.92	5.03			98
nga151	5	4.669.932	H	L	H	C	L	C	H	L	H	C	48.98	5.05	114.37	124.99	98
nga76	5	10.418.614	H	L	H	C	H	C	C	H	C	H	31.63	4.70	37.29	7.83	98
AthPHYC	5	14.025.127	H	L	C	C	H	C	C	H	H	H	34.69	4.81	42.77	9.27	98
SO191	5	15.021.915	H	L	H	C	H	C	C	H	C	H	31.91	4.81	37.76	8.11	94
MNF13	5	16.359.785	H	L	H	C	H	C	C	H	H	H	33.70	4.93	40.89	9.03	92
MCL19	5	18.753.210	H	L	H	C	H	C	C	H	H	H	36.96	5.03	47.43	11.09	92
MNB8	5	21.549.766	H	L	H	H	H	H	H	H	H	H	48.91	5.21			92
MUA2	5	23.396.513	H	H	H	C	L	C	H	H	H	H	43.88	5.01	68.25	21.80	98
K8K14	5	26.947.209	H	H	H	C	L	C	H	H	H	H	51.02	5.05			98

Genotyping results obtained for a mapping population of F2 plants derived from a cross involving the wild-type L*er* and a mutant with a Col-0 genetic background. For the sake of clarity, the genotypes of only 10 of the 50 F2 plants analyzed are shown (columns headed as F2 plants 1–10). Values in all the remaining columns correspond to calculations made for the whole mapping population of 50 plants. C: homozygote for the Col-0 allele (to be scored as 0 recombinant chromosomes). H: heterozygote (1 recombinant chromosome), L: homozygote for the L*er* allele (2 recombinant chromosomes), **–** : no amplification obtained for a given marker. **r**: recombination frequency, obtained as 100·recombinant chromosomes/total chromosomes studied. **Sr**: Standard error of recombination frequency calculated as $[r\ (100 - r)/n]^{1/2}$. D: map distance in cM, calculated using the mapping function of Kosambi ***(8)*** as $25 \cdot \ln[(100 + 2r)\ /\ (100 - 2r)]$. **Sd**: Standard error of map distance, calculated as $(2500Sr)\ /\ (2500 - r^2)$. ***n***: total number of chromosomes genotyped. Physical map positions of the markers were obtained from the Munich Information Center for Protein Sequences (MIPS) *Arabidopsis thaliana* database.

Data corresponding to recombination frequencies smaller than 30% are highlighted in bold and were considered unequivocal evidence of linkage. Analysis of the table indicates that the mutation under study is flanked by the T32N15 and F28P10 markers. In addition, recombination frequencies indicate that the mutation is linked to AthGAPab, T32N15, F28P10 and nga6. More specific information can be obtained from the genotypes of plants, which suggest that the mutation is located above F28P10 (plants 2, 4, 5, and 8) and below T32N15 (plants 7 and 9).

2. A very rough, but in many cases good enough, map position can be obtained by genotyping only 18 to 20 F2 plants (*4,7*).
3. As an alternative to using a Mixer Mill, plants can be individually ground using autoclavable plastic pestles attached to a cordless motor grinder. In this case, add 250 µL of DNA extraction buffer to each tube, completely grind the tissue using one pestle per sample. Add another 250 µL of extraction buffer and mix by vortexing.
4. Many of the primer sequences used here are different from those described in some databases for a given marker, and have been designed to obtain PCR amplification products of nonoverlapping sizes.
5. A 0.020–0.025 µmol oligonucleotide synthesis scale should be sufficient to perform a very large number of PCR amplifications. We purchase labeled and nonlabeled oligonucleotides from Applied Biosystems and Sigma, respectively.
6. Less complex mixes than those used here can be obtained by using only a subset of the primer pairs shown in **Table 1**. Some markers of chromosome 5, for instance, can be omitted.
7. We recommend using lyophilized primers, which can be dissolved in water to reach the required concentration. To avoid contamination, we use filtered micropipet tips for the preparation and manipulation of oligonucleotide stock solutions.
8. Some oligonucleotides should be used at a 75 n*M* final concentration in PCR mixes, as we have found that 50 n*M* is not sufficient to ensure a good yield of some amplification products.
9. Although we use a polymerase provided by a local supplier (Ecotaq; Ecogen cat. no. ETAQ500), any other *Taq* polymerase should provide efficient amplification.
10. Although we routinely use ABI thermocyclers, many others can be used to perform the amplifications described here. Nevertheless, some optimization of the thermocycling program may be necessary.
11. DNA sequencers other than those from ABI can be used with the appropriate software for data collection and analysis. Check the list of fluorophores compatible with your sequencer before ordering labeled oligonucleotides.
12. In a multiplex PCR amplification, it is common to obtain different yields of different amplification products. In addition, the intensity of the signal of the amplification products labeled with PET is lower than that of the remaining fluorophores at similar concentrations. Although the yield of amplification products for a few markers may be below the detection threshold in some reactions, the data collected after genotyping 50 plants will be sufficient to find linkage to one or more markers.
13. To map a large number of mutations, we recommend using the Genotyper program rather than manual genotyping.

Acknowledgments

The authors wish to thank V. Quesada for comments on the manuscript, and J. M. Serrano for technical assistance. This work was supported by grants GEN2001-4890-C07-07 and BMC2002-02840 (to J.L.M.), and BMC2003-09763 (to M.R.P.), from the Ministerio de Ciencia y Tecnologia of Spain.

References

1. Lukowitz, W., Gillmor, C. S., and Scheible, W. R. (2000) Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. *Plant Physiol.* **123,** 795–805.
2. Bell, C. J. and Ecker, J. R. (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19,** 137–144.
3. Jander, G., Norris, S. R., Rounsley, S. D., Bush, D. F., Levin, I. M., and Last, R. L. (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol.* **129,** 440–450.
4. Ponce, M. R., Robles, P., and Micol, J. L. (1999) High-throughput genetic mapping in *Arabidopsis thaliana*. *Mol. Gen. Genet*. **261,** 408–415.
5. Cho, R. J., Mindrinos, M., Richards, D. R., et al. (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* **23,** 203–207.

6. Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., and Newburg, L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1,** 174–181.
7. Robles, P. and Micol, J. L. (2001) Genome-wide linkage analysis of *Arabidopsis* genes required for leaf development. *Mol. Genet. Genomics* **266,** 12–19.
8. Kosambi, D. D. (1944) The estimation of map distance from recombination values. *Ann. Eugen.* **12,** 172–175.

10

Gene Identification and Cloning by Molecular Marker Mapping

Georg Jander

Summary

In the course of map-based cloning, mutant genes are identified through linkage to a sufficiently small region of the genetic map and subsequent DNA sequencing. This process has become fairly straightforward for *Arabidopsis* mutations, owing to the completed genome sequence and the discovery of many thousands of molecular markers. Initially, plants with the desired phenotype are identified in populations treated with ethylmethanesulfonate or other mutagens. Once the mutant phenotype has been discovered and confirmed, map-based identification of the mutated gene is a four-stage process: (1) mutant plants are crossed to another *Arabidopsis* ecotype and F2 seeds are generated; (2) approx 50 homozygous mutant F2 plants are genotyped to determine linkage to four or five molecular markers on each of the five chromosomes; (3) A larger F2 population (1000 to 2000 plants) is grown, genotyped, and phenotyped to determine fine-scale genetic linkage, ideally narrowing to a chromosomal region of about 40 kbp; and (4) sequencing of mutant and wild-type DNA is used to verify the identity of the mutated gene. Given a mutant phenotype that can be determined unambiguously in a single F2 plant, it is possible to complete an *Arabidopsis* map-based cloning project in about 1 yr.

Key Words: Map-based cloning; linkage; mutation; mapping; gene identification; markers; recombination.

1. Introduction

Map-based cloning, also called positional cloning, is the process of identifying the location of a gene or mutation by measuring genetic linkage to markers whose physical position in the genome is known. All map-based cloning strategies rely on the fact that, as physical distance between a gene of interest and a molecular marker decreases, so does the genetic recombination frequency. A succession of progressively closer flanking markers on either side of a mutation are identified and, once the physical distance between two flanking markers is small enough, the intervening DNA is sequenced to determine the genetic basis of the mutant phenotype.

The effort involved in map-based cloning has been greatly reduced by recent advances in *Arabidopsis* genetics, including near-complete sequencing of the Columbia (Col-0) genome ***(1)***, the identification of tens of thousands of molecular markers through partial sequencing of the genomes of Landsberg *erecta* (L*er*) ***(2,3)*** and other ecotypes ***(4)***, and improvements in the methods used for detecting molecular markers. As a result, the minimal start-to-finish time for a mapping project has been shortened considerably. Several recent articles describe genotyping methods and iterative approaches that can result in map-based identification of *Arabidopsis* genes in less than one year ***(2,4–11)***.

The development of approaches such as TILLING (**t**argeted **i**nduced **l**ocalized **l**esions **in** **g**enomes) ***(12,13)*** and large collections of T-DNA insertions ***(14–16)*** make it possible to selectively identify mutations in almost any *Arabidopsis* gene. However, in most cases, using these

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

approaches requires some prior knowledge that the protein product of the targeted gene is involved in the plant phenotype that is being investigated. In contrast, map-based cloning requires no prior assumptions, other than that some lesion in the genome causes the phenotype of interest. This makes it possible to associate phenotypes with mutations in previously uninvestigated genes, thereby helping to define the function to a 40% of all *Arabidopsis* genes that have no resemblance to genes with known or inferred activities from plants or other organisms ***(1)***.

Most *Arabidopsis* genetic mapping is done to identify mutations selected from populations of ethyl methanesulfonate (EMS) mutagenized plants. There are several advantages to performing such mutant screens on populations mutagenized with EMS rather than by insertions (T-DNA or transposons) or radiation (fast neutron, gamma ray, or x-ray): (1) EMS induces a high frequency of mutations, typically one every 200 to 300 kbp ***(12,17)***, reducing the number of plants needed for a saturation mutagenesis; (2) the spectrum of EMS-induced mutations is broader, making it possible to find hypomorphic knockdown rather than knockout mutations (particularly important for studying genes with essential function), and dominant mutations; and (3) it is relatively easy to apply EMS mutagenesis to any plant line, including nonstandard ecotypes and cases in which it is desired to find suppressors of already existing mutations.

When contemplating a genetic mapping project, it is important to consider mapping resolution, the physical distance in DNA base pairs to which the position of a mutation can be narrowed by genetic mapping. Given unlimited resources, it is possible to map a mutation to single-gene or even finer resolution. However, in practice, plant growth space and the cost of phenotyping and genotyping limit *Arabidopsis* mapping resolution to about 40 kbp, which typically encode 10 or more genes. With a mapping population of a given size, the resolution to which a gene can be mapped depends on two factors: (1) the number of available genetic markers and (2) the meiotic recombination frequency in the region of the mutation.

Marker density is almost never limiting for mutation mapping projects involving the Col-0 and L*er* ecotypes. More than 50,000 Col-0/L*er* sequence differences have been identified, on average one single-nucleotide polymorphism (SNP) every 3.3 kb and one insertion/deletion (InDel) polymorphism every 6.6 kb throughout the genome ***(2)***. Because most polymorphisms are unique to each individual ecotype ***(4)***, Col-0/L*er* polymorphisms are not necessarily useful for genetic mapping using other pairs of ecotypes. More recently, however, there have been efforts to find polymorphisms among other ecotypes. More than 8600 polymorphisms were identified by sequencing 12 ecotypes ***(4)***. An ongoing project to sequence 1% of the genomes of each of 96 ecotypes (http://walnut.usc.edu/2010/2010.html) will result in the identification of many thousands of polymorphisms that can be used for mapping in crosses between many different ecotypes. These marker collections will open up many new opportunities for gene discovery through genetic mapping.

Given the essentially unlimited supply of *Arabidopsis* genetic markers, meiotic recombination limits mapping resolution. On average, one centiMorgan (cM), the genetic distance that results in one recombination in every 100 meiotic products, corresponds to about 250 kbp of physical distance in the *Arabidopsis* genome ***(5)***. To determine the relative positions of a mutation and a genetic marker in the genome, there needs to be at least one meiotic recombination event between the two. As the distance between the marker and the mutation decreases (i.e., higher mapping resolution), the number of plants needed to have a high probability of finding a recombinant increases exponentially (**Fig. 1**). To map a mutation to a given interval of DNA, at least two recombination events are needed, between the mutation and each of two flanking markers on either side. In an *Arabidopsis* mapping population of 1000 to 2000 plants, there is a high probability of having at least two meiotic recombination events in a given interval of 25 to 50 kbp of DNA (**Fig. 1**). However, with larger mapping populations, the genotyping effort increases linearly with the number of plants, but the mapping resolution is not greatly improved.

A typical mutation-mapping project has four phases (**Fig. 2**). First, the mutant line is crossed to another *Arabidopsis* ecotype, preferentially one with known DNA differences. Phenotypic analysis of the F1 and F2 plants is used to determine whether the mutation is dominant or

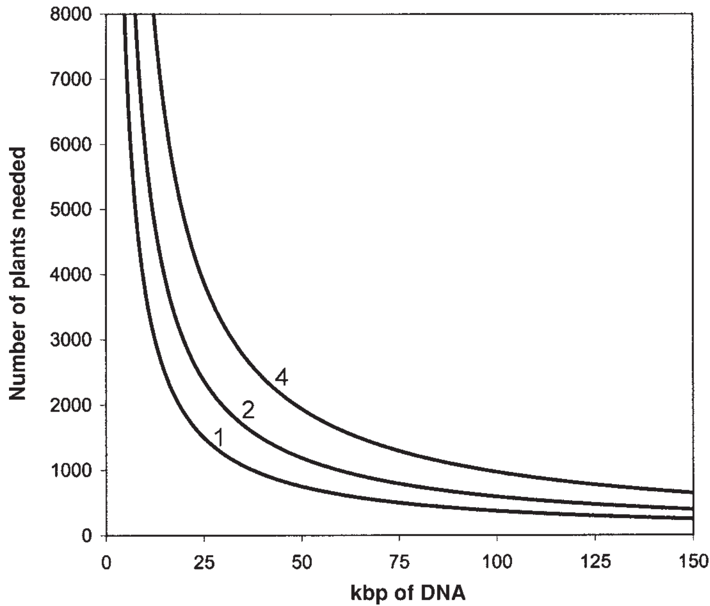


Fig. 1. Number of plants needed to find recombinants in a given DNA interval. The curves show the number of F2 plants needed to have a 95% chance of finding at least one plant (1), at least two plants (2), or at least four plants (4) with recombination events in a given physical interval of DNA. The calculations assume an average 250 kb/cM for *Arabidopsis* (5). The possibility of multiple recombination events in one individual plant has a negligible effect and is not included in the calculations.

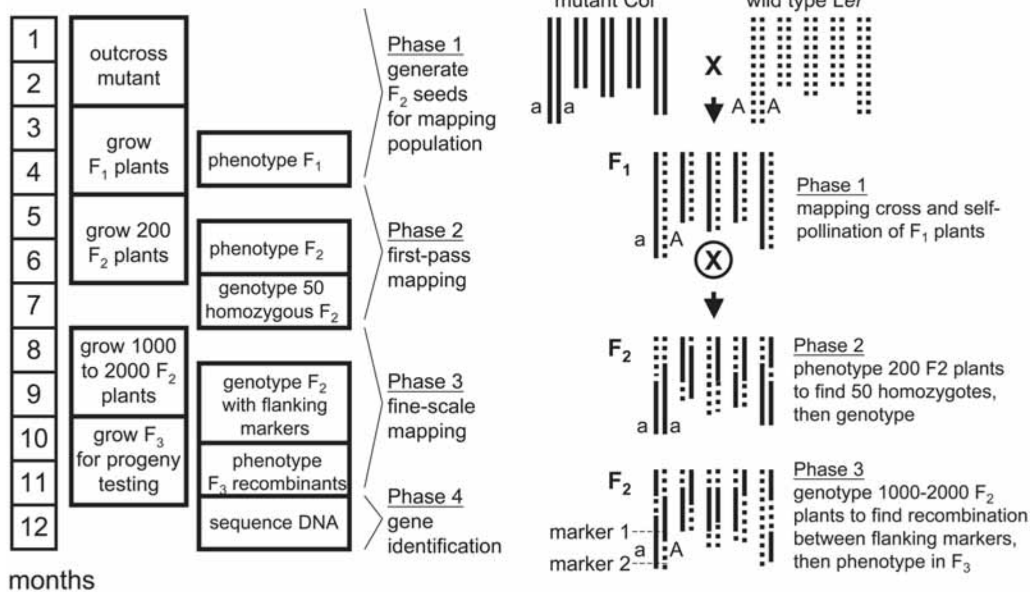


Fig. 2. Left: phases of a map-based cloning project, with possible timeline leading to gene identification in 1 yr. Right: schematic of the five pairs of *Arabidopsis* chromosomes during critical stages of a sample mapping of a recessive mutation on chromosome 1 in the Col background, crossed to wild-type L*er*.

recessive. Second, approx 50 homozygous mutant (for recessive mutations) or homozygous wild-type (for dominant mutations) F2 plants are identified based on the phenotype. These F2 plants are genotyped with markers on each of the five chromosomes to narrow the position of the mutation to about 20 cM resolution. Additional markers in the 20 cM interval are used to narrow the position as much as possible, ideally to less than 5 cM. Third, a larger population of 1000 to 2000 F2 plants is generated and genotyped with markers flanking the 5 cM or smaller interval containing the mutation. The phenotypes are tested for plants that are recombinant in this interval to infer the genotype at the site of the mutation and to determine on which side the mutation the meiotic crossovers occurred. Additional DNA markers are used to narrow down the position of the crossovers and thereby also the site of the mutation. Fourth, once the position of the mutation has been narrowed to less than 50 kbp, the DNA sequence is analyzed to determine the underlying genetic lesion. Because the entire *Arabidopsis* genome sequence is known, it is often possible to pick out candidate genes that might be mutated to cause a given phenotype. Therefore, the iterative mapping process can be short-circuited at any stage to sequence likely candidate genes and identify the mutation.

2. Materials

2.1. Phase 1: Generation of F2 Seeds for Mapping Population

1. Standard plant growth materials.
2. Materials for crossing *Arabidopsis*, as described in Chapter 6 of this book.
3. Other materials needed to analyze the phenotype of interest.

2.2. Phase 2: First-Pass Mapping

2.2.1. Identification of Plants for Genotyping

1. Standard plant growth materials.
2. Other materials necessary for assessing the mutant phenotype.

2.2.2. DNA Preparation Using CTAB Method

1. Materials for cetyl trimethylammonium bromide (CTAB) protocol, as described in Chapter 14 of this book.

2.2.3. Determining Genetic Linkage Using 20 to 25 Markers

1. Polymerse chain reaction (PCR) materials: *Taq* polymerase, buffer, dNTPs.
2. Restriction enzymes, if needed.
3. Thermal cycler.
4. Tubes for thermal cycler.
5. Agarose, Sigma IIA.
6. Agarose, Fisher low-melt.
7. TAE agarose gel buffer (0.04 *M* Tris-acetate, 0.001 *M* EDTA, pH 8.0).
8. Agarose gel box and power supply.
9. Computer or pencil and paper for data analysis.

2.2.4. Refining Map Position With Additional Markers

1. Same materials as in **Subheading 2.2.3.**
2. Restriction enzymes for cleaved amplified polymorphic sequence (CAPS) markers.
3. Computer access to The Arabidopsis Information Resource (TAIR).

2.3. Phase 3: Fine-Scale Mapping

Same materials needed as in **Subheading 2.2.**

2.4. Phase 4: Gene Identification

1. Computer DNA sequence analysis software.
2. Thermal cycler and PCR reagents.
3. DNA sequencing machine, ABI 3700 or similar.

3. Methods

3.1. Phase 1: Generation of F2 Seeds for Mapping Population

As the first step to mapping a mutation, it is necessary to cross the mutant line to another *Arabidopsis* ecotype. For instance, a mutation in the Col-0 background can be crossed to L*er* or vice versa (*see* **Notes 1–3**).

1. Choose parent plants for crosses (*see* **Note 4**) and cross as described in Chapter 6 of this book.
2. Harvest, then plant F1 seeds (*see* **Note 5**).
3. If possible, phenotype F1 plants to determine whether the mutation is dominant or recessive.
4. Collect mature F2 seeds from F1 plants (*see* **Note 5**).

3.2. Phase 2: First-Pass Mapping

3.2.1 Identification of Plants for Genotyping

1. Plant a set of at least 200 F2 seeds from the crosses in Phase 1.
2. Analyze the phenotype of interest to determine whether the mutation is dominant or recessive (*see* **Note 6**).
3. Choose approx 50 plants that are either homozygous mutant (recessive mutations) or homozygous wild-type (dominant mutations) for genotyping (*see* **Note 7**).

3.2.2. DNA Preparation Using CTAB Method

The hexadecyltrimethyl ammoniumbromide (CTAB) method is one of several effective protocols for isolating PCR-quality DNA (*see* **Note 8**).

1. Harvest 50 to 100 mg of tissue from each of the 50 F2 plants into 1.5-mL microcentrifuge tubes (*see* **Note 9**).
2. Prepare DNA using the CTAB protocol described in Chapter 14 of this book.

3.2.3. Determine Genetic Linkage Using 20 to 25 Markers

1. Choose 20 to 25 codominant markers to use for genotyping, e.g., those listed in **Fig. 3** (*see* **Notes 10** and **11**).
2. Amplify marker DNA from F2 progeny and parental ecotypes by PCR. The markers listed in **Fig. 3** can be amplified with the following protocol:
   a. Reaction ingredients: 0.5 µL *Taq* polymerase, 1 µL PCR buffer with magnesium, 1 µL 25 m*M* dNTP, 1 µL template DNA, 0.1 µL of each primer (100 µ*M*), 6.3 µL water (10 µL total volume).
   b. PCR cycles for amplification are: step 1, 94°C for 3 min; 30 cycles (step 2, 94°C for 15 s; step 3, 55°C for 20 s; step 4, 72°C for 30 s), step 5, 72°C for 10 min; step 6, 4°C hold.
3. If using CAPS markers, cut DNA with appropriate restriction enzymes.
4. Separate amplified PCR products on an agarose gel and score genotypes. A 4% agarose gel (2.5% Sigma IIA and 1.5% Fisher low-melt) works for the markers in **Fig. 3** (*see* **Note 12**).
5. Identify the markers that are most closely linked to the mutation (*see* **Note 13**).
6. Determine order of the mutation and the markers using a three-point cross (**Fig. 4**) (*see* **Notes 14** and **15**).
7. Look for candidate genes in the area to which the mutation was mapped.

Chr.	Position (Mbp)	BAC	Marker Name	Forward Primer	Reverse Primer	Col-0 Size (bp)	L*er* Size (bp)
1	0.8	F21B7	CER450629	TGAATTTGCAAGACAAAGCAGGAGTA	AACCAAGAAAGACTTCACCGATCTCC	147	118
1	5.2	F10B6	CER448485	AAATGGGAGATGTTGATGTTCTTC	CCAAAGGGTCACGAGTGAAGTCTAG	230	209
1	11.0	F17F8	CER449723	TTGGATTATCTCTTGAATGGGTTTGG	CCGTCTTTAGCATTAGGAACTCTGGTG	216	164
1	16.2	F1I21	CER450299	AGGGCGGATAACAAACTGTTGTGGAC	GATCACCCATCTGAAAGCATAACACA	104	89
1	25.7	T26J14	CER470898	GGATTAAAAAGGTGAAATGAAAATGTGTGA	AGATCTGACGATATCCTTGCAACCAA	149	136
2	0.8	T20F6	CER459234	TCCTTTCTTAGCAGCCTTCACAGATG	TCATGAAGAAGATGTTCGTGTAGGAGC	228	179
2	14.3	T1B8	CER459006	GTCTGGATCTTGTATCGCAAACCAAA	TGCTTCAGAGCTGATGAACAAAAGAC	199	154
2	19.5	F17A22	CER449664	TCACATGCATTCATAACGTTTTCAC	TGTTATTACCACGGTCACCCATACAA	147	135
3	5.3	MSJ11	CER457071	TAACCACACACATCGTGTTTTTGTCC	GGGTCTGCTCATTCTTCAGTTCTTGT	174	153
3	15.0	F18P9	CER477689	AACAGATACCAACTCCACAGGGACAA	TCAATGGCAGCAAAGTCCATTAAATC	126	105
3	19.4	T25B15	CER470856	ATACTTGGGTCCCTTCGTCC	GTCGCGAGTTCTCTCCAATC	110	92
3	23.3	T20O10	CER479303	CGATCATGCGTTTACCATACATCTAA	TCATTTACGCTGTAGCAAACGTGGTA	125	100
4	2.7	C6L9	CER465774	TGAAGCGGTTTCAAGATTGTT	CAAATTTAGCGACACGACCA	143	118
4	8.0	F18A5	CER466221	TCAAACGCCAACTGAATCAA	CCTGGTGTTGTGTGTGGACT	216	191
4	12.8	F617	CER452834	GAAAATGCTTCCAGCTCGTT	AAGCGTATACGACATGTATAAGTAGC	151	113
4	16.2	T16L1	CER458597	GGCAATGTATTTTACACATACTTGATT	TTTCACTACTTCCAGCACCAG	129	88
5	0.3	T20L15	CER478436	GGACGACTATAAACCTACTTTTGCGTCA	CTGTTTTAATGAGTTGTCGGCATCG	222	205
5	9.4	F2P16	CER451893	AATTGTGGGAAGGACAACAACCAAA	GAGAGAGCACGTGAGATGTCACAGA	163	130
5	17.1	MFO20	CER462550	AAACATTTGGGCAGGGGTGGATA	GTGGCACTCACTGGTGTGATGA	237	195
5	23.4	MRI1	CER456965	CCGACTACCGCAAATAGAATACGAAA	GTCAACCACATACACGCACCATACATAA	168	154

Fig. 3. Col-0/L*er* InDel markers from the Monsanto/Cereon collection (www.arabidopsis.org/Cereon/index.jsp), which can be used to determine linkage of mutations to the five chromosomes of *Arabidopsis*. Each pair of primers will amplify a chromosomal segment where there is a DNA insertion in Col-0 relative to L*er.* Amplified products with the indicated sizes can be differentiated on agarose gels. These markers also work reasonably well for mutation mapping by bulked segregant analysis (*see* **Note 11**). Marker confirmation, primer design, and testing were done by Angela Peragine and Scott Poethig, University of Pennsylvania.

F_2 line #	Marker 1		Mutation		Marker 2	Mutation		Marker 1		Marker 2	Mutation		Marker 2		Marker 1
1	Het	X	Col		Col	Col	X	Het	X	Col	Col		Col	X	Het
2	Col		Col	X	Het	Col		Col	X	Het	Col	X	Het	X	Col
3	Col		Col	X	Het	Col		Col	X	Het	Col	X	Het	X	Col
4	Het	X	Col		Col	Col	X	Het	X	Col	Col		Col	X	Het
5	Ler	XX	Col		Col	Col	XX	Ler	XX	Col	Col		Col	XX	Ler
6	Col		Col	X	Het	Col		Col	X	Het	Col	X	Het	X	Col
7	Col		Col	X	Het	Col		Col	X	Het	Col	X	Het	X	Col
8	Het	X	Col		Col	Col	X	Het	X	Col	Col		Col	X	Het
9	Het	X	Col		Col	Col	X	Het	X	Col	Col		Col	X	Het
10	Col		Col	X	Het	Col		Col	X	Het	Col	X	Het	X	Col
11-50	Col		Col		Col	Col		Col		Col	Col		Col		Col
	Order A 11 crossovers					Order B 17 crossovers					Order C 16 crossovers				

Fig. 4. A typical three-point cross to determine the order a mutation and two nearby molecular markers. It is assumed that a recessive mutation in the Col-0 background is being mapped using a cross to L*er,* and that only homozygous mutant lines are genotyped. Likely genotypes of 50 F2 lines are presented and crossovers are indicated with *X*'s. Order A is the most likely, as it assumes the fewest meiotic crossovers. Het = heterozygous allele, Col = homozygous Col-0, and L*er* = homozygous L*er.*

3.2.4. Refine Map Position with Additional Markers

1. Identify additional markers between the flanking markers identified in **Subheading 3.2.3.** (*see* **Note 16**).
2. Genotype recombinants with additional markers (*see* **Note 16**).
3. Identify the closest markers that flank either side of the mutation (*see* **Note 17**).
4. Look for candidate genes in the area to which the mutation was mapped.

3.3. Phase 3: Fine-Scale Mapping

1. Plant a population of 1000 to 2000 F2 lines (*see* **Notes 18** and **19**).
2. Harvest tissue and prepare DNA as in **Subheading 3.2.2.**
3. Genotype plants with the closest flanking markers identified in **Subheading 3.2.4.**
4. Identify genetic recombination between the flanking markers (*see* **Note 20**).
5. Phenotype recombinants in the F3 generation (*see* **Note 21**).
6. Genotype with progressively closer markers until no further recombinants are found (*see* **Notes 22** and **23**).
7. Look for candidate genes in the area to which the mutation was mapped.

3.4. Phase 4: Mutation Identification

The experiments in Phase 3 should narrow the physical position of the mutation to a region of 50 kbp or less. At this point, additional genetic mapping is unlikely to be cost-effective. If no obvious candidate genes in the identified DNA interval can account for the mutant phenotype, it is necessary to do some sort of direct analysis of the DNA sequence to identify the mutation. Given the ever-decreasing costs of PCR primers and sequencing, complete sequencing of the 50 kbp or smaller interval may be the fastest and most cost-effective approach for identifying the mutation (*see* **Note 24**).

1. Divide the region containing the mutation into overlapping 600- to 1000-bp fragments using DNAStar (www.dnastar.com/), or a similar DNA sequence manipulation program.
2. Design primers and amplify the DNA fragments from a homozygous mutant line (*see* **Note 25**).
3. Sequence the DNA fragments in both directions and align sequences using Phred and Phrap (www.phrap.org/index.html), or similar base calling and contig assembly software.
4. Identify differences between the mutant and wild-type sequences using Consed (www.phrap.org/consed/consed.html), or similar software.
5. Resequence mutant and wild-type DNA to confirm the mutation (*see* **Note 26**).

4. Notes

1. It is uncommon that more than one induced mutation affects a given phenotype. Therefore, it is generally not necessary to backcross a mutant to “clean up” the genetic background prior to outcrossing to another ecotype to generate a mapping population.
2. Sometimes natural variation between ecotypes affecting the phenotype of interest will make it difficult or impossible to map an induced mutation. Such natural variation may become apparent only in progeny from the cross (transgressive segregation). Therefore, it may be desirable to cross the mutant line to more than one divergent ecotype in the hope that, in at least one cross, the genetic background affecting the phenotype of interest is similar to that of the mutagenized line.
3. Methods similar to those used for mapping induced mutations can be used to identify the genetic basis of quantitative trait loci (QTLs) that affect natural variation among ecotypes ***(18,19)***. Typically, pairs of *Arabidopsis* ecotypes have tens of thousands of DNA polymorphisms. For instance, comparison of the Col-0 and L*er* DNA sequences shows an average of one polymorphism every few hundred bp ***(3)***. In theory, mapping QTLs is not different from mapping induced mutations. However, unlike in the case of induced mutations, it is quite common that several QTLs influence a given phenotype. Therefore, to facilitate fine-scale mapping, it is often necessary to first obtain a rough map position for a QTL and introgress it into a common genetic background (*see* Chapter 7 in this book).

4. If it is already known whether the induced mutation is dominant or recessive, it is advantageous to use the line with the recessive allele (either mutant or wild-type) as the female parent in the cross. This will make it apparent if some F1 progeny are the result of self-pollination. It is also possible to use other known recessive mutations in the ecotype background, e.g., the *erecta* mutation in L*er*, to ensure that no self-pollination has occurred. A further consideration is that highly mutagenized plants that have not been backcrossed are often less robust than wild-type, making them less desirable as pollen recipients.
5. Do not pool F1 or F2 seeds from different crosses. If progeny from one cross are later discovered to be undesirable due to self-pollination or other problems, these lines can be eliminated from further analysis. If the mutant line is in an early stage of being characterized, it may be heterozygous for a dominant mutation. By not pooling F2 seeds, it is possible to continue later work only with progeny from F1 plants that received the mutant allele.
6. In an ideal case, the segregation pattern will be 3:1 mutant:wild-type for a dominant mutation and 1:3 mutant:wild-type for a recessive mutation. In practice, many factors can act to skew this ratio. The mutation may be incompletely penetrant or influenced by environmental factors. Inheritance of the mutant allele through one or the other gamete may be incomplete. Other induced mutations linked to the mutation being mapped may have deleterious effects and thereby cause a skewed segregation ratio. QTL segregating in the mapping population may affect the phenotype of interest. If the phenotype is difficult to determine accurately, it may be necessary to analyze plants in the F3 generation to confirm the phenotype of an individual F2 plant. Traits affected by the maternal genotype, e.g., seed phenotypes, also have to be analyzed in the F3 generation.
7. Only plants whose homozygous genotype can be inferred from the phenotype are chosen for genotyping. This reduces the amount of work, because it is not necessary to analyze the phenotype in the F3 generation to infer the genotype of individual F2 plants. For instance, in the case of a recessive mutation, F2 plants that do not show the phenotype have an indeterminate genotype, either heterozygous or homozygous wild-type. Therefore, only plants that show the phenotype and are therefore homozygous mutant are genotyped.
8. The CTAB protocol and variations thereof produce DNA that can be used reliably for at least 50 PCR reactions. Because some degradation of DNA occurs with repeated freezing and thawing, DNA should be frozen as several individual aliquots if it will be used repeatedly. Other published protocols, some of which lend themselves more readily to automation in microtiter plates than the CTAB method, can also be used to prepare PCR-quality DNA ***(20–23)***. Commercially available kits (e.g., Qiagen) are relatively expensive, but produce high-quality DNA that stores well.
9. This is two or three medium-sized leaves of *Arabidopsis*. Younger leaves tend to result in more DNA than ones that are starting to senesce. Use of flowers for DNA prep also results in a relatively large amount of clean DNA.
10. The InDel markers in **Fig. 3** are spread across the genome, making it possible to measure linkage of almost any mutation to one or more of these markers. These InDel polymorphisms have been chosen because PCR amplification yields different-sized bands for Col-0 and L*er,* which can easily be differentiated on agarose gels. The Mbp position of the markers (column 2 in **Fig. 3**) is only approximate and may change somewhat if the physical map of the *Arabidopsis* chromosomes is updated. Other sets of PCR-based markers for first-pass mapping, e.g., simple sequence repeat (SSR) markers ***(5)***, CAPS markers ***(10,24)***, or primer extension assays ***(9)*** have also been published. All of these marker types are codominant, meaning that it is possible to differentiate heterozygotes from either parent allele. Many additional verified markers are available through The Arabidopsis Information Resource (TAIR), www.arabidopsis.org/.
11. The protocol described here assumes that each F2 plant–marker combination will be amplified by PCR and analyzed individually. It is possible to reduce the amount of work involved considerably by pooling DNA from individual phenotyped plants for bulked segregant analysis ***(5)*** or performing several marker amplification reactions in the same tube ***(9,11)***. However, there are also disadvantages to these approaches: Information on individual plants is lost with bulked segregant analysis; misscoring of the phenotypes of individual plants is difficult to trace; and it is not possible to identify recombination between closely linked loci. Moreover,

sometimes it can be difficult to make multiplexed PCR reactions work with equal efficiency. In the future, if costs go down, microarray-based genotyping may be the most efficient for this phase of mutation mapping projects.

12. It is important to amplify DNA from the parental lines and run this DNA on the same 1% agarose gel for comparison. Amplified DNA markers from F2 lines can then be scored as homozygous for either parental allele (one band) or heterozygous (two bands).
13. Recombination frequency between the mutation and each of the 25 markers in the 50 F2 plants is calculated to determine linkage. For instance, if a recessive mutation in the Col-0 background is mapped in a cross to L*er*, the recombination frequency (F) for any one marker is calculated as: F = (no. heterozygotes + 2 × no. homozygous L*er*) / (2 × no. plants). For markers on other chromosomes and for markers that are distant on the same chromosome, the recombination frequency should be approx 0.5, i.e., there is an equal probability of having the Col-0 or the L*er* allele of a given marker. Given the spacing of the markers in **Fig. 3** and the meiotic recombination frequency in *Arabidopsis*, one or more of the markers should have a significantly reduced recombination frequency, i.e., an overabundance of the Col-0 allele. These markers are genetically linked to the mutation, which, in this particular case, is assumed to be homozygous Col-0.
14. As shown in **Fig. 4**, if two markers show linkage to a mutation, these three loci must have one of three possible orders in the genome. The most likely order is determined with a three-point cross-analysis, which is based on the fact that double meiotic recombination events in a short region of DNA are much less common than single meiotic recombination events. As in **Note 13**, mapping of a recessive mutation in the Col-0 background is assumed for this example. For each of the 50 F2 lines, the three possible locus orders are considered, and it is determined how many meiotic recombination events would have to occur for the loci to be in this order. For each of the three possible locus orders, the number of recombination events in the 50 F2 lines is added together, and the order that requires the fewest recombination events is assumed to be correct. The closest two markers for which the correct order in the genome is [marker 1–mutation–marker 2] are identified as flanking markers for the mutation. A special case occurs if the mutation is between a marker and a telomere, in which case there will be only one flanking marker. Three-point cross analysis can be done with pencil and paper or by computer. For instance, it is fairly easy to create a template or a macro in Microsoft Excel for doing this analysis.
15. When identifying flanking markers, it is common to find individual F2 lines that do not fit the model, i.e., given the more likely locus order that is determined with a three-point cross, one has to assume two meiotic recombination events in these lines. Although some fraction of plants will indeed be double recombinants, the most likely cause is that the phenotypes of these F2 lines were misscored. If there are not too many double recombinants, the easiest solution is to go with the "preponderance of evidence" and leave out the double recombinants from further analysis. Otherwise, it is possible to analyze the mutant phenotype in the F3 generation and thereby infer the mutant genotype in the F2 generation (*see also* **Note 21**). In this way, no mapping data are lost.
16. Additional markers, roughly one every 250 kbp, can be identified through the TAIR database, www.arabidopsis.org/. For researchers who work at not-for-profit institutions, the Monsanto/Cereon marker collection (www.arabidopsis.org/Cereon/index.jsp) provides a source of more than 50,000 polymorphisms that can be used to map mutations in F2 populations derived from Col-0 and L*er*. Although these polymorphisms are predicted only from sequence data, more than 90% are confirmed in subsequent analyses *(3)*. At this stage in the mapping process, it is likely that there are many available markers, and it is possible to choose those that can be analyzed the most easily with the available resources. For many labs, this means identifying codominant markers that can be analyzed as different-sized fragments on gels. InDel and SSR markers can be analyzed by PCR size differences as described in **Subheading 3.2.3.** SNP polymorphisms that either create or remove a restriction site can be utilized as CAPS markers *(25)*. For CAPS markers, DNA surrounding the polymorphism is amplified by PCR (similar to InDel and SSR markers), but, prior to running the PCR product on a gel, it is cleaved with a restriction enzyme that cuts in only one of the two alleles. CAPS markers are codominant, making it possible to differentiate heterozygotes from either homozygote.

17. Only plants that have genetic recombination between the flanking markers identified in **Subheading 3.2.3.** need to be genotyped with additional markers. Given the spacing of markers used for first-pass mapping, this is likely to be approx 10 plants. As molecular markers progressively closer to the mutation are analyzed, they will all eventually be homozygous for the same genotype as that which is inferred for the 50 F2 plants in **Subheading 3.2.1.** Three-point cross analysis (as in **Note 18**) can be used to define the relative order of the markers and the mutation. The markers closest to the mutation, but for which there is still at least one recombination event between the marker and the mutation, will be the new flanking markers. With a population of 50 F2 plants, there is a high probability that these final flanking markers are less than 5 cM apart. Since the accuracy of this result is important, DNA from these lines should be reisolated and the flanking marker genotypes confirmed. It is also worthwhile to confirm the mutant phenotype of these plants in the F3 generation. Since hundreds of plants will be genotyped with flanking markers in Phase 3 of the mapping process, it is important that the chosen flanking markers are robust and that the alleles are easy to differentiate.
18. With an F2 population of 1000 to 2000 plants, there is a high probability that it will be possible to map the mutation of interest to a region of less than 50 kbp (**Fig. 1**) if the local meiotic recombination frequency is 1/250 kbp. It is possible to start this phase of the mapping project with a smaller F2 population. However, this risks having to plant another F2 population later, thereby extending the timeline shown in **Fig. 2**.
19. In many cases, genotyping is faster and more accurate than phenotype determination, and hence should be done first in this stage of mapping. However, if the phenotypic assay is easy, nondestructive, and can be done during the growth of the plant (for instance, the mutation causes a dwarf phenotype), it is possible to phenotype F2 plants at this stage, and only genotype those that are homozygous, as in the Phase 1 stage of mapping. This approach will require planting four times as many F2 plants. Only plants that are recombinant between the flanking markers will eventually provide additional mapping information, and genotyping only homozygotes will not increase the odds of finding recombination near the mutation.
20. The goal is to identify plants that have meiotic recombination near the mutation. In most cases, these will be plants that are heterozygous for one flanking marker and homozygous for the other, although double recombinants will also occur, but infrequently (roughly as the product of the frequency of each single recombination event). The number of recombinants that are identified will depend on the genetic distance between the flanking markers. For instance, if the flanking markers are 5 cM apart, about 10% of all plants will be recombinant.
21. If the genotype of plants at the site of mutation was not determined prior to identifying recombinants with flanking markers, it will need to be inferred based on segregation of the phenotype in the F3 generation. Heterozygous F2 plants will produce both mutant and wild type progeny, whereas plants homozygous for either allele will produce only homozygous progeny. In the final stages of the mapping process, the predicted location of the mutation will depend on a small number of plants with nearby meiotic recombination. Therefore it is important to be *absolutely certain* of the phenotype and inferred mutant genotype of these lines.
22. Additional markers can be identified through the TAIR database for genotyping the recombinant plants. Three-point crosses are used to determine the relative order of the mutation and the markers. Unlike in the analysis described in **Note 18**, not all the plants will be homozygous at the site of the mutation, but the genotype will have been determined as described in **Note 21**. Progressively closer markers near the mutation are identified in this manner. In each cycle, the number of genotyping reactions is reduced, because only the plants that were previously found to be recombinant between one of the markers and the mutation need to be analyzed. As the interval containing the mutation is narrowed, the number of available markers will decrease, and it may not be possible to find InDel or CAPS markers that lend themselves to convenient gel-based analysis. Some SNP polymorphisms can be analyzed as derived CAPS (dCAPS) ***(26)*** or single nucleotide amplified polymorphism (SNAP) ***(27)*** markers through appropriate choice of PCR amplification primers. Other methods for determining the genotype of SNPs or small InDel polymorphisms, e.g., pyrosequencing ***(28)***, denaturing HPLC ***(29)***, or Taqman assays ***(30)***, are not feasible for all laboratories. If the number of remaining recombinants is small, the fastest and most cost-effective approach may be to PCR-amplify a segment of DNA containing the polymorphism and sequence it using standard methods to determine the genotype.

23. It should be possible to identify markers less than 50 kbp apart that flank the mutation. However, the curves in **Fig. 1** are based on average meiotic recombination frequencies, and the ratio of genetic and physical distance varies across the genome. Many mutations have been mapped to high resolution with fewer plants than indicated in **Fig. 1** *(5)*. Conversely, meiotic recombination frequency is reduced near the telomeres, near centromeres, and in regions of the genome that represent large-scale DNA inversions or translocations in one ecotype of a mapping cross relative to the other. If a mutation is in such a region, it may not be possible to grow and genotype enough plants to map a mutation to a region smaller than several hundred kbp.
24. If direct sequencing is not an option, other methods can be employed to narrow down the DNA region predicted to contain the mutation. Most commonly, individual genes or groups of genes from wild-type plants are cloned into *Agrobacterium* T-DNA vectors to complement the mutant *Arabidopsis* line, if the mutation is recessive. Conversely, in the case of dominant mutations, DNA cloned from the mutant line can be transformed into wild-type plants. Approaches that detect mismatched bases, e.g., denaturing HPLC *(31)*, can also be used to identify regions of DNA containing polymorphisms, but this requires specialized equipment that is not available to all researchers.
25. Depending on the size of the amplified fragments and the amount of overlap, somewhere between 100 and 200 primer pairs will be necessary to amplify and sequence 50 kbp. Obviously, fewer primers will be needed if the mutation is mapped to higher resolution. If the mutation was originally isolated in the Col-0 background, it is not necessary to sequence the wild-type allele, as this information is already available through TAIR or GenBank. In the case of a mutation in another genetic background, both mutant and wild-type alleles need to be amplified and sequenced. In addition, since primers for PCR amplification will most likely be designed based on the known Col-0 sequence, some may fail due to ecotype-specific polymorphisms in the primer binding site.
26. In any sequencing project of this size, it is likely that there will be apparent polymorphisms that are due to sequencing errors, either in the mapping project, or in the original wild-type DNA sequence. Therefore, all potential mutations should be confirmed by resequencing both the mutant and wild-type DNA.

References

1. The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815.
2. Jander, G., Norris, S. R., Rounsley, S. D., Bush, D. F., Levin, I. M., and Last, R. L. (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol.* **129,** 440–450.
3. Rounsley, S. (2003) Sharing the wealth. The mechanics of data release from industry. *Plant Physiol.* **133,** 438–440.
4. Schmid, K. J., Sorensen, T. R., Stracke, R., Torjek, O., Altmann, T., Mitchell-Olds, T., and Weisshaar, B. (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Res.* **13,** 1250–1257.
5. Lukowitz, W., Gillmor, C. S., and Scheible, W. R. (2000) Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. *Plant Physiol.* **123,** 795–805.
6. Peters, J. L., Constandt, H., Neyt, P., Cnops, G., Zethof, J., Zabeau, M., and Gerats, T. (2001) A physical amplified fragment-length polymorphism map of *Arabidopsis*. *Plant Physiol.* **127,** 1579–1589.
7. Peters, J. L., Cnops, G., Neyt, P., Zethof, J., Cornelis, K., Van Lijsebettens, M., and Gerats, T. (2004) An AFLP-based genome-wide mapping strategy. *Theor. Appl. Genet.* **108,** 321–327.
8. Peters, J. L., Cnudde, F., and Gerats, T. (2003) Forward genetics and map-based cloning approaches. *Trends Plant Sci.* **8,** 484–491.
9. Torjek, O., Berger, D., Meyer, R. C., et al. (2003) Establishment of a high-efficiency SNP-based framework marker set for *Arabidopsis*. *Plant J.* **36,** 122–140.
10. Baumbush, L. O., Sundal, I. K., Hughes, D. W., Galau, G. A., and Jakobsen, K. S. (2001) Efficient protocols for CAPS-based mapping in *Arabidopsis*. *Plant Mol. Biol. Rep.* **19,** 137–149.
11. Ponce, M. R., Robles, P., and Micol, J.L. (1999) High-throughput genetic mapping in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **261,** 408–415.
12. Greene, E. A., Codomo, C. A., Taylor, N. E., et al. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* **164,** 731–740.

13. Till, B. J., Reynolds, S. H., Greene, E. A., et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res*. **13,** 524–530.
14. Rosso, M. G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003) An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol. Biol*. **53,** 247–259.
15. Alonso, J. M., Stepanova, A. N., Leisse, T. J., et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301,** 653–657.
16. Sessions, A., Burke, E., Presting, G., et al. (2002) A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell*. **14,** 2985–2994.
17. Jander, G., Baerson, S. R., Hudak, J. A., Gonzalez, K. A., Gruys, K. J., and Last, R. L. (2003) Ethylmethanesulfonate saturation mutagenesis in *Arabidopsis* to determine frequency of herbicide resistance. *Plant Physiol*. **131,** 139–146.
18. Remington, D. L., Ungerer, M. C., and Purugganan, M. D. (2001) Map-based cloning of quantitative trait loci: progress and prospects. *Genet. Res*. **78,** 213–218.
19. Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C. J., and Koornneef, M. (1999) Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **96,** 4710–4717.
20. Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983) A plant DNA minipreparation: version II. *Plant Mol. Biol. Rep*. **1,** 19–21.
21. Weigel, D. and Glazebrook, J. (2002) *Arabidopsis: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Klimyuk, V. I., Carroll, B. J., Thomas, C. M., and Jones, J. D. G. (1993) Alkali treatment for rapid preparation of plant material for reliable PCR analysis: technical advance. *Plant J*. **3,** 493–494.
23. Li, J. and Chory, J. (1998) Preparation of DNA from *Arabidopsis*, in *Arabidopsis Protocols* (Martinez-Zapater, J. and Salinas, J., eds.) Humana Press. Totowa, NJ, pp. 55–60.
24. Glazebrook, J., Drenkard, E., Preuss, D., and Ausubel, F. M. (1998) Use of cleaved amplified polymorphic sequences (CAPS) as genetic markers in *Arabidopsis thaliana*, in *Arabidopsis Protocols* (Martinez-Zapater, J. and Salinas, J., eds.) Humana Press, Totowa, NJ, pp. 173–182.
25. Konieczny, A. and Ausubel, F. M. (1993) A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J*. **4,** 403–410.
26. Neff, M. M., Neff, J. D., Chory, J., and Pepper, A. E. (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant J*. **14,** 387–392.
27. Drenkard, E., Richter, B. G., Rozen, S., et al. (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. *Plant Physiol*. **124,** 1483–1492.
28. Ahmad, M., Jarillo, J. A., Klimczak, L. J., Landry, L. G., Peng, T., Last, R. L., and Cashmore, A. R. (1997) An enzyme similar to animal type II photolyases mediates photoreactivation in *Arabidopsis*. *Plant Cell* **9,** 199–207.
29. Spiegelman, J. I., Mindrinos, M. N., and Oefner, P. J. (2000) High-accuracy DNA sequence variation screening by DHPLC. *Biotechniques* **29,** 1084–1090, 1092.
30. Livak, K. J. (1999) Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet. Anal*. **14,** 143–149.
31. O'Donovan, M. C., Oefner, P. J., Roberts, S. C., et al. (1998) Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics* **52,** 44–49.

11

High-Throughput TILLING for *Arabidopsis*

Bradley J. Till, Trenton Colbert, Christine Codomo, Linda Enns, Jessica Johnson, Steven H. Reynolds, Jorja G. Henikoff, Elizabeth A. Greene, Michael N. Steine, Luca Comai, and Steven Henikoff

Summary

Targeting induced local lesions in genomes (TILLING) is a general strategy for identifying induced point mutations that can be applied to almost any organism. In this chapter, we describe the basic methodology for high-throughput TILLING. Gene segments are amplified using fluorescently tagged primers, and products are denatured and reannealed to form heteroduplexes between the mutated sequence and its wild-type counterpart. These heteroduplexes are substrates for cleavage by the endonuclease CEL I. Following cleavage, products are analyzed on denaturing polyacrylamide gels using the LI-COR DNA analyzer system. High-throughput TILLING has been adopted by the *Arabidopsis* TILLING Project (ATP) to provide allelic series of point mutations for the general *Arabidopsis* community.

Key Words: TILLING; mutation; SNP; CEL I; reverse genetics; functional genomics.

1. Introduction

Targeting induced local lesions in genomes (TILLING) is a reverse genetic method that combines random mutagenesis with polymerase chain reaction (PCR)-based screening of gene regions of interest ***(1,2)***. This provides a range of allele types, including missense and knockout mutations, which are potentially useful in a variety of gene function and interaction studies. TILLING is especially suitable for plants, even for those that lack well-developed genetic tools. We have developed a TILLING protocol that achieves high throughput using gel-based screening of heteroduplex PCR products that have been preferentially cleaved at mismatches ***(3)***. Based on this protocol, we developed a service for the *Arabidopsis* community known as the *Arabidopsis* TILLING Project (ATP), where users can request allelic series of ethyl methanesulfonate (EMS)-induced point mutations in target regions of interest ***(4)***.

The general scheme for high-throughput TILLING is outlined in **Fig. 1**. DNAs from mutagenized individuals are first arrayed in a 96-well format. Samples are then pooled to increase throughput. The initial screening procedure consists of (1) setting up and running the PCR on pooled DNAs using IRD700 and IRD800 primers for IR^2 gel analysis (LI-COR) ***(5)***; (2) heat inactivation of polymerase and annealing to create hetereoduplexes; (3) CEL I digestion of heteroduplexes; (4) sample cleanup on G50 spin plates; (5) loading and running the gels; (6) processing and examining the gel images to identify mutations; and (7) repeating steps 1–6 on individuals identified in the pooled screen. This is followed by sequencing the mutant region to ascertain the mutation. The screening strategy of labeling at both ends provides confirmation of each band detected, as its complement is detected independently (**Fig. 2**). In addition, running two channels on the same gel simplifies comparisons and helps to identify artifactual primer–dimer bands, which appear in both channels. Analysis of the first ~1900 mutations that ATP delivered to its users indicates that high-throughput TILLING is a robust method for mutation discovery in pooled samples ***(6)***.

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

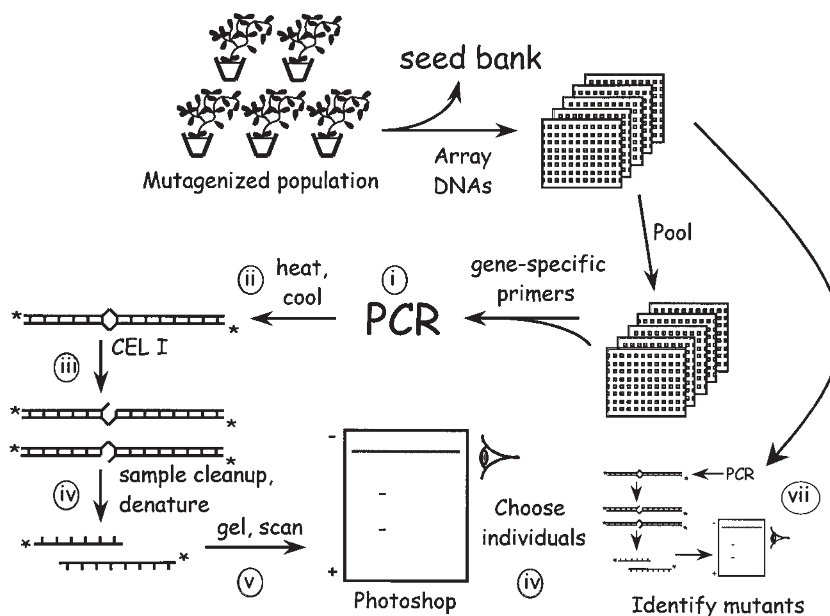


Fig. 1. Outline of high-throughput TILLING. DNA from individuals from a mutagenized population are first arrayed in a 96-well format. Seeds from these individuals are collected and stored for later analysis. DNAs are then pooled up to eightfold to increase screening throughput. High-throughput TILLING can be separated into seven steps as noted here: (*i*) PCR is performed on pooled populations with target-specific primers labeled with fluorescent dyes; (*ii*) an extended incubation at 99°C both kills *Taq* and denatures PCR products, followed by a slow cooling step, in which PCR products reanneal forming heteroduplexes; (*iii*) heteroduplexes are digested with the nuclease CEL I; (*iv*) samples are passed through Sephadex G50 spin plates to remove salts and buffer components that are inhibitory to gel runs and laser detection; (*v*) samples are loaded onto 100-tooth membrane combs, and samples are electrophoresed; (*vi*) gel images are analyzed for mutations in pools; and (*vii*) mutations are tracked down to the individual.

2. Materials

2.1. PCR and Heteroduplex Formation

1. Ex-*Taq* DNA polymerase, Hot Start Version (Takara). Store at –20°C.
2. 10x Ex-*Taq* PCR buffer (Takara) supplied with Ex-*Taq*. Store at –20°C.
3. 2.5 m*M* (each) dNTPs (Takara) supplied with Ex-*Taq*. Store at –20°C.
4. 25 m*M* $MgCl_2$.
5. TE: 10 m*M* Tris-HCl, pH 7.4, 1 m*M* ethylene diamine tetraacetic acid (EDTA).
6. Left primer (melting temperature [T_m] 67–73°C) labeled 5' with IRD700 (MWG) 100 µ*M* in TE. Store at –80°C.
7. Left primer (T_m 67–73°C) unlabeled (MWG) 100 µ*M* in TE. Store at –80°C.
8. Right primer (T_m 67–73°C) labeled 5' with IRD800 (MWG) 100 µ*M* in TE. Store at –80°C.
9. Right primer (T_m 67–73°C) unlabeled (MWG) 100 µ*M* in TE. Store at –80°C.
10. Primer mixture: 3 µL IRD700 left primer, 2 µL unlabeled right primer, 4 µL IRD800 right primer, 1 µL unlabeled right primer. Store at 4°C and discard after 1 wk (*see* **Note 1**).
11. PCR mixture for a 96-well plate: 360 µL water, 57 µL 10X Ex-*Taq* buffer, 68 µL 25 m*M* $MgCl_2$, 92 µL 2.5 m*M* dNTP mixture, 4 µL primer mixture, 6 µL Ex-*Taq*. Mix, adding polymerase last. Use immediately and discard remainder after use.
12. Genomic DNA (*see* **Notes 2** and **3**).

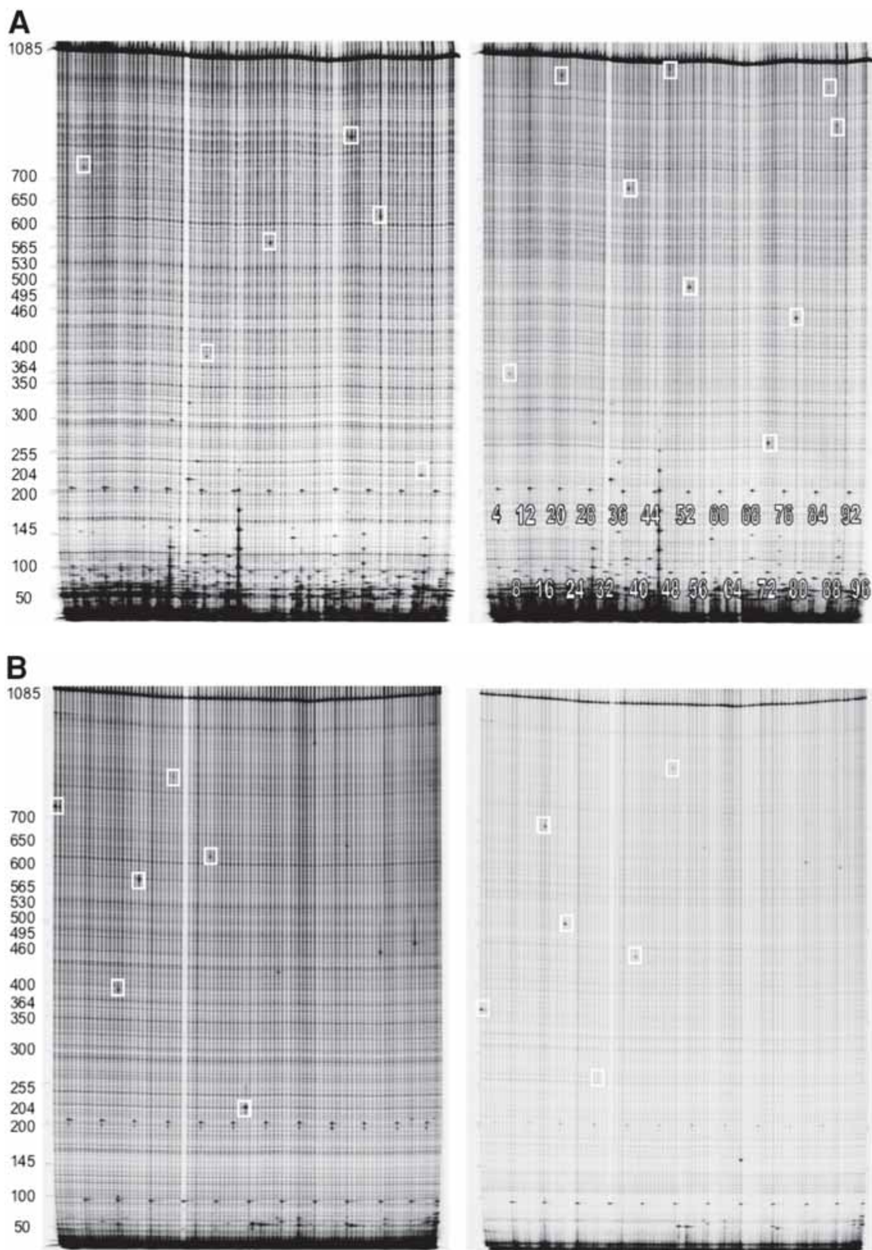


Fig. 2. Complementary strand images of an eightfold pool screening gel **(A)** and an unpooled individual gel **(B)**. **(A)** Nine mutations (white boxes) were identified in this pool screening gel. The IRD700 (left) and IRD800 (right) channels are shown. Six of nine mutations were confirmed by the appearance of bands in both channels that add up to the full-length product of 1085 bp. The other three mutations are of lower confidence as the IRD800 band is of high molecular weight and the complementary fragment in the IRD700 channel cannot be found, presumably because the band is obscured by the primer noise near the bottom of the gel. Note that some fast-migrating bands are ignored—these are detected in both channels and likely represent PCR primer dimers. Lanes were identified using the 95- and 200-bp markers present in every fourth lane, as indicated on the IRD800 image in panel A. Full-length product and marker lengths are indicated to the left of the IRD700 image. **(B)** Unpooled individual gel images. Boxed bands represent mutations identified in the pool screen shown in panel A. Other mutations were identified in other pool screens. For each mutation, all eight individual samples comprising a pool were subjected to TILLING analysis after mixing with an equal amount of wild-type DNA. Mutations were confirmed by a complement band in the IRD800 channel.

2.2. CEL I Digestion

1. CEL I enzyme. Prepared as previously described *(7)* or available commercially in the SURVEYOR mutation detection kit (Transgenomic).
2. 10X CEL I buffer: 5 mL 1 *M* $MgSO_4$, 5 mL 1 *M* 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5, 2.5 mL 2 *M* KCl, 100 µL 10% Triton® X-100, 5 µL 20 mg/mL bovine serum albumin, 37.5 mL water. Store aliquots at –20°C.
3. Cel I reaction mixture for a 96-well plate: 2.4 mL water, 420 µL 10X CEL I buffer, 36 µL CEL I. The amount of CEL I may vary based on the preparation. Mix on ice, use immediately, and discard remainder after use.
4. Stop solution: 0.15 *M* EDTA, pH 8.0.

2.3. Spin Plate Cleanup

1. 96-well membrane plates (Millipore).
2. Sephadex® G50 medium (Amersham Pharmacia Biotech).
3. Sephadex spin plates (*see* **Note 4**).
4. Deionized formamide (Sigma).
5. Formamide load buffer: 38.5 mL deionized formamide, 1.5 mL 0.5 *M* EDTA, pH 8.0, 2.3 mg bromophenol blue. Store 1.8-mL aliquots at –20°C.
6. 95- and 200-bp markers (*see* **Note 5**).

2.4. Electrophoresis and Gel Image Analysis

1. Ammonium persulfate (APS): dissolve APS 10% (w/v) in water. Store at 4°C in small aliquots.
2. Combine 20 mL gel mixture (6.5% acrylamide, 7 *M* urea; LI-COR), 15 µL N,N,N',N'-tetramethyl ethylenediamine (TEMED), and 150 µL APS. Use immediately.
3. 0.8X Tris-borate-EDTA (TBE) buffer: dissolve 89.2 g Tris-base and 45.8 g boric acid in water; add 68 mL 0.25 *M* EDTA, pH 8.0; bring volume to 10 L.
4. IR^2 gel analyzer, 25-cm glass plates, and 25-mm spacers (LI-COR).
5. 100-tooth membrane combs (The Gel Company).
6. Size Standard IRDye™ 700 and IRDye 800 molecular weight markers (LI-COR).
7. Isopropanol.
8. 1% Ficoll®.

3. Methods

3.1. PCR and Heteroduplex Formation

Perform these steps using equipment and consumables that are segregated from PCR products to avoid contamination. Multipipettors are used to reduce the number of pipetting steps.

1. In advance, array 5 µL of pooled genomic DNA per well in 96-well microtiter plates. These plates can be stored for more than 4 wk at –20°C in a sealed container. The final concentration of genomic DNA and the maximal allowable pooling may vary depending on the organism and DNA extraction method used (*see* **Notes 2**, **3**, and **6**). For *Arabidopsis*, the concentration of DNA is 0.015 ng/µL, and samples are pooled eightfold.
2. Add 5 µL freshly made PCR mixture. Immediately after addition, centrifuge briefly, and place in thermal cycler.
3. Run the following thermal cycler program: 95°C for 2 min; loop 1 for 8 cycles (94°C for 20 s, 73°C for 30 s, reduce temperature 1°C per cycle, ramp to 72°C at 0.5°C/s, 72°C for 1 min); loop 2 for 45 cycles (94°C for 20 s, 65°C for 30 s, ramp to 72°C at 0.5°C/s, 72°C for 1 min); 72°C for 5 min; 99°C for 10 min; loop 3 for 70 cycles (70°C for 20 s, reduce temperature 0.3°C per cycle); hold at 8°C. After thermal cycling store samples in the dark at 4°C for use within 1 wk.

3.2. CEL I Digestion

1. Place PCR samples on ice and add 20 µL CEL I reaction mixture to each sample. Mix by pipetting up and down 2 to 5 times. The same tips can be reused if rinsed with water between pipetting steps. Incubate at 45°C for 15 min.

2. Place samples on ice and stop reaction by adding 5 µL 0.15 *M* EDTA. Store samples in the dark at 4°C, for use within 1 wk.

3.3. Spin Plate Cleanup

1. Assemble hydrated spin plate and an empty 96-well catch plate for centrifugation.
2. Spin hydrated plate containing Sephadex G50 for 2 min at 440*g*.
3. Remove the catch plate and insert a sample receptacle plate containing 1.5 µL formamide load dye in each well. Load the CEL I digestion products onto the Sephadex plate within 10 min (*see* **Note 7**).
4. Spin for 2 min at 440*g*.
5. Add markers prior to reducing the volume. We add a 200-bp marker onto row D of a 96-well plate and a 95-bp marker in row H (approx 0.5 ng), thus labeling every fourth lane, which also facilitates lane identification (*see* **Note 5**).
6. Reduce the volume at 90°C to approx 1.5 µL (this takes approx 45 min), leaving formamide–bromophenol blue solution ready for loading. Transfer to ice until ready to load. Samples can be stored in the dark at 4°C for up to 4 wk prior to use.

3.4. Electrophoresis

3.4.1. Preparing Gels

1. Assemble 25-cm glass plates, 25-mm spacers, and casting rails. Plates can be preassembled and stored in a dust-free environment for weeks in advance.
2. Pour gels. For each 25-cm plate assembly, fill a 20-mL syringe with freshly prepared acrylamide mixture, then dispense along the top, avoiding bubbles by rapping continuously on the plate just above the liquid edge. If any bubbles appear, remove them quickly with a thin wire tool after the gel is poured.
3. Leaving a little excess acrylamide at the well, insert the top spacer all the way into the glass, making sure spacer is centered horizontally.
4. Insert the Plexiglas pressure plate between the glass plate and casting rails. Tighten the top screws as soon the spacer is inserted, slightly compressing the rubber pads on the pressure plate.
5. Add acrylamide to the top glass edge where the comb is inserted and on the edges to ensure that polymerization is not inhibited within the gel.
6. Let the gel set at least 90 min before placing in gel box. Gels can be stored wrapped in plastic wrap at 4°C for up to 24 h prior to use (*see* **Note 8**).
7. Prior to placing the gel in the gel box, wash the plates with distilled water, removing the comb spacer and excess polyacrylamide at the top edge. Dry the plates and wipe with isopropanol, making sure that the back plate is spotless where the laser shines through.
8. Insert the top buffer reservoir between the glass plate and the casting rails. If this is tricky, moisten the gasket with buffer and remove one casting rail in order to fit the top reservoir. Fill the lower buffer reservoir to the fill line with 0.8X TBE (approx 500 mL), and insert the gel.
9. Tighten the screws to seal the upper reservoir and fill with buffer. Rinse the slot vigorously using a large syringe without a needle (*see* **Note 9**).

3.4.2. Loading Samples onto Membrane Combs

1. Load samples onto a 100-tooth membrane comb, such that position A1 on the plate represents lane 4, position B1 equals lane 5, A2 equals sample 12, etc. (*see* **Note 10**).
2. Use a pipettor to add 0.25 to 0.5 µL IRD700 plus 800 molecular weight markers to lanes 1, 3, and 100.
3. Spot the IRD700 ladder alone to tooth 2. This asymmetry assures that if the comb is inserted inadvertently in reverse, then the A1 lane is always next to the doubled markers and the H12 lane next to the single marker.

3.4.3. Electrophoresis

1. Access the user controls (LI-COR) using an internet browser.
2. Provide a gel run name; hit "Create Run" (*see* **Notes 11** and **12**).

3. Start the prerun (20 min), waiting for the all-ready signal from the scanner before proceeding. The prerun can be started while samples are being applied to the comb.
4. After the prerun, clean the slot out with a syringe and drain the top buffer reservoir until the level is below the glass edge. Wick out the remaining buffer, first with a paper towel and then with a 6-in.-wide strip of Whatman 1 paper, sliding it into the slot left by the spacer.
5. Using a 1-mL pipettor, fill the slot with 1% Ficoll, leaving just a thin bead approx 1 mm above the slot.
6. Hold the comb at a 45°C vertical angle with lane 1 on the left, aim for the slot and insert rapidly by pushing gently (*see* **Note 13**). Push the comb down until it just touches the gel surface.
7. Gently fill the reservoir to the fill line, insert the electrode cover, close the top, and then click on "Collect image." From the time the comb touches the slot until the time the current is applied should be no more than about 20 s or so to prevent diffusion.
8. After 10 min, open the lid (be sure that you hear the "pling" signal and that the high-voltage light goes off), remove the comb, and gently rinse the slot with buffer (*see* **Note 14**). Replace the top electrode and close the lid. You should hear the "pling" and see the laser and high-voltage lights go on. The gel can be monitored from a browser (*see* **Fig. 2**).

3.5. Gel Image Analysis

Gel images are saved on the LI-COR as tagged image format file (TIFF) images. For visual analysis, the quality of a default JPEG image is sufficient. The program "grab" transfers these images to another server via file transfer protocol (ftp) and converts them to JPEG format (*see* **Note 15**). Once grab is done, the layered image can be created.

1. Use Fetch (for a Macintosh®) or ftp (from a Windows® PC) to place the two JPEG files onto the desktop or into a local directory.
2. Start up Adobe Photoshop®, then File > Open the 700 and 800 channel files (**Fig. 2**). Move the 800 channel image to one side, then click on Image > Adjust > Levels, and move the leftmost slider arrowhead towards the right until a midtone image is obtained for the 800 channel image (usually when the arrowhead is just at the point at which the density begins to increase; be sure that Preview is active).
3. Click on OK when the image is optimized. Click on the 700 channel image and repeat the level adjustment procedure. You may want to enlarge the images for setting the levels, which can be done by holding down the Control (or for a Macintosh, the Command) key and press +.
4. Go to Select > All and Edit > Copy, then click on the 800 channel image and Edit > Paste. The 700 image will be precisely superimposed over the 800 image. Close the window to exit from the 700 channel file. You will need the Layers palette, which can be opened by clicking on Window > Show Layers. If your version of Photoshop does not show rulers on the top and left, click on File > Preferences > Units and Rulers (Edit > Preferences > Units and Rulers on the Mac) and choose percent. If the rulers are not visible, click on View > Show Rulers.
5. Click on Image > Size and change to 2500 (width) × 1750 (height) pixels (uncheck "constrained proportions"), then hold down the Control (or for Macintosh, the Command) key and press + repeatedly until the image is at 100%. You can tell if it is, because the ruler will show numbers at 5-U intervals (such as 45…50). Click on View > Fit to screen and adjust the image dimensions as desired (with Image > Size) if needed.
6. Using the Rectangular Marquee tool, draw a rectangle that encompasses the image from the edge of the electrophoresis front (at bottom) to the full-length product (dark band at top), and from the outside edge of lane 1 to the outside edge of lane 100. Click Image > Crop.
7. In the Layers box, click repeatedly on the eye icon to switch back and forth between the superimposed images. Look for bands present on one image but absent on the other. Ignore bands that are present on both images, as we expect that these are primer dimers resulting from mispriming, not from CEL I cleavage. An exception is a singular position midway in the fragment, where the 700-labeled band coincides in size with the 800-labeled band, and their sum equals the size of the full-length product. In each case, where a band of a specific molecular weight is detected in one channel but absent in the other, look for the complement band in the

other channel. Mark each positive lane by double-clicking in the vertical ruler area and pulling a vertical line over to the right of the mutation. Because size is nearly directly proportional to vertical distance, it should be relatively easy to anticipate where to look: for instance, a band one-third of the distance from the front at the bottom of the gel should be paired with a band in the other channel that is one-third of the distance from the full-sized band.

8. Run the program "squint" (*see* **Note 15**). You will be asked a series of questions about sizes and locations of bands on the image that you are viewing in Photoshop. To ascertain the size corresponding to a band, place the horizontal crosshair over the band and look at the ruler on the side: the guideline will indicate the distance migrated as a percentage of the distance to the full-sized band. Be sure to compensate for any mobility differences across the gel, such as "frowning" or "smiling," using the background bands as guides. Enter this distance into squint.
9. To determine the lane location, find a favorable vertical position (usually the full-size band) and count to determine the lane position using the markers. When migration distances and lane numbers have been entered, squint returns approximate molecular weights and their sum, which is compared to the molecular weight of the full-length PCR product. If these numbers are nearly equal, then this is almost certainly a mutation. Squint also returns a plate position of the pool given the lane position.

3.6. Analyzing Mutant Individuals

Once squint entries are entered for eightfold pools, individuals used to make the pools can be efficiently screened to track down each mutation. It is most efficient to screen for individuals once 12 mutations have been entered (if pooling eightfold), as the screen can then proceed with a full 96-well plate of samples.

1. Run the program "pick" (*see* **Note 15**). The pick program takes squint output from multiple pool plates and returns a list of rows from plates containing individual DNAs arrayed in an 8×8 grid on a 96-well plate. All members of a single pool are present in one row on this plate. These rows will be deposited into successive columns numbered 1 through 12 on the pick list. This list is the template for the new screening plate to be made containing individuals from the identified mutant pools. Each well in a single column of this plate will contain one individual in the eightfold pool.
2. From the pick output, take out the individual plates for the first 12 mutations from a single set of oligonucleotides. Rotate the individual plates such that position A1 is in the upper right corner. Using an eight-channel pipettor, transfer 10 µL from each corresponding row of the individual plates to a column in a new plate.
3. Once this plate has been created from 12 rows of individual plates, transfer 5 µL to a new plate containing 5 µL of wild-type DNA (*see* **Note 16**). The original plate can be stored at –20°C for later amplification and sequencing. TILLING is performed as described in **Subheadings 3.1.–3.5.**, and results should resemble that shown in **Fig. 2B**.

4. Notes

1. Special care should be taken when using primers labeled with IRD700 and IRD800. When possible, avoid prolonged exposure of labeled primers and PCR products to fluorescent lights. Primer stocks should be diluted to no more than 100 µ*M*, aliquoted, and stored at –80°C. Primer mixtures are used for no more than about 1 wk. Over time, we see the amount of labeled PCR product decrease dramatically when using old primer mixtures or primer stocks that were stored at 4°C or that have undergone repeated freeze–thaw cycles. IRD-tagged oligonucleotides do not prime as well as untagged oligonucleotides, presumably because of the hydrophobic group at the 5' end. To obtain consistently high PCR product yield, we add a mixture of both tagged and untagged primers. Using the CODDLE program for primer design (www.proweb.org/input/), which runs the Whitehead Primer3 program, we have found that >90% of our primer pairs with Tm approx 70°C are successful in amplifying approx 1-kb fragments from *Arabidopsis* DNA samples and providing adequate TILLING results. Because IRD800 gives a weaker signal than IRD700, it fails more frequently, and mutations might be overlooked when there is only a single channel for detection. This is especially a problem for

mutations that are distant from the tagged oligonucleotide priming site, because the large molecular weight strand produced by CEL I digestion has reduced signal and band resolution.

2. The quality and quantification of genomic DNA starting material is crucial. DNA samples for the *Arabidopsis* TILLING Project (http://tilling.fhcrc.org:9366/) are prepared using the FastDNA® kit (Bio101), and samples are electrophoresed on agarose gels to equalize concentrations before arraying and pooling. A single sharp band of high molecular weight indicates high-quality DNA.
3. DNA pooling provides higher throughput by allowing less machine time per sample for mutation discovery. For pooling, DNA quantification between samples is very important. As eightfold pooling approaches the limit of robust discovery for a heterozygous mutation (one-sixteenth) in a high-throughput production setting, any sample whose concentration is lower than others in the pool may escape discovery. Before proceeding with higher pooling and sample preparation from thousands of individuals, we suggest trying several different levels of pooling on a small subset of samples. This will also determine the robustness of amplification.
4. Use a Sephadex loading device (MultiScreen 45-µL column loader; Millipore; cat. no. MACL 096 45) to fill all wells of a 96-well membrane plate with approx 0.03 g of G50. Hydrate with 300 µL water. Allow the plate to stand for 1 h at room temperature. Plates may be stored at 4°C in a sealed container to prevent evaporation for up to 1 wk.
5. Perform PCR (**Subheading 3.1.**) and spin plate cleanup (**Subheading 3.3.**) with primers designed to yield a 95- or 200-bp fragment.
6. Before proceeding with fluorescently tagged primers in PCR, it might be desirable to perform trial reactions with unlabeled primers. A yield of 7 to 10 ng/µL of final PCR product is required for robust and consistent identification of mutations on gel analyzers when pooling samples eightfold. Primer trials are discontinued when consistent amplification is achieved, and ATP's primer failure rate of $<10\%$ is tolerated to maximize high throughput.
7. It is important to deposit sample directly over the center of the Sephadex column, thus avoiding any disturbance of the column, such as touching the column with pipet tips.
8. Gel plates may be prepoured and stored for up to several days at 4°C covered in plastic wrap. Each 25-cm plate requires approx 20 mL. To assemble new plates, clean the plates with dilute liquid detergent (i.e., 2% Tween-20) and a soft scrub brush. Rinse plates with distilled water, wipe down with 0.2 *N* HCl, rinse with distilled water again, and wipe with isopropyl alcohol. Spacers are cleaned by wiping with a wet tissue. Assemble the pieces with the screws backed off and align the pieces by standing the assembly vertically while tightening the screws. Tighten just beyond where you begin to feel resistance (overtightening will crack plates). Place on a horizontal support. When handling acrylamide or polyacrylamide gels, always wear gloves, as unpolymerized acrylamide is a nerve toxin.
9. It is important that the slot be clean, as any loose acrylamide will inhibit insertion of the comb. For increased visibility of the loading well, insert a background card wedged behind the back plate such that it is centered between the ears of the front plate. Two vertical marks can be made on the card that are a comb-width apart and will provide a guide for later inserting the comb precisely in the middle of the gel, which is necessary so that all of the lanes are scanned.
10. A variety of options are available for loading samples. Samples can be loaded onto the membrane comb with the aid of robotics using the comb load robot (MWG). Alternatively, samples can be loaded manually directly onto the comb using a pipettor. We have found that a sample vol of 0.5 µL is optimal for our assays. Manual loading can be aided by using a membrane loading tray (The Gel Company) and an adjustable-width multichannel pipetor (Matrix Technologies, Lowell, MA). Combs can be preloaded and stored at 4°C, but as samples are more stable in formamide load buffer, it is suggested that combs be used within 2 h of loading. It is not advised to use loaded combs that have been stored for more than 2 d.
11. Gels can be run twice, even after a day. The prerun is necessary if the plate has been moved, because it is needed for focusing the lasers. After runs are complete, and a new gel is ready, remove the old gel, pour out the buffer from both reservoirs, and clean plates as described in **Note 8**.

12. For a 1-kb fragment, enter the following settings: collect time 3 h 45 min run at 1500 V, 40 ma, 40 W, 50°C. (Be sure that the current is off before touching a buffer chamber.) Other parameters are pixel size, 16; bin size, 8; and motor speed, 3.
13. Practice this step with used combs. Take special care to avoid bending any teeth. If a tooth sticks to the plate, it may not be possible to save it without moving the comb around and thus losing sample from all teeth.
14. Occasionally LI-COR gel images look fuzzy with diffuse and ill-defined lanes, for any of the following reasons:
    a. Urea is not thoroughly rinsed from the well before adding Ficoll.
    b. The comb is pushed into acrylamide when loaded.
    c. The comb moves laterally once inserted into the well.
    d. The upper buffer chamber is filled too rapidly after the comb is inserted.
    e. Ficoll is not rinsed out of well after the comb is removed.
    f. Ficoll is rinsed too thoroughly from the well after the comb is removed.
15. Current versions of the UNIX® programs grab, squint, and pick are available upon request. The TILLING Web site (http://tilling.fhcrc.org:9366/) provides links to CODDLE, PARSESNP, and SIFT, which facilitate fragment choice, primer design, and mutation analysis.
16. In order to identify individual mutants that are homozygous, samples must be doped with wild-type DNA to generate heteroduplexes that are the substrate for CEL I.

Acknowledgments

The methods described here were developed with support from the National Science Foundation Plant Genome Research Program. We thank Faith Hassinger for the images shown in **Fig. 2**.

References

1. McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000) Targeted screening for induced mutations. *Nat. Biotechnol.* **18,** 455–457.
2. McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol.* **123,** 439–442.
3. Colbert, T., Till, B. J., Tompa, R., et al. (2001) High-throughput screening for induced point mutations. *Plant Physiol.* **126,** 480–484.
4. Till, B. J., Reynolds, S. H., Greene, E. A., et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res.* **13,** 524–530.
5. Middendorf, L. R., Bruce, J. C., Bruce, R. C., et al. (1992) Continuous, on-line DNA sequencing using a versatile infrared laser scanner/electrophoresis apparatus. *Electrophoresis* **13,** 487–494.
6. Greene, E. A., Codomo, C. A., Taylor, N. E., et al. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* **164,** 731–740.
7. Till, B. J., Colbert, T., Tompa, R., et al. (2003) High-throughput TILLING for functional genomics. *Methods Mol. Biol.* **236,** 205–220.

12

Genotyping and Mapping With High-Density Oligonucleotide Arrays

Justin Borevitz

Summary

This chapter discusses the specific details of how oligonucleotide arrays originally designed for expression analysis can be used for high-density genotyping to reveal tens of thousands of single-feature polymorphisms (SFPs). Cluster of SFPs reveal potential natural deletion polymorphisms. SFPs can be readily used for mapping mutations using bulk segregant analysis. An example is given using the Columbia and Landsberg *erecta* strains of *Arabidopsis thaliana*.

Key Words: Single-feature polymorphisms (SFPs); high-density oligonucleotide arrays; bulk segregant analysis.

1. Introduction

The completion of the *Arabidopsis* genome sequence has facilitated an increase in functional characterization of genes via forward genetics. Several studies of genome-wide polymorphism have provided markers to rapidly map and chromosome walk to induced mutations. Genome-wide markers also make it possible to investigate the phenotypic differences between wild accessions, using natural variation for gene discovery (reviewed in refs. ***1*** and ***2***).

Traditionally mutations are mapped in *Arabidopsis* by testing a handful of molecular markers for cosegregation with the mutant phenotype. This involves individual genotyping of >20 mutant plants at usually >10 individual markers. When and if a particular marker is identified as linked to the phenotype, additional markers are typed on additional plants to narrow the interval. This process is repeated each time, scaling up the number of plants to identify new recombination events and to identify and test new molecular markers to delimit these events. This process, well known as "chromosome walking," is straightforward but tedious especially when large numbers of mutants need to be mapped. The most common markers for this purpose are traditional gel-based cleaved amplified-length polymorphisms (CAPS) ***(3)*** or simple sequence length polymorphisms (SSLPs) ***(4)***.

High-density oligonucleotide arrays, originally designed to measure gene expression, can be used for large-scale polymorphism discovery and genotyping ***(5)*** by extending a method originally developed in yeast ***(6)***. Single-feature polymorphisms (SFPs) are revealed as significant hybridization intensity differences on 25mer oligonucleotide features. The exact sequence change within or near the 25mer is unknown for SFPs, as is the case for restriction fragment length polymorphisms (RFLPs) and amplified fragment length polymorphisms (AFLPs). In **Subheading 3.**, I specifically address the steps required for identification and genotyping of SFPs in *Arabidopsis*. I then discuss how to cluster SFPs to reveal modest-sized deletions between natural accessions or in mutagenized lines. Finally I show how segregating populations, such as F2s, can be used to rough map induced mutations or modest effect quantitative trait loci (QTLs), with bulk segregant analysis or eXtreme Array Mapping.

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

2. Materials

1. Three or more plants of the reference Columbia (Col) genotype (*see* **Note 1**).
2. Three or more plants of another *Arabidopsis* accession used for mapping, Landsberg *erecta* (L*er*) in this case (*see* **Note 1**).
3. Fifty to 100 F2 plants with a mutant phenotype (may be dominant/recessive or additive) (*see* **Note 1**).
4. Fifty to 100 F2 plants with a wild-type phenotype (*see* **Note 1**).
5. Qiagen DNA Easy Plant prep kit.
6. Invitrogen Bioprime labeling kit (biotin-dCTP).
7. *Arabidopsis* ATH1 GeneChip arrays from Affymetrix.
8. Standard reagents for Affymetrix hybridization, including oligo B2, and washing protocols (*see* **Note 2**).
9. Software R (www.r-project.org) free statistical software for all platforms—Mac, PC, Linux, etc.
10. Bioconductor "affy" package (www.bioconductor.org) free array analysis packages for R 512MB RAM.

3. Methods

3.1. Labeling

1. Extract DNA from a single leaf of each of 3 or more Col and L*er* plants using Qiagen DNA Easy Plant prep kit.
2. Resuspend DNA in 100 µL water.
3. Take 1 leaf (or equal amount of tissue) from each of 50 to 100 mutant or wild-type F2 plants and freeze together in liquid nitrogen.
4. Grind tissue together and use ~100 mg powder in the Qiagen DNA Easy Plant prep kit.
5. Resuspend DNA in 100 µL water.
6. To 30 µL (~300 ng) of plant DNAs add 60 µL 2.5X random primers (Bioprime kit) and 42 µL of water.
7. Denature DNAs at >95°C for 5 to 10 min.
8. Cool on ice.
9. To each denatured DNA add 15 µL 10X dNTPs mix with Biotin dCTP (Bioprime kit) and complete to 150 µL with 3 µL Klenow polymerase (Bioprime kit).
10. Incubate overnight at room temperature.
11. Add 15 µL 3 *M* NaOAc and 400 µL cold 100% EtOH.
12. Spin, remove supernatant, and wash with 500 µL cold 75% EtOH.
13. Dry DNA pellets and resuspend in 100 µL water.
14. Use 5 µL to check yield and quality on a gel (**Fig. 1**) (*see* **Note 3**).

3.2. Hybridization

Labeled DNAs are treated the same as labeled cRNA in standard hybridization protocols for gene expression (*see* **Note 2**).

1. To 95 µL of labeling reaction add 3.3 µL control oligo B2 (3 n*M*), 2 µL herring sperm DNA (10 mg/mL), 2 µL acetylated bovine serum albumin (BSA) (50 mg/mL), 110 µL 2X hybridization buffer.
2. Denature probes at >95ºC for 5 min.
3. Cool on ice.
4. Prehybridize arrays with 1X hybridization buffer at 45ºC for 10 min.
5. Discard the prehybridization solution.
6. Hybridize with 200 µL of denatured hybridization cocktail overnight.
7. Wash as in standard eukaryotic hybridizations and stain with double antibody stain (*see* **Note 2**).

3.3. Single-Feature Polymorphism Identification

3.3.1. Required Files

The arrays are scanned and the .CEL files are saved in ASCII text format, which may require a free translation tool available from Affymetrix.com. The .CEL files contain the coordinate

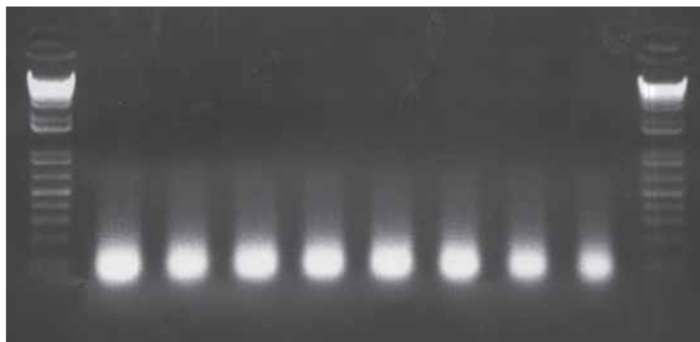


Fig. 1. 5 µL of 100 µL was loaded from each of 8 bioprime random labeling reactions. Lanes 1–3 are independent Col reactions, lanes 4–6 are independent L*er* reactions, lane 7 is from a pool of wild-type F2 plants, and lane 8 is from a pool of mutant F2 plants. Photo courtesy Todd Michael.

and intensities for each 25mer feature on the array. For the current *Arabidopsis* expression array named ATH1, this corresponds to 712 × 712 features. The .CEL files should be put together in a common folder. Sample Col and L*er* .CEL files are available as "ColLerCEL.zip" (downloaded from http://naturalvariation.org/methods/). Other required files for SFP identification are "readcel.R," which contains the R function to read in .CEL files. "ath1V5.RData" is loaded into R and contains a data.frame that describes the Perfect Match features on the ATH1 *Arabidopsis* array. This is the result of a remapping of 25mer sequences to the *Arabidopsis* Version 5 annotation and selection of features that have one unique location on a chromosome. Features with a second partial match of 18 bp or greater (megablast cutoff e = 0.01, word size = 12) are removed. ATH1 arrays also have a horizontal strip of saturation 16 features wide. These data are also removed. Three scripts are provided on our web site for download (http://naturalvariation.org/methods/). Results of "SFP.R" are discussed in the next section. "deletions.R" and "Map.R" are discussed in the following two sections.

3.3.2. Quality Control

The raw .CEL file intensity data are read into R using the read.cel() function. The arguments to this function allow settings for spatial correction and whether correction images should be saved. The spatial correction is performed via a sliding square window of specified size, which can reveal local array artifacts ***(5)***. An example is shown in **Fig. 2**. This background is then subtracted from the raw data on the log scale. Other possibilities for background correction are available in the bioconductor "affy" package ***(7–9)***. Next, the arrays are quantile normalized so that they all have the same distribution ***(7)***. Finally an important quality control is to examine pair plots and images of the correlation matrices to confirm that the replicates are more closely correlated to each other than they are to arrays of different genotypes (**Fig. 3**).

3.4. Calculating SAM d-Statistics and False Discover Rates

A matrix is created of spatially corrected normalized Perfect Match (PM) oligo data, which contain 8 columns (1 for each array) and 202,882 rows (one for each feature). The means and sum of squared deviations are calculated for each feature. The SAM *d*-statistic is similar to a *t*-statistic except that a constant (s0) is added to the denominator ***(10)***. This guards against large *d*-statistics for small mean differences between genotypes when the error is also small. This is especially helpful for small sample sizes when the individual feature error is artificially small due to sampling. The *d*-statistic can be thought of as a hybrid between an absolute fold change threshold and a strict relative change *t*-statistic threshold.

To estimate the false discovery rate (FDR), one needs to determine the null distribution of the ordered *d*-statistics. Permutations allow one to use the data within an experiment to esti-

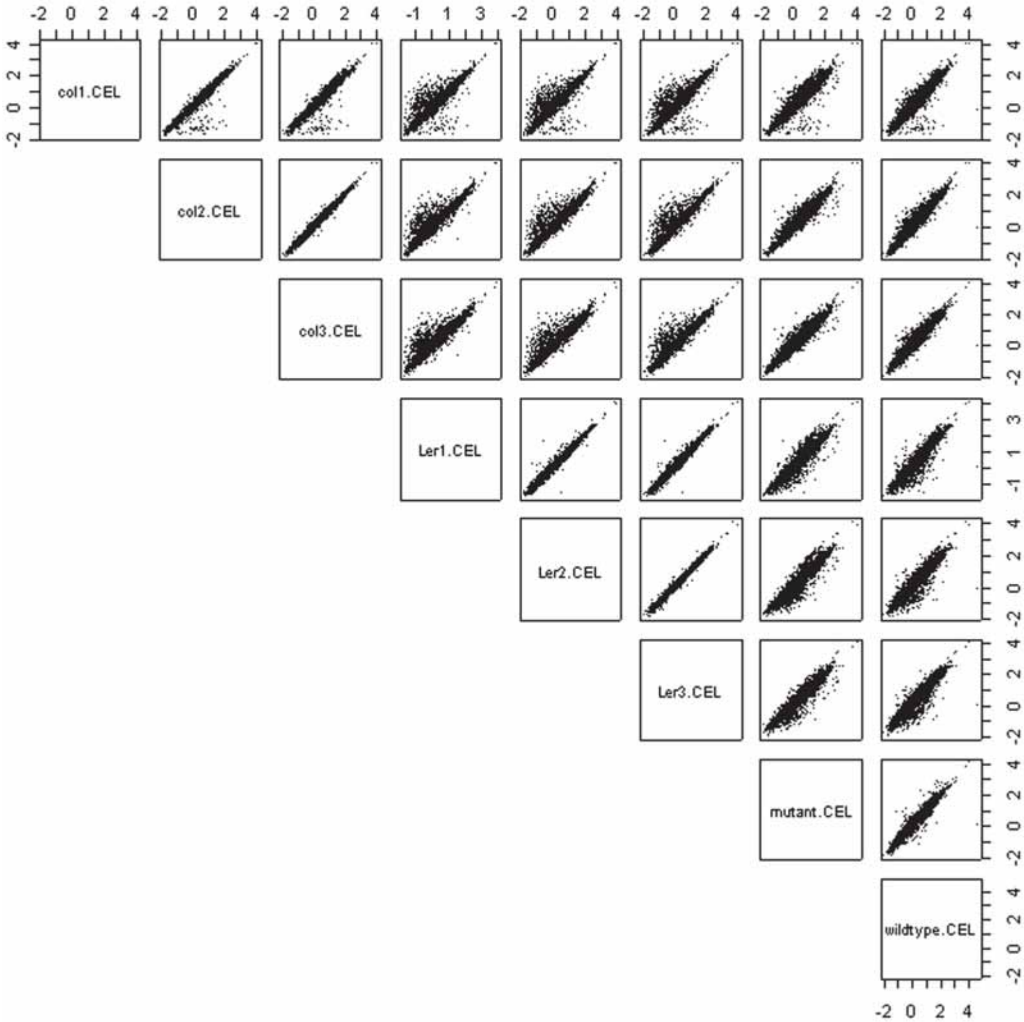


Fig. 3. Pairs plots of normalized arrays. Six thousand randomly chosen features are plotted against each other for all pairwise array combinations. The units for the *x* and *y* axes are the background-corrected log intensity values. Replicates are more closely related to each other as points fall along a diagonal line. SFPs fall above the diagonal line as Col features often hybridize with greater intensity.

mate the null distribution without assuming any underlying distribution. To do this, labels on the 3 Col and 3 L*er* arrays are shuffled as if the experiment was performed without keeping track of the genotypes. In each permutation, *d*-statistics are calculated for all features. For small numbers of replicates, all permutations are enumerated including the original labeling and its reverse. With 6 arrays, 3 for each genotype, there are 20 permutations. A minimum FDR is thus 1/20, as one permutation is from the original data. With 8 arrays, 4 for each genotype, there are 70 permutations, which allows a better estimate of the FDR. The *d*-statistics are sorted for each permutation and averaged to estimate the null distribution. The assumption here is that the ordered d-statistics for each feature come from the same ordered distribution. The FDR is then calculated at different thresholds by counting the number of ordered *d*-statistics greater than a threshold in both the original data and in each permuted data set. The average number of

d-statistics that exceed a threshold among the permuted data sets is an estimate of the number of false positives. The FDR is just this number divided by the number of SFPs identified (*d*-statistics exceeding a threshold) in the original data. At more stringent thresholds, fewer SFPs are identified; however, the proportion of falsely discovered SFPs decreases. This is the standard trade-off between specificity and sensitivity. As SFPs are truly quantitative, the threshold can be adjusted depending on the need for specificity or sensitivity. A simple way to improve both is to increase the number of replicates. The FDR accounts for the large number of related tests and because it uses the actual data to estimate the null distribution, the FDR also accounts for the specifics of the experiment. Thus the FDR is an appropriate error measure that considers the proportion of the total called SFP that may be miscalled rather than a pointwise estimate of a false positive rate across the total of hundreds of thousands of features ***(11)***.

Tens of thousands of SFPs can be identified at $\leq 5\%$ FDR. These can be confirmed by sequence analysis if necessary. We find that SNPs more often fall within the central bases of the 25mer; however, SFPs can also be detected when SNPs fall at the edge of the 25mer ***(5)***. This is consistent with position-dependent nearest neighbor stacking energies seen for RNA hybridization ***(12)***.

3.5. Detecting Potential Deletions via Clustering

The principle behind identifying potential deletions is that multiple adjacent SFPs may not be independent. They may in fact be the result of a single deletion polymorphism that simultaneously disrupts binding to many features. The power to detect deletions depends on the quality of the signal (number replicates), the density of array features, and the size of the mutation. Thus large deletions spanning an entire gene or more may be detected by dozens of adjacent SFPs (**Fig. 4**). Smaller deletions covering only a few features that may extend into uninterrogated intergenic regions may not be identified. Duplications are also possible to detect; however, it is unknown from hybridization data alone where these duplications map.

Natural deletions, such as transposon excisions, may be small and frequent, while induced fast neutron deletions may be large and less common. Our example will explore natural deletion polymorphisms between the Col and L*er* ecotypes; however, we have been successful in localizing many induced deletions. The same code with minor adjustments (noted within) can be used. We start by clustering the d-statistics using a fast linear clustering program, lcluster. This will join adjacent features that have similar *d*-statistics until all features are joined. For induced deletions two arrays maybe used. In this case the intensity difference between the arrays is clustered. The first clusters to form will be investigated as potential deletions by satisfying several criteria. They must have a mean d-statistic above a threshold, have a size of greater then 4 features and 100 bp, and be supported by 1 feature every 350 bp. These settings are suitable to probe many types of deletions and avoid too many false positives. They can certainly be adjusted. We find hundreds of potential deletions between two accessions. An important test is to randomly shuffle the positions of the features along the chromosome and count how many false potential deletions are called. This is rarely greater than 20. The generated output file “deletions.csv” shows the genes that are detected as potentially deleted, the interval of the deletions, and the annotation of the genes in the interval. In addition a graphical .pdf file is created so one can visualize particular deletions or probe certain regions. Examples are found on the web site http://naturalvariation.org/methods.

3.6. Array Mapping

Bulk segregant analysis is a simple method to identify a region linked to a particular trait by analyzing only two DNA samples ***(13,14)***. To map mutation(s) using array genotyping a direct comparison is made between mutant and wild-type pools of F2 seedlings. In a segregating population that has been split according to phenotype, the chromosome linked to the gene causing the phenotype will be fixed for alternative alleles between the two pools. Unlinked chromosomes will be at approximately equal frequency in each pool. Therefore, we can identify the

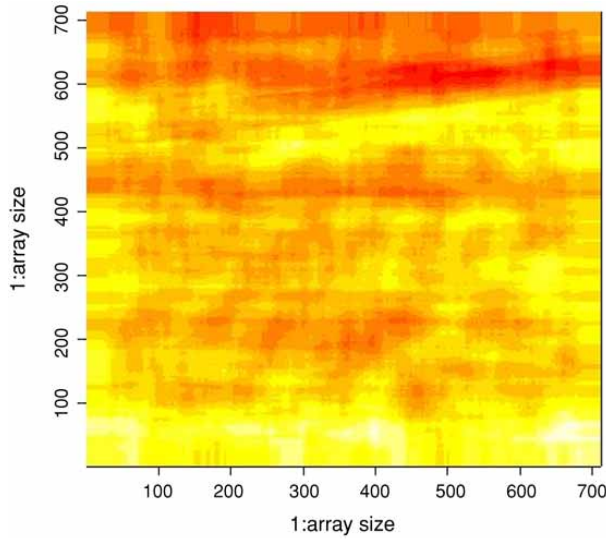


Fig. 2. Spatial correction of ATH1 array "col1DNA.CEL". The mean of the log intensities for features within a square sliding window of size 51 is shown in false color. The window means are calculated by moving vertically then horizontally in 1-feature steps. The resulting background matrix of window means is then subtracted from the original data.

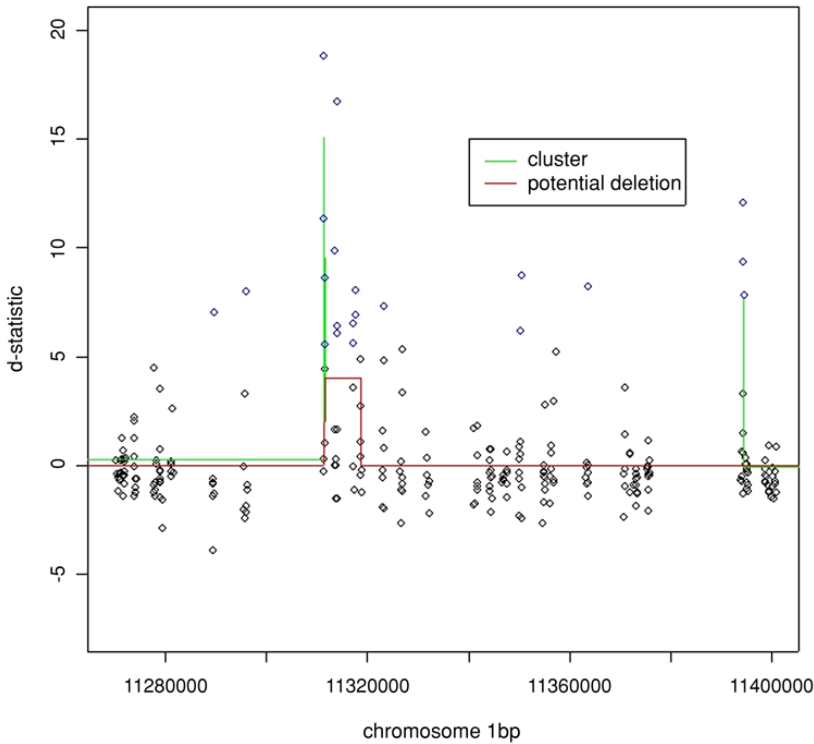


Fig. 4. Example of a natural deletion in L*er* removing At1g31580, At1g31600, At1g31620, At1g31630. Array features are shown as black circles, SFPs are shown in blue. Clusters are shown in green and potential deletions in red.

region of interest as the region of difference in allele frequency between the two pools. The accuracy of this approach is a function of the type of population, population size, and trait heritability. Although the quality of individual SFP genotypes is limited, their abundance can be used to make an accurate prediction about the allele frequency in a chromosome region; thus the error contributed by SFP genotyping to mapping is minimal.

Included in the supplemental data are two .CEL files from mutant and wild-type pools of 100 seedlings each. They have been read in, spatially corrected, and normalized along with the parental arrays. The intensity data at the SFPs have been saved. The first step of bulk segregant mapping is to scale each SFP so that the difference between Col and L*er* genotypes is 1 and the mean is 0. With each SFP on the same scale, homozygous Col genotypes = 0.5, heterozygous = 0, and L*er* = –0.5. The allele frequency of any single SFP is noisy; however, an accurate estimate can be made by considering multiple SFPs along a region. We are interested in deviations from equal proportions of Col and L*er* genotypes in a pool, an allele frequency of 0 on our scale. Importantly, the deviation in allele frequency should be in opposite directions between the mutant and wild-type pools. To identity these regions the genotype of the mutant pool is subtracted from the wild-type pool. The resulting vector of allele frequency differences for each chromosome is then smoothed by loess. For additive (or semidominant) mutations, where both homozygous classes can be separated, the allele frequency difference approaches 1 at the location of the mutation. Elsewhere it bounces around 0. If the mutation is in the Col background the mutant–wild type allele frequency difference will have a positive sign [0.5 – (–0.5) = 1]. If the mutation is in the L*er* background then the allele frequency difference will approach –1 [–0.5 –0.5 = –1]. Less separation is possible for dominant (or recessive) mutations where a single homozygous class cannot be distinguished from the heterozygous. In the case of a recessive mutation in L*er* (**Fig. 5**) the mutant pool will have a –0.5 allele frequency at the location of the mutation. The wild-type pool will be made up of one-third homozygous Col plants, with genotype 0.5, and two-thirds heterozygous plants with genotype of 0. The net allele frequency in the wild-type pool will be $\frac{1}{3} \times 0.5 + \frac{2}{3} \times 0 = 0.17$. The allele frequency difference will approach –0.5 – 0.17 = –0.67 at the site of the L*er* mutation.

Unlinked chromosomes may show a deviation from 0 due to random assortment of chromosomes. This variance decreases with the number of plants included in the pools. Simulation studies have been used to set appropriate thresholds for different genetic models and pool sizes *(**15**)*. A threshold of ±0.17 is appropriate for 60 plants per pool. Allele frequency differences on unlinked chromosomes fall within this range 95% of the time. The chromosome containing the mutation was always detected above this threshold in 1000 simulations. To estimate the precision of mapping via this approach we looked at the range of predicted locations in the simulations. Ninety-five percent of the predicted locations fell within a 7-cM interval while 80% fell within a 4-cM interval.

3.7. Extensions of Array Mapping

The concept of bulk segregant mapping using pools of plants that have been selected for certain phenotypes can easily be extended to non-Mendelian or quantitative traits. In this approach, which we call eXtreme Array Mapping (XAM), plants are chosen from the extreme tails of a distribution. For example, an F2 population may segregate with a continuous distribution for flowering time owing to the modest effects of several genes. Here the tails of the distribution are selected to be contrasted in an early-flowering pool versus a late-flowering pool. Late-flowering alleles will be enriched in the late pool and early alleles at the same loci will be enriched in the early pool. Again, unlinked loci will be represented in approximately equal allele frequencies of parental alleles in each pool. We recommend that a minimum of 50 plants be selected in each pool. Small effect genes will require larger starting populations of, say, 500 F2 plants so that more extreme tails can be selected, which are more likely to contain genetically early and late plants relative to environmental effects. Multiple loci can be simultaneously mapped using our approach. This can reveal multiple modest-effect QTL or double

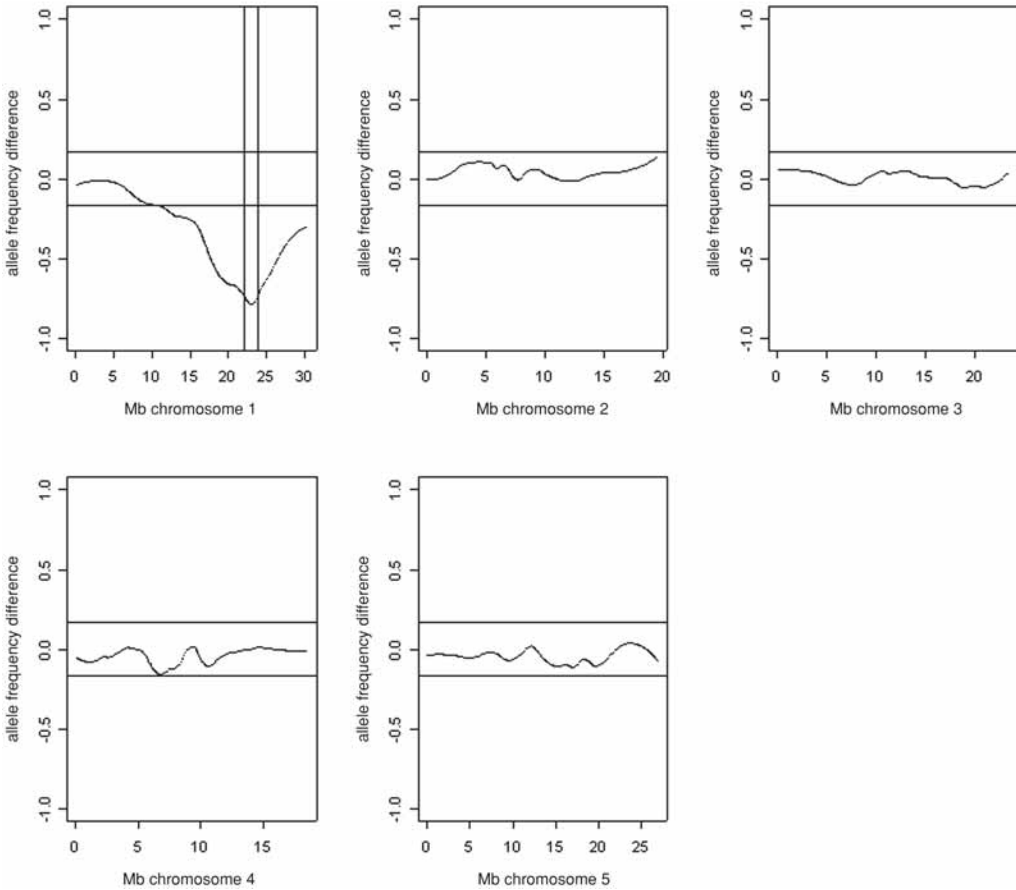


Fig. 5. Mapping of an L*er* mutation to chromosome 1 using bulk segregant analysis and array genotyping. Vertical bars delimit the likely position of the mutation here on chromosome 1. Negative values indicate that the mutation is in L*er*. Horizontal bars represent thresholds for detection. (Data courtesy Sarah Liljegren.)

mutants such as suppressor or enhancer mutations. It is important to always contrast two phenotypic pools. Segregation distortion results from the unequal contribution of allele or allele combinations to the next generation. If only one pool was genotyped on an array, segregation distortion would falsely look like the location of the mutation due to distorted allele frequency. Thus contrasting allele frequencies across two pools controls for this distortion, which is likely independent of the phenotype of interest (*see* **Note 4**).

4. Notes

1. Plants should all be grown together and tissue collected at the same developmental stage.
2. Standard reagents for Affymetrix hybridization and washing protocols can be found at www.affymetrix.com/auth/support/downloads/manuals/expression_ever_manual.zip.
3. Labeled products should be of about 50 bp and all bands should be close to the same intensity. Dilute samples if necessary so that each reaction has the same amount of labeled product.
4. In conclusion, array genotyping is a rapid method to identify and genotype several hundred thousand loci in a single assay. As with RFLPs and AFLPs, the exact nature of the polymorphism is unknown; however, SFPs make for excellent molecular markers. Clusters of SFPs can reveal natural deletions that may serve as candidate genes for QTLs. Similarly, clusters of

features with altered hybridization can reveal the location of induced mutations. Finally, bulk segregant mapping and XAM can quickly identify the rough map position(s) of Mendelian loci and moderate-effect QTLs. Once a rough map position is obtained, candidate genes can be identified and directly sequenced.

Potential problems using array mapping may arise if experiments are not performed in parallel. Subtle differences in the DNA preparation and labeling will result in different SFPs being called between experiments. Plant DNA should be collected at the same developmental stage. It is possible to use SFPs defined from parental hybridizations in one experiment for mapping mutations performed in another experiment. In this case it is crucial to have the wild-type pool as a control. It is also important to load a very similar amount of labeled DNA onto each array.

Acknowledgments

Thank you to Sam Hazen and Yunda Huang for comments on the manuscript. Thanks to Tom Gal for help with programming and the Helen Hay Whitney Foundation for funding.

References

1. Borevitz, J. O. and Chory, J. (2004) Genomics tools for QTL analysis and gene discovery. *Curr. Opin. Plant Biol.* **7,** 132–136.
2. Borevitz, J. O. and Nordborg, M. (2003) The impact of genomics on the study of natural variation in *Arabidopsis*. *Plant Physiol.* **132,** 718–725.
3. Konieczny, A. and Ausubel, F. M. (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4,** 403–410.
4. Bell, C. J. and Ecker, J. R. (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19,** 137–144.
5. Borevitz, J. O., Liang, D., Plouffe, D., et al. (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res.* **13,** 513–523.
6. Winzeler, E. A., Richards, D. R., Conway, A. R., et al. (1998) Direct allelic variation scanning of the yeast genome. *Science* **281,** 1194–1197.
7. Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19,** 185–193.
8. Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31,** e15.
9. Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004) Affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20,** 307–315.
10. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98,** 5116–5121.
11. Storey, J. D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100,** 9440–9445.
12. Zhang, L., Miles, M. F., and Aldape, K. D. (2003) A model of molecular interactions on short oligonucleotide microarrays. *Nat. Biotechnol.* **21,** 818–821.
13. Churchill, G. A., Giovannoni, J. J., and Tanksley, S. D. (1993) Pooled-sampling makes high-resolution mapping practical with DNA markers. *Proc. Natl. Acad. Sci. USA* **90,** 16–20.
14. Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. USA* **88,** 9828–9832.
15. Wolyn, D. J., Borevitz, J. O., Loudet, O., et al. (2004) Light-response quantitative trait loci identified with composite interval and eXtreme array mapping in *Arabidopsis thaliana*. *Genetics* **167,** 907–917.

13

Forward Genetic Screening of Insertional Mutants

Anja Schneider and Dario Leister

Summary

Insertional mutagenesis has contributed to the success of forward genetics in *Arabidopsis thaliana*. The availability of large collections of lines mutagenized by either transposon or T-DNA insertions, in combination with the systematic sequencing of insertion/genome junctions, enables the identification of mutations for any given gene in the nuclear genome of this plant species. Protocols for the identification and confirmation of mutations by forward genetics, the isolation of insertion/genome junctions, and the stabilization of transposon-induced mutations are provided, together with an overview of available insertional mutant collections and of their characteristics.

Key Words: Amplification of insertion mutagenized sites (AIMS); flanking DNA; footprint; forward genetics; insertion; mutant; PCR; segregation; transposon; T-DNA.

1. Introduction

For unclear reasons, knockout of nuclear genes by homologous recombination is still not practicable in *Arabidopsis thaliana* and in other higher plants ***(1,2)***. On the contrary, analysis of gene functions by antisense, cosuppression, or RNA interference, as well as chemical or physical mutagenesis, is feasible in *A. thaliana* ***(3,4)***. In addition, in this species, mobilization of retrotransposons has been reported ***(5,6)***, which might facilitate the generation of nontransgenic insertional mutants at a genomic level, such as in rice ***(7,8)***. The most important contribution to the systematic and large-scale analysis of nuclear gene functions in *Arabidopsis* has been made possible by insertional mutagenesis. Large collections of insertional mutants, based on gene disruptions by transposons or T-DNAs, are available. The latter ones, in particular, predominate in *Arabidopsis*, because of (1) the ease with which this species can be transformed via *Agrobacterium* and (2) the compactness of its genome, which ensures a high rate of genic insertions. *Arabidopsis* insertional mutant collections can be systematically searched for mutations in genes of interest ("reverse genetics") by PCR, by hybridization of gene-specific probes to insertion-flanking DNA fragments spotted on arrays, or—most efficiently—by querying databases of the flanking sequences of the insertions (*see*, e.g., refs. ***9–14***). Complementary phenotypic screens result in the systematic identification of insertional mutants exhibiting phenotypes of interest; the mutated genes can be relatively easily identified starting from the insertion as a "molecular tag" ("forward genetics").

Based on statistical calculations ***(10)***, ~190,000 independent and random insertions seem to be necessary to knockout, with a probability of 95%, an *Arabidopsis* gene of average size (~ 2 kbp) (**Fig. 1**; *see* **Note 1**), and a 99% chance would require ~290,000 insertions. However, the probability of finding an insertion within a given gene is also a function of gene length: smaller genes are less likely to be targets of insertions than larger ones (**Fig. 1**). Furthermore, insertions of T-DNA or transposons are not randomly distributed in the genome. Some transposons are known to preferentially insert into genes (*see*, e.g., ***15*** and ***16***), whereas T-DNA integrations

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

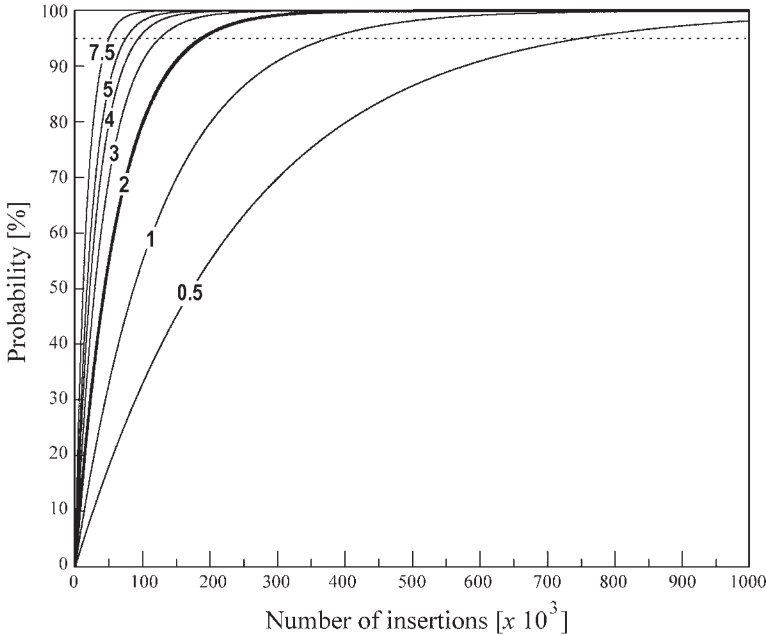


Fig. 1. The probability of finding an insertion within a given gene is shown as a function of gene length. Curves are drawn for different gene lengths from 0.5 to 7.5 kbp. Calculations were performed as detailed in **Note 1**. The curve for a gene length of 2 kbp, the average gene length in *A. thaliana*, is indicated by a doubled thickness. A probability of 95% is indicated by a dotted line. Most insertional mutant collections contain one or few insertions per line. Exceptions are multicopy transposon lines ***(9)***. Whereas the ease of gene isolation is the clearest advantage in using insertional mutants, the incomplete coverage of the entire genome by insertions in all populations available is a drawback, in particular if the isolation of saturated mutant collections is intended.

occur less frequently in *Arabidopsis* in the pericentromeric chromosomal regions, compared with the rest of the genome. In this model plant species, T-DNA insertions occur more likely in intergenic regions, promoters or untranslated regions (UTRs) than in exons or introns ***(12,13)***. Currently, the most advanced resource for the screening of insertional mutants in *Arabidopsis* are collections of insertion lines with sequenced insertion/genome junction sequences (*see* **Subheading 2.1.1.**). The SALK T-DNA collection alone was estimated to have a probability of 96.6% of obtaining an insertion in an *Arabidopsis* gene of average length ***(13)***. Taking all flanking-sequence-indexed populations together, more than 350,000 insertion/genome junctions have been sequenced so far—a number that should enable the identification of insertional mutants for almost any given gene in this species (*see* **Fig. 1**). For most phenotypic screens—in particular the more sophisticated ones involving extensive manual work or elaborate instrumentation—screening of such a large number of lines is not feasible. In the case that screens shall result into the isolation of saturated mutant collections, heavily mutagenized populations (e.g., EMS lines) or certain multicopy insertional lines (e.g., ref. ***9***) are the approach of choice to minimize the number of lines to be screened.

In this chapter we present protocols for the identification and confirmation of insertional mutants in the course of forward genetics experiments. Also, protocols for the isolation of flanking sequences of insertions, as well as for the stabilization of transposon-induced mutations, are supplied.

2. Materials

2.1. Forward Genetic Screens

2.1.1. Resources for Mutant Identification by Forward Genetic Screens

1. Insertion collections (**Table 1**).
2. Stock centers:
   a. Arabidopsis Biological Resource Center (ABRC; www.arabidopsis.org/abrc/).
   b. Nottingham Arabidopsis Stock Centre (NASC; http://nasc.nott.ac.uk/).
   c. Sendai Arabidopsis Seed Stock Center (SASSC; www.shigen.nig.ac.jp/arabidopsis/).

2.1.2. DNA Extraction

1. 1.5-mL tubes.
2. Liquid nitrogen.
3. Extraction buffer: 0.3 *M* NaCl, 50 m*M* Tris-HCl, pH 7.5, 20 m*M* EDTA, 2% (w/v) sarkosyl, 0.5% (w/v) SDS, 5 *M* urea, 5% (v/v) phenol.
4. Phenol/chloroform/isoamylalcohol (25:24:1).
5. Isopropanol.
6. 70% Ethanol.
7. Solution of 10 µg/mL RNAse A in TE (10 m*M* Tris-HCl, pH 7.4, 1 m*M* EDTA, pH 8.0).
8. λ-DNA.
9. 1% (w/v) agarose gel in TAE buffer (40 m*M* Tris-Ac, 1 m*M* EDTA, made from 50X stock solution) containing 1/10,000 vol of ethidium bromide (5 mg/mL).

2.1.3. Southernblot Analysis

1. Genomic DNA.
2. DNA restriction enzyme and appropriate stocks of 10X reaction buffer.
3. Loading buffer: 15% (w/v) Ficoll, 50 m*M* EDTA, pH 8.0, 0.5% (w/v) sodium dodecyl sulfate (SDS), 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF.
4. 0.8% agarose gel in TAE buffer (*see* **Subheading 2.1.2.**).
5. Horizontal electrophoresis chamber and power supply.
6. Denaturation solution: 1.5 *M* NaCl, 0.5 *M* NaOH.
7. Neutralization solution: 1 *M* NH_4Ac, 10 m*M* NaOH.
8. Blotting buffer: 20X SSC (3 *M* NaCl, 0.3 *M* Na-Citrate, pH 7.0).
9. Filter paper (3MM, Whatman), paper towels and nylon membrane (Hybond N+, Amersham).
10. UV-crosslinker (Stratagene).
11. (Pre-)hybridization solution: 6X SSC, 5X Denhardt's (made from a 100X Denhardt's stock: 2% [w/v] Ficoll, 2% [w/v] polyvinylpyrrolidon, 2% [w/v] bovine serum albumin), 0.5% (w/v) SDS.
12. Sonicated nonhomologous DNA (1 mg/mL) (e.g., herring sperm DNA).
13. Hybridization tubes and hybridization oven.
14. Border-specific probe labeled by any method described in ref. ***29***.
15. Washing solutions: (a) 2X SSC, 0.1% (w/v) SDS and (b) 0.2X SSC, 0.1% (w/v) SDS.
16. Saran wrap.
17. Detection system (phosphoimager or x-ray film).
18. 0.4 *M* NaOH.
19. 2X SSPE made from a 20X SSPE stock (3.6 *M* NaCl, 0.2 *M* sodium phosphate, 0.02 *M* EDTA, pH 7.7).

2.2. Isolation of Insertion/Genome Junction Sequences by AIMS

2.2.1. Plant Material, DNA Extraction, and Restriction

1. F2 plants.
2. DNA extraction as in **Subheading 2.1.2.**
3. DNA restriction enzymes and appropriate stocks of 10X reaction buffers.

Table 1
Collections of Insertional Mutant Lines

Collection	Genetic background	Type of insertion	Number of lines	Mode of distribution	Web page/reference
Feldmann lines	Ws	T-DNA: 3850:1003 Ti plasmid	~10,000	Stock centers: pools of 20 or 100 lines	(*17*)
Sainsbury Laboratory *Arabidopsis Thaliana* Transposants (SLAT) Collection	Col-0	Transposon: *Spm/dSpm*	~58,000	Stock centers: pools of 9600 or 50 lines	http://www.jic.bbsrc.ac.uk/sainsbury-lab/jonathan-jones/SINS-database/index.htm (*18*)
Cold Spring Harbor Laboratory (CSHL) *Arabidopsis* Genetrap Collection	L*er*	Transposon: *Ds*[a]	~37,000	Stock centers: 338 lines are available for ordering. Seed from individual sequence-indexed lines are available on request.	http://genetrap.cshl.org/
RIKEN BioResource Center Collection	Nossen	Transposon: *Ds*	~12,000	Available through the RIKEN Bio-Resource Center	http://rarge.gsc.riken.go.jp/dsmutant/index.pl (*19*,*20*)
Salk Institute Genome Analysis Laboratory (SIGnal)	Col-0	T-DNA: pROK2	~150,000	Stock centers: individual and bulks of sequence-indexed lines; 40,000 lines are available as pools of 100	http://signal.salk.edu/cgi-bin/tdnaexpress (*13*)
Syngenta *Arabidopsis* Insertion Library (SAIL)	Col-0	T-DNA: pDAP101 and pCSA110	~100,000	Stock centers: ~45,000 sequence-indexed lines can be individually ordered	www.nadii.com/pages/collaborations/garlic_files/GarlicDescription.html; from 2004: www.arabidopsis.org/abrc/ (*12*)
Kölner *Arabidopsis* T-DNA Lines (GABI-Kat)	Col-0	T-DNA: pAC106[b] and pAC161[b]	~60,000	Seed from individual sequence-indexed lines are available on request	www.mpiz-koeln.mpg.de/GABI-Kat/ (*21*,*22*)
Versailles-INRA	Ws	T-DNA: pGKB5[c]	~43,000	Stock centers: ~10,000 lines are distributed in pools of 20 or 100. Seed from individual sequence-indexed lines are available on request	http://flagdb-genoplante-info.infobiogen.fr/projects/fst/DocsIntro/introCollection.html (*23*,*24*)
Weigel lines	Col-7	T-DNA: pSKI15[b]	~23,000	Stock centers: lines are distributed in pools of 12, 96, or 100	(*25*)

Syngenta (previously Zeneca Mogen) Promoter Trap Lines	C24	T-DNA: pMOG553[a]	~1000	Stock centers: pools of 20	(26)
AKF lines (Sussman and Amasino lines)	Ws	T-DNA: pD991-AP3[d]	~60,000	Stock centers: pools of 9 or 225	www.biotech.wisc.edu/Arabidopsis/Index2.asp
Scheible and Somerville lines	Col-2	pSKI15[b]	~63,000	Stock centers: pools of various sizes (100 to 350)	www.arabidopsis.org/abrc/Scheible.jsp
Jack lines	Col-6, *gl1-1*	T-DNA: pD991[d]	~11,000	Stock centers: pools of 10 or 100	(27)
Yokoi, Koiwa, Bressan lines	C24 RD29A-LUC	T-DNA: pSKI15[b]	~24,000	Stock centers: pools of 10 or 100	www.arabidopsis.org/abrc/bressan.jsp
LeClere and Bartel lines	Col	T-DNA: 35SpBARN	~33,000	Stock centers: pools of 100	(28)
John Innes Centre Gene Trap lines (FGT collection)	L*er*	Transposon: *Ds*[a]	~23,000	Stock centers: individual lines	www.jic.bbsrc.ac.uk/science/cdb/exotic/

[a]Gene trap.
[b]Activation tag.
[c]Promoter trap.
[d]Enhancer trap.

2.2.2. Adapter Ligation

1. Primer sequences for adapters:
   LR32: ACTCGATTCTCAACCCGAAAGTATAGATCCCA
   APL16: *P*-TATGGGATCACATTAA-NH_2
   APL17: *P*-CGTGGGATCACATTAA-NH_2
   LR26: ACTCGATTCTCAACCCGAAAGTATAG.
2. Adapter annealing buffer: 10 m*M* Tris-HCl, pH 7.5, 10 m*M* $MgCl_2$.
3. AIMS adaptors:
   APL1632 (LR32 + APL16) with AT overlap for ligation to *Bfa*I- or *Csp*6I-digested DNA,
   APL1732 (LR32 + APL17) with GC overlap for ligation to *Hin*6I-restricted DNA.
4. T4 DNA Ligase and 10x Ligase buffer (Boehringer-Roche).

2.2.3. Linear and Exponential PCR

1. 0.1 m*M* stocks of "inside" insertion-specific primer (must be positioned "inside" the insertion relative to the nested primer used in the subsequent exponential PCR), of nested insertion-specific primer and of LR26 (*see* **Subheading 2.2.2.**). "Inside" and nested primers should have annealing temperatures around 64°C.
2. 10X PCR buffer (Clontech).
3. *Taq* polymerase (Advantage 2, Clontech).
4. 2 m*M* stock of dNTPs (Boehringer-Roche).

2.2.4. Polyacrylamide Gel Electrophoresis, Staining, and Elution of Bands

1. 4.5% (w/v) denaturing polyacrylamide gel.
2. Vertical electrophoresis chamber and power supply.
3. 1-kb ladder (Invitrogen).
4. Loading dye: 98% (v/v) deionized formamide, 10 m*M* EDTA, pH 8.0, 0.05% (w/v) bromophenol blue, 0.05 % (w/v) xylene cyanol FF.
5. Fixing solution (for 3 L): 315 mL 96% EtOH, 15 mL 100% acetic acid, 2670 mL water.
6. Staining solution (for 1 L): 2 g $AgNO_3$ in 1 L of fixing solution.
7. Developing solution (for 1 L): 30 g NaOH, 5 mL formaldehyde, 965 mL water.
8. Extraction buffer for gel elution: 50 m*M* KCl, 10 m*M* Tris-HCl, pH 9.0, 1% (v/v) Triton.
9. High-pure PCR product purification kit (Boehringer-Roche).

2.3. Confirmation and Initial Molecular Analysis of Mutation–Phenotype Relationships

2.3.1. In Silico Identification, Confirmation, and Genotyping of Additional Insertional Mutant Alleles and Plant Propagation

1. Insertion–specific primers for some widely used populations:
   SIGnAL/SALK collection: the two left border primers used for sequencing were LBa1 (5'-TGGTTCACGTAGTGGGCCATCG-3') and the nested primer LBb1 (5'-GCGTGGACCGCTTGCTGCAACT-3'). For the right border, the primer RB438as (5'-GTGCCCAGTCATAGCCGAATAGC-3') can be used.
   GABI-Kat collection: several primers were used for sequencing. Left border: 8409 (5'-ATATTGACCATCATACTCATTGC-3'), oriL (5'-ATATTGACCATCATACTCATTGC-3'); right border: 3144 (5'-GTGGATTGATGTGATATCTCC-3'), 35St (5'-GTGGATTGATGTGATATCTCC-3').
   Versailles collection/FlagDB lines: for the right border the primer Tag3 (5'-CTGATACCAGACGTTGCCCGCATAA-3') was used for sequencing, for the left border Tag5 (5'-CTACAAATTGCCTTTTCTTATCGAC-3').
   For primer specific for other populations and further protocols, *see* **Note 2**.
2. Freely available primer design programs:
   iSect Primer Design: http://signal.salk.edu/isectprimers.html.
   Primer3: www.basic.nwu.edu/biotools/Primer3.html.
3. DNA extraction as in **Subheading 2.1.2.**

4. PCR:
   a. Genomic DNA.
   b. PCR tubes.
   c. Water.
   d. 10X PCR buffer.
   e. 1 m*M* dNTPs.
   f. 0.1 m*M* stock of each primer.
   g. *Taq* polymerase.
   h. 1.5% (w/v) agarose gel in TAE buffer (*see* **Subheading 2.1.2.**).

2.3.2. RNA Extraction

1. Mortar, pestle, liquid nitrogen, 2-mL tubes.
2. RNA extraction buffer: dissolve 50 g guanidinium thiocyanate and 0.5 g sodium lauryl sarcosinate in 80 mL water, add 830 µL 3 *M* NaAc, pH 5.2, adjust to pH 7.0 with approx 160 µL 10 *M* NaOH, fill up to a total volume of 100 mL with water, autoclave, add 0.7 mL β-mercaptoethanol, and mix the solution with 1 vol of water-saturated phenol.
3. Chloroform/isoamylalcohol (24:1).
4. 1 *M* acetic acid.
5. 100% and 80% EtOH.
6. 3 *M* NaAc.
7. Diethylpyrocarbonate (DEPC)-treated water.
8. Photometer.

2.3.3. RT-PCR

1. DNase (10 U/µL, RNase free) (Invitrogen).
2. 25 m*M* EDTA, pH 8.0.
3. Oligo-dT18 primer (500 ng/µL).
4. 2.5 m*M* dNTPs.
5. 0.1 *M* DTT.
6. Superscript II Reverse Transcriptase and 5X first-strand buffer (Invitrogen).
7. 0.1 m*M* stock of gene-specific primer.
8. *Taq* polymerase and 10X PCR buffer.
9. 1.5% agarose gel in TAE (*see* **Subheading 2.1.2.**).

2.4. Identification of Transposon Footprints by PCR

2.4.1. Plant Material

1. Around 50 to 100 plants of the progeny of a homozygous mutant (*see* **Note 3**) and WT control plants at the 6-leaf stage.
2. DNA isolation as in **Subheading 2.1.2.**

2.4.2. PCR

1. 0.1 m*M* stocks of primers flanking the transposon insertion site (*see* **Note 4**).
2. 10X PCR buffer.
3. *Taq* polymerase.
4. 2 m*M* stock of dNTPs.

2.4.3. Polyacrylamide Gel Electrophoresis, Staining, and Elution of Bands

As in **Subheading 2.2.4.**, except that 6% (w/v) instead of 4.5% (w/v) denaturing polyacrylamide gel is used.

2.4.4. Identification of Stabilized Mutant Alleles

As in **Subheadings 2.4.1.–2.4.3.**

3. Methods

3.1 Forward Genetic Screens

3.1.1. Mutant Identification and Segregation Analysis

1. Identify mutant lines from collections of insertional mutant lines (for suitable populations, *see* **Table 1**) with a phenotype of interest (*see* refs. ***30–37*** for some examples).
2. Backcross mutant to wild-type and collect F1 seeds. Sow out F1 seeds in the greenhouse and allow them to self. Collect F2 seeds. Check the phenotype of both F1 and F2 plants (*see* **Note 5**).
3. Grow 60 individual plants of a segregating F2 population to the 8-leaf stage and analyze segregation pattern (*see* **Note 5**). Harvest leaf material from each plant.

3.1.2. DNA Extraction According to Ref. 38

1. Grind 3 to 4 leaves in a 1.5-mL tube in liquid nitrogen.
2. Add 500 µL extraction buffer and vortex.
3. Add 400 µL phenol/chloroform/isoamylalcohol.
4. Centrifuge for 15 min at 20°C at maximum speed in a bench-top centrifuge.
5. Transfer supernatant to a fresh tube, add 0.8 vol of isopropanol, and incubate for 10 min at room temperature.
6. Centrifuge for 10 min at room temperature at maximum speed in a bench-top centrifuge.
7. Wash the precipitate with 70% ethanol, repeat washing step.
8. Collect precipitate, dry it, and dissolve in 50 µL TE containing RNAse A.
9. Check integrity and concentration of DNA by agarose gel electrophoresis using λ-DNA as a standard.

3.1.3. Southern Blot Analysis

1. Digest each plant DNA (32 µL, *see* **Subheading 3.1.2.**) with an appropriate restriction enzyme (*see* **Note 6**) in a 40-µL reaction containing 4 µL 10X buffer and 4 µL (10 U/µL) restriction enzyme. Mix well and incubate for 4 h at 37°C.
2. Stop the restriction reaction by adding 8 µL loading buffer. Load the reaction on an agarose gel and run the gel for 5 h at 60 to 70 V.
3. Denature the gel for 30 min with gentle shaking. Rinse the gel with tap water and put in neutralization solution for 20 min.
4. Set up a capillary blot as follows: Fill tray with blotting buffer (20X SSC) and make a platform with filter paper, saturated with blotting buffer. Place the gel on the filter paper and put a sheet of Hybond N+ membrane on top of the gel. Place three sheets of filter paper, wetted with blotting buffer, and a stack of absorbent paper towels on top of the membrane. Put a 0.5-kg weight on top and allow to transfer for 16 h. After blotting fix the membrane by ultraviolet (UV) crosslinking.
5. Add the membrane to 25 mL prehybridization buffer in a hybridization tube. Denature 0.5 mL of sonicated nonhomologous DNA by heating to 100°C for 5 min, add to prehybridization solution, and incubate in a hybridization oven at 65°C for 1 h.
6. Denature the labeled border-specific probe by heating to 100 °C for 5 min, add to prehybridization solution, and incubate at 65°C for at least 12 h.
7. Following hybridization wash the filter by incubating in 2X SSC, 0.1% (w/v) SDS at 65°C for 30 min. Repeat washing. Replace the solution with 0.2X SSC, 0.1% (w/v) SDS. Incubate at 65°C for 10 min.
8. Remove filter, wrap in Saran wrap, and carry out autoradiography.
9. For removal of the border-specific probe, incubate the filter in 0.4 *M* NaOH at 65°C for 30 min. Replace the solution with 2X SSPE and incubate at 65°C for 10 min. Repeat the last step.
10. Repeat **steps 5** to **8** for a probe specific for the other border (optional).
11. Analyze the presence of the T-DNA or transposon insertion for cosegregation with the mutant phenotype (*see* **Notes 6** and **7**).

3.2. Isolation of Insertion/Genome Junction Sequences by AIMS (Amplification of Insertion Mutagenized Sites) (Modified From Refs. 39,40)

3.2.1. Plant DNA Extraction and Restriction

1. Extract DNA from F2 plants as in **Subheading 3.1.2.**
2. Digest 50 to 100 ng of genomic DNA for 2 to 3 h at 37°C. Use a total volume of 20 µL 1X restriction buffer and 2 U of restriction enzyme (*Bfa*I or *Csp*6I to generate AT overlaps; *Hin*6I for GC overlaps).
3. Inactivate restriction enzyme by incubating for 20 min at 80°C (*Bfa*I) or at 65°C (*Csp*6I, *Hin*6I).

3.2.2. Adapter Ligation

1. Adaptors are generated by diluting oligo stocks (250 m*M*) in annealing buffer (10 m*M* Tris-HCl, pH 7.5, 10 m*M* $MgCl_2$) to a final concentration of 25 m*M*. This mixture is warmed up to 80°C for 5 min, allowed to cool down slowly, and can be stored at –20°C.
2. For the ligation, take 0.5 µL of the appropriate adapter (APL1632 for AT-overhangs, APL1732 for GC-overhangs), 5 µL of the restricted DNA, 2 µL of 10X ligation buffer, 1.5 U T4 ligase, and fill up to a total volume of 20 µL with water.
3. Ligate overnight at 16°C.

3.2.3. Linear and Exponential PCR

1. Into a PCR tube, add 14.1 µL water, 1 µL of the ligation reaction, 2 µL of 10X PCR buffer, 2 µL of 2 m*M* dNTP, 0.4 µL of the "inside" primer (0.1 m*M*), and 0.5 µL of Advantage 2 *Taq* polymerase (Clontech).
2. Run the reaction in a PCR machine with heated lid with the following program: initial denaturation for 2 min at 94°C, followed by 30 cycles of 30 s denaturation at 94°C, 1 min annealing at 64°C, and 1 min 30 s elongation at 73°C.
3. Into a PCR tube, add 34.5 µL water, a 1-µL aliquot of the linear PCR, 5 µL of 10X PCR buffer, 5 µL of 2 m*M* dNTPs, 2 µL of the "nested" primer (0.1 m*M*), 2 µL of LR26, and 0.5 µL of Advantage 2 *Taq* polymerase (Clontech).
4. Run the reaction as for the linear PCR but for 40 cycles.

3.2.4. Polyacrylamide Gel Electrophoresis, Elution, and Sequencing of Bands

1. Mix 5 µL of the PCR reaction 1:1 with loading dye and load on a 4.5% polyacrylamide gel; as a size marker, load 250 ng of 1 kbp ladder.
2. Run for 3 to 4 h at 58 V.
3. For silver staining, remove "nonstick" plate and incubate the "stick" plate with the gel for 3 min in fixing solution; then transfer for 5 min in staining solution, 20 s in water, exchange water, and incubate for another 2 min. Then incubate for around 5 min in developing solution, followed by 5 min in fixing solution and a final 10-min incubation in water.
4. Excise band(s) of interest (*see* **Note 8**) and add 150 to 200 mL of extraction buffer to the gel slice.
5. Incubate at 95°C for 20 min, centrifuge for 30 s in a tabletop centrifuge at maximum speed, and rescue the supernatant.
6. Use 1 mL of this solution for an exponential PCR as described in **steps 3** and **4** in **Subheading 3.2.3.**
7. Purify the PCR product and sequence it.

3.3. Confirmation and Initial Molecular Analysis of Mutation–Phenotype Relationships

3.3.1. In Silico Identification, Confirmation, and Genotyping of Additional Insertional Mutant Alleles and Plant Propagation

1. Search databases of sequence-indexed T-DNA/genome junctions (**Table 1**) for additional insertions in your gene of interest (*see* **Note 9**) and order electronically seeds of suitable insertional mutant lines (*see* **Note 10**) from stock centers, or directly from the supplier if seeds have not been deposited at stock centers.

2. Grow 12 to 36 individual plants of each line to be analyzed in the greenhouse to the 8-leaf stage. Check phenotype. Harvest leaf material from each plant for DNA preparation (as in **Subheading 3.1.2.**), but leave plants otherwise intact to obtain seeds.
3. Perform two PCR reactions for each plant: one to detect the mutant allele and a second one for the wild-type allele (*see* **Note 11**). For each reaction, add into a PCR tube 14.1 μL water, 1 μL of DNA, 2 μL of 10X PCR buffer, 2 μL of 1 m*M* dNTPs, 0.4 μL of each primer (0.1 m*M*), and 0.1 μL (0.5 U) of *Taq* polymerase.
4. Run the reaction in a PCR machine with heated lid using the following program: initial denaturation for 2 min at 94°C, followed by 35 cycles of 15 s denaturation at 93°C, 45 s annealing at 55°C, and 30 s elongation at 72°C.
5. Separate PCR products on an agarose gel. Purify the amplicon of the insertion/genome junction and sequence it.
6. Repeat **steps 1** to **3** for the other border of the insertion (*see* **Note 12**).
7. Genotype F2 plants by PCR and analyze genotype–phenotype relationships to confirm that the phenotype is caused by the insertion (*see* **Note 7**).

3.3.2. RNA Extraction

1. Grow 2 to 4 progeny of a selected homozygous mutant plant and 2 to 4 wild-type plants to the desired stage (*see* **Note 13**).
2. Harvest 300 mg of plant material and freeze in liquid nitrogen.
3. Grind the plant material in a mortar in liquid nitrogen, add 2 ml RNA extraction buffer, and grind for another 2 min.
4. Distribute the homogenate into 2-mL tubes, add 0.2 mL chloroform/isoamyalcohol, mix, and incubate for 30 min on ice.
5. Centrifuge for 10 min at 12,000*g* at 4°C.
6. Transfer supernatant to a fresh tube; add 1/20 vol of 1 *M* acetic acid and 0.7 vol of 100% ethanol. Incubate 30 min on ice.
7. Centrifuge for 10 min at maximum speed in a bench-top centrifuge.
8. Wash the precipitate with 0.4 mL 3 *M* NaAc.
9. Centrifuge for 15 min at maximum speed in a bench-top centrifuge.
10. Wash the precipitate with 80% ethanol; repeat wash step.
11. Collect precipitate, dry it, and dissolve in 50 μL water.
12. Check integrity of RNA by agarose gel electrophoresis and determine concentration by absorption at 260 nm.

3.3.3. RT-PCR

1. Perform a DNase treatment for each RNA sample. Add into a 1.5-mL tube 1 μg of total RNA, 5 μL 5X first-strand buffer, 1 μL DNase (10 U/μL); fill up to a total volume of 25 μL with water.
2. Incubate the reaction for 15 min at 37°C.
3. Add 2.5 μL of 25 m*M* EDTA and incubate for 10 min at 65°C. Put the reaction on ice.
4. Split the reaction into two separate 1.5-mL tubes. For each reaction, add 2 μL oligo dT primer (500 ng/μL), 7.5 μL 5X first-strand buffer, 10 μL dNTPs (2.5 m*M* each dNTP), 5 μL 0.1 *M* DTT, and 11.5 μL water. Add 1 μL Superscript II Reverse Transcriptase to one reaction. Add 1 μL water to the control reaction.
5. Incubate 60 min at 42°C.
6. Inactivate the enzyme for 15 min at 70°C.
7. Perform a PCR with 2 μL of first-strand cDNA and 2 μL of control reaction as follows: For each reaction, add into a PCR tube 2 μL template, 2 μL of 10X PCR buffer, 2 μL of 1 m*M* dNTPs, 0.5 μL of each primer, 0.1 μL (0.5 U) of *Taq* polymerase, and 12.9 μL water (*see* **Note 14**).
8. Run the PCR reaction as in **Subheading 3.3.1.** Separate PCR products on an agarose gel and compare the intensity of the wild-type amplicon with the one from the mutant (*see* **Note 10**).

3.4. Stabilization of Transposon-Induced Mutations: Identification of Footprints by PCR (Note 15)

3.4.1. Plant DNA Extraction

1. Prepare DNA of two leaves each of mutant plants (*see* **Note 3**) and the WT control according to **Subheading 3.1.2.**

3.4.2. Amplification of Empty Donor Sites

1. To amplify empty donor sites, add into a PCR tube 14.1 µL water, 1 µL of DNA, 2 µL of 10X PCR buffer, 2 µL of 1 m*M* dNTPs, 0.4 µL of each primer (0.1 m*M*), and 0.1 µL (0.5 U) of *Taq* polymerase (*see* **Note 4**).
2. Run the reaction in a PCR machine with heated lid using the following program: initial denaturation for 2 min at 94°C, followed by 35 cycles of 15 s denaturation at 93°C, 45 s annealing at 55°C, and 30 s elongation at 72°C.

3.4.3. Polyacrylamide Gel Electrophoresis, Elution, and Sequencing of Bands

1. Mix 5 µL of the PCR reaction 1:1 with loading dye and load on a denaturing 6% polyacrylamide gel; as a size marker load 250 ng of 1-kbp ladder.
2. Run for 3 to 4 h at 58 V.
3. Perform silver staining as in **Subheading 3.2.4.**
4. Excise bands with a different size than in the wild-type (*see* **Note 15**); extract DNA and reamplify as in **Subheading 3.2.4.**
5. Purify the PCR product and sequence it.

3.4.4. Identification of Stabilized Mutant Alleles

1. Select plants that contain a footprint of interest in both leaves (*see* **Note 16**), allow them to self, and let them produce seeds
2. Analyze around 12 plants of each offspring as described in **Subheadings 3.4.1.–3.4.3** to identify germinally transmitted footprints. Control the presence of the transposon insertion by a PCR employing an insertion site flanking primer and a transposon-specific primer.
3. Reconfirm the footprint by sequencing the PCR-amplified empty donor site.

4. Notes

1. The probability (P) of finding an insertion within a given gene is based on the formula: $P = 1 - (1-[x/125{,}000])^n$, where x is the length of the gene in kbp, and n is the number of insertions present in the genome. The average size of a gene in *Arabidopsis* is ~2 kbp and the total size of the genome is 125 Mbp ***(41)***. Note that this formula depends on a random distribution of insertions in the genome, which is actually not the case in *A. thaliana* (*see* **Subheading 1.**).
2. Primer sequences and protocols for sequencing of insertion/genome junction sequences of the following collections are available: SLAT collection/SINS database: http://arabidopsis.info/info/SINSmap.html; http://arabidopsis.info/info/IPCRprotocol.html; CSHLArabidopsis Genetrap Collection: http://genetrap.cshl.org/traps.html; IMA collection: http://arabidopsis.info/imainfo.html.
3. Genotyping of insertional mutants is difficult when the insertion is due to an autonomous transposon (such as in ref. ***9***; *see also* **Note 15**): PCR assays based on primers, flanking the insertion site and that should only lead to amplification of a genomic fragment from the wild-type allele, can result also in PCR products in homozygous mutants due to somatic reversions.
4. Primers flanking the insertion sites should span a region between 100 and 200 bp to allow optimal separation of transposon footprints on polyacrylamide gels.
5. If the mutation is recessive, F1 plants resulting from a cross between the mutant and the wild-type will appear like the wild-type. When mutations are dominant, as is desired during activa-

tion tagging, F1 plants exhibit the mutant phenotype (in the case the original mutation was heterozygous, F1 plants segregate the mutant phenotype in 1:1 fashion). A Mendelian distribution (1:2:1) is expected in the F2 population for the different genotypes concerning the insertional event. If the mutation is recessive, the phenotypic segregation will be 3:1 (wild-type:mutant), in case of a dominant mutation a 1:3 ratio is expected.

6. A 6-cutter restriction enzyme should be chosen that generates, on average, 3-kbp fragments. Because most insertional mutagens are quite large, this restriction enzyme should cut within the insertion, around 500 to 1000 bp apart from a border. At best (in a single-copy insertion line), after hybridization only one band, present in all mutant plants and in about two-thirds of wild-type plants, should be detectable. If additional bands are present, they could arise from the same T-DNA/transposon locus (in case of a complex T-DNA/transposon insertion), or from multiple, independent insertions. In the case of a complex insertion locus it is not possible in F2 individuals to separate the hybridizing bands. When multiple independent insertions exist, a number of F2 individuals should be checked for single-copy lines. If this fails, an additional round of backcrossing to the wild-type has to be performed.
7. In the course of either forward or reverse genetics a segregation analysis is required to establish whether a mutant phenotype is linked to the T-DNA/transposon insertion. A minimum of 59 plants should be analyzed for a >95% confidence level (i.e., an α value of 5%) of linkage within 5 centiMorgan, according to the formula $n = \log P/\log f$, where n is the number of individuals, P is the α value, and f is the probability that the crossover event does not occur. The equation is as follows: $59 > \log 0.05/\log 0.95$. Once linkage is established to a certain degree, additional experiments might be required to prove that the mutant phenotype is caused by a T-DNA/transposon insertion. This can be done by analyzing a second independent mutant allele and/or complementation the mutation with the wild-type gene (*see* **Subheading 3.3.**).
8. At best, only one band will be obtained after PCR amplification. In multicopy insertion lines, as well as because of unspecific priming, however, several bands can be obtained during AIMS analysis. In such cases, only bands that are present in all mutant plants, but only in about two-thirds of plants exhibiting wild-type phenotypes, are candidates for the searched insertion/genome junction.
9. In the MAtDB (MIPS *Arabidopsis Thaliana* DataBase; http://mips.gsf.de/proj/thal/db/index.html) database, for each gene direct links to T-DNA insertions of the SALK and GABI-KAT collections are available.
10. Lines with an insertion within the translated region of the gene have the highest probability to generate a null allele. Nevertheless, insertions in the 5'- or 3'-untranslated region (UTR) of the gene can also affect the gene function, or even induce a complete knockout of the gene function. The following terminology for the outcome of insertional mutagenesis has been proposed ***(10)***: “knockout” for an insertion within the gene or the promoter that results in a null allele; “knockdown” for insertions in the promoter or 3' UTR that result in reduced expression; and “knockon” for lines in which increased expression of the gene results from the activity of a strong promoter or enhancer in the insertional mutagen (activation tagging).
11. To assess the presence of the insertion, PCR assays specific for both wild-type and mutant alleles must be developed. Presence of a T-DNA or transposon insertion is detected by PCR based on gene-specific and insertion-specific primers. At this stage of the analysis, the border of the insertion used to generate the database sequence must be confirmed. The wild-type allele is identified using gene-specific primers flanking the insertion site. In most cases, the flanking region of only one border of the insertion has been sequenced and deposited. Typically, the primers used for sequencing the insertion/genome junction can be used as insertion-specific primers. The design of gene-specific primers should consider a position relative to the insertion suitable for efficient PCR (250 to 750 bp), as well as a base composition and length of the primer facilitating the annealing for both gene- and insertion-specific primers. For some populations, direct links to primer design programs are available (for instance, SIGnAL iSect Primer Design; http://signal.salk.edu/isectprimers.html). Insertion-specific primers for the SALK and GABI-KAT population are listed in **Subheading 2.**
12. Insertion events can lead to chromosomal rearrangements. During the integration process, parts of the insertion can be lost and the DNA fragment of concern can lack one of its original borders, or it consists of concatamers of multiple insertional sequences. It is also possible that

regions of the flanking genomic sequence are rearranged or deleted. When the second border cannot be amplified by PCR, it is recommended to test whether the two borders are identical (such as in the case of a head-to-head type of concatamer); to employ insertion-specific primers that anneal more distantly to the border of the insertional mutagen (in the case of partial deletions of the T-DNA or transposon); or to use gene-specific primers that anneal more distantly from the putative insertion site (in the case of genomic deletions).

13. Before analyzing the expression of the gene of interest in the homozygous mutant plant, the expression pattern in the wild-type should be determined. For optimal results the tissue with highest expression in wild-type should be chosen. The protocol given should work for all tissues; however, for expression analysis in siliques, it might be necessary to perform an additional $LiCl_2$ precipitation.
14. The position of RT-PCR primers depends on the position of the T-DNA insertion within the gene. In general, PCR primers flanking the T-DNA insertion should be used and the RT-PCR carried out under saturating conditions. If the T-DNA insertion is in the 5' or 3'-UTR, or very close to the start or stop codon, both PCR primers should be positioned 3' or 5', respectively, of the T-DNA insertion. Because in these cases residual transcript accumulation can be expected, PCR should be performed under nonsaturating conditions (i.e., fewer cycles).
15. Upon transposon insertion, a short target sequence of specific length is duplicated (*En/Spm*: 3 bp, *Ac*: 8 bp, and *Mutator*: 9 bp). Upon excision, additional rearrangements at the original insertion site can occur, so that a spectrum of different footprints for a particular insertion can be obtained, ranging from deletions to the addition of several nucleotides. This particular feature of transposons can be used to generate a series of mutant alleles starting from the original insertion allele. Moreover, because some transposons tend to insert into physically linked sites, such "starter lines" can be used to saturate a genomic region with insertions (the "launch pad" approach). The same feature, however, can be also disadvantageous if the population is built on an autonomous transposon: in such lines, germinal and somatic revertants can occur that hamper the physiological analysis of the concerned knockout phenotype. In such cases it is favorable to stabilize the mutation by isolating footprints that result in a stable frameshift mutation. Suitable footprints are possible only when the original transposon insertion is located in an exon or nearby intron–exon junctions.
16. A footprint present in two different leaves is indicative of a large somatic event that likely also affects the reproductive organs and is transmitted to the next generation.

References

1. Hanin, M. and Paszkowski, J. (2003) Plant genome modification by homologous recombination. *Curr. Opin. Plant Biol.* **6,** 157–162.
2. Puchta, H. (2002) Gene replacement by homologous recombination in plants. *Plant Mol. Biol.* **48,** 173–182.
3. Britt, A. B. and May, G. D. (2003) Re-engineering plant gene targeting. *Trends Plant Sci.* **8,** 90–95.
4. Li, X. and Zhang, Y. (2005) Novel reverse genetics tools in plant functional genomics, in *Plant Functional Genomics* (Leister, D., ed.), The Haworth Press Inc., Binghampton, pp. 37–53.
5. Pauls, P. K., Kunert, K., Huttner, E., and Grandbastien, M. A. (1994) Expression of the tobacco Tnt1 retrotransposon promoter in heterologous species. *Plant Mol. Biol.* **26,** 393–402.
6. Hirochika, H., Okamoto, H., and Kakutani, T. (2000) Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* **12,** 357–369.
7. Hirochika, H. (1997) Retrotransposons of rice: their regulation and use for genome analysis. *Plant Mol. Biol.* **35,** 231–240.
8. Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., and Kanda, M. (1996) Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. USA* **93,** 7783–7788.
9. Wisman, E., Hartmann, U., Sagasser, M., et al. (1998) Knock-out mutants from an *En-1* mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proc. Natl. Acad. Sci. USA* **95,** 12,432–12,437.
10. Krysan, P. J., Young, J. C., and Sussman, M. R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* **11,** 2283–2290.
11. Steiner-Lange, S., Gremse, M., Kuckenberg, M., et al. (2001) Efficient identification of *Arabidopsis* knock-out mutants using DNA-arrays of transposon flanking sequences. *Plant Biol.* **3,** 391–397.

12. Sessions, A., Burke, E., Presting, G., et al. (2002) A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell* **14,** 2985–2994.
13. Alonso, J. M., Stepanova, A. N., Leisse, T. J., et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301,** 653–657.
14. Sessions, A. (2005) T-DNA mutagenesis: from tagging to insertion sequence databases, in *Plant Functional Genomics* (Leister, D., ed.), The Haworth Press Inc., Binghampton, pp. 23–35.
15. Zhang, Q., Arbuckle, J., and Wessler, S. R. (2000) Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize. *Proc. Natl. Acad. Sci. USA* **97,** 1160–1165.
16. Cresse, A. D., Hulbert, S. H., Brown, W. E., Lucas, J. R., and Bennetzen, J. L. (1995) *Mu1*-related transposable elements of maize preferentially insert into low copy number DNA. *Genetics* **140,** 315–324.
17. Forsthoefel, N. R., Wu, Y. W., Schulz, B., Bennett, M. J., and Feldmann, K. A. (1992) T-DNA insertion mutagenesis in *Arabidopsis*—prospects and perspectives. *Aust. J. Plant Physiol.* **19,** 353–366.
18. Tissier, A. F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M. A., Murphy, G. and Jones, J. D. (1999) Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics. *Plant Cell* **11,** 1841–1852.
19. Kuromori, T., Hirayama, T., Kiyosue, Y., et al. (2004) A collection of 11800 single-copy *Ds* transposon insertion lines in *Arabidopsis*. *Plant J.* **37,** 897–905.
20. Ito, T., Motohashi, R., Kuromori, T., et al. (2002) A new resource of locally transposed *Dissociation* elements for screening gene-knockout lines in silico on the *Arabidopsis* genome. *Plant Physiol.* **129,** 1695–1699.
21. Strizhov, N., Li, Y., Rosso, M. G., Viehoever, P., Dekker, K. A., and Weisshaar, B. (2003) High-throughput generation of sequence indexes from T-DNA mutagenized *Arabidopsis thaliana* lines. *Biotechniques* **35,** 1164–1168.
22. Li, Y., Rosso, M. G., Strizhov, N., Viehoever, P., and Weisshaar, B. (2003) GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in *Arabidopsis thaliana*. *Bioinformatics* **19,** 1441–1442.
23. Samson, F., Brunaud, V., Balzergue, S., et al. (2002) FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucl. Acids Res.* **30,** 94–97.
24. Brunaud, V., Balzergue, S., Dubreucq, B., et al. (2002) T-DNA integration into the *Arabidopsis* genome depends on sequences of pre-insertion sites. *EMBO Rep.* **3,** 1152–1157.
25. Weigel, D., Ahn, J. H., Blazquez, M. A., et al. (2000) Activation tagging in *Arabidopsis*. *Plant Physiol.* **122,** 1003–1013.
26. Goddijn, O. J., Lindsey, K., van der Lee, F. M., Klap, J. C., and Sijmons, P. C. (1993) Differential gene expression in nematode-induced feeding structures of transgenic plants harbouring promoter–gusA fusion constructs. *Plant J.* **4,** 863–873.
27. Campisi, L., Yang, Y., Yi, Y., et al. (1999) Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence. *Plant J.* **17,** 699–707.
28. LeClere, S. and Bartel, B. (2001) A library of *Arabidopsis* 35S-cDNA lines for identifying novel mutants. *Plant Mol. Biol.* **46,** 695–703.
29. Sambrook, J., Fritsch, E. F., and Maniatis, T. (eds.) (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
30. Caryl, A. P., Jones, G. H., and Franklin, F. C. (2003) Dissecting plant meiosis using *Arabidopsis thaliana* mutants. *J. Exp. Bot.* **54,** 25–38.
31. Lim, P. O., Woo, H. R., and Nam, H. G. (2003) Molecular genetics of leaf senescence in *Arabidopsis*. *Trends Plant Sci.* **8,** 272–278.
32. Leister, D. and Schneider, A. (2003) From genes to photosynthesis in *Arabidopsis thaliana*. *Int. Rev. Cytol.* **228,** 31–83.
33. Chory, J., Peto, C. A., Ashbaugh, M., Saganich, R., Pratt, L., and Ausubel, F. (1989) Different roles for phytochrome in etiolated and green plants deduced from characterization of *Arabidopsis thaliana* mutants. *Plant Cell* **1,** 867–880.
34. Schwechheimer, C. and Deng, X. W. (2000) The COP/DET/FUS proteins-regulators of eukaryotic growth and development. *Semin. Cell Dev. Biol.* **11,** 495–503.
35. Glazebrook, J. (2001) Genes controlling expression of defense responses in *Arabidopsis*—2001 status. *Curr. Opin. Plant Biol.* 4, 301–308.
36. Lohmann, J. U. and Weigel, D. (2002) Building beauty: the genetic control of floral patterning. *Dev. Cell* **2,** 135–142.

37. Schaller, H. (2003) The role of sterols in plant growth and development. *Prog. Lipid Res.* **42,** 163–175.
38. Liu, Y. G., Mitsukawa, N., Oosumi, T., and Whittier, R. F. (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8,** 457–463.
39. Yephremov, A. and Saedler, H. (2000) Technical advance: display and isolation of transposon-flanking sequences starting from genomic DNA or RNA. *Plant J.* **21,** 495–505.
40. Frey, M., Stettner, C., and Gierl, A. (1998) A general method for gene isolation in tagging approaches: Amplification of insertion mutagenised sites (AIMS). *Plant J.* **13,** 717–721.
41. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815.

14

PCR-Based Screening for Insertional Mutants

Anna N. Stepanova and Jose M. Alonso

Summary

In *Arabidopsis*, where homologous recombination is highly inefficient, insertional mutagenesis has provided a great alternative means for identifying loss-of-function mutants in genes of interest. Many collections of T-DNA and transposon-tagged plants have been generated and several have been made available to the research community. With recent sequencing of the insertion sites and the creation of public databases, identification of knockouts in most of the genes of interest has become a matter of a few minutes. However, because of the random selection of the insertion lines to be sequenced, knockouts of the remaining several thousand genes may not be easily found by the mere continuation of these types of approaches. Therefore, alternative methods such as polymerase chain reaction (PCR)-based screening of pooled genomic DNA from mutant lines using a combination of tag- and gene-specific primers, once again, are becoming valuable for obtaining loss-of-function alleles in the genes "missed" by the sequencing projects. In this chapter, we provide a detailed description of the design and implementation of a PCR-based screen for insertional knockouts using T-DNA-mutagenized lines as an example.

Key Words: PCR-based screening; knockout mutant; T-DNA insertion; loss of function; multi-dimensional pooling.

1. Introduction

Analysis of knockout mutants is believed to be one of the most powerful approaches to studying gene function ***(1)***. In the absence of an efficient homologous recombination system to specifically and systematically generate knockouts for all *Arabidopsis* genes, generation of large collections of more or less random insertions has proven to be a very useful alternative ***(2)***. The facile and highly efficient transformation of *Arabidopsis* with *Agrobacterium* harboring a Ti plasmid, along with the semirandom translocation of highly active transposable elements in the *Arabidopsis* genome, represent two main approaches for high-throughput insertional mutagenesis.

To screen these collections for a specific gene knockout, two different strategies are commonly employed: a gene-specific PCR-based approach ***(3–7)***, and systematic sequencing of insertion sites. In the past few years, several collections of sequence-indexed insertions have become available ***(6,8–12)***. With these remarkable tools at hand, identification of knockouts in the gene of interest has become as simple as searching a database and ordering the seeds for the corresponding mutant from a stock center. Unfortunately, there are several thousand genes for which knockouts are not available through these databases. Due to the random nature of insertional mutageneses, identification of knockouts for the remaining genes by continuing systematic sequencing would be highly inefficient. Nevertheless, only a fraction of the insertions present in these collections have been sequenced, and, therefore, the same collections can be utilized for more directed searches.

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

In this chapter, we present the methodology of screening a T-DNA insertional mutant collection using a combination of a multidimensional pooling strategy with gene-specific polymerase chain reactions (PCRs). Although there are some particularities about using T-DNA-versus transposon-mutagenized plants (see below), the logic described herein can be adapted to either type of collections. Before the methodology is presented, two important points will be discussed that need to be considered prior to starting the pooling and screening processes: the type of mutagenesis (transposon versus T-DNA) and the initial propagation of the generated mutants.

As indicated above, both transposons and T-DNA have been widely used to obtain insertional mutants ***(2)***. The main differences between these two systems are as follows: (1) Distribution of the transposons in the genome is more biased than that of the T-DNAs. The bias is due to the high levels of local transposition, as well as a strong preference for specific genomic context in the case of transposons (telomeres, parts of the genes, etc). Although, in general, a bias is considered disadvantageous, it may be desirable in some instances (for example, to identify knockouts in genes physically linked to a known insertion) ***(11,12)***. (2) If plants with active transposons are used, experimental approaches to eliminate potential false positives due to somatic transpositions should be considered ***(13)***. (3) The structure of an insertion is much more complex in the case of the T-DNAs, involving multiple copies of the T-DNA, often accompanied by deletions, rearrangements, and/or duplications of genomic regions ***(14,15)***. This could result in false positives during the screening process unless special precautions are taken (*see* **Notes 11**, **13**, and **14**).

Once the type of mutagenesis has been chosen, the next critical decision is whether to grow the primary mutants (T1 plants) in pools or individually. Although the propagation in pools might seem more efficient, it implies that down the road additional steps in the screening process will be unavoidable, and this extra work will have to be repeated over and over for each mutant of interest (*see* **Note 15**). The strategy described herein not only reduces the amount of labor required for the identification of homozygous knockouts, but also prevents any type of bias in the identification of mutations in plants with low fertility often found among the T1 plants and eliminates the number of false positives after just one round of screening.

2. Materials

2.1. Pooling and Arabidopsis DNA Preparation

1. Aluminum custom-made spoons for seeds (**Note 1**).
2. Disposable 1.5-mL microcentrifuge and 50-mL conical tubes.
3. 250-mL flask.
4. 5-L glass desiccator chamber.
5. Fume hood.
6. Laminar hood.
7. Bleach.
8. HCl.
9. Parafilm.
10. 150 × 15 mm Petri dishes.
11. MS agar plates: 4.3 g/L Murashige & Skoog salt mixture (Caisson Laboratories, Inc.), 10 g/L sucrose, adjust pH 6.0 with 1 *M* KOH, 15 g/L agar.
12. 4°C dark incubator/cold room.
13. 22°C dark incubator.
14. Liquid nitrogen.
15. Aluminum foil.
16. Fine-pointed forceps for tissue harvesting.
17. Mortar and pestle.
18. CTAB buffer: 1.4 *M* NaCl, 20 m*M* EDTA, 100 m*M* Tris-HCl, pH 8.0, 3% (w/v) cetyl trimethylammonium bromide (CTAB), 1% (v/v) β-mercaptoethanol.

19. 250-mL and 38-mL Nalgene centrifuge tubes.
20. J2-HC centrifuge (Beckman Coulter) or equivalent.
21. JA14 and JA20 rotors (Beckman Coulter) or equivalents.
22. 24:1 chloroform:isoamyl alcohol mix.
23. 2-propanol.
24. 100% and 70% ethanol.
25. 10 mg/ml RNase stock (DNase-free).
26. Gel electrophoresis setup.

2.2. PCR Screening

1. Gene-specific and T-DNA-specific oligonucleotides.
2. General PCR reagents: dNTPs, 10X PCR buffer, *Taq* polymerase.
3. Thermocycler.
4. Gel electrophoresis setup.
5. 10 mg/mL ethidium bromide.
6. Hybond N+ (Amersham Pharmacia) membrane or equivalent.
7. Gel-transfer setup.
8. 0.4 *N* NaOH.
9. Radioactive isotope and a labeling kit.
10. Hybridization oven.

3. Methods

3.1. Construction of DNA Pools

Mutagenesis, selection, propagation, and harvesting of individual mutant plants are performed using standard protocols *(16)*. Seeds of each mutant line are kept separately. A three-dimensional pooling of mutant seeds is performed prior to DNA preparation, presenting several important benefits: the number of required DNA extractions is significantly reduced, representation of each mutant line in the pools is equalized, and the number of screening steps to identify a knockout plant is lowered.

1. Combine equal amount of seeds from 10 different original T1 transformants (about 200 seeds per line) in a 1.5-mL microcentrifuge tube to generate each one of the primary pools (P) (**Fig. 1A** and **Note 1**).
2. Generate row, column, and plate superpools (RS, CS, and PS) by combining equal amounts (300 to 400 seeds) of 100 corresponding primary pools (**Fig. 1B** and **Note 2**). In the super pools, each original line is represented on average by 30 to 40 seeds.
3. Generate number superpools (NS) by combining equal amount of seeds from the original mutant lines whose numbers ended with the same digit (**Note 3**).
4. Transfer about 30,000 to 40,000 seeds (0.6 to 0.8 g) of each one of the superpools to individual 50-mL conical plastic tubes for sterilization.
5. Place the uncovered tubes with the seeds into a 5-L glass desiccator chamber. Perform **steps 6** to **8** in the fume hood.
6. Place a 250-mL glass flask containing 100 mL of bleach into the dessicator chamber.
7. Add 4 mL of concentrated HCl to the bleach and immediately close and seal the dessicator with parafilm (**Note 4**).
8. After ~1 h, open the desiccator in the fume hood and let the chlorine vapor dissipate for several minutes (**Note 5**).
9. Transfer the tubes with the seeds to a laminar hood and keep them uncovered for at least 12 h to eliminate any residual chlorine gases.
10. Spread the dry seeds on MS agar plates (~3000 seeds per 15 × 150 mm Petri dish) without antibiotic/herbicide supplementation (**Note 6**).
11. Stratify the seeds for 3 d in the dark at 4ºC.
12. Grow the plants for 3 to 4 d in the dark at 22ºC.

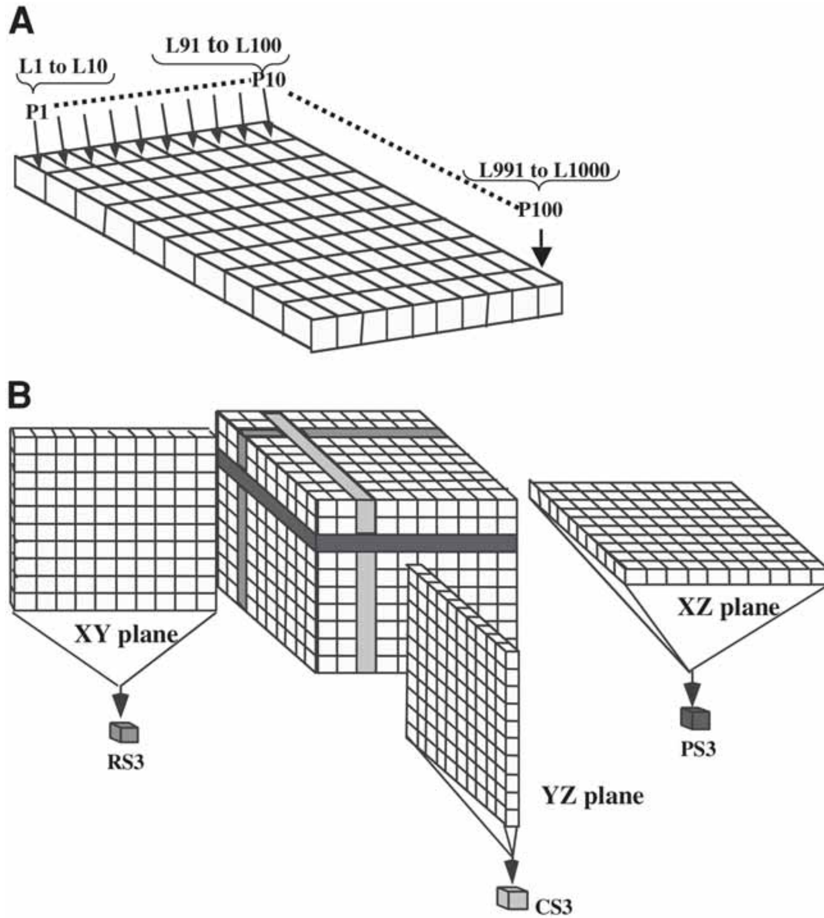


Fig. 1. Schematic representation of the pooling strategy. **(A)** Generation of the primary (P) pools. Each one of the cubes in the plate represents a pool of 10 mutant lines (L). **(B)** Generation of the superpools. Aliquots of seeds of 100 P pools from each XY ("row") plane are combined to generate the corresponding RS superpools. Similarly 100 P pools from the XZ ("plate") or YZ ("column") planes are combined to generate the PS or CS superpools, respectively.

13. Harvest and combine the seedlings from all of the plates from the same super pool and immediately freeze them in liquid nitrogen (**Note 7**).
14. Grind each sample to a fine powder using a mortar and a pestle in the presence of liquid nitrogen.
15. Homogenize the tissue in 125 mL of CTAB extraction buffer and transfer to a 250-mL Nalgene centrifuge tube.
16. Incubate at 65°C for 30 min with occasional shaking.
17. Cool to room temperature (about 20 min), then add 125 mL of chloroform:isoamyl alcohol (24:1), and mix the sample by inverting the tube several times.
18. Centrifuge at 4000*g* for 15 min.
19. Transfer the aqueous upper phase to a new 250-mL Nalgene centrifuge tube, add 1 vol of isopropanol, and mix by inverting the tube several times.
20. Centrifuge at 4000*g* for 15 min.
21. Pour off and discard supernatant, and air-dry the pellet for 10 min at room temperature (**Note 8**).
22. Resuspend the pellet in 6 mL of sterile DI water containing 10 mg/mL of RNase (DNase-free).
23. Transfer the DNA to a 38-mL Nalgene centrifuge tube, add 6 mL of chloroform:isoamyl alcohol (24:1), and mix the sample by inverting the tube several times.

24. Centrifuge at 10,000*g* for 5 min and transfer upper phase to a new 38-mL Nalgene centrifuge tube.
25. Precipitate DNA with 15 mL (2.5 vol) of 100% ethanol and 0.6ml (0.1 vol) of 5 *M* sodium acetate, pH 5.2. Mix by inverting the tube several times.
26. Centrifuge for 15 min at 10,000*g* at 4ºC.
27. Pour off and discard the supernatant, and wash the pellet with 10 mL of 70% ethanol.
28. Centrifuge for 15 min at 10,000*g* at 4ºC.
29. Pour off and discard the supernatant, air-dry the pellet at room temperature (**Note 8**), and resuspend in 2 mL of sterile DI water.
30. Check the quality of the genomic DNA by running 1 µL of each sample on a 1% agarose gel. The amount of DNA can be estimated by running in parallel an aliquot of genomic DNA of known concentration (**Note 9**).

3.2. PCR-Based Screening

Screening is done by PCR-amplifying T-DNA/genomic DNA junction fragments using pooled mutant DNA as templates and the T-DNA/gene-specific primers, separating PCR products by gel electrophoresis, and then performing conventional Southern blotting and hybridization of the membranes with gene-specific probes.

1. Design a set of primers specific to the T-DNA borders and a set of primers specific to the gene of interest (**Note 10**).
2. Test each one of the DNA superpools with the following four primer combinations: gene forward (GF) + T-DNA left border (LB), gene reverse (GR) + T-DNA right border (RB), GF + RB, and GR + LB (**Note 11**)
3. Place the plates into the thermocycler and run the PCR amplification program (**Note 11**).
4. Analyze the PCR products on a 1% agarose gel, stain with EtBr, and photograph the gel.
5. Transfer the DNA products to a nylon membrane in 0.4 *N* NaOH for 12 h to overnight. After transfer, mark the wells with a pencil (**Note 12**).
6. Air-dry the membrane for 1 h.
7. Hybridize the membrane with a radioactively labeled gene-specific probe overnight.
8. Wash the membrane and expose it to a film or a PhosphoImager cassette for 4 h to overnight (**Note 13**).
9. Score the results and identify the corresponding knockout insertion line (**Note 14**).
10. Identify homozygous mutant plants (**Note 15**).
11. Confirm and analyze knockouts (**Note 16**).

4. Notes

1. Each primary (P) pool corresponds to 10 individual T1 plants (i.e., pools P1 to P1000 represent the first 10,000 T1 plants). To generate the pool P1, for example, combine approx 200 seeds from each one of the lines 1 to 10. Similarly, pool P2 will be made using lines 11 to 20, and so on. These primary pools are arranged in a 10 × 10 × 10 cubic array (**Fig. 1A**). For pooling purposes, to consistently pick approx 200 seeds from each T1 plant, a “microspoon” can be custom-made. A solid aluminum bar (~20 cm long and ~4 mm in diameter) with a concavity (~2 mm in diameter and ~1.5 mm deep) drilled on either end makes a perfect “pooling tool.” By changing the size of the drilled holes, microspoons of different capacities can be created.
2. To generate the superpools, equal amounts (300 to 400 seeds) of each one of the 100 P pools from the first cube’s XY “row” plane (pools P1 to P10, plus P101 to P110, … plus P901 to P910) should be pooled together to generate the RS1 superpool. Similarly, the RS2 superpool will be obtained by combining seeds from the pools P11 to P20 plus P111 to P120, etc. The YZ “column” superpools can be generated in a comparable way, where CS1 will contain seeds for the pools P1, P11, P21, … P91; CS2 will include pools P2, P12, P22, … P92, etc. Finally, the XZ “plate” superpools will be obtained by mixing seeds from P1 to P100 (plate superpool PS1), P101 to P200 (PS2), … P901 to 1000 (PS10). Note that each one of the superpools corresponds to 100 P pools from the same row (XY), column (YZ), or plate (XZ) plane (**Fig. 1B**), or to 1000 original T1 plants, with each one of the T1 plants being represented in three different superpools, one RS (XY), one CS (YZ), and one PS (XZ).

3. To generate the fourth type of superpools (referred to as the "number" superpool [NS]), seeds from the original T1 plants (rather than from the primary pools) are used. To obtain the NS1, about 200 seeds from each one of the T1 lines 1, 11, 21, 31, 41, … 9991 should be combined. Similarly, the super pool NS2 will contain seeds from lines 2, 12, 22, 32, … 9992, etc.
4. A long thin-neck flask instead of a beaker should be used to prevent small drops of HCl/bleach produced during the reaction of these two compounds jumping onto the seeds. The chlorine fumes produced are highly toxic and corrosive and should be handled with extreme care.
5. Although 1 h usually works well, each batch of seeds may respond differently to this treatment. Some batches of seeds may be very sensitive to this treatment and may die. It is therefore recommended to test the effect of the treatment on each new batch using a small aliquot before sterilizing all of the seeds.
6. This will ensure that lines in which the resistance gene has undergone silencing will still be represented in the DNA pools. This silencing phenomenon is very common in the progeny of T-DNA-transformed plants.
7. Whole seedlings are harvested from the agar plates using forceps. Fifty to 100 seedlings can be pooled at a time and then frozen by being placed onto aluminum foil cups floating on the surface of liquid nitrogen. Once all of the seedlings from one pool have been harvested, the tissue can be stored wrapped in aluminum foil at –80°C until needed.
8. To eliminate any residual ethanol from the pellets, the tubes can be inverted and left upside down at a 45° angle for 10 to 15 min or until no ethanol smell can be detected.
9. Typically, these approaches are sufficient to estimate the concentration and quality of the genomic DNA. The concentration and quality of the DNA to be used as a template in a PCR reaction are not critical, so it is not necessary to use more accurate approaches such as measurement of OD at λ 260/280 nm and/or digestion with restriction enzymes followed by Southern analysis using a radiolabeled probe.
10. The quality of the primers is the single most important factor in the success of the screen. The T-DNA border-specific primers should be designed using the following criteria. A melting temperature of 65°C or higher is desirable. Primers should be complementary to the sequence about 150 bp internal to the border, ensuring that insertion accompanied by a small deletion of the border could still be recovered. Low sequence complexity primers should be avoided. Primers ending with G or C bases in the 3'-end are recommended. Screens using a gene-specific primer in a combination with a left border (LB) primer typically identify more insertions than those with the gene-specific primer plus the right border (RB) primer. However, not all insertions can be identified using only the LB, so both LB and RB should be used for exhaustive screening.

    Two different strategies can be used for designing gene-specific primers, dependent on the size of the gene of interest. For smaller genes (<2 kb), two primers facing each other and complementary to the 5'- and 3'-ends of the gene can be used for screening (**Fig. 2A**). For medium-size genes (2 to 5 kb), a pair of primers centered in the middle of the gene and positioned ~500 bp apart is used. This same approach can be used for bigger genes (>5 kb) using several primer pairs, each pair covering a region of about 4 to 5 kb (**Fig. 2A**). The melting temperature of the gene-specific primers should be comparable to that of the T-DNA primers. The gene-specific primers should be tested using wild-type genomic DNA as a template and the actual thermocycler/cycling protocol (**Note 11**) to be used in T-DNA screening.
11. The number of PCRs can be reduced by combining several primer combinations in one PCR reaction. We do not, however, recommend, combining LB with RB in the same PCR tube, as this could result in product amplification in lines with tandem T-DNA repeats. When more than one pair of primers is needed to screen one gene (for genes bigger than 5 kb), all forward primers could be used together in the same PCR reaction with one of the border-specific primers; the same could be done with all reverse primers. The typical PCR reaction will contain 1 μL of 10X PCR buffer, 5 U of *Taq* polymerase, 0.25 μL of 2 m*M* dNTPs, 0.25 μL of each one of the primers (20 μ*M*) and 1 μL of genomic DNA (10–20 ng) in a 10-μL reaction. The following amplification cycle can be used: 94°C for 1 min, 56°C for 1 min, and 72°C for 3 min (repeat this cycle 30–40 times).

Fig. 2. Schematic representation of the PCR-screening strategy. **(A)** Primer design for small (<2 kb) and large (>5 kb) genes. Gray arrows symbolize genes. Black arrows indicate the recommended positioning and orientation of the gene-specific primers. **(B)** Simulation of an autoradiograph resulting from a typical PCR screen of 4 × 10 DNA superpools (10,000 original lines). Crossed bands represent false positives (i.e., PCR products that are not present once in each of the four kinds of the superpools). The four bands labeled as "1" represent the first true insertion, whereas the four products labeled as "2" indicate the presence of a second positive mutant line in the 10,000 lines screened. The de-convolution of the positive pool numbers reveals that the lines 4075 (positive 1) and 7728 (positive 2) contain insertions in or near the gene of interest (see text for details).

12. It is critical in this procedure to determine the number of the superpool in which a PCR product was obtained (positive superpool). Marking the position of the wells on the membrane helps with the tracking process.
13. A small drop (1 µL) of ink spiked with P32 can be used to mark the wells and to track the positive pools in the autoradiography. Although in many cases the positive bands in the Southern blot correspond to bands detectable in the EtBr-stained gels, this is not always the case, as autoradiography is by far more sensitive. On the other hand, not all PCR products visible on the EtBr staining are due to instertions in the gene of interest (but rather due to annealing of one or more primers in other parts of the genome), and therefore Southern blotting followed by autoradiography is strongly recommended.
14. A positive will be represented by four hybridizing bands of the same size on the Southern blot (**Fig. 2B**). Each one of the four hybridizing bands will correspond to the same positive line in each of the four different types of superpools (i.e., one band in the "plate" superpool, one in the "row" superpool, one in the "column" superpool, and one in the "number" super pool). This high degree of redundancy in the screen not only facilitates the identification of the individual line (primary transformant) with the insertion, but also eliminates false positives. In some cases, only three bands of the same size will be observed in three of the four different types of the superpools, making it necessary to rescreen the 10 superpools of the a particular type (RS, CS, PS, or NS) that failed to reveal the positive. Using this system, in which each 10,000 primary transformants are screened in a set of 40 PCRs, it is not uncommon to identify more than one insertional mutant. In this case, the positive pools for each one of the mutants can typically be distinguished by the size of the bands (*see* **Fig. 2B**). The individual line(s) with the

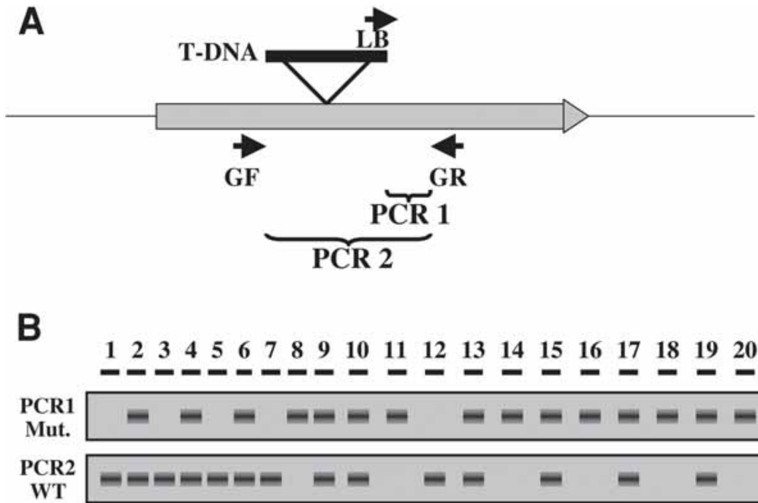


Fig. 3. Identification of homozygous mutant plants. **(A)** Schematic representation of a T-DNA insertion. The position of the T-DNA insertion is shown as a black box above the gene that is symbolized by the gray arrow. The primers LB and GR are used to detect the presence of the insertion in the progeny of the original mutant (PCR1). Similarly, the primers GF and GR are used to determine the presence of the wild-type version of the gene (PCR2). **(B)** Simulation of a gel picture resulting from the PCR analysis of 20 T2 plants. Presence of a PCR product in PCR1, but not in PCR2, indicates that the corresponding plant is a homozygous mutant (plants 8, 11, 14, 16, 18, 20). Amplification of a fragment in PCR2, but not in PCR1, implies that the corresponding individual is homozygous wild-type (plants 1, 3, 5, 7, 12). Finally, amplification of products in both PCR1 and PCR2 suggests that the corresponding plant is heterozygous for the insertion (plants 2, 4, 6, 9, 10, 13, 15, 17, 19).

insertion can be easily determined by the numbers of each of the four positive super pools. In the example 2 of **Fig. 2B**, the positive in the PS8 indicates that the line with the insertion should be among L7001 to L8000, the positive in RS8 narrows this down to the interval between L7701 to L7800, the positive in CS3 points to a line between L7721 to L7730, and finally, the positive in NS8 indicates that the line with the insertion is L7728.

15. To identify a line homozygous for the insertion of interest, plant to soil about 20 T2 or T3 seeds and grow mutants for 3 to 4 wk. (If the original mutants have been grown in pools, the number of plants to be genotyped from each pool of interest should be increased proportionally to the size of the pool, i.e. ~200 plants need to be tested for each pool of 10 original transformants). Clip a leaf of each individual plant, extract DNA, and perform a PCR (PCR1) using the same combination of gene-specific and T-DNA border-specific primers that identified positives on the original Southern. In parallel, set up a second PCR (PCR2) on each of the 20 DNA samples to amplify the genomic DNA fragment using two gene-specific primers that anneal 5' and 3' to the expected T-DNA insertion site and are facing the insertion (**Fig. 3**). PCR1 (amplification of a product) allows to identify plants with the insertion, whereas PCR2 enables one to distinguish between homozygous mutant (no product), heterozygous mutant (product), and wild type (product). To determine the exact position of the T-DNA insertion, cut the PCR1 band from the gel, purify DNA from it using conventional methods, and sequence the insertion site using the T-DNA-specific primer.

    In recent literature, there have been several cases described in which the T-DNA was found to be integrated in a duplicated copy of a gene ***(14)***. In such cases, no homozygous mutants can be identified using the PCR1/PCR2 approach above, because the original wild-type copy of the gene is always present and will always give a product in the PCR2. To differentiate between

this phenomenon and a recessive lethal mutation, several plants apparently heterozygous for the insertion need to be propagated and allowed to self-pollinate. In the case of recessive lethality, each putative heterozygote will produce one-third wild-type and two-thirds heterozygous plants. Conversely, in the case of a T-DNA-associated duplication, upon selfing, some of the apparent "heterozygotes" (which are, in fact, homozygous both for the wild-type and the disrupted copy of the gene of interest, and hence behave like heterozygotes in the genotyping process) will produce 100% of such false "heterozyogotes." Unfortunately, such lines do not represent useful mutants.

16. Once the homozygous knockout plants have been identified and before making any conclusions about the phenotypic consequences of the mutation of interest, two important questions should be addressed. First, is the mutant a transcriptional null? And second, is the observed phenotype (if any) caused by the insertion in the gene of interest rather than by some unrelated mutation? To answer the first question, the levels of expression of the tagged gene should be examined, which is especially important for the cases where insertions are located in an untranslated region of the gene (*see* Chapter 13 in this book).

    Typically, the identification of the mutant is followed by its careful phenotypic analysis. The presence of a specific phenotype in the homozygous knockout plants does not necessary imply a causal relationship between the gene and the phenotype. Additional insertions or unrelated mutations may be present in these plants and be responsible for the phenotype. The identification of two or more independent mutant alleles and/or complementation of the mutant phenotype with the genomic fragments containing the gene of interest would provide the final confirmation. Strong linkage between the insertion and the phenotype is also a good indication, although inconclusive, of the causal relation between the insertion and the phenotype. Either a T2 or an F2 population (from a cross between the mutant and the parental wild-type strain) can be used in these linkage experiments. The genotype of the T2/F2 plants can be scored as described in **Fig. 3**. A complete cosegregation between the phenotype and the homozygous insertion in 100 to 200 mutant plants is indicative of a strong genetic linkage between the phenotype and the insertion of interest. It should, however, be emphasized that these types of experiments do not provide definite answers, and additional alleles and/or complementation experiments are required to obtain an unequivocal evidence.

References

1. Carpenter, A. E. and Sabatini, D. M. (2004) Systematic genome-wide screens of gene function. *Nat. Rev. Genet.* **5,** 11–22.
2. Parinov, S. and Sundaresan, V. (2000) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. *Curr. Opin. Biotechnol.* **11,** 157–161.
3. Winkler, R. G., Frank, M. R., Galbraith, D. W., Feyereisen, R., and Feldmann, K. A. (1998) Systematic reverse genetics of transfer-DNA-tagged lines of *Arabidopsis*. Isolation of mutations in the cytochrome p450 gene superfamily. *Plant Physiol.* **118,** 743–750.
4. Galbiati, M., Moreno, M. A., Nadzan, G., Zourelidou, M., and Dellaporta, S. L. (2000) Large-scale T-DNA mutagenesis in *Arabidopsis* for functional genomic analysis. *Funct. Integr. Genomics* **1,** 25–34.
5. Sussman, M. R., Amasino, R. M., Young, J. C., Krysan, P. J., and Austin-Phillips, S. (2000) The *Arabidopsis* knockout facility at the University of Wisconsin-Madison. *Plant Physiol.* **124,** 1465–1467.
6. Alonso, J. M., Stepanova, A. N., Leisse, et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301,** 653–657.
7. Rios, G., Lossow, A., Hertel, B., et al. (2002) Rapid identification of *Arabidopsis* insertion mutants by non-radioactive detection of T-DNA tagged genes. *Plant J.* **32,** 243–253.
8. Sessions, A., Burke, E., Presting, G., et al. (2002) A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell* **14,** 2985–2994.
9. Rosso, M. G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003) An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol. Biol.* **53,** 247–259.
10. Szabados, L., Kovacs, I., Oberschall, A., et al. (2002) Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. *Plant J.* **32,** 233–242.
11. Kuromori, T., Hirayama, T., Kiyosue, Y., et al. (2004) A collection of 11 800 single-copy *DS* transposon insertion lines in *Arabidopsis*. *Plant J.* **37,** 897–905.

12. Parinov, S., Sevugan, M., De, Y., Yang, W. C., Kumaran, M., and Sundaresan, V. (1999) Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in *Arabidopsis*. *Plant Cell* **11,** 2263–2270.
13. Wisman, E., Hartmann, U., Sagasser, M., et al. (1998) Knock-out mutants from an En-1 mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proc. Natl. Acad. Sci. USA* **95,** 12,432–12,437.
14. Tax, F. E. and Vernon, D. M. (2001) T-DNA-associated duplication/translocations in *Arabidopsis*. Implications for mutant analysis and functional genomics. *Plant Physiol.* **126,** 1527–1538.
15. Nacry, P., Camilleri, C., Courtial, B., Caboche, M., and Bouchez, D. (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics* **149,** 641–650.
16. Alonso, J. M. and Stepanova, A. N. (2003) T-DNA mutagenesis in *Arabidopsis*. *Methods Mol. Biol.* **236,** 177–188.

15

Cytogenetic Analyses of *Arabidopsis*

Martin Lysak, Paul Fransz, and Ingo Schubert

Summary

For a long time, *Arabidopsis thaliana* was not a favorite of cytogeneticists because of its small chromosomes. Nevertheless, most cytogenetic approaches have become applicable to this model plant during the past decade. Recently, the very small genome, the low content, and strong clustering of repetitive sequences even turned out to be advantageous for the first establishment of chromosome painting in a euploid plant. Chromosome painting together with the high-resolution fluorescence *in situ* hybridization (FISH) mapping on *Arabidopsis* pachytene chromosomes has opened new perspectives for plant cytogenetics.

In this chapter, we describe standard cytogenetic methods including preparation of chromosomes and chromatin fibers, isolation of interphase nuclei, FISH and immunostaining, painting of *Arabidopsis* chromosomes, and comparative chromosome painting in other *Brassicaceae* species.

Key Words: *Arabidopsis*; cytogenetics; chromosome preparation; mitosis; meiosis; isolation of interphase nuclei; fluorescence *in situ* hybridization (FISH); chromosome painting; DNA-fiber FISH; immunostaining.

1. Introduction

The complete genomic sequence of *Arabidopsis thaliana* ***(1,2)***, the public availability of a BAC tiling path for each chromosome, and the steadily growing collections of mutants and transgenic lines make this species a model of still growing importance for plant research. The very small genome (~157 Mb) ***(3)*** is represented by five chromosome pairs with most of the repetitive sequences (~15% of the genome) ***(4)*** clustered within the pericentromeric heterochromatin and within the distal nucleolus organizers (NORs) on the short arms of chromosomes 2 and 4 (**Fig. 1**). These heterochromatic regions form conspicuous chromocenters in interphase nuclei (at maximum 14: 10 pericentromeres and 4 NORs; but because of fusion tendency, rarely more than 10), distinguishable by high chromatin density ***(5)*** and strong DNA fluorescence intensity after 4,6-diamidino-2-phenylindole (DAPI) staining (**Fig. 1**).

The *Arabidopsis* mitotic chromosomes are rather small (~1.5 µm) and therefore hardly accessible to detailed cytogenetic analysis. Nevertheless, chromosome C-banding ***(6)*** and localization of repetitive sequences by fluorescence *in situ* hybridization (FISH) ***(7,8)*** have been performed early. Chromosome counts in aneuploids and polyploids were conducted by conventional staining (e.g., ***9,10)*** and by flow-cytometric analysis ("flow-karyotyping") ***(11)***. Since the late 1990s, preparation of meiotic (pachytene) chromosomes became feasible ***(12,13)*** and allowed high-resolution physical mapping of DNA sequences by FISH (***14,15***; for review *see* ref. ***16***). The improvement was based on the advantage that *Arabidopsis* chromosomes are 25-fold longer during pachytene than during mitotic metaphase. Meiotic chromosome synapsis and recombination can be studied by FISH and immunostaining ***(17)***. A further improvement of

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

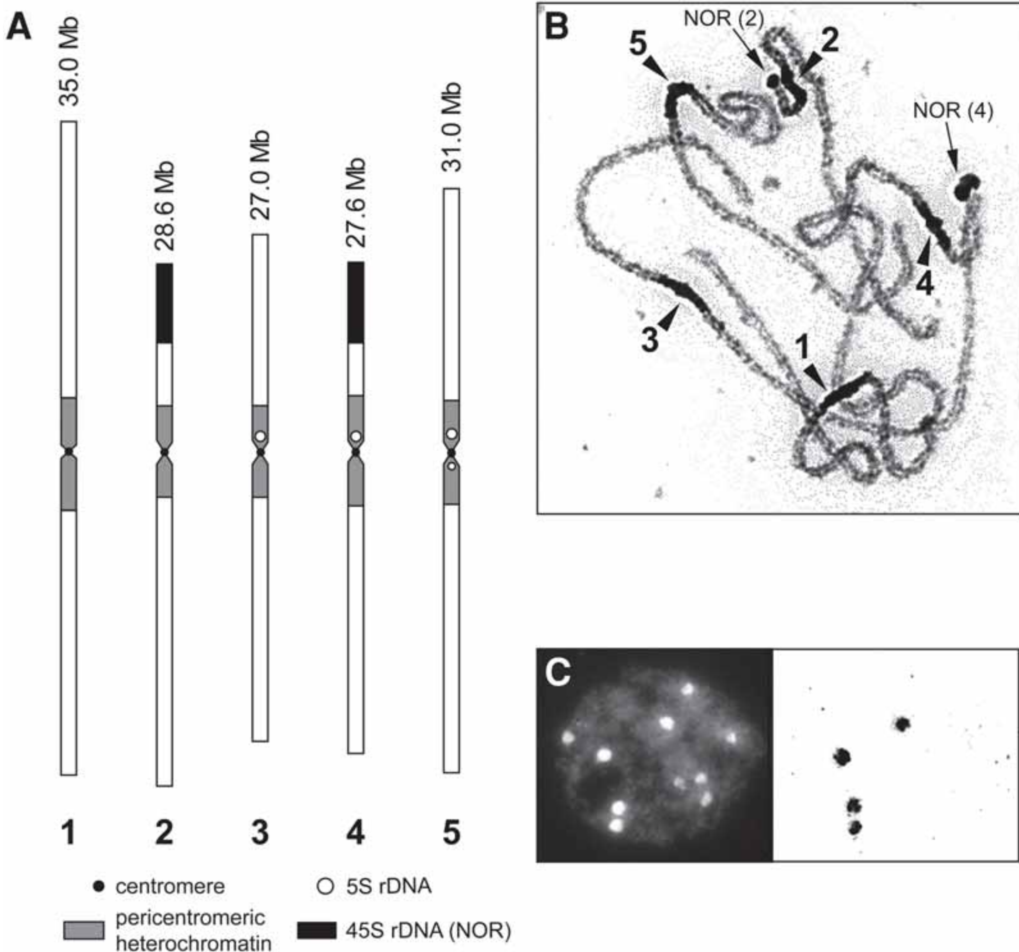


Fig. 1. Idiogram, pachytene chromosomes, and interphase nucleus of *Arabidopsis thaliana.* **(A)** Idiogram of ecotype Columbia. The chromosome size estimation is based on data of refs. *2,3,13*. **(B)** Reverse image of DAPI-stained chromosomes at pachytene stage (prophase I) isolated from a young anther of accession C24. Homologous chromosomes form five bivalents with heterochromatic centromeres (arrowheads) and terminal NORs on chromosomes 2 and 4. **(C)** Leaf nucleus of accession Col showing 10 DAPI-positive heterochromatic chromocenters (left); 4 chromocenters of the same nucleus revealed after FISH 45S rDNA as one component (right, reverse image).

the mapping resolution was accomplished by the FISH on extended chromatin fibers, which has been adapted to plants using *Arabidopsis* nuclei ***(18)***.

Recently, cytogenetic analysis of chromatin alterations in interphase nuclei became possible due to the availability of antisera against chromatin epitopes and the application of nuclei isolated from different organs. This approach allowed to trace the dynamics of epigenetic modifications, such as DNA methylation, methylation, and acetylation of histones, within distinct chromatin domains (addressed by FISH) along cell cycle and developmental stages ***(19–21)***.

With transgenically expressed and fluorescently (e.g., by green fluorescent protein [GFP]) labeled fusion proteins, the position and movement of such proteins could be monitored *in situ* and in vivo ***(22,23)***.

Furthermore, the low DNA content and the relatively simple organization of the *Arabidopsis* genome enabled for the first time to paint all chromosomes of a euploid plant species individually using pools of BAC contigs as chromosome-specific probes for FISH *(24,25)*. By chromosome painting the interphase organization of chromosome territories and their potential dynamics can be studied and chromosome rearrangements can be visualized. In addition, painting probes specifically designed for *Arabidopsis* chromosomes proved to be suitable to detect homologous chromosome regions within the pachytene karyotypes of several other Brassicaceae species and to reconstruct the evolutionary history of chromosome complements for differently related species of this family *(25)*.

Labeling of the entire chromosome complement by genomic *in situ* hybridization (GISH) is difficult with small genomes because usually these show only heterochromatic blocks labeled after GISH. However, under appropriate conditions GISH turns out to be applicable in interspecific hybrids of *Arabidopsis (26)*.

To study genotoxic impacts cytologically, the "comet-assay" with isolated nuclei has been adopted for *Arabidopsis (27)*. For detection of somatic recombination via sister chromatid exchanges, *Arabidopsis* chromosomes are too small. However, the involvement in rearrangements of NOR-bearing chromosomes could be demonstrated by FISH in anaphase nuclei of pistil cells from telomerase-deficient mutants *(28)*. Applying multicolor painting of all chromosomes simultaneously, reciprocal translocation between any chromosomes should become identifiable.

Thus, almost the entire spectrum of cytogenetic approaches is now applicable to this model species.

This chapter describes (1) the preparation of mitotic and meiotic chromosomes and of chromatin fibers, (2) the isolation of interphase nuclei from different organs, (3) FISH and immunostaining on meiotic chromosomes and interphase nuclei (**Fig. 2**), (4) FISH on extended chromatin fibers (**Fig. 2**), and (5) painting of *Arabidopsis* chromosomes and comparative chromosome painting in other Brassicaceae species using BAC contigs specific for *Arabidopsis* chromosomes as probes (**Fig. 2**).

2. Materials

2.1. *Fluorescence* In Situ *Hybridization*

2.1.1. *Chromosome Preparation by Spreading*

1. Freshly prepared ice-cold Carnoy's fixative (ethanol:glacial acetic acid, 3:1) (*see* **Note 1**).
2. 70% ethanol.
3. Citrate buffer: 10 m*M* sodium citrate/citric acid in distilled water, pH 4.5.
4. Pectolytic enzyme mixture: 0.3% (w/v) cellulase, 0.3% (w/v) pectolyase, 0.3% (w/v) cytohelicase (all from Sigma) in citrate buffer (*see* **Note 2**).
5. 60% acetic acid.
6. Stereo microscope.
7. Light microscope with phase contrast.
8. Microscopic slides.
9. Dissection needles and fine forceps.
10. Small Petri dishes.
11. Moist chamber for enzyme digestion.
12. Glass capillary tubes.
13. Heating block.
14. Incubator (37°C).

2.1.2. *Probe Labeling*

2.1.2.1. Probe Labeling Using Nick Translation Mix

1. Nick Translation Mix (Roche).
2. Nucleotides: dATP, dCTP, dGTP, dTTP and either biotin-dUTP or digoxigenin-dUTP (all from Roche), DNP-dUTP (PerkinElmer), Cy3-dUTP (Amersham), or DEAC-dUTP (Perkin Elmer).

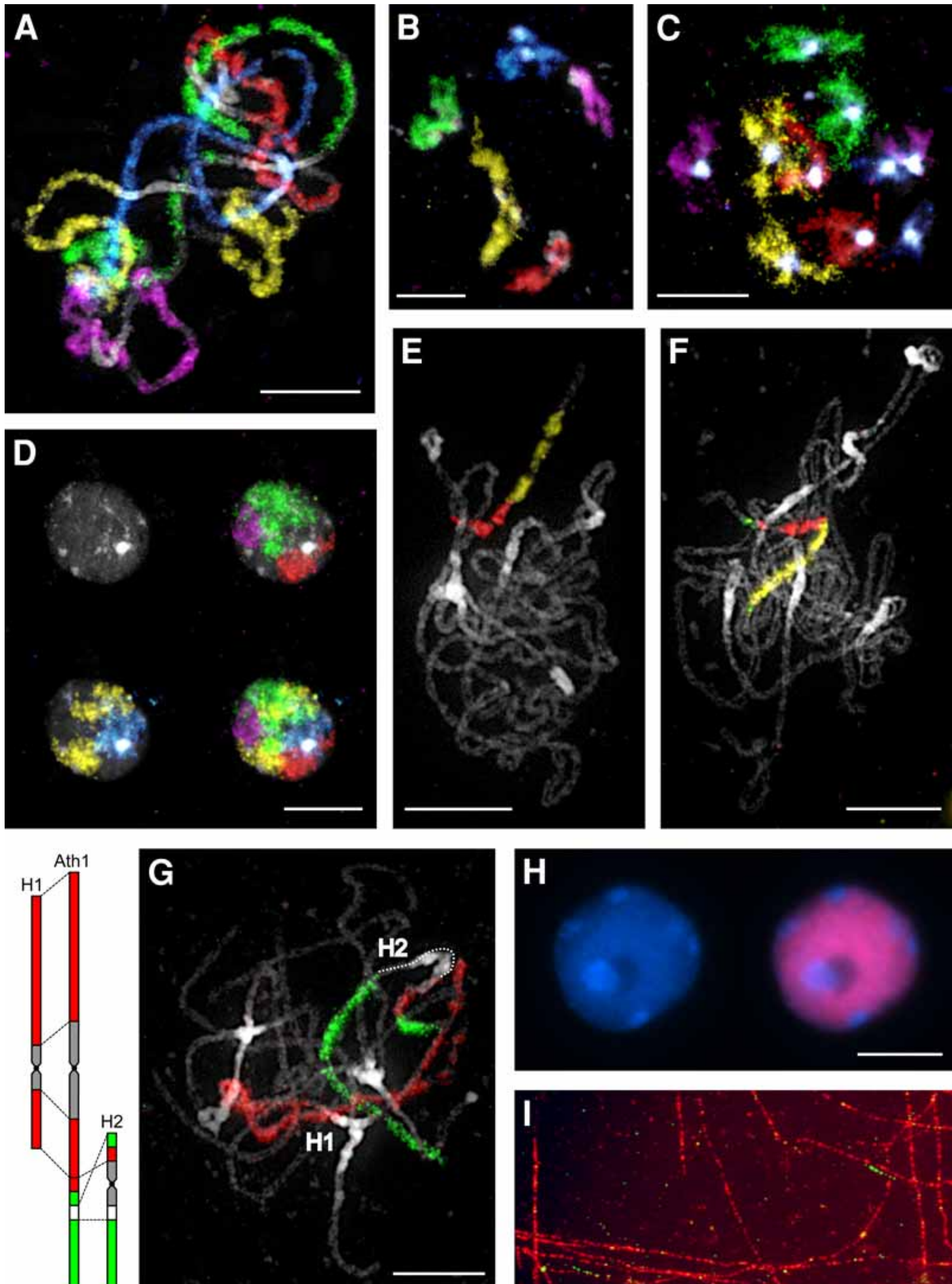


Fig. 2. Applications of fluorescence *in situ* hybridization and immunostaining techniques. **(A–D)** Multicolor painting of the entire *Arabidopsis* chromosome complement at pachytene **(A)**, diplotene **(B)**, mitotic prometaphase **(C)**, and interphase **(D)** of accession C24. Chromosome-specific BAC contigs of chromosomes 1 (yellow) and 3 (blue) have been directly labeled by Cy3-dUTP and DEAC-dUTP, respectively. Probes specific for chromosomes 2 (magenta), 4 (red), and 5 (green) were labeled by DNP-dUTP, biotin-dUTP and digoxigenin-dUTP, and sub-

3. 3 *M* sodium acetate, pH 5.2.
4. 70% and 96% ice-cold ethanol.
5. 0.5 *M* EDTA, pH 8.0.
6. Eppendorf tubes.
7. Water baths (15°C, 65°C).
8. Centrifuge for Eppendorf tubes.

2.1.2.2. Alternative Single-Component Approach for Nick Translation

1. All materials from **Subheading 2.1.2.1.** except **1**.
2. 10X NT buffer: 0.5 *M* Tris-HCl, pH 7.5, 50 m*M* $MgCl_2$, 0.05% bovine serum albumin (BSA). Store in aliquots at –20°C.
3. 0.1 *M* β-mercaptoethanol. Store in aliquots at –20°C.
4. DNase I (Roche). Use a 1:250 dilution of a 1 mg/mL DNase stock in 0.15 *M* NaCl, 50% glycerol (*see* **Note 3**).
5. DNA polymerase I (10 U/µL, MBI Fermentas).
6. Electrophoresis system.
7. 1% agarose gel.
8. 100-bp ladder DNA marker.

2.1.3. In situ *Hybridization*

1. Labeled probe(s).
2. Hybridization buffer (HB50): 50% deionized formamide, 2X SSC (1X SSC = 0.15 *M* NaCl/ 0.015 *M* Na_3-citrate), 50 m*M* sodium phosphate, pH 7.0.
3. 20% dextran sulfate in HB50.
4. 3 *M* sodium acetate, pH 5.2.
5. 2X SSC.
6. 100 µg/mL RNase (Roche) in 2X SSC.
7. 1X phosphate-buffered saline (PBS) (10 m*M* sodium phosphate, pH 7.0, 143 m*M* NaCl) (conveniently from 10X stock).
8. 4% formaldehyde in 1X PBS (*see* **Note 4**).
9. Ethanol series (70%, 90%, 100%).

Fig. 2. *(continued)* sequently immuno-detected using Cy5, Texas Red, and Alexa 488, respectively. **(D)** Individual *Arabidopsis* chromosomes form distinct chromosome territories filling the entire nuclear volume (top left: DAPI-stained nucleus; top right: the same nucleus with FISH signals that mark the territories of chromosomes 2, 4, and 5; bottom left: signals mark territories of chromosomes 1 and 3; bottom right: territories of all five chromosomes differently labeled). **(E)** Pachytene complement of accession Col with the top arm of chromosome 3 painted by two adjacent BAC contigs occupying a proximal (red) and more distal position (yellow), respectively. **(F)** A paracentric inversion of the entire region labeled in (E) had occurred during the integration of a transgene (green) in positions flanking the painted region. **(G)** Comparative chromosome painting with four BAC contigs specific for *Arabidopsis* chromosome 1 (Ath1) to pachytene chromosomes of *Neslia paniculata* (L.) Desv. (2n = 14). Two chromosomes (H1 and H2) of the target species revealed extended homeology to *A. thaliana* chromosome 1. **(H)** Immunostaining of an *Arabidopsis* leaf nucleus (left: DAPI-stained) with rabbit anti-histone H3 dimethylated at lysine 4 detected with goat anti-rabbit~Alexa 488 (right). Euchromatin appears uniformly enriched in H3 dimethylated at lysine 4, while the nucleolus (nu) and the heterochromatic chromocenters are free of signals. **(I)** FISH on extended DNA fibers with 5S rDNA (red) and a cosmid clone (green) containing a unique genomic sequence of 30 Kb used here as an internal scale corresponding to 10 µm. Counterstaining with DAPI in (A–H). Bars = 5 µm. *See* color version in insert following p. 239.

10. Moist chamber.
11. Eppendorf tubes.
12. Cover slips 24 × 50 and 22 × 22 or 32 × 24 mm.
13. Heating block (80°C).
14. Incubators (60°C, 37°C).
15. Centrifuge.
16. Vacuum centrifuge.

2.1.4. Fluorescent Detection of Hybridized Probes (see **Note 5***)*

1. 2X SSC.
2. SF50 buffer: 50% formamide in 2X SSC, pH 7.0.
3. Water bath of 42°C (in fume hood for incubation in SF50 buffer).
4. Coplin jars.
5. 4T buffer: 4X SSC, pH 7.0, 0.05% (v/v) Tween-20.
6. Blocking solution: 5% BSA, 0.2% (v/v) Tween-20 in 4X SSC.
7. TNT buffer: 100 m*M* Tris-HCl, pH 7.5, 150 m*M* NaCl, with 0.05% (v/v) Tween-20.
8. TNB buffer: 100 m*M* Tris-HCl, pH 7.5, 150 m*M* NaCl, 0.5% Boehringer blocking reagent.
9. Avidin~Texas Red (Vector Laboratories).
10. Goat anti-avidin~biotin (Vector Laboratories).
11. Mouse anti-digoxigenin (Jackson Laboratories).
12. Rabbit anti-DNP (Sigma).
13. Goat anti-mouse~Alexa 488 (Molecular Probes).
14. Goat anti-rabbit~Cy5 (Jackson Laboratories).
15. Vectashield antifade mounting medium (Vector Laboratories) with 2 µg/mL DAPI.
16. Ethanol series (70%, 90%, 100%).
17. Cover slips (24 × 50 mm).
18. Moist chamber.
19. Incubator (37°C).
20. Fluorescence microscope equipped with optical filters for DAPI, FITC/Alexa 488, Texas Red, Cy5, DEAC and Cy3 fluorochromes (all filters from AHF Analysentechnik).
21. Digital charge-coupled device (CCD) camera.
22. Image acquisition software (e.g., MetaVue, Spot).

2.2. Isolation of Interphase Nuclei

1. Razor blade or mechanical homogenizer (e.g., Polytron PT 1200).
2. Petri dishes.
3. Falcon tubes.
4. Nylon mesh (pore size 35–50 µm).
5. Filter paper.
6. Microscopic slides and cover slips.
7. Tris buffer: 10 m*M* Tris-HCl, pH 7.5, 10 m*M* Na_2-EDTA, 100 m*M* NaCl.
8. 4% formaldehyde in Tris buffer.
9. LB01 lysis buffer: 15 m*M* Tris-HCl, pH 7.5, 2 m*M* Na_2-EDTA, 0.5 m*M* spermine·4 HCl, 80 m*M* KCl, 20 m*M* NaCl, 0.1% Triton X-100.
10. 0.5% (w/v) agar in distilled water.
11. DAPI stock in water (diluted to final concentration of 1 µg/mL).
12. Fluorescence microscope.
13. Eppendorf centrifuge.

2.3. In Situ *Immunolocalization of Modified DNA*

This technique is exemplified for cytosine-methylated DNA but can be adapted easily to other DNA modifications (e.g., bromodeoxyuridine that substitutes thymidine) using appropriate antibodies.

1. Standard chromosome preparations from either Carnoy's fixed tissue (*see* **Subheading 3.1.1.**) or formaldehyde-fixed isolated nuclei (*see* **Subheading 3.2.**) on microscopic slides.
2. 1X PBS: 10 m*M* sodium phosphate, pH 7.0, 143 m*M* NaCl.
3. 4% formaldehyde in 1X PBS (*see* **Note 4**).
4. Hybridization buffer HB50: 50% deionized formamide in 2X SSC and 50 m*M* sodium phosphate, pH 7.0.
5. 1% BSA in 1X PBS.
6. Mouse anti-5-methylcytosine (Eurogentec).
7. Donkey anti-mouse~Cy3 (Dianova).
8. Vectashield with 1 µg/mL DAPI.
9. Ethanol series (70%, 90%, 100%).
10. Cover slips 22 × 22, 22 × 32, and 24 × 50 mm.
11. Heating block (80°C).
12. Moist chamber.
13. Incubator (37°C).

2.4. In Situ *Immunolocalization of Chromatin Proteins*

This technique is examplified for histone H3 methylated at lysine 4.

1. Formaldehyde-fixed nuclei on microscopic slides (*see* **Subheading 3.2.**).
2. 1X PBS: 10 m*M* sodium phosphate, pH 7.0, 143 m*M* NaCl.
3. 4% formaldehyde (*see* **Note 4**) in 1X PBS.
4. Primary antibody, e.g., rabbit anti-histone H3 dimethylated at lysine 4 (K4) (Upstate).
5. 1% BSA in 1X PBS.
6. Secondary antibody, e.g., goat anti-rabbit~Alexa 488 (Molecular Probes).
7. Vectashield with 1 µg/mL DAPI.
8. Moist chamber.
9. Incubator (37°C).

2.5. FISH to Extended DNA Fibers

2.5.1. Isolation of Nuclei

1. Nuclei isolation buffer (NIB), freshly prepared and kept on ice: 10 m*M* Tris-HCl, pH 9.5, 10 m*M* EDTA, 100 m*M* KCl, 0.5 *M* sucrose, 4 m*M* spermidine, 1.0 m*M* spermine, 0.1% (v/v) β-mercaptoethanol.
2. 10% Triton X-100 in NIB.
3. Glycerol.
4. Razor blade.
5. Microscopic slides.
6. Eppendorf tubes.
7. Nylon mesh filters (120, 50 µm).
8. Vectashield with 1 µg/mL DAPI.
9. Eppendorf centrifuge.
10. Fluorescence microscope.

2.5.2. Preparation of DNA Fibers

1. 1X PBS: 10 m*M* sodium phosphate, pH 7.0, 143 m*M* NaCl.
2. STE buffer: 0.5% (w/v) SDS, 5 m*M* EDTA, 100 m*M* Tris-HCl, pH 7.0.
3. Carnoy's fixative: ethanol:glacial acetic acid, 3:1.
4. Microscopic slides boiled for 20 min in deionized water (alternative: ethanol-cleaned) and air-dried.
5. Vectashield with 1 µg/mL DAPI.
6. Eppendorf centrifuge.
7. Coplin jars.

8. Incubator (60°C).
9. Fluorescence microscope.

3. Methods

3.1. Fluorescence In Situ *Hybridization*

3.1.1. Chromosome Preparation by Spreading

The most crucial step in the procedure is the preparation of high-quality chromosome spreads. Because the anthers containing pollen mother cells at pachytene are hardly visible, even with a standard binocular, entire flower buds are used for preparation of pachytene chromosome spreads. A major advantage of using immature flower buds is the presence of meiotic and mitotic cells as well. The following protocol is applicable also for leaf or root tissue and for chromosome preparation from other Brassicaceae taxa as well.

1. Fix young inflorescences in freshly prepared ice-cold Carnoy's fixative (*see* **Note 1**) for at least 3 × 30 min. Fixed material can be stored in fixative or in 70% ethanol at 4°C or –20°C for several months.
2. Wash fixed inflorescences with distilled water 2 × 10 min in a small Petri dish.
3. Replace the water by citrate buffer and wash 3 × 5 min; remove unwanted parts of the inflorescences under a stereomicroscope.
4. Incubate three inflorescences in ~1 mL of pectolytic enzyme mixture for 3 h in a moist chamber at 37°C.
5. Take an inflorescence and select a flower bud. Usually, the large yellowish flower buds and the first white flower bud contain anthers with pollen and microspores or tetrad stages, respectively. The second white bud has pollen mother cells (PMCs) in meiosis II. PMCs at prophase I, including diakinesis and pachytene, can be found in the third or fourth white flower bud. Put the flower bud on a clean microscopic slide.
6. Remove excess of fluid (by capillary forces of a fine forceps or a glass capillary tube). Keep ~4 μL of liquid.
7. Tap the flower bud with a dissection needle until a fine suspension has formed. If there are still large tissue fragments, then extend the enzyme incubation time for further digestion. If a fine suspension has been obtained, replace the enzyme mixture in the Petri dish by citrate buffer and keep on ice or at 4°C.
8. Add about 15 μL of 60% acetic acid to the suspension on slide to clear the fine suspension and to make cells adhere to the slide. Place the slide on a heating block at 45 to 50°C, add twice ~15 μL of 60% acetic acid, and spread the drop by careful stirring for 15 s (*see* **Note 6**).
9. Precipitate the cells by careful pipetting ~100 μL of Carnoy's fixative around the drop with cleared cell suspension and wait until the fixative covers the whole slide. Discard excessive solution by tilting the slide.
10. Air-dry the preparation (*see* **Note 7**).
11. Examine the preparations for specific meiotic stages in the phase contrast microscope. (*see* ref. ***12*** for a detailed microscopic atlas of meiosis in *Arabidopsis*). Check that cells are spherical or oval, without remnants of cell wall material. Mitotic metaphase chromosomes should appear as dark gray or black structures (not white). Nuclei look dark gray with black conspicuous spots (i.e., heterochromatic chromocenters); cytoplasm is almost invisible; pachytene chromosomes should be free of cytoplasm. If chromosomes and nuclei appear to be covered by cytoplasm, preparations aimed for FISH should be treated with pepsin (**Note 8**).
12. Store selected and dried slides in a dust-free box at 4°C. Let stored slides adapt to room temperature before further use.

3.1.2. Probe Labeling

Molecular biology laboratory manuals contain detailed protocols for isolation of DNA from living plant tissues, bacteria, and yeast. For DNA isolation from BAC clones we recommend to use the plasmid isolation protocol of Sambrook et al. ***(29)*** and to scale up the volumes to 50 mL of bacterial culture (*see* **Note 9**).

3.1.2.1. Probe Labeling Using Nick Translation Mix

1. Combine in an Eppendorf tube:
   a. 1 µg DNA in 12 µL sterile distilled water.
   b. 4 µL nucleotide mix (250 µ*M* dATP, dCTP, dGTP; 170 µ*M* dTTP and 80 µ*M* of either biotin-dUTP or digoxigenin-dUTP, DNP-dUTP, Cy3-dUTP or DEAC-dUTP).
   c. 4 µL Nick Translation Mix.
2. Mix carefully and centrifuge briefly.
3. Incubate for 90 min at 15°C.
4. Stop the reaction by adding 1 µL 0.5 *M* EDTA, pH 8.0, and heating to 65°C for 10 min.

3.1.2.2. Alternative Single-Component Approach for Nick Translation (*see* **Note 10**)

1. Combine in an Eppendorf tube:
   a. 1–2 µg DNA.
   b. 5 µL nucleotide mix (2 m*M* dATP, dCTP, dGTP, 400 µ*M* dTTP).
   c. 5 µL 10X NT-buffer.
   d. 1–4 µL 1 m*M* X-dUTP (in which X is either biotin, digoxigenin, DNP, Cy3, or DEAC).
   e. 5 µL 0.1 *M* β-mercaptoethanol.
   f. Fill up with distilled water to 47 µL.
   g. 2 µL DNase I.
   h. 1 µL DNA polymerase I, mix gently, and centrifuge briefly.
2. Incubate for 90 min at 15°C.
3. Transfer the tubes to ice and load 1/10 of the reaction volume on a 1% agarose gel together with a 100-bp ladder DNA marker. When the bulk of labeled fragments is ~400 to 500 bp long, stop the reaction by adding 1 µL 0.5 *M* EDTA, pH 8.0, and heat the probe to 65°C for 10 min (*see* **Note 11**).
4. For purification it is recommended to precipitate the probe by adding 1/10 vol of 3 *M* sodium acetate, pH 5.2, and 2.5 vol ice-cold 96% ethanol and keep at –70°C for 30 min or at –20°C overnight. Centrifuge the precipitate with 13,000*g* for 30 min at 4°C, wash pellet with 70% ethanol, centrifuge again for 15 min, air-dry, and resuspend the pellet in 50 µL of sterile distilled water.

3.1.3. In Situ *Hybridization*

1. To apply one to three labeled probe(s) for FISH, add 1 to 3 µL of the probe(s) to 2 µL distilled water in an Eppendorf tube and centrifuge in a vacuum centrifuge at medium speed for 12 to 15 min. To the vacuum-dried probe add 10 µL HB50 and 10 µL 20% dextran sulfate in HB50 (*see* **Note 12**).
   If more BAC probes must be applied simultaneously, e.g., for chromosome painting, it is recommended to label individual BAC clones separately and pool them for hybridization. Pipet 3 to 5 µL of each labeled probe into an Eppendorf tube. Precipitate the DNA with 1/10 volume of 3 *M* sodium acetate, pH 5.2, and 2.5 vol of ice-cold 96% ethanol. Mix well and keep on ice or at for –20°C at least 30 min. Spin down at 13,000*g* for 30 min at 4°C. Discard the supernatant, dry the pellet, dissolve it in 10 µL HB50 at 37–42°C, and add 10 µL 20% dextran sulfate in HB50 (*see* **Note 12**).
2. Bake the slides at 60°C for 30 min.
3. Pipet 100 µL RNase solution on the slides and cover with a 24 × 50 mm cover slip and incubate at 37°C for 60 min.
4. Rinse in 2X SSC at RT for 2 × 5 min. All washing steps are performed in Coplin jars (*see* **Note 8**).
5. Rinse in 1X PBS for 5 min.
6. Post-fix in 4% formaldehyde in 1X PBS (*see* **Note 4**) at room temperature for 10 min.
7. Rinse in 1X PBS for 2 × 5 min.
8. Dehydrate the slides through an ethanol series (70%, 90%, 100%), each step 1 to 3 min.
9. Air-dry the slides.
10. Add 20 µL probe and cover with a 22 × 22 or 32 × 24 mm cover slip.

11. Denature the probe DNA and chromosomal DNA together on the slides at 80°C for 2 min on a heating block.
12. Put the slides in a moist chamber and incubate overnight at 37°C (*see* **Note 13**).

3.1.4. Fluorescent Detection of Hybridized Probes

Signal detection and amplification are exemplified for hapten-labeled probes (biotin-, digoxigenin-, DNP-dUTP) visualized by indirect immunofluorescence via Texas Red, Alexa 488, and Cy5. (If the first detection step for hapten-labeled probes yields a sufficiently strong signal, amplification can be omitted.) This protocol can be combined with the application of probes containing directly fluorochrome-labeled nucleotides (e.g., Cy3- or DEAC-dUTP) that require only posthybridization washing prior to microscopic evaluation. FISH can be performed with single probes labeled according to one of the above options or with various combinations of differently labeled probes. Avoid drying of slides during the entire procedure.

1. Wash slides in a Coplin jar with 2X SSC at 42°C for 2 min. Keep the slides in the dark as much as possible when fluorochrome-labeled probes are used (valid for all following steps).
2. Wash slides in SF50 at 42°C 3 × 5 min (*see* **Note 14**).
3. Wash in 2X SSC at 42°C for 2 min. If only fluorochrome-labeled probes are used, rinse the slides in 2X SSC at RT and proceed with **step 14**.
4. Rinse the slides briefly in 4T at 42°C.
5. Pipet 100 µL blocking solution onto slide, put cover slip on slide, and incubate in a moist chamber at 37°C for 30 min.
6. Rinse in 4T at 42°C for 2 × 5 min.
7. Rinse in TNT at 42°C for 5 min.
8. For detection of biotin-labeled probes pipet 100 µL avidin~Texas Red in TNB (1:1000) onto slide, put cover slip on slide and incubate in a moist chamber at 37°C for 30 min. Keep the slides in the dark as much as possible during the following steps.
9. Rinse in TNT at 42°C for 3 × 5 min.
10. Mix goat anti-avidin~biotin, mouse anti-digoxigenin, and rabbit anti-DNP in TNB that the final concentration is 1:200, 1:250, and 1:400, respectively, and pipet 100 µL onto slide, put cover slip on slide and incubate in a moist chamber at 37°C for 30 min.
11. Rinse in TNT at 42°C for 3 × 5 min.
12. Mix avidin~Texas Red, goat anti-mouse~Alexa 488, and goat anti-rabbit~Cy5 in TNB (final concentrations: 1:1000, 1:200, and 1:100), pipet 100 µL onto slide, put cover slip on slide, and incubate in a moist chamber at 37°C for 30 min.
13. Rinse in TNT at 42°C for 3 × 5 min.
14. Dehydrate in an ethanol series of 70%, 90%, and 100%, 1 to 3 min each, and air-dry.
15. Mount in Vectashield with DAPI.
16. Check the slide under a fluorescence microscope using the appropriate filters. In particular the fluorescence signals of far-red fluorochromes (e.g., Cy5) require a CCD camera and image acquisition software for visualization.

3.2. Isolation of Interphase Nuclei

For the detection of chromatin proteins, in most cases isolated nuclei fixed in formaldehyde are required to preserve the targeted antigenic epitopes (*see* **Note 15**).

1. For isolation of nuclei from young seedlings germinate ~2000 seeds (~20 mg) arranged in parallel rows in a Petri dish on a layer of 0.5% agar covered by filter paper. Keep the Petri dish in an almost vertical position for 2 to 5 d at room temperature (*see* **Note 16**).
2. Dissect ~10 mg tissue from entire seedlings, from ~0.5-cm-long roots, or from young leaflets, and fix the material in 10 mL of ice-cold 4% formaldehyde in Tris buffer for 20 min.
3. Wash the sample in ice-cold Tris buffer for 2 × 10 min.
4. Chop the material with a razor blade in 400 µL LB01 buffer in a small Petri dish on ice to a fine suspension (*see* **Note 17**).

5. Filter the suspension through nylon mesh (pore size 35–50 μm) into a new tube.
6. Stain 1 to 2 μL of the sample with DAPI (final concentration 1 μg/mL).
7. Check a drop of the nuclear suspension under fluorescence microscope for yield and density of nuclei. If you find several nuclei in one field (×63 objective), the density is satisfactory (*see* **Note 18**).
8. Pipet 2 to 10 μL of the suspension onto a microscopic slide and air-dry.
9. Store the slides until use at –20°C. Before using, let stored slides adapt to room temperature.

3.3. In Situ *Immunolocalization of Modified DNA*

The immunodetection of modified DNA requires a denaturation step to increase the accessibility of the modified bases to antibodies. Since the epitope is a DNA constituent and not a peptide, the plant material can be fixed either in Carnoy's fixative or in formaldehyde. Follow **Subheading 3.1.1.** for spreading of cells or **Subheading 3.2.** for isolation of nuclei.

1. Bake the slides at 60°C for 30 min.
2. Rinse in 1X PBS for 2 × 5 min.
3. Postfix in 4% formaldehyde in 1X PBS (*see* **Note 4**) for 10 min at room temperature.
4. Rinse in 1X PBS for 2 × 5 min.
5. Dehydrate in 70%, 90%, 100% ethanol, each step 1 to 3 min, and air-dry.
6. Pipet 20 μL HB50 on the slide, cover with a 22 × 22 mm cover slip and denature at 80°C for 2 min on a heating block.
7. Rinse in ice-cold 1X PBS for 2 × 5 min. During the first rinse the cover slip is removed.
8. Block nonspecific binding sites by incubating in 1% BSA in 1X PBS in a moist chamber at 37°C for 30 min. Use 100 μL per slide and a 24 × 50 mm cover slip.
9. Rinse in 1X PBS at RT for 3 × 5 min.
10. Mix 1 μL mouse anti-5-methylcytosine (primary antibody) with 40 μL 1% BSA and pipet the solution onto slide, put a 22 × 32 mm cover slip on the slide, and incubate in a moist chamber at 37°C for 30 min or at 4°C overnight.
11. Rinse in 1X PBS at RT for 3 × 5 min.
12. Mix 1 μL donkey anti-mouse~Cy3 (secondary antibody) with 500 μL 1% BSA in 1X PBS and pipet 50 μL onto slide, put a 22 × 32 mm cover slip on the slide, and incubate in a moist chamber at 37°C for 30 min.
13. Rinse in 1X PBS at RT for 3 × 5 min.
14. Mount in Vectashield with 1 μg/mL DAPI.

3.4. In Situ *Immunolocalization of Chromatin Proteins*

The immuno-detection of nuclear proteins is usually performed on isolated nuclei (*see* **Subheading 3.2.**).

1. Postfix the preparation on slides in 4% formaldehyde in 1X PBS at room temperature for 30 min.
2. Rinse in 1X PBS for 2 × 5 min.
3. Incubate in 1% BSA in 1X PBS in a moist chamber at 37°C for 30 min. Use 100 μL per slide and a 24 × 50 mm cover slip.
4. Rinse the slides in 1X PBS for 5 min.
5. H3dimetK4 detection: Mix 1 μL rabbit anti-H3 dimethylated K4 (H3dimetK4) with 100 μL 1% BSA, pipet the solution on a slide, cover with a 24 × 50 mm cover slip, and incubate in a moist chamber at 37°C for 2 h or at 4°C overnight.
6. Rinse in 1X PBS at RT for 3 × 10 min.
7. Mix 5 μL goat anti-rabbit~Alexa 488 with 1 mL 1% BSA in 1X PBS. Use 100 μL per slide and a 24 × 50 mm cover slip and incubate at 37°C for 30 min.
8. Rinse in 1X PBS at RT for 3 × 5 min.
9. Mount in Vectashield with 1 μg/mL DAPI.

3.5. FISH to Extended DNA Fibers

3.5.1. Isolation of Nuclei

The protocol is similar to the one described in **Subheading 3.2.**, but does not include a fixation step.

1. Collect ~1 g young leaf material.
2. Place tissue and ~0.5 mL NIB into a Petri dish and chop the tissue to a fine suspension with the a razor blade.
3. Filter the suspension through a 120-μm mesh filter into an Eppendorf tube.
4. Filter the suspension through a 50-μm mesh filter into a new Eppendorf tube. Plastids may be removed by adding 1/20 vol of 10% Triton-X100 in NIB.
5. Spin down at 300*g* (Eppendorf centrifuge) for 3 min at 4°C.
6. Remove the supernatant and resuspend the (white) pellet with nuclei in ~20 μL NIB.
7. Inspect the nuclei by mixing 2 μL of suspension with 5 μL Vectashield containing 1 μg/mL DAPI on a microscope slide. Examine under a fluorescence microscope. Nuclei should have a spherical appearance.
8. Mix the suspension with an equal volume of glycerol.
9. Store at –20°C until use.

3.5.2. Preparation of DNA Fibers

1. Spin down up to 30 μL of the suspension containing the nuclei at 950*g* for 5 min.
2. Resuspend the pellet in 20 μL 1X PBS.
3. Pipet 2 droplets of 1 μL of the suspension onto one end of a clean microscopic slide and air-dry for 2 to 3 min.
4. Lyse the nuclei by adding 30 μL of STE and incubate for 4 min.
5. Tilt the slide at an angle of ~45° to let the buffer float downward. Free ends of DNA fibers will stretch and move downward out of the nuclei.
6. Air-dry the slides and fix the DNA fibers in Carnoy's fixative in a Coplin jar for 2 min.
7. Air-dry the slides. Examine the fibers under the fluorescence microscope in 5 μL of DAPI per slide. A dense concentration of stretched fluorescent fibers should be visible under a ×63 objective.
8. Bake the slides at 60°C for 30 min.
9. Store the preparations in a dry box until use.

3.5.3. FISH to DNA Fibers

Follow the protocol as described in **Subheadings 3.1.3.** (from **step 10**) and **3.1.4.**

4. Notes

1. Alternatively, 6:3:1 fixative (ethanol:chloroform:glacial acetic acid, 6:3:1) can be used.
2. A stock solution containing 1% of each enzyme in citrate buffer is stored at –20°C.
3. Dilution should be prepared freshly and the stock can be stored at –20°C.
4. For postfixation on slides, either 4% formaldehyde prepared from commercially available 37% formaldehyde or 4% paraformaldehyde prepared from solid substance by stirring for ~30 min on a heating plate (~60ºC) under a fume hood and subsequent filtration through filter paper is suitable. Formaldehyde is hygroscopic; therefore, once opened, the concentration will be, after longer storage, less than 37% in the original vessels.
5. When using differently hapten-labeled probes, select antibodies for detection/amplification of FISH signals in a way that will prevent unintentional cross-reaction.
6. In this step and the following one the acid-soluble proteins and various cytoplasmic components will dissolve in the acetic acid and clear up the spread preparation. Omitting this step will result in chromosome spreads that are still surrounded by cytoplasm. Too little acetic acid or too high a density of cells will result in an undesirable layer of cytoplasmic debris. Monitor the result under a phase contrast microscope. If needed add either more acetic acid or extend exposure time to acetic acid. The cells move in the droplet and settle down at the periphery of the drop-

let, forming a ring of cells. To prevent disposition of too many cells at the same ring position, the droplet should be stirred now and then with a needle without touching the glass surface.

7. For unknown reasons the standard spreading protocol sometimes may yield faint chromosome/nuclei preparations. This phenomenon becomes visible on slides after denaturation at high temperature. It can be prevented by formaldehyde postfixation immediately after spreading and precipitation in Carnoy's fixative. To this, after **step 9** rinse the slide in water and postfix the material in 4% freshly prepared formaldehyde in 1X PBS (10 m*M* sodium phosphate, pH 7.0, 143 m*M* NaCl) or in distilled water for 10 min and air-dry.
8. If chromosomes/nuclei appear to be covered by cytoplasm after checking of dry slides in a phase contrast microscope, pipet 100 µL pepsin (100 µg/mL in 0.01 *N* HCl) on the slide, cover with a 24 × 50 mm cover slip, and incubate at 37°C until cytoplasm is no longer microscopically detectable. Too long a treatment with pepsin may degrade chromatin structures.
9. Occasional milky appearance of the DNA solution after isolation, indicating contamination with proteins and polysaccharides, does not impair the hybridization efficiency of the probe.
10. The single-component approach for nick translation allows adjusting the desired probe length.
11. If the fragments are clearly longer than 500 bp, extend the incubation at 15°C for another 30 min. Additional DNase can be added as well. If the probe is too short, DNase can be reduced in a repeated experiment. The suitable incubation time and DNase concentration should be determined empirically. The conditions may slightly differ from probe to probe.
12. Insufficiently dissolved DNA may cause a high background.
13. Hybridization of *Arabidopsis* probes to chromosomes of other Brassicaceae species requires longer hybridisation times (48–72 h) to ensure proper hybridisation to the homeologous target sequences. Sealing of cover slips by rubber cement (e.g., Fixogum) avoids drying out of the probe and/or its dilution by condensing water during long hybridization times.
14. If applying *Arabidopsis* probes to chromosomes of other Brassicaceae species, wash under less stringent conditions (20% formamide in 2X SSC, pH 7.0) at 42°C for 3 × 5 min.
15. If needed, suspensions of isolated nuclei can be used for sorting of nuclei according to their DNA content into different cell cycle/ploidy fractions by means of a flow cytometer equipped with an ultraviolet laser and a sorting device (for flow-cytometric analysis and sorting *see*, e.g., ref. ***30***).
16. To prepare nuclear suspension from adult aerial tissues, e.g., from young healthy leaves, grow plants under standard conditions in the greenhouse.
17. Alternatively, chop with a razor blade in 200 µL LB01 buffer, pipet the suspension into a Falcon tube containing another 200 mL LB01, and homogenize with a homogenizer at 11,000 rpm for 30 s.
18. To increase the concentration of nuclei, one may spin the suspension down at ~500*g* for 3 min at 4°C. Then, remove the supernatant and resuspend the pellet with nuclei in ~40 µL LB01 buffer.

Acknowledgments

We thank Joerg Fuchs, Ales Pecinka, and Joachim Bruder for helpful comments and Zuzana Jasencakova, Ales Pecinka, and Joerg Fuchs for providing images integrated into **Figs. 1** and **2**.

References

1. The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815.
2. Hosouchi, T., Kumekawa, N., Tsuruoka, H., and Kotani, H. (2002) Physical map-based sizes of the centromeric regions of *Arabidopsis thaliana* chromosomes 1, 2, and 3. *DNA Res.* **9,** 117–121.
3. Bennett, M. D., Leitch, I. J., Price, H. J., and Johnston, J. S. (2003) Comparisons with *Caenorhabditis* (~100 Mb) and *Drosophila* (~175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus ~25% larger than the *Arabidopsis* genome initiative estimate of ~125 Mb. *Ann. Bot.* **91,** 547–557.
4. Leutwiler, L. S., Hough-Evans, B. R., and Meyerowitz, E. M. (1984) The DNA of *Arabidopsis thaliana*. *Mol. Gen. Genet.* **194,** 15–23.
5. Laibach, F. (1907) Zur Frage nach der Individualität der Chromosomen im Pflanzenreich. *Beih. Bot. Centralbl. Abt. 1* **22,** 191–210.

6. Ambros, P. and Schweizer, D. (1976) The Giemsa C-banded karyotype of *Arabidopsis thaliana* (L.) Heynh. *Arabidopsis Inf. Serv.* **13,** 167–171.
7. Maluszynska, J. and Heslop-Harrison, J. S. (1991) Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. *Plant J.* **1,** 159–166.
8. Maluszynska, J. and Heslop-Harrison, J. S. (1993) Molecular cytogenetics of the genus *Arabidopsis*: in situ localization of rDNA sites, chromosome numbers and diversity in centromeric heterochromatin. *Ann. Bot.* **71,** 479–484.
9. Steinitz-Sears, L. M. (1963) Chromosome studies in *Arabidopsis thaliana*. *Genetics* **48,** 483–490.
10. Röbbelen, G. and Kribben, F. J. (1966) Erfahrungen bei der Auslese von Trisomen. *Arabidopsis Inf. Serv.* **3,** 16–17.
11. Samoylova, T. I., Meister, A., and Miséra, S. (1996) The flow karyotype of *Arabidopsis thaliana* interphase chromosomes. *Plant J.* **10,** 949–954.
12. Ross, K. J., Fransz, P., and Jones, G. H. (1996) A light microscopic atlas of meiosis in *Arabidopsis thaliana*. *Chromosome Res.* **4,** 507–516.
13. Fransz, P., Armstrong, S., Alonso-Blanco, C., Fischer, T. C., Torres-Ruiz, R. A., and Jones, G. (1998) Cytogenetics for the model system *Arabidopsis thaliana*. *Plant J.* **13,** 867–876.
14. Fransz, P. F., Armstrong, S., de Jong, J. H., et al. (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* **100,** 367–376.
15. Fransz, P., de Jong, J. H., Lysak, M., Ruffini Castiglione, M., and Schubert, I. (2002) Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. USA* **99,** 14584–14589.
16. Koornneef, M., Fransz, P., and de Jong, H. (2003) Cytogenetic tools for *Arabidopsis thaliana*. *Chromosome Res.* **11,** 183–194.
17. Jones, G. H., Armstrong, S. J., Caryl, A. P., and Franklin, F. C. H. (2003) Meiotic chromosome synapsis and recombination in *Arabidopsis thaliana*; an integration of cytological and molecular approaches. *Chromosome Res.* **11,** 205–215.
18. Fransz, P. F., Alonso-Blanco, C., Liharska, T. B., Peeters, A. J. M., Zabel, P., and de Jong, J. H. (1996) High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibres. *Plant J.* **9,** 421–430.
19. Soppe, W. J. J., Jasencakova, Z., Houben, A., et al. (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J.* **21,** 6549–6559.
20. Jasencakova, Z., Soppe, W. J. J., Meister, A., Gernand, D., Turner, B. M., and Schubert, I. (2003) Histone modifications in *Arabidopsis*—high methylation of H3 lysine 9 is dispensable for constitutive heterochromatin. *Plant J.* **33,** 471–480.
21. Probst, A. V., Fransz, P. F., Paszkowki, J., and Mittelsten Scheid, O. (2003) Two means of transcriptional reactivation within heterochromatin. *Plant J.* **33,** 743–749.
22. Chytilova, E., Macas, J., and Galbraith, D. W. (1999) Green fluorescent protein targeted to the nucleus, a transgenic phenotype useful for studies in plant biology. *Ann. Bot.* **83,** 645–654.
23. Kato, N. and Lam, E. (2001) Detection of chromosome tagged with green fluorescent protein in live *Arabidopsis thaliana* plants. *Genome Biol.* **2,** research 0045.
24. Lysak, M. A., Fransz, P. F., Ali, H. B. M., and Schubert, I. (2001) Chromosome painting in *Arabidopsis thaliana*. *Plant J.* **28,** 689–697.
25. Lysak, M. A., Pecinka, A., and Schubert, I. (2003) Recent progress in chromosome painting of *Arabidopsis* and related species. *Chromosome Res.* **11,** 195–204.
26. Ali, H. B. M., Lysak, M. A. and Schubert, I. (2004) Genomic *in situ* hybridization in plants with small genomes is feasible and elucidates the chromosomal parentage in interspecific *Arabidopsis* hybrids. *Genome* **47,** 954–960.
27. Menke, M., Chen, I. P., Angelis, K. J., and Schubert, I. (2001) DNA damage and repair in *Arabidopsis thaliana* as measured by the comet assay after treatment with different classes of genotoxins. *Mutat. Res.* **493,** 87–93.
28. Siroky, J., Zluvova, J., Riha, K., Shippen, D. E., and Vyskot, B. (2003) Rearrangements of ribosomal DNA clusters in late generation telomerase-deficient *Arabidopsis*. *Chromosoma* **112,** 116–123.
29. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual. 2nd ed.* Cold Spring Harbor Laboratory. Press, Cold Spring Harbor, NY.
30. Dolezel, J., Lysak, M. A., Kubaláková, M., Simková, H., Macas, J., and Lucretti, S. (2001) Sorting of plant chromosomes. *Methods Cell Biol.* **64,** 3–31.

16

Using Information From Public *Arabidopsis* Databases to Aid Research

Margarita Garcia-Hernandez and Leonore Reiser

Summary

The volume of *Arabidopsis* information has increased enormously in recent years as a result of the sequencing of the genome and other large-scale genomic projects. Much of the data are stored in public databases, where data are organized, analyzed, and made freely accessible to the research community. These databases are resources that researchers can utilize for making predictions and developing testable hypotheses. The methods in this chapter describe ways to access and utilize *Arabidopsis* data and genomic resources found in databases.

Key Words: Data mining; database; genomics; gene expression; bioinformatics; computational biology; *Arabidopsis*.

1. Introduction

Technological advances have fostered a new era in *Arabidopsis* research, giving us a well-annotated, sequenced genome ***(1)*** complementing a rich body of literature. More recently, international functional genomics initiatives (http://arabidopsis.org/info/workshop2010.jsp) have ignited a new wave of research to decipher the functions of every gene in the genome, and eventually, to understand what it takes to make a flowering plant. Large amounts of data about gene expression, metabolism, and protein and gene interactions are being generated by these projects. To accomplish the task of organizing and managing the data, groups and individual labs have created databases to store the information generated and make it available to the research community.

Scientists doing research in this "postgenomic" area are compelled to know how to make use of databases to extract the relevant information needed to further their research. The protocols in this chapter describe how to use databases to find what is known about *Arabidopsis* and to make inferences and predictions that can later be tested experimentally. Each protocol includes a summary of the rationale, a brief description of the database/tool(s), and the specific steps for querying, retrieving, and interpreting the data. The protocols, along with the corresponding databases and tools, are outlined in **Table 1**. This table of contents can be used to find specific protocols of interest within the chapter. When the methods described can be applied to databases other than the ones described below, the database is shown in the table along with a brief description. The databases described here represent a small portion of the vast collection of databases and bioinformatics resources available on the Internet. The protocols described here use tools and data available in databases as of the spring of 2004.

2. Materials

Programming experience is an asset to a scientist who wishes to analyze and manipulate complex, large datasets, but it is not essential to effectively mine databases. Anyone with access

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

Table 1
Databases and Tools that Can Be Used to Perform Methods Described in This Chapter

Database: Tool	URL	Protocol/Description
Searching literature databases		
NCBI PubMed Database	www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed	Finding relevant articles in the NCBI PubMed database
TAIR Publication Search	http://arabidopsis.org/servlets/Search?action=new_search&type=publication	Finding Arabidopsis publications using TAIR's Publication Search
Finding information about genes in Arabidopsis genome databases		
TAIR: Gene Search	www.arabidopsis.org/servlets/Search?action=new_search&type=gene	Finding gene information by name
TAIR GO Annotation Bulk Download and Analysis	www.arabidopsis.org/tools/bulk/go/index.jsp	Finding functional information about genes
TIGR: Gene Search	www.tigr.org/tdb/e2k1/ath1/ath1.shtml	Gene search by locus identifier, description, name
MatDB: Gene Search	http://mips.gsf.de/proj/thal/db/search/search_frame.html	Gene search using locus identifiers, BAC or cosmid clone names, keyword, MIPS codes, and BAC-based name or description
NCBI: Gene Search	www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene and www.ncbi.nlm.nih.gov:80/mapview/map_search.cgi?taxid=3702	Gene search can be limited by name, locus identifier, and description and species. Or search and browse the Arabidopsis reference genome database.
MIPS FuncCat Database	http://mips.gsf.de/proj/funcatDB/search_main_frame.html	Find genes with similar functions using MIPS functional categories
Gene Ontology Consortium	www.geneontology.org	Find genes with similar functions in Arabidopsis and other organisms
Searching DNA microarray data		
TAIR: Microarray Expression Search	www.arabidopsis.org/servlets/Search?action=new_search&type=expression	Finding the expression patterns of genes in different microarray experiments

(continued)

NASCArrays: Spot History	http://affymetrix.arabidopsis.info/narrays/spothistory.pl	Finding the expression history of a gene across all microarray experiments
TAIR: Microarray Experiment Search	http://arabidopsis.org/servlets/Search?type=expr&search_action=new_search	Search experiments and download data
TAIR: Microarray Element Search	http://arabidopsis.org/tools/bulk/microarray/index.jsp	Search array elements.
		View precomputed gene clusters
NASC Arrays	http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl	Search experiments and download data from experiments conducted at the NASC microarray facility. Several data mining tools
NCBI Gene Expression Omnibus	www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=geo	Find microarray and SAGE experiments and gene expression profiles. Clustering analysis, and visualization tools
Array Express	http://www.ebi.ac.uk/arrayexpress/	Search microarray experiments, protocols and arrays
Stanford Microarray Database	http://genome-www.stanford.edu/microarray/	Find, analyze, and download microarray data
Genevestigator	www.genevestigator.ethz.ch/	Displays patterns of expression of genes from Arabidopsis Affymetrix datasets in the context of plant organ, growth stage, or stress response
MapMan	http://gabi.rzpd.de/projects/MapMan/	Displays Arabidopsis Affymetrix datasets onto diagrams of metabolic pathways or other processes
Searching massively parallel signature sequencing data		
MPSS: Simple Search	http://mpss.udel.edu/at/	Searching massively parallel signature sequencing data
Searching expressed sequence tags		
NCBI UniGene Search	www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene	Using NCBI's UniGene to find ESTs for a sequenced gene
TIGR Gene Index Database	www.tigr.org/tigr-scripts/tgi/libtc.pl?db=atest	Using ESTs to find genes that are differentially expressed
AtGDB	www.plantgdb.org/AtGDB/prj/ZSB03PP/	EST mapping and clustering
Arabidopsis metabolic pathways		
AraCyc	www.arabidopsis.org/tools/aracyc	Finding metabolic pathways, reactions, enzymes, and compounds

(continued)

Table 1 ***(Continued)***
Databases and Tools that Can Be Used to Perform Methods Described in This Chapter

Database: Tool	URL	Protocol/Description
AraCyc Expression Viewer	www.arabidopsis.org:1555/expression.html	Detecting changes in the expression of genes involved in metabolism
Kegg	www.genome.ad.jp/kegg/pathway.html	Search and browse metabolic pathways, regulatory pathways, and molecular complexes
Finding related protein sequences in Arabidopsis		
TAIR: WU-BLAST	www.arabidopsis.org/wublast/index2.jsp	Finding similar protein sequences
TAIR: Bulk Protein Download	http://tigrblast.tigr.org/er-blast/	Finding similar sequences
MIPS: BLAST	http://mips.gsf.de/proj/thal/db/search/search_frame.html	Finding similar sequences
NCBI: BLAST	www.ncbi.nlm.nih.gov/BLAST/	Finding similar sequences
NCBI: Blink	www.ncbi.nlm.nih.gov/	Precomputed pairwise similarity search of Arabidopsis proteins against all proteins in GenBank
From genotype to phenotype—finding mutants for analysis of gene function		
SIGnAL T-DNA Express	http://signal.salk.edu/cgi-bin/tdnaexpress	Finding knockout mutations in a gene
TAIR: Germplasm Search	www.arabidopsis.org/servlets/Search?action=new_search&type=germplasm	Finding other types of mutations and mutants with similar phenotype
NASC: Catalogue Search	http://nasc.nott.ac.uk/catalogue.html	Finding mutants in a gene of interest
TAIR: WU-BLAST	www.arabidopsis.org/wublast/index2.jsp	Using BLAST to find T-DNA/transposon mutations in a specific gene or genes of interest
NASC: Insert BLAST	http://atensemble.arabidopsis.info/Multi/blastview	Using BLAST to find T-DNA/transposon mutations in a specific gene or genes of interest
AtIDB	http://atidb.org/cgi-perl/blast	Using BLAST to find T-DNA/transposon mutations in a specific gene or genes of interest
From phenotype to genotype—databases and tools for map-based cloning		
TAIR: Genetic Marker Search	www.arabidopsis.org/servlets/Search?action=new_search&type=marker	Finding and downloading sets genetic markers for mapping
TAIR: Polymorphism/Allele Search	www.arabidopsis.org/servlets/Search?action=new_search&type=polyallele	Finding polymorphisms between two ecotypes for generating new markers
TAIR: SeqViewer	www.arabidopsis.org/servlets/sv	Finding candidate genes in a genetically defined interval
TAIR: Monsanto Polymorphism Collection	http://arabidopsis.org/Cereon/index.jsp	SNPs and In Dels between Col and Ler Ler sequence
MASC SNP database	www.mpiz-koeln.mpg.de/masc/	SNP data for 12 ecotypes

The third column contains the subheading for the exact's protocol name described in detail in the text, or a brief description of other alternative tools that can be used to perform similar tasks.

to the Internet and a reasonably up-to-date computer should be able to perform all the steps in the protocols. A basic familiarity with computers, Internet browsers, and commonly used bioinformatics tools such as BLAST is assumed. There are a wide variety of textbooks, manuals and Web-based tutorials available for learning the basics of bioinformatics.

2.1. Computer Hardware and Software for Database Mining

The minimum requirements for database mining are a personal computer (PC), an Internet connection, and Web browsing software. Internet connection speeds are usually the rate-limiting step in web browsing. Because the amount of data being accessed from databases may be very large, it is desirable to have access through a high-speed network such as Ethernet, cable, or digital subscriber line (DSL) connection. To access databases, Web browser software, such as Internet Explorer, Netscape, or Safari, is required. Database interfaces should behave the same regardless of what operating system or browser is used. However, some functions may not work properly on older browsers. If possible, you should upgrade your browser to the most recent version available that can run on your operating system.

2.2. Databases

Databases are information storage and retrieval software systems. Typically, databases have three components: the database software for storing data, software that translates and executes requests (queries), and software applications that allow users to view data. *Arabidopsis* researchers have access to myriad databases that are either entirely devoted to *Arabidopsis* ("*Arabidopsis*-specific") or include *Arabidopsis* data along with information about other organisms. **Table 2** lists some of the main *Arabidopsis* databases and the types of information they contain. These databases are described briefly in the next section (**Subheading 2.2.1.**). There are many more *Arabidopsis* databases containing a variety of data types that are not included in the table. Links to these resources can be found at The Arabidopsis Information Resource (TAIR) (http://www.arabidopsis.org/info/2010_projects/Resources.jsp and http://www.arabidopsis.org/links/index.jsp). Some of these databases are about specific types of information, such as *cis*-regulatory elements or gene and enhancer traps, whereas others focus on specific classes of genes or disseminate data from a functional genomics project. A significant amount of *Arabidopsis* data can also be found in databases that contain information on other organisms, such as the National Center for Biotechnology Information's (NCBI) GenBank (http://www.ncbi.nlm.nih.gov/), the European Bioinformatics Institute's (EBI) InterPro (http://www.ebi.ac.uk/interpro/), PlantGDB (http://www.plantgdb.org/), and UK CropNet (http://ukcrop.net/).

2.2.1. Arabidopsis *Databases*

The databases described below are some of the primary sources of information about *Arabidopsis* and seed and DNA stocks. All these databases contain gene and protein data; however, they differ significantly in the breadth of information stored, the annotation methods employed, and the search tools and formats for displaying data. **Table 2** summarizes the main types of data found in each of these databases. Choosing which database to use depends on the task at hand and how well the database meets these needs in terms of content and accessibility. It can be both valuable and frustrating to have multiple places and ways to access data, particularly when different sources provide overlapping and sometimes inconsistent or conflicting information (*see* **Subheading 3.1.2.**).

2.2.1.1. Munich Information Center for Protein Sequences Arabidopsis thaliana Database (MatDB)

MatDB (www.mips.biochem.mpg.de/proj/thal/db/index.html) ***(2)*** contains structural and functional data about genes and proteins. Genes can be searched by locus identifier, Open Reading Frame (ORF) name (a nomenclature based on BAC or cosmid names), key words, and sequence similarity (BLAST). Results can be displayed on the graphical genome browser or in

Table 2
Types of Data and Tools Found at Some Primary Public Arabidopsis Resources

Types of Data/Databases	TAIR/ ABRC	TIGR	MIPS MatDB	AtGDB	AtIDB	SIGnAL	NASC
Alleles and polymorphisms	X						
Clones	X						
DNA stocks	X						
Full-length cDNAs	X	X	X	X		X	X
Gene expression—microarrays	X	X	X	X		X	X
Gene expression—Northerns, *in situ* localization	X						
Gene families	X	X	X	X			X
Gene structural annotations	X	X	X	X	X	X	X
Gene ontology annotations	X	X			X	X	
Genetic maps	X						X
Genetic markers	X						X
Genome browser	X		X	X	X	X	X
Metabolic pathways	X						
Nucleotide sequences	X	X	X	X	X	X	X
People and labs	X						
Protein localization	X	X	X				
Protein sequences	X	X	X	X		X	X
Protein structure annotation	X	X	X	X			X
Publications	X						
Seed stocks	X						X
Sequence analysis tools	X	X	X	X	X	X	X
T-DNA/transposon insertions	X				X	X	X

The left column shows a list of some major data types and tools that can be found in the public *Arabidopsis* databases listed in the top row.

table format. MatDB uses Functional Catalogue (FunCat) controlled vocabulary to classify gene functions. Protein families can be found using the MIPS in-house algorithm, SESAM, which can detect distantly related proteins.

2.2.1.2. The Institute for Genome Research (TIGR) *Arabidopsis thaliana* Annotation Database

TIGR's Arabidopsis Annotation Database (www.tigr.org/tdb/e2k1/ath1/) includes structural and functional annotations of all sequenced genes. Genes can be searched by keyword (description), locus name, location on a BAC or cosmid clone, or by sequence similarity (Wu-BLAST). Gene data are summarized in text and graphical formats. TIGR presents alignments supporting gene structures (primarily full-length cDNAs and expressed sequence tags [ESTs]) along with the annotation for validating gene structures. Along with TAIR, TIGR has been annotating gene functions using controlled vocabularies developed by Gene Ontology (GO) Consortium. TIGR has also analyzed and classified the Arabidopsis proteome into families (TIGRFams) that include newly identified groupings that may be specific to Arabidopsis or plants in general (*3*) (www.tigr.org/TIGRFAMs/index.shtml).

2.2.1.3. The *Arabidopsis* Information Resource and the *Arabidopsis* Biological Resource Center (ABRC)

TAIR strives to be a comprehensive resource for *Arabidopsis* (*4–6*). Like TIGR and MIPS, TAIR includes data for sequenced genes and also includes information about nonsequenced genes (genetic loci) and a wider variety of other data types (**Table 2**). As part of the GO Con-

sortium, TAIR collaborates in developing the vocabularies to accommodate plant genes and annotates *Arabidopsis* gene products ***(7)***. Data can be accessed through a variety of text-based searches and visualization tools such as the SeqViewer Genome browser. Sequence analysis tools are also provided. TAIR is also the gateway for finding and ordering seed and DNA stocks from the ABRC (www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/abrchome.htm).

The ABRC distributes seed in North America and DNA stocks worldwide. Stocks can be searched and ordered using the DNA and Germplasm search tools, or browsed through the Web catalog (http://arabidopsis.org/servlets/Order?state=catalog).

2.2.1.4. Nottingham *Arabidopsis* Stock Center (NASC)

NASC (http://nasc.nott.ac.uk/) provides seeds within the European community. NASC also maintains the recombinant inbred (RI) map for the Lister and Dean Columbia-Landsberg *erecta* mapping population, which is used for genetic mapping of mutations. Seed stocks are accessible via text searching or through a catalog browser (http://nasc.nott.ac.uk/catalogue.html). NASC has recently expanded the scope of its database to include a wide variety of genomic data (*see* **Table 2**).

2.2.1.5. Salk Institute Genomic Analysis Laboratory (SIGnAL) T-DNA Express Gene Mapping Tool

SIGnAL T-DNA Express Gene Mapping Tool (http://signal.salk.edu/) ***(8)*** was initially developed as a tool for researchers to quickly find genes with T-DNA insertions in transgenic lines generated by the SSP T-DNA project. SIGnAL has become a more general resource for reverse genetics and now incorporates flanking sequences from insertion lines developed by other projects, full-length cDNAs and other data (**Table 2**). Genes can be searched by name, sequence, description, or GO annotation, and are displayed in a whole-genome view based on TIGR's tiling path. Gene structural annotations are imported from TIGR. Links for ordering T-DNA line stocks are also provided, as well as tools for designing primers for sequencing the T-DNA insertions.

3. Methods

A primary objective of database mining for most researchers is to find out everything that is known about a specific gene or set of genes. Some of the basic questions are: What does my gene encode? In what biological processes is it involved? With what other genes/proteins does it interact? In what tissues is it located and how is it regulated? In order to generate a testable hypothesis and design meaningful experiments, the current available knowledge must be obtained and analyzed.

The published literature continues to be the primary medium for reporting experimental data and is the most comprehensive resource. The peer review process employed by journals and the inclusion of experimental methods along with the results ensures that data quality can be easily assessed and the experiments can be reproduced if necessary. Publishing in journals is a slow process, articles are typically accessible only to subscribers, not all data produced can be published, and the sheer volume of information makes it difficult to synthesize. Public databases fill an important niche in the chain of data and hypothesis-driven experimentation. Databases can provide large amounts of data quickly and without restrictions. Many include data that would otherwise not be published. The nature of the Internet and databases also facilitates the integration of information from multiple sources through hypertext links. A disadvantage of databases is that data may be incomplete and data accuracy and quality can sometimes be difficult to assess. The distinction between information that is well supported by experimental evidence and what is inferred or predicted is not always clearly presented in genomic databases. To be maximally useful, databases must always include enough information so that a researcher can evaluate the quality and methods behind the data. Likewise, researchers need to approach the data with a healthy skepticism and consider the methods employed in order to evaluate the data with the appropriate confidence.

3.1. Searching Literature Databases

Researchers have published a wealth of data about all aspects of *Arabidopsis* physiology, biochemistry and development. Databases such as PubMed, Agricola, or BIOSIS index articles from a wide variety of journals and can be used to find citations and articles in electronic or print format.

The National Center for Biotechnology Information (NCBI's) PubMed (www4.ncbi.nlm.nih.gov/PubMed/) is the primary database for life-science literature. At the beginning of 2004 the number of *Arabidopsis* publications in PubMed totaled 12,798. PubMed has a powerful search interface and links to the rest of databases within the NCBI system, such as sequence and expression databases. Other useful features include MyNCBI (Cubby), which can store searches that users run regularly to check for new items retrieved. Full-text copies can be ordered, provided one is a registered user. PubMed records are linked to publishers' sites for access to the full text of the article. For help using the resource refer to the PubMed tutorial (http://www.nlm.nih.gov/bds/pubmed_tutorial/m1001.html).

TAIR compiles bibliographic records about *Arabidopsis* from PubMed, BIOSIS, and Agricola. In addition, TAIR includes publications not found in these databases, such as abstracts from the *Arabidopsis* Conferences, defunct *Arabidopsis* electronic journals (The *Arabidopsis* Information Service and Weeds World), books, and dissertations. In early 2004, publications in TAIR totaled 21,982 records. Publication records may be linked to authors' community profiles at TAIR, which facilitates finding contact information.

The following protocols describe how to find *Arabidopsis* publications in PubMed and TAIR.

3.1.1. Finding Relevant Articles in the NCBI PubMed Database

1. Start at the PubMed search page (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed).
2. Enter the desired term(s) in the text input box. Searches can be restricted using the Boolean operators AND, OR, and NOT to combine terms. To search for a phrase, it must be enclosed in quotes (e.g., "transcription regulation"), or with a special flag "[tw]" (e.g., "transcription factor [tw]"). Use wild-card characters (*) for inexact matching. For example, to find all the articles about all the Agamous-like genes, type in "AGL*." For more refined searching, the Preview/Index function provides a guided search that can be used to match each term to specific fields such as author, title, or abstract. After making your selections, click on "Preview" to view the number of articles that match your query. If too many results are found, additional limits can be imposed. To complete the search click on the "Go" button.
3. Finding the article text and saving citations: The default display format is a summary of the citation. The complete citation, including available abstracts, can be viewed by clicking on the author's names. Articles that are available online are linked to the publisher's Web sites, which may be freely accessible or require a subscription. To modify the display of results, select the appropriate option from the display menu. For example, to import a citation into reference management software, choose MEDLINE format. References can be saved into a file for downloading or sent to an e-mail address. After selecting the articles by clicking on the checkboxes alongside the citations, choose the desired option under the "Send to" menu and click on the "Send to" button.

3.1.2. Finding Arabidopsis Publications Using TAIR's Publication Search

1. Start at TAIR's Publication Search page (http://arabidopsis.org/servlets/Search?action=new_search&type=publication).
2. To search with a specific name or phrase, enter the desired terms in the text query boxes and choose the field to search from the drop-down menu (author, title, abstract or title, URL for electronic publications, journal, or book title). For example, to search for all publications about oxidative stress, type the phrase into the text box and select "Title/Abstract" in the drop-down menu. Unlike the PubMed search, quotes are not required; all text in a single box is treated as a phrase. To restrict the search by publication dates or publication type, fill in the corresponding boxes.
3. Click on the "submit" button to start the search.

Results are displayed in a summary format including the title, journal, authors, and year. The title is hyperlinked to a page containing the complete citation, links to authors' TAIR profiles, the abstract, if available, and a list of associated key words and genes.

3.2. Finding Information About Genes in Arabidopsis Genome Databases

When searching for genes it is helpful to be aware of the different types of gene names and problems associated with *Arabidopsis* nomenclature. In *Arabidopsis*, experimentally defined genes have been named using gene symbols based on the gene product's function (e.g., DFR for dihydroflavanol 4-reductase) or mutant phenotype (e.g., TT3 for Transparent Testa3). In addition, all the genes identified by the genome sequencing effort have been assigned a systematic name based on the chromosomal location (e.g., AT2G27150; www.arabidopsis.org/info/guidelines.jsp). These AGI codes, or locus identifiers, uniquely define a locus. One source of frustration is that different databases can sometimes have inconsistent or conflicting information about the same gene. After the *Arabidopsis* genome sequence was completed, TIGR was funded to reannotate the complete genome. This effort resulted in the consecutive release of five genome versions. In each release new loci have been added and many existing loci have been modified (e.g., gene models have been merged or split into two gene models) ***(3,9,10)***. TAIR has incorporated each different genome version released by TIGR, and has now replaced TIGR in the task of maintaining and updating the annotation of the genome. MIPS has also been involved in the annotation of the *Arabidopsis* genome independently using its own computational methods for genome and proteome analysis (PEDANT) ***(2,11,12)***. As a result, there have been some gene structure and naming discrepancies among TIGR, TAIR, and MIPS. To help follow the history of each gene identified by TIGR and MIPS use the Locus History tool available at TAIR (http://arabidopsis.org/tools/bulk/locushistory/index.jsp).

3.2.1. Finding Gene Information by Name

This protocol describes how to use TAIR's gene search to find genes by name, GenBank accession or description.

1. Start at TAIR's Gene Search: www.arabidopsis.org/servlets/Search?action= new_search&type=gene.
2. Define the name search criteria. Select name from the "Search Name" drop-down menu. This option is used to search by symbolic names (e.g., ABI3), full names (e.g., Abscisic Acid Insensitive 3) or locus identifier (e.g., AT3G24560). The name search also includes aliases.
3. Choose an exact or inexact search mode. When searching with a gene symbol choosing the "starts with" option is a way to find similarly named genes, such as members of a gene family. When searching with a GenBank accession, it is better to use an exact match in order to avoid retrieving spurious results. To search for a word or phrase within a gene description, choose the "contains" option.
4. Select the output format. The default values are 25 records, sorted by name. The position option can be used when finding genes by location.
5. Click "Submit Query" to start your search.
6. Finding information about the gene: The results display locus and gene model names that link to detailed information pages. The locus detail page collects and displays all the information associated to each locus and should be the starting point for finding comprehensive information about a gene.

Locus information includes a description, list of alternative names (aliases) for the locus, the chromosomal coordinates, and the date the record was last updated. The last item is useful as a means to assess the currency of the data. Locus information includes sequences and gene/protein structure and function. Clicking on the link from any of the sequences in the "Nucleotide Sequence" section will retrieve the sequence. Functional information about the gene may include Gene Ontology annotations, metabolic pathways, gene expression, and protein localization data. Other associated information may include genetic markers, alleles, polymorphisms,

transcript clones that map to the locus, publications, and people working on the gene or contributing data. The structure of the gene, along with any objects that map to the locus, such as transcripts, T-DNA/transposon insertions, markers, and polymorphisms can be displayed graphically by clicking on the link to the SeqViewer genome browser. Detailed information about all associated data can be obtained by clicking on the name of the object (e.g., polymorphism/allele name).

3.2.2. Finding Functional Information About Genes

To make data about a gene's function more amenable to computational methods of querying and analysis, many databases use structured controlled vocabularies for annotating gene products. As described above, both TIGR and TAIR have actively annotated *Arabidopsis* genes using vocabularies developed by the Gene Ontology (GO) Consortium (www.geneontology.org). The GO vocabularies describe three aspects of a gene product: molecular function (the biochemical activity of a gene product), biological process (the ordered assembly of more than one molecular function), and cellular component (location within the cell) ***(13)***. These vocabularies have been widely adopted by many model organism databases and are considered to be the standard for functional annotation. At TAIR, each annotation to a GO term includes a description of the evidence for the association and links to reference(s) supporting the annotation ***(7)***. Genes for which no experimental or computational data are available are annotated to the term "unknown" to distinguish them from genes that have not yet been annotated. As of January 2004, more than 27,000 loci (approx 78% of the genome) have been annotated to one or more GO terms ***(7)***. Of those, only about 3600 genes have been annotated based on experimental information extracted from the literature. The rest of the annotations are based on sequence similarity to known genes or from inferences based on computational predictions.

GO annotations provide an efficient mechanism for finding functional information for a gene or to find relationships within groups of genes, such as members of a gene family or clusters of genes with similar expression profiles. The following protocols describe how to retrieve GO annotations to classify sets of genes and find genes with related functions.

1. Start at the TAIR GO Annotation Bulk Download and Analysis tool (www.arabidopsis.org/tools/bulk/go/index.jsp)
2. Input the locus identifiers in the query box. Type, paste, or upload a file containing your list of locus identifiers.
3. Define the output options. Select HTML to view the results hyperlinked. Choose text for saving the results as a file.
4. To obtain a list of annotations click on the "Get all GO Annotations" button at the bottom of the page. From this page it is possible to find proteins annotated similarly in other organisms. To find other *Arabidopsis* genes with similar annotations and to see a term's definition click on the name of the locus, scroll to the Annotation section, and click on the term's name. The GO vocabularies are extremely detailed and include well over 12,000 terms that are organized into hierarchies. A simpler ontology (GO Slim), consisting only of the higher-level terms that places genes into broader categories, is also available.
5. To retrieve a group of genes classified into broader categories, click on the "Functional Categorization" button. Genes with specific functions, such as negative regulation of transcription, will be grouped according to broader category of the process of transcription. Keyword category refers to the type of GO term (Biological Process, Cellular Component, and Molecular Function). The second column corresponds to the GO Slim term that defines the broad category. Frequency refers to the total number of annotations (associations of genes to terms) that appear in the set of genes.
6. Visualizing the data as a graph: The functional classification table data can be transformed into a pie chart by clicking on the "Pie Chart" button at the top of the results page. The results are displayed as three separate pie charts, each grouping annotations for one of the three GO categories. The chart includes a key that shows the individual GO Slim terms, the percentage of total annotations represented by the term, and the total number of annotations to the term (raw value).

3.3. Finding Information About Gene Expression

An important tool for finding functional information comes from the analysis of gene expression. There are many reasons to analyze expression data, such as finding the pattern of expression of a gene in an organism, determining the effect of the environment on the expression of particular genes, or understanding how the expression of one gene affects the expression of other genes.

In *Arabidopsis*, several methods have been applied to study gene expression. Besides the traditional gene-by-gene approach using Northern, reverse transcription-polymerase chain reaction (RT-PCR), *in situ* hybridization, and β-glucoronidase/green fluorescent protein (GUS/GFP) reporter methods, the *Arabidopsis* community has invested substantially in applying large-scale methods that allow monitoring the expression of thousands of genes at once. Among those methods are DNA microarray, massively parallel signature sequencing (MPSS), and serial analysis of gene expression (SAGE). In addition, a large collection of expressed sequence tag (EST) sequences and gene trap lines are available, which can also be used in the analysis of gene expression. Expression data obtained by the use of traditional methods can be found mainly in the literature, whereas data from high-throughput methods are for the most part stored in databases. In this section we describe how to find expression information using public repositories that contain DNA microarray, MPSS, and EST data.

3.3.1. Searching DNA Microarray Data

DNA microarrays are one of the most powerful tools for investigating the expression patterns of thousands of genes in parallel, and it is now a common technique in many *Arabidopsis* labs. The widespread use of this technology has been facilitated by the existence of several public and commercial array designs for *Arabidopsis* and by the establishment of publicly funded projects, such as Arabidopsis Functional Genomics Consortium (AFGC) ***(14)*** and NASCArrays ***(15)***, which offer microarray services at affordable prices. As a result, a vast amount of microarray data has been generated, and a considerable portion of it can be found in public repositories. Some of the databases containing *Arabidopsis* microarray data are shown in **Table 2**. Public *Arabidopsis*-specific microarray databases are found in NASC and TAIR. In addition, some laboratories or projects have developed private databases to store the data they produce, and in some cases have made their data available on the Web. For links to some of those projects, and related microarray information, see http://arabidopsis.org/info/expression/index.jsp. Other microarray public repositories, such as NCBI's Gene Expression Omnibus (GEO), the Stanford Microarray Database (SMD), or ArrayExpress, include microarray data from several organisms, including *Arabidopsis*.

The TAIR microarray database supports both spotted and chip array technologies. TAIR currently houses two-channel cDNA-based arrays from the AFGC project ***(14)*** (115 experiment sets with 515 slides), and is adding Affymetrix-based arrays from AtGenExpress (http://arabidopsis.info/info/masc_annual_june03.pdf) and NASCArrays ***(15)*** (around 3000 chips total). In the near future TAIR will integrate *Arabidopsis* data from ArrayExpress (www.ebi. ac.uk/arrayexpress/) and individual users. TAIR microarray data are subjected to extensive curation and annotation. The result is a consistent representation of experiment and sample information. Also, raw data are renormalized at TAIR to filter out low-quality spots and to facilitate data comparison. TAIR provides the arithmetic mean and standard error measures of gene expression for each gene from replicated spots per array (if available), replicated hybridizations (if available), and across all arrays. Also, mapping of array elements to genes based on BLAST analysis is also provided (ftp://ftp.arabidopsis.org/home/tair/home/tair/Microarrays/AFGC/; ftp://ftp.arabidopsis.org/home/tair/home/tair/Microarrays/Affymetrix/). The renormalized data are accessible through text searching, and both the original and the TAIR-normalized datasets are available for download. Moreover, the AFGC dataset has been subjected to two clustering analyses at TAIR. One analysis clustered all slides against all slides, and the whole dataset is available for download, along with the accompanying visualization software VxInsight (ftp://ftp.arabidopsis.org/home/tair/

home/tair/Software/VxInsight/). In the second analysis, each slide was clustered against all the slides belonging to the same experimental category. These data can be visualized using the Expression Viewer tool.

NASCArrays (http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl) is a public repository for Affymetrix-based microarray data generated by the NASC facility as a public service *(15)* to the community. It provides access to experiment information and complete datasets are available for download. Query for specific gene expression profiles are not possible, but they can be viewed through a series of data mining tools. "Spot history" allows one to see the pattern of gene expression for each gene over all slides in the database. "Two-gene Scatter" allows the user to see the pattern of gene expression over all slides for two genes as a scatter plot. "Gene Swinger" looks at the experiments in which the expression of a given gene varied most. Samples and experiments are described in detail. At the time of writing, March 2004, NASC Arrays contained 58 experiments, and it is anticipated to produce up to 1000 chips.

3.3.1.1. Finding Expression Patterns of Genes in Different Microarray Experiments

This protocol shows how to use TAIR microarray database to find the expression profiles of a gene or genes in specific experiments.

1. Start at the TAIR Microarray Expression Search (http://arabidopsis.org/servlets/Search?action=new_search&type=expression).
2. Enter the gene name of interest in the query text box and select the appropriate type from the tab. Gene symbols (e.g., ap2); locus identifiers (e.g., At5g01810); array element names (e.g., 39B5T7 or at_34561), or GenBank accessions (e.g., T13762) can be used. To search for more than one gene at a time, upload or paste a list of locus identifiers or array element names in the bigger query text box.
3. Select the output options (optional). This can be used to select the format in which the results will be presented. The default options are 25 records per page, sorting results by fold change, and green/red color scheme to display expression values. Once the results are returned to the browser it is possible to re-sort by other criteria.
4. Limit search by Expression Parameters (optional). Click on the plus sign alongside this heading to display the available parameters. This is an advanced option to restrict the search to only certain expression values. "Absolute expression" allows selection of qualitative absolute expression values, such as "expressed," "not expressed," or "absent." The "expressed" and "not expressed" option retrieves all the signal intensity values higher or lower, respectively, than an arbitrary expression value (350 in both channels for dual-channel arrays). The "absent" option retrieves the unreliable values only. "Relative expression" allows selection of array elements/genes that are differentially expressed, either upregulated or downregulated. The default is any relative value. "Std error" allows selection of the confidence level of the measurement. The lower the value, the most reliable is the expression signal. "Analysis level" allows you to retrieve only the expression measurements of replicate arrays ("replicate sets"), or all the arrays ("slide/chip").
5. Limit search by Experiment Parameters (optional). This is an advanced option to restrict a search to only certain experiments. Click on the plus sign alongside this heading to display the available parameters. Experiments can be searched by name, author, description and key word. Key word searches include experimental variables, plant tissue, experiment category (e.g., biotic treatment), and experiment goal. If no limits are imposed, all the experiments in the database are searched.
6. Limit search by Array Design (optional). This option is used to restrict the search to only certain array designs (e.g., Affymetrix 25K). The default option includes all array designs in the database.
7. Submit the query.

The results display the expression measurements (average fold change and standard errors) for each array element in each slide and/or replicate set. The values are color-coded according to expression level and whether the change in expression was positive or negative. Each record

also includes links to pages with detailed information about the experiments and array elements. The results can be downloaded in text format by clicking the check boxes for the records of interest, or selecting all records by pointing to the "Check All" button, and then clicking "Download."

When interpreting the results obtained from microarray analysis, several issues should be taken into consideration. The first one is related to the complexity of the microarray technology. All the numerous steps involved in microarray analysis, from array elements preparation to array manufacture, RNA preparation, hybridization, image processing, and data normalization, are potential sources of variability that can affect the accuracy of the final observation ***(16)***. Many systematic errors are removed by data transformation and normalization ***(17)***, but issues related to sensitivity and background noise, among others, are inherent to the technology and cannot be resolved. An additional error factor is the potential for cross-hybridization. cDNA-based microarrays have higher potential for cross-hybridization, and thus the intensity signal associated with a spot may be a composite of signals from related genes ***(14)***. Also, because hybridization properties vary from sequence to sequence, averaging results from different sequences (array elements) to obtain expression summaries for one gene can be misleading. Additionally, not all the clones spotted on the arrays are fully sequenced, which together with potential mislabeling of clones, results in the measurements associated to a spot not corresponding to the real gene ***(14)***. For example, 1.5% of EST clones used in the AFGC arrays were found to be mislabeled ***(14)***. Moreover, gene-specific dye effects can also add variability to the dual-channel arrays ***(18)***.

3.3.1.2. Finding Expression History of a Gene Across All Microarray Experiments

One way to estimate the variability of a gene's expression profile across all experiments is by plotting the distribution of expression values in those experiments. Both NASCArrays (http://affymetrix.arabidopsis.info/narrays/spothistory.pl) and SMD (http://genome-www5.stanford.edu/) offer tools designed for this purpose. This protocol demonstrates how to use the NASCArrays Spot History tool to find the expression history of a gene across all experiments in the database.

1. Start at the NASC Spot History page (http://affymetrix.arabidopsis.info/narrays/spothistory.pl).
2. Input the gene of interest in the query box and select the type of name from the drop-down menu (AGI code, symbolic name, or Affymetrix probe set name). If desired, check the corresponding box to plot the *x* axis in log scale.
3. Submit your query by clicking on the "Plot!" button.

This search will return a histogram of the frequency of each signal value bin.

To find detailed information about the experiments that have this expression level, click on a bar in the histogram.

3.3.2. Searching Massively Parallel Signature Sequencing Data

Massively parallel signature sequencing (MPSS) is a powerful technique to quantitatively measure gene expression ***(19)***. The relative abundance of the signatures in a given library represents a quantitative estimate of expression of that gene. The advantages of MPSS is the high precision and sensitivity of the expression levels, allowing the detection of very low expression levels not detectable with microarray nor EST methods. However, the cost is very high, so it is expensive to analyze many treatments. In *Arabidopsis*, only one public project developed in the Meyers lab has used this technique so far, and the data have been made public on the Web (http://mpss.udel.edu/at/). The *Arabidopsis* MPSS data set includes the signatures obtained from sequencing 14 different libraries from different tissue types (e.g., leaf, root, silique) and treatments (e.g., salicylic acid). The data can be queried by locus identifier, key word, and signature sequence, as well as by chromosomal position. Advanced query tools allow restricting the

search to one or more libraries, and a visualization tool is also provided to see pairwise comparisons of five of the libraries. This protocol shows how to search MPSS data with locus identifiers.

1. Start at the MPSS database Simple Query section (http://mpss.udel.edu/at/).
2. Type either the AGI code name (e.g., At2g34560) or paste the sequence in the appropriate box.
3. Submit the query by clicking on the "Get Data" button.

The results show a graphical representation of the gene structure overlaid with all the matching signatures. The signatures are classified according to their position on the DNA strands and the direction and gene features they fall in, and are color-coded accordingly to the class. A table showing all the signatures that matched the gene is also shown. The columns in the table include the signature sequence, class, number of hits, DNA strand, and abundance (transcripts per million [TPM] values) in each of the 12 libraries assayed. Gray TPM values indicate a lack of significant expression. A weakly expressed gene value is in the range of 1–10 TPM, while a very strongly expressed gene may be more than 1000 TPM.

When interpreting MPSS data, it is important to note the signature's class. Class 1 signatures fall anywhere within a transcript or the 3' untranslated region on the genes defined according to the available genome version. If a signature matching an intron (class 5) shows evidence of expression, the best explanation is that there is an alternative splice site, which means that is most likely part of an unannotated exon. Antisense signatures (class 3) may indicate the presence of an antisense transcript, although they could also result from mispriming of the oligo-dT during first-strand cDNA synthesis. Signatures outside a gene region may correspond to an unannotated gene. Moreover, not all signatures are unique in the genome. Around 10% occur two or more times (http://mpss.udel.edu/at/java.html?). Signatures that are not unique may result from duplications of genomic regions or of individual genes.

3.3.3. Searching Expressed Sequence Tags

Expressed sequence tags (ESTs) provide researchers with a quick route for discovering new genes, for obtaining data on gene expression and regulation, and for structural annotation. For many genes, the existence of EST matches is the only evidence for their expression. Information about the source of the cDNA library provides clues on expression localization and regulation of the gene. For example, a cDNA cloned from a leaf library indicates that the corresponding gene is expressed in leaves. In some cases, the number of EST clones in a library can be used to infer transcript levels, but EST data are not generally considered to be quantitative.

UniGene clusters (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene) are built using pairwise similarity searches with nonredundant EST sequences from the dbEST division of GenBank. Each cluster contains sequences that represent a unique locus, as well as library information and map location. For a sequence to be included in a UniGene cluster, the insert sequence must have at least 100 bp of high quality sequence, not be repetitive, and have a 50-bp overlap in the 3' UTR with 100% identity. The UniGene Web site allows the user to view UniGene information on a per-cluster, per-sequence, or per-library basis ***(20)***.

TIGR has assembled *Arabidopsis* ESTs into tentative consensus (TC) sequences and provides the results as a service to the community (www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=arab). Assembly was made by clustering the EST sequences and then assembling the clusters into TCs. EST clusters and transcripts are clustered together only if they meet a certain criteria (minimum of 40 bp match, greater than 94% identity in the overlap region, and maximum unmatched overhang of 30 bp). AtGI allows searching by sequence using WU-BLAST, or using text-based searches, including GB identifiers, tissue and library names, and GO functional categories.

3.3.3.1. Using NCBI's UniGene to Find ESTs for a Sequenced Gene

1. Start at the NCBI's UniGene Search page (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene).

2. Type in the gene name or locus identifier in the text box. To restrict the search terms to specific fields, click "Preview/index." For example, to search for EST data on the Arabidopsis COP9 gene type in "COP9." Click the AND button and choose "organism" in the fields tab menu, and type in "Arabidopsis thaliana." The resulting query will read "COP9 AND organism [Arabidopsis thaliana]."
3. Submit the query.

The results will display a list of gene clusters that match your criteria. Clicking on the clusters brings a page with information on the libraries where the sequences were derived, as well as protein similarities with other organisms, mRNA and EST sequences, and links for downloading sequences. The EST library information is not curated, and as such may not be correct or may be described inconsistently. Users should keep in mind that not all the *Arabidopsis* genes are represented by a UniGene cluster.

3.3.3.2. Using ESTs to Find Genes That Are Differentially Expressed

This protocol describes how to use TIGR's AtGI database to find genes that are differentially expressed in two tissues or libraries. The program identifies TC sequences having statistically significant differential expression in two libraries.

1. Start at the AtGI Library Expression Search (www.tigr.org/tigr-scripts/tgi/libtc.pl?db=atest).
2. To compare tissues select the desired tissue by clicking on its name from the "A" tissue menu box. Press the Ctrl key while clicking to select more than one tissue, and then select the second tissue from the "B" menu box. To compare libraries, use the drop-down menus on the right, selecting a name from each box. To find information about the libraries click on the "Library Description" link.
3. Restrict the search to libraries with a certain number of ESTs (optional). Some libraries contain a small number of EST sequences, and the comparison against bigger libraries may be biased.
4. Press the "Get Expression" button.

The results display a list of TCs matching your query, along with the number of ESTs in the TC and the number of ESTs in each library or tissue compared. Significant differential expression is identified using the "R statistic." A large R indicates that is a significant bias toward one or more libraries in that TC. TCs with R-values in the top 5% are indicated with an asterisk and highlighted in red.

3.4. **Arabidopsis** ***Metabolic Pathways***

Of the 26,207 protein coding genes included in the latest release of the genome sequence (TIGR version 5.0, January 2004) about 8170 unique loci can be classified as enzymes based on their GO annotations. To date, approx 1200 have been associated to metabolic pathways in the AraCyc Pathways Database (www.arabidopsis.org/tools/aracyc/) developed at TAIR in collaboration with SRI International ***(21)***. AraCyc contains information about *Arabidopsis* enzymes, biochemical pathways, reactions, substrates, and product compounds. The database was initially built using computational methods to generate pathways based on the correlation of predicted enzymes from the *Arabidopsis* genome with components of known pathways in other organisms ***(21)***. AraCyc pathways are then manually curated to correct, modify, and otherwise update existing pathways, and to add new pathways. When examining metabolic pathways in AraCyc it is important to bear in mind that computationally predicted pathways may be inaccurate and should be considered only as a starting point for experimentation. In addition, some pathways may also be incomplete, reflecting the current state of knowledge. There are several ways to search and browse AraCyc. The first protocol (**Subheading 3.4.1.**) illustrates how to find information about enzymes, compounds, and reactions for a specific pathway. The second protocol (**Subheading 3.4.2.**) describes how AraCyc can be used to analyze changes in the expression of genes involved in metabolism.

3.4.1. Finding Metabolic Pathways, Reactions, Enzymes, and Compounds

1. To search for information about a specific pathway, start at the main AraCyc query page at www.arabidopsis.org:1555/index.html.
2. Finding pathways: To search for pathways, choose "Pathway (by name)" and type the name of the pathway of interest into the input box. For example, type in "Lignin biosynthesis" to view this pathway. The results are displayed in a graphical format. Compounds are shown in red and reactions are indicated by blue arrows. The default view shows the broadest outline of the pathway with the least detail. A brief description of the pathway is shown in the comments section. The pathway evidence glyph summarizes the evidence supporting the reactions in the pathway. If an enzyme is present in Arabidopsis and the reaction is unique to the pathway, the reaction is well supported for that pathway. Conversely, a reaction in the pathway may not be well supported because the enzymatic function has not been identified in Arabidopsis or the reaction is present in multiple pathways.
3. Viewing the pathway in more detail: Pathways can be displayed in greater detail by clicking on the "more detail" button at the top of the diagram. At the highest level of resolution the display includes E.C. number, enzymes, locus, individual reactions, and compounds.
4. Displaying individual reactions: Clicking on the blue arrows will display a new page showing details about the specific reaction, including catalytic enzymes, a reaction diagram showing the structure of the compounds, a list of all pathways that include this reaction, and a gene-reaction schematic showing all enzymes/loci that catalyze the reaction.
5. Displaying information about specific enzymes: Clicking on the name of the enzyme will display a new page with a summary of the reactions catalyzed by the enzyme and all pathways in which the enzyme is known (or is predicted) to participate. To view the TAIR locus detail page for the enzyme, click on the symbolic gene name or locus identifier.
6. Displaying individual compounds: Clicking on the compound will display the compound formula, SMILES string, molecular weight, and a list of reactions for which the compound is either a reactant or a product.

3.4.2. Detecting Changes in Expression of Genes Involved in Metabolism

The Pathway Omics Viewer allows displaying expression data on the pathway diagrams, and thus can be used as a tool to examine how enzymes may be regulated in response to experimental treatments in a pathway context. Expression values are displayed on the pathway overview map as color-coded reaction lines indicating the expression level of the enzyme that catalyzes the reaction. For time-course experiments, data can be displayed as an animation, and the changes in gene expression are easily distinguished by following the color changes in a particular pathway for each time point. The animation can be stopped at any point to facilitate scanning and querying of the pathways.

1. Start at the AraCyc Omics Viewer at www.arabidopsis.org:1555/expression.html.
2. Click on "Browse" to find the expression data file in your local computer. The file should be a tab-delimited text file. Word or other word processing software files cannot be used. The first column must be the AGI locus identifier. The remaining columns should contain the expression values (absolute or relative) from each treatment. To display data from a time series, the expression values corresponding to consecutive time point should be included in the same order in consecutive columns (e.g., values for time point 1 in column 1, values for time point 2 in column 2).
3. Select the types of expression values present in your file (absolute or relative) and the number of data columns to be analyzed. If the data contain absolute values in a single column, choose "absolute" and "single" column display. If the values in the columns are already relative values, select only a single column to display and select relative expression.
4. Specify whether the data file has log values. Choose this option to display log values. If not, negative values will be discarded.
5. To display a single data column or ratio of two data columns, select the number of columns to display. By default, the first column of locus identifiers is 0 and the first data column is 1. Columns that are not selected will not be displayed on the Overview diagram. For example, to

display only the first data column, type in 1 or to show three time points in a file containing absolute values, type in 1, 2, and 3 in the single column data input box.

6. Specify a value for the maximum cutoff. By default the cutoff values are determined by the data values. To compare expression from different experiments, select the same maximum cutoff value for each display.
7. Click "Submit." The results will be displayed in the Pathway Overview Omics Viewer, which shows the metabolic pathway diagram colorized with expression data according to the relative or absolute expression level of the gene that codes for the enzyme that catalyzes that reaction step. The display includes a histogram showing the distribution of values in the dataset and a key to the diagram. Statistics, including minimum and maximum values and missing information, are shown on the bottom of the display, including the loci in the expression file that were not found in AraCyc.

3.5. Finding Related Protein Sequences in Arabidopsis

For sequenced genes for which limited experimental data are available, one of the first steps toward understanding a gene's function is to search for evolutionarily related proteins and conserved motifs. The function of an unknown gene may be inferred from its similarity to a well-characterized homolog or from the presence of conserved domains. Other motifs, such as transmembrane domains or signal peptides, can be used to infer protein localization. Structural classification based on protein folding is another way of finding more distantly related proteins. The assumption is that protein topology is functionally important and proteins with similar structures may have similar molecular activities.

A comprehensive analysis of protein families and domains typically requires queries of multiple databases and algorithms and can be very time-consuming. To facilitate gene family and domain analysis, TIGR, TAIR, and MIPS databases routinely analyze the entire *Arabidopsis* proteome as part of their annotation "pipelines" ***(2–4)***. Such a catalog of structural features facilitates the classification of domains that are over- or underrepresented in a genome. It also makes possible the grouping of members of gene families according to domain composition and order.

3.5.1. Finding Similar Protein Sequences

Searching for similar protein sequences in *Arabidopsis* using local sequence alignment methods can be performed at TAIR, TIGR, MIPS, and NCBI **(Table 2)**. These groups have overlapping *Arabidopsis* datasets; TAIR has some other *Arabidopsis*-specific datasets not found at TIGR, MIPS, or NCBI (www.arabidopsis.org/help/helppages/BLAST_help.jsp#datasets). These datasets are used by all of TAIR's sequence similarity programs (FASTA, WU-BLAST, and NCBI BLAST and PatMatch).

This protocol illustrates how use TAIR's WU-BLAST tool to identify similar genes in *Arabidopsis*.

1. Start at the WU-BLAST search page. Point your browser to the URL www.arabidopsis.org/wublast/index2.jsp.
2. Select the appropriate BLAST program. Five different algorithms are available to match amino acid or nucleotide sequences. The choice of the program depends on the type of sequence to be queried and the query database. For example, when comparing a protein sequence to other protein sequences choose the BLASTP program (www.arabidopsis.org/help/helppages/BLAST_help.jsp#methods).
3. Input your query sequence. The tool accepts sequences or locus identifiers as input. To use a sequence as input, paste in the sequence as raw text or in FASTA format, or upload it from a file. Sequences pasted directly from GenBank records can also be used. To use a locus identifier as input, choose the locus name option under the input header, and type in the name of the locus, or upload it from a file. When using locus identifiers as input the program retrieves the coding sequence (CDS) for the locus; therefore it cannot be used with the BLASTP or TBLASTN options. To perform a search using more than one query sequence, submit multiple

sequences as a list of locus identifiers or as a set of FASTA formatted sequences, each sequence having its own FASTA header.

4. Define the dataset to search against. For example, to find homologous proteins in Arabidopsis choose the AGI protein dataset. This dataset is a nonredundant set of all known Arabidopsis proteins.
5. Customize the BLAST search parameters. The default parameters are filtering on an expect threshold (cutoff) of 10. The default S value is calculated based on the E value and represents the single high-scoring pair (HSP) score that satisfies the expect threshold.
6. Submit the query. Click on the "RUN BLAST" button. If you have chosen an inappropriate combination of query sequence and database, an error will be returned to your browser.

Results from the WU-BLAST search are presented in a graphical format that can be used to rapidly assess the significance of the results. The graph displays the query sequence in red and the HSP matches below. The length of the bar corresponds to the length of the HSP and the color of the bar indicates the range of expected values (the probability of finding the sequence match by random chance). The direction of the bar indicates whether the match is on the forward or reverse strand. Pointing the mouse over the HSP markers will display the description line of the matched sequence. Clicking on the HSP will display the selected sequence alignment. For AGI genes and loci, the name in the alignment is hyperlinked to the TAIR locus detail page (*see* **Subheading 3.2.1.**). Other similarity groupings can be found using the External links from locus detail pages. Both TIGR and MIPS databases include precomputed matches to similar proteins in their locus records. NCBI Blink shows similar proteins in GenBank. The Blink results show BLASTP searches against all proteins in GenBank in a graphical format along with the similarity score. To find the best matches for each taxonomic group, click on the button labeled "Best Hits." To access the GenBank record for each sequence, click on the sequence accession number.

3.5.2. Finding Similar Structures and Domains in Proteins

Comprehensive protein structural data can be found at TIGR, MIPS, NCBI, and TAIR (**Table 2**). The protocol described here demonstrates how to use TAIR's Bulk Protein Download tool to obtain a list of structural, physical, and chemical properties for a set of proteins.

1. Start at the Bulk Protein Download. Point the browser to the URL www.arabidopsis.org/tools/bulk/protein/index.jsp.
2. Choose the output display. The output options include molecular weight, isoelectric point, transmembrane domains, SCOP structural class, domains, and hyperlinked SwissProt IDs. Selecting the HTML option will display links to TAIR locus detail pages, protein sequence, SeqViewer graphical display, the protein record in SwissProt, and INTERPRO. The last two links are shown only if domains and Swiss Prot IDs are included in the output. Choose "text" output if you wish to download the data into your computer. Queries that return more than 1000 results will be returned as text-only format.
3. Limit the search by protein properties. For example, to obtain a list of proteins with a given range of molecular weights.
4. Submit the query by clicking on the "Get Protein Data" button.

Protein structural annotations may not be constant from database to database because different analysis methods or sequences are used. Domain databases are also updated frequently as new domain structures are identified. Frequent checks of genome databases should be done to determine whether new domains have been identified.

3.6. From Genotype to Phenotype—Finding Mutants for Analysis of Gene Function

By analyzing the effects of mutations it is possible to infer the function of a gene and, through epistasis analysis, define networks and interactions among genes. Different classes of mutations provide diverse tools for genetic analysis of gene function. Typically, when trying to ascertain a gene's function, knockouts or null mutations are preferred to determine the effect of

a complete loss of gene function on the phenotype of an organism. In some cases, a knockout does not give a detectable phenotype. There are many reasons why this may occur. If the gene is a member of a closely related gene family, the effect of the mutation may be masked by the activity of homologous genes. In this case double, triple, or even quadruple mutants may be needed before a phenotype can be detected. Loss-of-function mutations may also result in embryonic lethality or sterility, thus complicating the analysis. Another possibility is that the phenotype may not be observable under the growth conditions used. Correlation with functional annotations and gene expression data may be useful in determining when, where, and under what conditions a phenotype can be observed. Dominant gain-of-function or knock-on mutations can be useful in revealing the functions of genes that do not have a phenotype when knocked out, by causing over or ectopic gene expression ***(22)***. Analyzing the phenotypes of point mutations can provide important information about specific motifs ***(23,24)*** and can be a preferred source material for dissecting genetic pathways through enhancer/suppressor screens. This section describes how to find knockouts and other types of mutations in a gene or genes of interest.

3.6.1. Finding Knockout Mutations in a Gene

There are several large scale functional genomics projects that utilize modified T-DNA or transposons aimed to generated knockout mutations in every *Arabidopsis* gene ***(25)***. TAIR, AtIDB, SIGNAL T-DNA Express, and NASC have identical or at least largely overlapping datasets of the major T-DNA and transposon stocks (**Table 2**). See the appropriate database release notes to determine which sets of insertions are stored. The following method describes how to use BLASTN at the SALK T-DNA Express Gene Finding Tool to locate insertions in a gene of interest. The gene sequence is used to query a database of genomic sequences flanking the site of an insertion. T-DNA and transposon insertion sites are estimated, based on matching the flanking sequences to the genome. This calculated position is not always correct due to sequence errors and incomplete matches to the genome (http://signal.salk.edu/tdna_FAQs.html).

1. Start at the T-DNA Express search page (http://signal.salk.edu/cgi-bin/tdnaexpress).
2. Input the query sequence and select the search parameters. Paste your sequence into the input box and select the number of results to display and a cutoff similarity score (E-value). Choose a smaller E-value to display the best matches and a larger one to find lower-quality matches. For best results use the cDNA or coding sequence and BLASTN and the default cutoff value. The best matches may still be inexact, as genomic sequences flanking insertions are generally of moderate quality, and may be derived from a different genetic background than the Col-0 ecotype used for the genomic sequence. Submit the query.
3. Find the insertions. At the top of T-DNA Express BLAST report will be a summary box showing the locations of the best match to your sequence. The alignments are shown below the summary. To display a graphical view in T-DNA express click “To TDNA Express” hyperlink in the summary section. This will open a genome browser centered on the gene, which matched your query sequence. Genes are located on the top of the chromosome bar and insertion flank sequences are annotated below. The starting point of the arrow indicates the position of insertion and the arrowhead indicates the direction of the sequence match. Zoom controls can be used to obtain close-up views in order to better see where the insertion flanks are located within the gene. In the gene annotation band, exons are identified as green bands and introns are clear/white bands. To view detailed information about any insertion, click on the name of the insertion as it appears in the lower band. From the detail page you can see the exact coordinates and determine whether the insertion is in an exon, intron, intergenic region, or splice junction. To order stocks, use the appropriate links to NASC, ABRC, or FLAGDB.

TAIR, AtIDB, and NASC all have similar functionalities, but there are a few important differences to note. NASC BLAST differs in returning only alignments to the insertion flanking sequences, which are not linked to a graphical genome browser. NASC also features the Insert Watch utility, which allows researchers to submit a sequence of interest and receive notification when a new insertion flank sequence has been found with a significant match to the query sequence of interest. Researchers are automatically notified via email of the existence of

a new insertion that is potentially located in their gene of interest. If using TAIR's BLAST, clicking on the T-DNA flank sequence name from the BLAST results will open a page displaying the stock details for the T-DNA insertion line. To see a graphical display of the insertion site in the genome, click on the locus link, scroll down to the Map Links, and click on the Sequence Viewer. Registered users can order stocks from ABRC or NASC directly from this page. TAIR BLAST also allows for batch BLAST queries to identify insertions in more than one gene (e.g., multiple members of a gene family).

3.6.2. Finding Other Types of Mutations and Mutants With Similar Phenotypes

Mutants with similar phenotypes might be new alleles of known genes or function in the same genetic pathways. Quantitative trait analysis of natural variants can be used to find genes that may not be identified using standard mutational analysis. Comparing the phenotypic descriptions of related mutants can reveal other associations that may not be obvious from the literature. For example, a mutation identified in a screen for hormone resistance may have also been found in a screen for floral mutants. Finding correlations from phenotypes of different alleles may suggest new avenues of experimentation.

TAIR's Germplasm Search can be used to find natural variants and different types of mutants including T-DNA insertions, TILLED substitutions, and other types of induced mutations. The following protocol describes a method for searching for any germplasm having a mapped polymorphism at a locus of interest, or having a specific phenotype.

1. Start at TAIR's Germplasm Search (www.arabidopsis.org/servlets/Search?action=new_search &type=germplasm).
2. To search for mutants in a specific gene, define the gene of interest. In the Search by Name, Description, or Stock Number section choose "Locus Name" from the drop-down menu. Type in the AGI locus identifier for your gene of interest.
3. To find mutants with similar phenotypes, choose "description" from the drop-down menu found under the Search by Name section at the top of the page. For the broadest search choose the "contains" option and type in the terms to search for in the descriptions. Entering different terms in each input box will return those germplasms that contain all the specified terms within their descriptions (e.g., LEAF and SERRATED). Additional options can be selected to further refine your search, such as selecting the background of the mutation, which can be useful when searching for mutations in the same genetic background.
4. Submit the query. The results will list the germplasm name(s), polymorphisms, locus, background strain, a description of the germplasm (e.g., phenotype), donor name, and stock number. If images of the germplasm are available, a camera icon will be displayed.
5. Find allele(s) of interest. Click on the name of the polymorphism to view associated phenotypes, physical location of the lesion, mutagen that caused the mutation, and other germplasms that contain the allele. For some genes, it may be possible to acquire a range of mutations from knockouts (e.g., T-DNA insertion) to substitutions, or overexpressers (e.g., activation tag lines). A graphical representation of the gene structure and all the associated polymorphisms can be obtained by clicking on the Sequence Viewer link in the Map Links section on any polymorphism detail page. This will open a view of the SeqViewer genome browser with the polymorphism highlighted. The Zoom to drop-down menu can be used to display a closeup view of the gene, and the global controls (top left box) can be set to display specific annotations such as genes, transcripts, polymorphisms, and T-DNA/Transposons (www.arabidopsis.org/seqViewer/help/sv_intro.jsp). A more precise view of the sites of the polymorphisms can be obtained by mousing over the polymorphism of interest and selecting the nucleotide sequence view from the pop-up window. The nucleotide sequence view displays 10 kb of sequence. In the upper right corner of the nucleotide view window choose the appropriate objects to display. For example, choose Genes, Markers, and Polymorphisms to show the exact location of the polymorphisms within the gene. To get to a 10-kb centered view of a locus from a close up view anywhere on the chromosome, first zoom to a 10-kb view, then type in the name of the locus in the upper left text box and click on "Find."

6. Finding out about the germplasm: In the results page, the name of the germplasm is linked to a detailed record containing a full description of the line; information about the origins of the strain, such as pedigree and genetic background, images, and known polymorphisms; and, for germplasms available from ABRC or NASC, a description of the seed stock.
7. To order the seed stock from the ABRC, check the box in the "order" column. For each page of results, check the stocks you wish to order. When you are done selecting stocks, go to the top of the results page and click on the button to order the checked stock. If you are not logged in you will be prompted to do so. If you are not registered, you will need to register and be affiliated with a lab (for billing purposes) before you can order the stocks.

3.7. From Phenotype to Genotype—Databases and Tools For Map-Based Cloning

A primary method for gene discovery is the analysis of mutants and natural variations and subsequent cloning of the corresponding locus controlling the trait(s) of interest. Map-based cloning is a method for isolating genes for quantitative trait loci (QTLs) and mutations that are not tagged with foreign DNA. Fine mapping is done to reduce the interval to a small region (about 0.16 cM or approx 40 kb) containing only a few genes and then identifying candidate genes within the mapped region. Typically this involves generating a mapping population between two ecotypes and analyzing a large number of recombinants (approx 3000 chromosomes). Confirming the identity of the gene requires rescuing the mutant phenotype with a transgenic construct containing the wild-type locus. The combination of a sequenced genome, a rich resource of genetic markers, and publicly available polymorphism data has substantially reduced the amount of time required for positional cloning ***(26,27)***. For QTLs, the process is more complex and usually requires creating a population of isogenic lines (NILs) in order to fine-map the QTL ***(28,29)***. The protocols in this section describe how to use information at TAIR to assist in genetically mapping a mutation, finding candidate genes, and obtaining biological materials for complementation.

3.7.1. Finding and Downloading Sets of Genetic Markers for Mapping

1. Start TAIR's Genetic Marker Search page at www.arabidopsis.org/servlets/Search?action=new_search&type=marker).
2. Choose the type(s) of markers you want to use. PCR-based markers such as CAPS and SSLPs are by far the most commonly used marker types. In the section marked "Restrict by Features" check the boxes marked SSLP and CAPS.
3. Choose the mapping (parental) ecotypes. You can select markers that are known to reveal polymorphisms between two ecotypes. In the "Restrict by Features" section use the drop-down menus to select the two parental ecotypes. For the broadest search, or if your parental lines are not shown, keep the default selection (any).
4. Select the region of the chromosome of interest. If you have already determined linkage to a genetic marker or gene you can use the Range option to further restrict your search. First choose the map you are using. As most PCR-based markers are located on the AGI sequence map, click on the button next to this option in the "Map" section. Then select the region of the chromosome. To find markers 100 kb distal and 100 kb proximal to an existing marker, choose the "Around" option and type in the name of the linked marker. Note that the second option is disabled.
5. Choose the output option. You can choose to have the results displayed by name, position, or type. For example, if you want to see SSLP and CAPS markers separated, choose the "type" option.
6. Submit the query.
7. Viewing the results: The results of the query are returned to your browser window. You can view each genetic marker record by clicking on the marker name link. Alternatively, you can download the entire result set as a tab-delimited file and save it on your computer.
8. Downloading a list of markers: To save one or more of the markers to a list, use the check boxes to select the markers you want to save. If more than one page of results is returned, go through each page and select the markers you want. Once you have selected the markers, scroll to the top of the page and click on the "Download" button. The text file will have many fields, including primer sequences (if known), polymorphisms, and corresponding ecotypes.

3.7.2. Finding Polymorphisms Between Two Ecotypes for Generating New Markers

To generate a high-resolution map, it is often necessary to narrow down the interval by generating new markers in the region of interest. Single-nucleotide polymorphisms (SNPs) and small insertion/deletion (In Del) polymorphisms can be used to generate additional markers. SNPs can be used to make CAPS markers if the substitution causes a restriction site polymorphism. If no site is present, degenerate (dCAPs) markers can be made by introducing a mismatch in one primer that creates a restriction site polymorphism ***(30)***. In Dels can be used as a basis for generating SSLP markers. SNPs and In Dels can be used with high-throughput SNP detection methods that allow for faster mapping using smaller populations. Several large SNP identification projects have made their data available to the public. Monsanto has made its database of more than 57,000 polymorphisms between the Landsberg *erecta* and Columbia strains available to the research community (www.arabidopsis.org/Cereon/index.jsp; ***26***). SNP data from the Stanford Genome Sequencing Center (SGC; ***31***) and the Max Planck Arabidopsis SNP Consortium (MASC) database ***(32,33)*** are searchable at TAIR and are displayed on the genome browser. The SGC SNPs represent polymorphisms between L*er* and Columbia; MASC SNPs are tested in 12 different ecotypes.

TAIR's Polymorphism Search can be used to find many types of polymorphisms using a variety of criteria such as gene or locus name, polymorphism type, mutation site, and chromosomal location. This method describes how to find SNP or In Del polymorphisms between two specific ecotypes and shows how this information can be used to design PCR primers for genetic markers.

1. Start at TAIR's Polymorphism/Allele Search (www.arabidopsis.org/servlets/Search?action=new_search&type=polyallele).
2. Select the type(s) of polymorphisms to retrieve. For designing PCR markers, choose substitution, SNP, or In Del. To select more than one type, hold down the CTRL key (PC) or Apple key (Mac) and select each type with a mouse click.
3. Select the parental ecotypes. Use the drop-down menus to choose the ecotypes in which the polymorphism should be present. The selection is based on the current information in the database. If only one ecotype is listed, choose this as the first and "any" as the second option. If neither ecotype is represented, leave the default (any) option selected for both.
4. Define the chromosome of interest. Assuming a rough map position has been established, you can choose to limit the search to polymorphisms on a specific chromosome and further refine this by selecting the physical boundaries. Because sequenced polymorphisms are mapped on the sequence (AGI) map, choose this as the map type. Having selected this option, you can restrict to a range of kilobases or use any gene, marker, or clone in TAIR to define the boundaries.
5. Define the output format. Use the drop-down menus at the top of the search page to select the number of records to display and the sorting order of the display. Sorting by position will list all polymorphisms first by chromosome (if none was selected) and then by position from the top to bottom of the chromosome. Other options include sorting by name or type. The results include some or all of the following: name of the polymorphism linked to a detailed record, alternate names (aliases), polymorphism type, chromosome, starting position of the polymorphism, links to the SeqViewer close-up view centered on the polymorphism, links to the locus detail page if the polymorphism is within a locus, and a description.
6. Finding information about the polymorphic sequences: Click on the name of the polymorphism to display the detailed information, including the sequences that differ between ecotypes, along with their length and all ecotypes that contain this variant, and approx 20 bp of sequences flanking the polymorphism. For many polymorphisms this information can be used as a starting point to design markers.
7. To generate CAPS or dCAPs markers, go to the dCAPs finder program (http://helix.wustl.edu/dcaps/dcaps.html) by clicking on the link to the dCAPs finder from the "External Links" section. This program can be used to find existing restriction site polymorphisms or to design primers that will introduce a restriction site in one of the two alleles. You will need to use the flanking sequence data and the polymorphic sequence data from the polymorphism detail page to create the dCAPs primer.

3.7.3. Finding Candidate Genes in a Genetically Defined Interval

After the locus has been fine-mapped, the next step is to determine which genes located in the interval may be responsible for the phenotype. TAIR's genome browser, called SeqViewer, can be used to view the *Arabidopsis* genome from a whole-genome view down to the nucleotide level. SeqViewer displays genes, annotation units (BAC clones used to generate and assemble the genome sequence), transcripts (including ESTs and full-length cDNAs), polymorphisms, T-DNA/transposon insertions, and markers. The tool allows searching with up to 250 names or four short (<150 nucleotide) nucleotide sequences. The locations of one or many search hits on the whole genome can be shown in a close-up view (zoomable from 50 megabases to 10 kilobases) or in a 10-kb-nucleotide window. This protocol describes how to use SeqViewer to find candidate genes within a region containing a genetically defined interval using flanking marker data.

1. Start at the TAIR SeqViewer (www.arabidopsis.org/servlets/sv).
2. To view a region containing two or more markers, in the search box below the whole-chromosome view, type in the names of the flanking markers or polymorphisms and click "Submit." The matched hits will be displayed as red tick marks on the whole-chromosome view. Point your mouse to the region between the red tick marks and click. This will open a new view of the chromosome in that region. Adjust the zoom levels until both highlighted markers are visible.
3. Create a view centered on the region of interest between the two flanking markers. Type in the leftmost coordinate of the left marker and rightmost coordinate of the right flanking marker in the "Select Range" input box. Click "Go" to display a custom view of the chromosome. If the markers you have used for mapping are not in TAIR's database, you can use the primer sequences to locate the region of interest. From the SeqViewer home page, choose "search by sequence" and paste in the primer sequences in the input box and submit the query. Exact matches will be displayed as red tick marks on the whole-genome view. Click on the red marks to display a closeup view. Use the controller to align the view to display the region of interest.
4. Obtain a list of genes located between the marker endpoints and identify appropriate candidates. To obtain a list of genes use the Close Up view controller; click on the button to "List Genes in Range." This will open a new browser window containing a list of all loci in the close-up view. Click on the locus name to view the detailed record for this locus in TAIR or follow the links to view the corresponding locus details in the TIGR or MIPS *Arabidopsis* databases. To save the list, go to the upper right corner of the list of genes and click on "Download as text file." This list of locus identifiers can be used as input for a variety of tools. For example, it can be used to obtain a list of GO annotations (**Subheading 3.2.**) that may suggest a function or role consistent with the mutant phenotype or to obtain expression data for these genes (**Subheading 3.3.**). The expression data may show correlations with the phenotype in question (expressed in the same tissues that exhibit an abnormal phenotype). Go to the TAIR locus page to find cDNA clones to use as probes for Northern and *in situ* hybridization, genomic clones for generating complementation constructs, and additional alleles to confirm the phenotypes.

Acknowledgments

We would like to thank Eva Huala, Peifen Zhang, Tanya Berardini, Suparna Mundodi, and Sue Rhee for reviewing the manuscript. This project was supported in part by the National Science Foundation (grant numbers DBI-9978564, DBI-0091471, and DBI-9813586) and by the National Institutes of Health (grant number HG-02273). This is the Carnegie Institution of Washington Department of Plant Biology Publication 1679.

References

1. Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815.
2. Schoof, H., Ernst, R., Nazarov, V., Pfeifer, L., Mewes, H. W., and Mayer, K. F. (2004) MIPS Arabidopsis thaliana Database (MAtDB): an integrated biological knowledge resource for plant genomics. *Nucleic Acids Res.* **32,** Database issue, D373–D376.

3. Wortman, J. R., Haas, B. J., Hannick, L. I., et al. (2003) Annotation of the *Arabidopsis* genome. *Plant Physiol.* **132,** 461–468.
4. Garcia-Hernandez, M., Berardini, T. Z., Chen, G., et al. (2002) TAIR: a resource for integrated *Arabidopsis* data. *Funct. Integr. Genomics* **2,** 239–253.
5. Huala, E., Dickerman, A. W., Garcia-Hernandez, M., et al. (2001) The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **29,** 102–105.
6. Rhee, S. Y., Beavis, W., Berardini, T. Z., et al. (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* **31,** 224–228.
7. Berardini, T. Z., Mundodi, S., Reiser, L., et al. (2004) Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiol.* **135,** 745–755.
8. Alonso, J. M., Stepanova, A. N., Leisse, T. J., et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301,** 653–657.
9. Anderle, P., Duval, M., Draghici, S., et al. (2003) Gene expression databases and data mining. *Biotechniques* Suppl, 36–44.
10. Haas, B. J., Volfovsky, N., Town, C. D., et al. (2002) Full-length messenger RNA sequences greatly improve genome annotation. *Genome Biol.* **3,** RESEARCH0029.
11. Frishman, D., Albermann, K., Hani, J., Heumann, K., Metanomski, A., Zollner, A., and Mewes, H. W. (2001) Functional and structural genomics using PEDANT. *Bioinformatics* **17,** 44–57.
12. Mewes, H. W., Amid, C., Arnold, R., et al. (2004) MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* **32** Database issue, D41–D44.
13. Ashburner, M., Ball, C. A., Blake, J. A., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29.
14. Wisman, E. and Ohlrogge, J. (2000) *Arabidopsis* microarray service facilities. *Plant Physiol.* **124,** 1468–1471.
15. Craigon, D. J., James, N., Okyere, J., Higgins, J., Jotham, J., and May, S. (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res.* **32** Database issue, D575–D577.
16. Wang, X., Hessner, M. J., Wu, Y., Pati, N., and Ghosh, S. (2003) Quantitative quality control in microarray experiments and the application in data filtering, normalization and false positive rate prediction. *Bioinformatics* **19,** 1341–1347.
17. Quackenbush, J. (2001) Computational analysis of microarray data. *Nat. Rev. Genet.* **2,** 418–427.
18. Kerr, K. M., Churchill, G. A. (2001) Statistical design and the analysis of gene expression. *Genet. Res.* **77,** 123–128.
19. Brenner, S., Johnson, M., Bridgham, J., et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18,** 630–634.
20. Wheeler, D. L., Church, D. M., Edgar, R., et al. (2004) Database resources of the National Center for Biotechnology Information: update. *Nucl. Acids. Res.* **32,** D35–D40.
21. Mueller, L. A., Zhang, P., and Rhee, S. Y. (2003) AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.* **132,** 453–460.
22. Ichikawa, T., Nakazawa, M., Kawashima, M., et al. (2003) Sequence database of 1172 T-DNA insertion sites in *Arabidopsis* activation-tagging lines that showed phenotypes in T1 generation. *Plant J.* **36,** 421–429.
23. Colbert, T., Till, B. J., Tompa, R., et al. (2001) High-throughput screening for induced point mutations. *Plant Physiol.* **126,** 480–484.
24. McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000) Targeted screening for induced mutations. *Nat. Biotechnol.* **18,** 455–457.
25. May, S. T., Clements, D., and Bennett, M. J. (2002) Finding your knockout: reverse genetics techniques for plants. *Mol. Biotechnol.* **20,** 209–221.
26. Jander, G., Norris, S. R., Rounsley, S. D., Bush, D. F., Levin, I. M., and Last, R. L. (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol.* **129,** 440–450.
27. Lukowitz, W., Gillmor, C. S., and Scheible WR. (2000) Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. *Plant Physiol.* **123,** 795–805.
28. Yano, M. (2001) Genetic and molecular dissection of naturally occurring variation. *Curr. Opin. Plant Biol.* **4,** 130–135.
29. Maloof, J. N. (2003) Genomic approaches to analyzing natural variation in *Arabidopsis thaliana*. *Curr. Opin. Genet. Dev.* **13,** 576–582.

30. Neff, M. M., Turk, E., and Kalishman, M. (2002) Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet.* **18,** 613–615.
31. Cho, R. J., Mindrinos, M., Richards, D. R., et al. (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* **23,** 203–207.
32. Schmid, K. J., Sorensen, T. R., Stracke, R., Torjek, O., Altmann, T., Mitchell-Olds, T., and Weisshaar, B. (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Res.* **13,** 1250–1257.
33. Torjek, O., Berger, D., Meyer, R. C., et al. (2003) Establishment of a high-efficiency SNP-based framework marker set for *Arabidopsis*. *Plant J.* **36,** 122–140.

17

In Planta Agrobacterium-Mediated Transformation by Vacuum Infiltration

Brian W. Tague and Joanna Mantis

Summary

In planta Agrobacterium-mediated transformation using vacuum infiltration results in transgenic *Arabidopsis thaliana* without the use of sterile conditions or plant regeneration. Plants are grown in pots, in standard potting mix. *Agrobacterium tumefaciens*, carrying an appropriate plant transformation vector, is suspended in an infiltration medium that contains, at a minimum, sucrose and the surfactant Silwet L-77. Flower buds are immersed in the suspension of *A. tumefaciens*. The application of a vacuum drives the bacteria into the intercellular air spaces. A portion of the *Agrobacterium* Ti plasmid known as the T-DNA region, which has been engineered to carry a selectable marker, becomes integrated into the plant genomic DNA. Plants are allowed to set seed. Seeds are germinated in selective conditions to recover transformants.

Key Words: *In planta* transformation; *Agrobacterium tumefaciens*; *Arabidopsis thaliana*; vacuum infiltration.

1. Introduction

The molecular transformation of plants is a crucial tool in all areas of higher plant biology. Standard plant molecular transformation entails delivery of the appropriate DNA to plant cells in tissue culture, followed by the selection and regeneration of whole plants from a single transformed cell. Most commonly, DNA delivery is accomplished by utilizing the natural transforming ability of the soil bacterium, *Agrobacterium tumefaciens*. In the wild the T-DNA, a portion of the *Agrobacterium* Ti plasmid, becomes integrated into the plant genome. In the laboratory, the T-DNA is engineered to carry a selectable marker and often some other gene of interest that serves as a reporter or whose expression alters some characteristic of the plant. As is the case in the wild, this DNA becomes integrated into the plant genome after *Agrobacterium* infection. Infected tissue is cultured under conditions that select for the transformed cell. Appropriate plant growth regulators are also included to regenerate whole plants from selected cells.

A difficulty in standard molecular transformation is that some plants and mutant lines are recalcitrant to regeneration. Regeneration can lead to uncovering somatic mutations, which can be increased by the presence of the plant growth regulators used during tissue culture. Finally, sterile conditions are required to regenerate plants.

In light of these difficulties, a number of *in planta* methods of plant transformation have been developed, most notably for the model plant *Arabidopsis thaliana*. It was first discovered that seeds treated with the appropriate *Agrobacterium* strain produce transformed progeny in the next generation. This method was instrumental in developing the "insertion mutagenized" lines of *A. thaliana* (*1*), but proved too inconsistent for general use.

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

The next improvement for *in planta* transformation was the treatment of adult *A. thaliana* plants. Based on the observation that wounded plant tissue is the preferred site of *Agrobacterium* infection in nature, reproductive inflorescences are cut off flowering *Arabidopsis,* and the bacteria is applied to the wound site *(2,3)*. New inflorescences arising from the rosette produce seed; this seed is screened for growth under selective conditions. This method gives transformation rates close to the traditional tissue culture method, but is still somewhat labor-intensive.

It was subsequently determined that the wounding associated with the above method was unnecessary. In the vacuum infiltration method, flowering *Arabidopsis* plants are uprooted, and the whole plants are submerged in a suspension of *Agrobacterium (4,5)*. The application of a vacuum causes the infiltration of the bacteria into the air spaces of the plant tissues. Plants are repotted and allowed to set seed. Transformation rates on the order of 1% are observed with this procedure. The removal and repotting of plants limits the utility of this method.

As it turns out, the uprooting and replanting of *Arabidopsis* is also unnecessary. It is sufficient to submerge the inflorescence of plants in the early stages of flowering in an *Agrobacterium* suspension. In fact, it is sufficient to dip flower buds in the suspension in the absence of vacuum infiltration *(6)*. The inclusion of the surfactant Silwet L-77 dramatically increases the transformation efficiency *(7)*. With this procedure, 0.5 to 5.0% of the seed is found to be transgenic. This simple technique, which avoids extensive labor and the problems associated with regeneration, has now been used on large number of *Arabidopsis* ecotypes and mutants, many of which are recalcitrant to transformation by tissue culture procedures *(8)*.

2. Materials

2.1. Growth Room and Greenhouse

1. 130-mm and 60-mm diameter standard plastic pots.
2. Trays: Standard flats with and without drainage holes, 8 cell inserts, and clear plastic domes.
3. Nylon window screen netting, available in most hardware stores.
4. Metromix 200 (a soilless mixture of vermiculite, perlite, peat moss, and sand; Scotts, Marysville, OH).
5. Elastic bands.

2.2. Equipment and Media

1. Rotary shaker.
2. Incubator for bacterial cultures.
3. Vacuum pump.
4. Vacuum desiccator, 240 mm interior diameter
5. Plastic Petri dishes: 100 × 15 mm for plating bacteria, 150 × 15 mm for screening seeds.
6. Luria-Bertani (LB) medium: 10 g/L Bacto tryptone, 5 g/L yeast extract, 10 g/L NaCl, adjusted to pH 8.0 with 1 *N* NaOH. For plates, add 15 g Bacto agar. Add appropriate antibiotic for *Agrobacterium* selection after sterilization and cooling of the medium to 65°C.
7. Pyrex crystallizing dish, 14.5 cm diameter × 7.5 cm deep.
8. Infiltration medium: 5% sucrose, 10 m*M* $MgCl_2$, 0.005% Silwet L-77 (Lehle Seeds, Round Rock, TX). The infiltration medium does not need to be sterilized, but should be prepared immediately before use.
9. Seed sterilization solution I: 10% commercial bleach, 0.02% Triton-X 100 (0.525% final sodium hypochlorite concentration). Make fresh as required.
10. Seed sterilization solution II: 70% ethanol.
11. Plates for selection of transformed seeds: 1X Murashige & Skoog (MS) basal salt mixture, 1.0% sucrose, 0.5 g/L MES (4-morpholine ethanesulfonic acid), pH 5.8 with 1 *N* KOH, 0.8% plant tissue culture agar. Filter sterilized kanamycin is added to a final concentration of 50 to 100 µg/mL for *npt*II-conferred resistance. Filter-sterilized mannose is added to 2 to 5 m*M* for *pmi*-conferred resistance. Add selective agents after autoclaving and cooling media to 65°C.
12. Micropore gas-permeable surgical tape (3M Health Care, St. Paul, MN).

13. Spray for selection of transformed plants grown in greenhouse: 600 mg/L phosphinothricin (sold under the trade names BASTA, Liberty, and Finale; stable at room temperature) or sterilized 50 m*M* mannose in distilled water.

2.3. Arabidopsis thaliana *Plants*

Arabidopsis thaliana (L.) Heyn., ecotypes Columbia (CO-0), Wassilevskija (WS), C-24, and Landsberg *erecta* are used routinely for vacuum infiltration with similar levels of transformation (*see* **Note 1**). *In planta* transformation has been successful with a wide variety of *Arabidopsis* ecotypes and mutants (*see*, for example, ***8***).

2.4. Agrobacterium tumefaciens *Strains and Vectors*

We have used strains LBA4404 (commercially available from Invitrogen, Carlsbad, CA) and GV3101 for the majority of our work. Successful transformation with a wide variety of *Agrobacterium* strains has been reported in the literature. In all cases, plasmids are introduced into *Agrobacterium* using standard electroporation techniques ***(9)***. Two alternative methods, not requiring costly equipment such as the electroporation device, are triparental mating and freeze-thaw transformation. Triparental mating is a lengthy procedure that requires more complex media and selection methods ***(10)***. A number of efficient, direct freeze-thaw methods are in use, all of which require minimal equipment ***(11)***.

3. Methods

3.1. Growth Conditions of Plant Material Before Infiltration

1. Plastic 130-mm diameter pots are filled with Metromix 200, which is drenched with 0.5X MS salts (*see* **Note 2**). The Metromix 200 is mounded above the rim of the pot (**Fig. 1A**).
2. Approximately 50 seeds are evenly distributed on the surface of the drenched growth medium (*see* **Note 3**).
3. Nylon netting (24 × 24 cm) is placed over the seeds and secured with an elastic band. Pull down the netting tightly to ensure good contact with growth media (*see* **Note 4**).
4. For even germination, place pots at 4°C for 48 h.
5. Transfer pots to a growth room or incubator on 54 × 26 cm trays. A 16-h photoperiod at 22°C will result in vigorous growth.
6. Top-water the plants for 3 d after placing the plants in growth conditions. After germination, bottom-water plants as needed (*see* **Note 5**).
7. Plants are ready for infiltration when the inflorescence has developed and there are many unopened floral buds, about 4 to 6 wk (**Fig. 1C**; *see* **Note 6**).

3.2. Agrobacterium *Culture and Preparation*

1. Inoculate a 5-mL culture of LB containing the appropriate selective agent with an *Agrobacterium* colony from a Petri dish. The Petri dish can be prepared from a glycerol stock or from a previous culture of *Agrobacterium*.
2. Incubate a 5-mL culture with good aeration, overnight at 30°C.
3. Inoculate a 1-L LB culture containing the appropriate selective agent with the 5-mL overnight culture. Grow with good aeration at 30°C until the culture reaches an OD_{600} of 0.7 to 0.9 (*see* **Notes 7**, **8**).
4. Harvest the *Agrobacterium* cells by centrifugation at 8000*g* for 10 min at room temperature in sterile 250-mL bottles. Resuspend the culture with 0.5 to 1.0 times the original volume of infiltration media (*see* **Notes 9**, **10**).

3.3. Infiltration and Harvesting

1. A Pyrex crystallizing dish is filled with about 500 mL of the *Agrobacterium* culture in infiltration medium.
2. A platform that can support an inverted pot of *A. thaliana* is placed in the *Agrobacterium* solution (**Fig. 1D**; *see* **Note 11**).

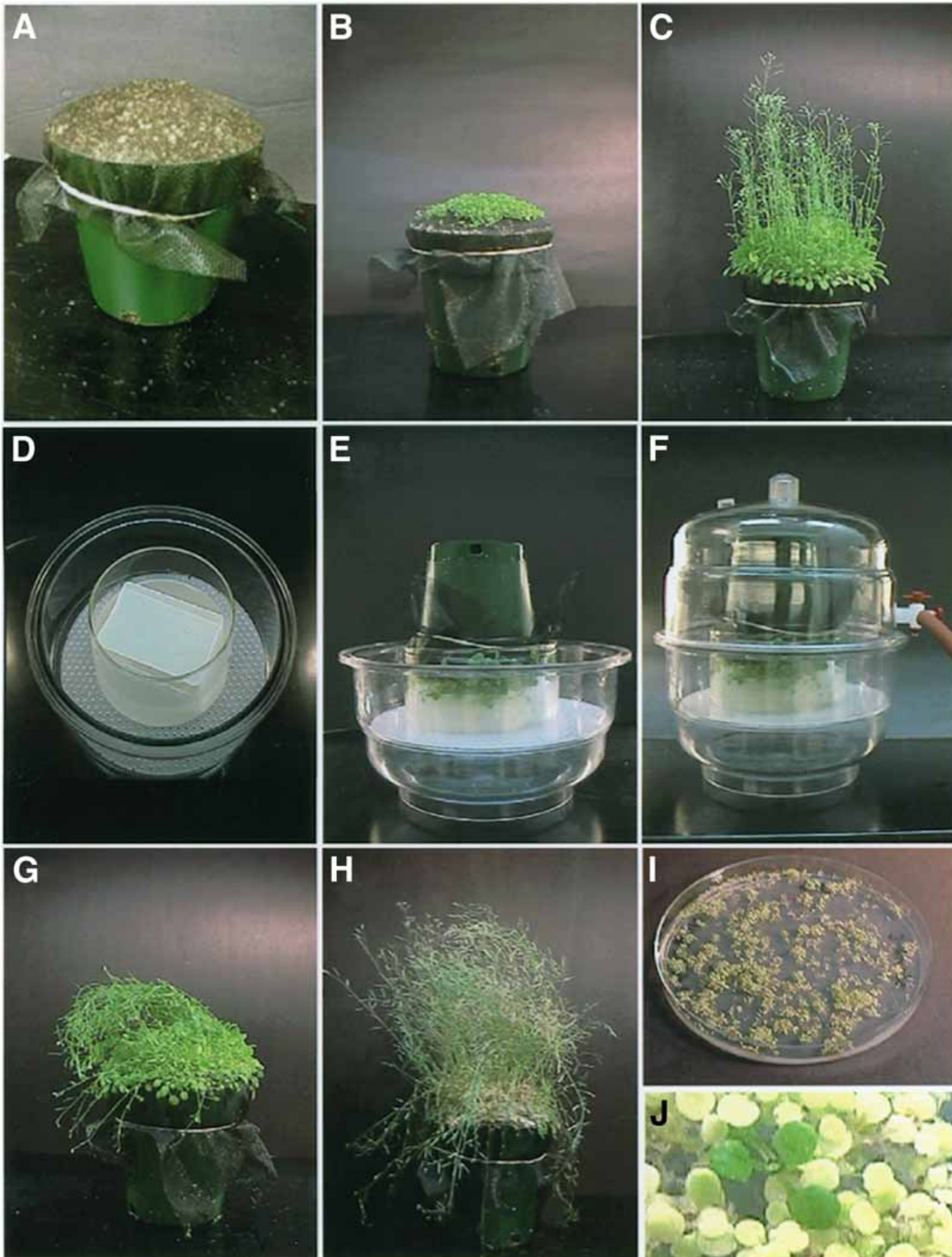


Fig. 1. Steps in the *in planta Agrobacterium*-mediated transformation by vacuum infiltration of *Arabidopsis thaliana*. **(A)** Growth medium is mounded up to ensure good contact with nylon netting. Seedlings **(B)** germinate through netting in 1 to 2 wk and inflorescences are ready in 5 or 6 wk **(C)**. A Pyrex dish is filled with an *Agrobacterium* suspension; a support for the inverted pot is placed in the dish **(D)**. The pot of *Arabidopsis* is inverted with the flower buds immersed in the *Agrobacterium* suspension **(E)**. The dish and pot are subjected to a vacuum in a vacuum flask **(F)**. After vacuum infiltration, the plants are waterlogged **(G)** but will recover. Seed is ready to harvest after a month **(H)**. Seeds can be sown densely on selective plates **(I)**. In about 10 d kanamycin-resistant transformants are evident, staying green and producing true leaves **(J)**. *See* color version in insert following p. 239.

3. A pot of *Arabidopsis thaliana* at the appropriate stage is placed upside down on the pipet tray, with the floral buds submerged in the *Agrobacterium* solution. Be careful not to let the soil contact the solution (**Fig. 1E**).
4. The Pyrex dish and plastic pot are placed in a vacuum chamber connected to a vacuum pump, and approx 100 to 400 mmHg of vacuum is applied for 3 min. Air bubbles released from the plant material should be evident (**Fig. 1F**; *see* **Note 12**).
5. Slowly release the vacuum and remove the pots (*see* **Note 13**).
6. Wrap the plants in plastic wrap to maintain humidity and place pots back in the growth room or incubator (**Fig. 1G**; *see* **Notes 14**, **15**).
7. After 24 h, remove the plastic wrap.
8. Water plants as necessary until seed is set, usually about 1 mo (**Fig. 1H**). Gradually allow plants to dry.
9. Harvest all of the inflorescences from one pot and place in a single envelope. Incubate at 30°C for 1 wk, and then harvest seed (*see* **Notes 16**, **17**).

3.4. Screening of Transformants

3.4.1. Screening in Sterile Conditions

1. Sterilize the seeds for plating. In an appropriately sized capped tube, add ~10 vol of sterile distilled water. Imbibe for ~30 min. Let seeds settle to bottom of tube. Aspirate off liquid.
2. Add ~10 vol 10% bleach, 0.02% Triton X-100. Shake for 5 min. Let seeds settle. Much of the chaff left in the seeds should collect in the foam at the top of the tube. Aspirate off chaff and remaining liquid.
3. Add ~10 vol 70% ethanol. Shake for 5 min (*see* **Note 18**). Let seeds settle. Aspirate off liquid.
4. Rinse seeds five times in sterile distilled water.
5. Disperse seeds evenly onto an appropriate plate (*see* **Note 19**). This can be done by adding about 1 vol of distilled water and using a wide-bore 200-μL pipet tip to deliver seeds to the plate (*see* **Note 20**). One thousand to 5000 seeds can be screened on each 150-mm plate.
6. Seal plates with gas-permeable surgical tape.
7. Germination can be synchronized by treatment of the plates at 4°C in the dark for 2 to 3 d, followed by placing plates in a growth room or incubator with 16-h light at 22°C.
8. All seeds will germinate and the cotyledons will expand. Only resistant seedlings will stay green and produce primary leaves. Selection should be evident within 10 d (**Fig. 1I**, **J**).
9. T1 (transformed generation 1) seedlings are carefully removed from the agar by using flat, wide, smooth forceps and pulling gently on the cotyledon or hypocotyl (*see* **Note 21**).
10. Prepare 60-mm diameter pots with fertilized Metromix 200. Poke a hole in the medium with a pencil and push down the root. Gently pack medium around the root and water around the plant with a squeeze bottle to ensure good contact between root and medium. Cover pot with plastic wrap or a plastic dome to increase humidity. These can be removed in a few days.
11. Let plants mature and set T2 seed. Collect T2 seed for subsequent analysis (*see* **Note 22**).

3.4.2. Screening in the Greenhouse

1. Trays with drainage holes, fitted with 8-cell inserts, are filled with Metromix 200 and watered well with 0.5X MS.
2. Dry T1 seeds are dispersed by hand-sowing or in a suspension of sterile, cool 0.1% agarose. Cover the flat with a plastic dome.
3. Germination is synchronized by incubation of the flats at 4°C, followed by placing flats in a greenhouse, growth room, or incubator. Remove dome when seedlings germinate. Bottom-water flats as necessary.
4. Phosphinothricin and mannose resistance are easily detected in soil-grown plants (*see* **Notes 19**, **23**). After germination, spray plants three times, 3 d apart with phosphinothricin solution (*see* **Note 24**) or with mannose solution. Susceptible plants will yellow and die quickly. Flag resistant plants to distinguish them from late germinators. Let plants reach maturity and set T2 seed. Harvest T2 seed for subsequent analysis (*see* **Note 22**).

4. Notes

1. Although *in planta* transformation works well in *Arabidopsis thaliana*, it has been less successful in other species. Published examples of Brassicaceae include *Brassica rapa* (pakchoi, ***12***), *Raphanus sativus* (radish, ***13***) and *Arabidopsis lasiocarpa* (***14***). A model leguminous species, *Medicago truncatula*, has also been successfully transformed *in planta* (***15***) as has the fiber crop, *Hibiscus cannabinus* (kenaf, ***16***). It is likely that there are many unreported failures. Our lab has attempted *in planta* transformation with *Arabidopsis griffithiana, A. lasiocarpa, A. petraea, Barbarea verna*, *Capsella bursa-pastoris*, and *Cochlearia officinalis* and has been successful only with *A. lasiocarpa*. A key consideration for *in planta* transformation is the number of seeds produced on an individual plant. Some crucifers, for example, produce only four or five seeds per silique and only 20 to 30 siliques per plant (e.g., *Lunaria annua*). *In planta* transformation efficiencies range from 0.1 to 5%. For a plant that produces only about 100 seeds per reproductive cycle, it may be more efficient to try standard regeneration procedures for transformation. Before attempting the *in planta* method with a new species, test the species for resistance to the selective agent. Some plant species (*C. officinalis*, for example) have naturally high levels of resistance to kanamycin. Seed set in other species (*C. bursa-pastoris*, for example) is inhibited by submersion in the infiltration media; changes to the media or using the "spray" method of inoculation (***17***) may ameliorate the problem. Finally, given the particular developmental requirements for *in planta* transformation (***18–21***), it might be worth attempting the procedure with plants at different developmental stages.
2. Fertilize with one-half strength Murashige & Skoog (MS) basal salts once, when preparing pots for planting seeds or transplanting seedlings. Commercially available nutrient fertilizer solutions can also be used, following directions given by the manufacturer.
3. For small plants such as *A. thaliana*, ~50 seeds will allow both vigorous growth and numerous inflorescences. For larger species, one or a few plants per pot are better and omitting the nylon netting may be necessary.
4. Some practice will ensure the right curve to the mounded soil to get good contact between the soil and the netting. If the netting does not contact the soil closely, the seedlings will germinate under the netting.
5. Before germination, be careful not to let water run over the sides, which may cause loss of seeds. Using a spray bottle to mist the top of the pots ensures good wetting with little runoff. After germination, let soil dry between waterings.
6. Infiltrating the primary inflorescence will result in good transformation. To increase the amount of transformation or to delay use of the plants, the primary bolts can be clipped to allow growth of secondary inflorescences. The secondary inflorescences can be infiltrated when there are substantial numbers of unopened floral buds.
7. It will take overnight growth (14–18 h) for most *Agrobacterium* strains to reach the appropriate density. This must be determined empirically for each strain.
8. Acetosyringone (3, 5-dimethoxy-4-hydroxyacetophenone), a phenolic inducer of the virulence genes of *Agrobacterium*, is often added to cultures used for tissue culture-based transformation protocols and can increase the transformation efficiency. There are no published data indicating whether acetosyringone can increase *in planta* transformation efficiency.
9. The *Agrobacterium* pellet after centrifugation is fairly tight. Gently resuspend the pellet in a small volume of infiltration media using a sterile transfer pipet. Many *Agrobacterium* strains (such as LBA4404) are very clumpy. Make sure the bacteria are evenly suspended before diluting down to the final volume.
10. Many different infiltration media have been reported in the literature. The key ingredients are sucrose and the surfactant, Silwet L-77. Silwet L-77 can be toxic to plants at high concentrations. If attempting to transform a new species, it may be worthwhile to try a range of concentrations of sucrose and surfactant.
11. Our lab uses an empty 200-µL pipet tip tray for this support. The holes in the tray allow in the bacterial solution, and the tray is narrow enough to support the pot. An alternative is to use a plastic box or a beaker that is smaller than the plastic pot of plants. Some labs use a number of rubber stoppers to support the pot. As long as the buds are submerged and the soil is not in contact with the solution, any number of setups will work.

12. One can use standard house vacuum (available in some laboratories) or use a vacuum aspirator attached to a water tap to produce a vacuum. Lower vacuum pressure may require slightly longer times to produce air bubbles from the plant material. The vacuum treatment can be skipped all together in favor of a simple "floral dip." The dipping alone may result in lower overall transformation rates but higher seed set. Some plant species may be more susceptible to damage from vacuum treatment. Spraying of the *Agrobacterium* may be a useful alternative ***(20)***, although greater care must be taken to contain the aerosol of genetically modified bacteria. A recent report indicates that multiple inoculations of flower buds with a more concentrated *Agrobacterium* culture result in increased transformation efficiencies in *Arabidopsis thaliana* ***(22)***.
13. The same *Agrobacterium* suspension can be used for multiple pots, even as it collects small bits of Metromix 200 and plant parts. Swirl the medium between infiltrations to resuspend the *Agrobacterium* that will tend to settle. When the infiltration media is very contaminated, change for fresh *Agrobacterium* suspension.
14. The plants can be considerably water-logged and damaged after vacuum infiltration. Some leaves may die back, especially if they were submerged. The inflorescences and flower buds are less susceptible to damage from this treatment. They will recover and produce a substantial amount of seed.
15. Wrap the plastic wrap length-wise around the pot, about 6 cm below the upper edge. The wrap can be taped at the top to create a dome surrounded the inflorescences.
16. Tape bottom of envelopes so that seeds are not lost. Keep seeds from individual pots separate through the harvesting and screening procedure. *A priori*, transformants from different pots will represent independent transformation events. Having multiple independently transformed lines is often crucial. To demonstrate the effect of a transgene on the characteristics of a plant, one must distinguish between the effects of the transgene and the effects of integration into the genome, which can be mutagenic. If independently transformed lines show the same phenotype, it is likely due to the transgene as integration of the T-DNA occurs randomly throughout the genome.
17. A handy device for seed collection is the Cellector tissue sieve from Bellco Glass (Vineland, NJ). Appropriately sized screens allow the rapid removal of seeds from chaff. Remove all chaff, as this material is often a major source of contamination during sterile screening.
18. Longer treatment in bleach or ethanol can damage seeds. Do not let seeds sit in the sterilization solutions for longer than 5 min.
19. A variety of selectable markers are available for plant transformation. Some consideration should be given to the ultimate goals of a transformation experiment before deciding on an appropriate selection scheme. For example, many of the insertion mutant lines of *Arabidopsis* have been generated with T-DNA conferring kanamycin resistance. If an experimental protocol calls for complementation of such a mutant, use of a plasmid carrying a different selectable maker is necessary. The majority of plasmids available for plant transformation carry the *npt*II gene and confer resistance to kanamycin. Kanamycin selection is inexpensive and works extremely well in plates. Selection in soil is more difficult as spraying of kanamycin results in necrotic lesions on sensitive plants, but not complete killing. Our lab uses pBIN-mgfp5-ER ***(23)***, which confers both kanamycin resistance and green fluorescent protein expression for attempts at transforming new species. Selection for resistance to the antibiotic hygromycin, conferred by the *hpt* gene, also works well in plates, although less well in our hands than kanamycin selection. Hygromycin is extremely toxic to humans and is not recommended for selection in soil by spraying. Phosphinothricin is a potent herbicide. Resistance is conferred by the *bar* gene. Phosphinothricin is an excellent selective agent for soil-grown plants, easily applied by spraying. Our lab has used a series of plasmids carrying the *bar* gene ***(24)***; many other such plasmids are available. A very promising selective agent for plant transformation is mannose, which is converted into an inhibitor of glycolysis in plants, but has no effect on animals ***(25)***. Resistance is conferred by the *pmi* gene, is easy to achieve in plates and in soil, and is very inexpensive. Our lab has used the pNOV022 plasmid ***(26)*** from Syngenta Seeds AG (Basel, Switzerland).

20. Some labs omit the water rinses at the end, add an additional ethanol rinse, and let the ethanol dry off in a laminar flow hood overnight. Dry seed is easily dispersed onto the surface of a Petri dish. Alternatively, after sterilization, seeds are resuspended in sterile cooled 0.1% agarose, which keeps seeds dispersed for pipetting onto surface of medium.
21. It is possible to remove the entire root system with little damage. Seedlings at this stage are very susceptible to drying out. Work with one seedling at a time. Keep the lid on the selection plate while transferring the seedling to the potting mixture.
22. Substantial amounts of T2 seed will be produced from T1 plants for further analysis. Most T1 plants (>50%) will contain an insert at a single Mendelian locus. The vast majority of the T1 plants will be heterozygous, as *Agrobacterium* preferentially transforms female reproductive tissue, so the T2 seed will contain homozygous untransformed, heterozygous, and homozygous transformed individuals. Sow ~100 T2 seeds under selective conditions and determine the ratio of resistant to sensitive seedlings. Use this data to calculate the number of segregating T-DNA loci. Six to 10 T2 seeds can be planted as individuals, and a proportion of the T3 seed examined for segregation of the selectable marker. When a homozygous transformed line is identified, some of the remaining T2 seed can be planted to bulk up that line. Alternatively, molecular techniques such as Southern hybridization and polymerase chain reaction can be used to identify transformed individuals.
23. An advantage to both phosphinothricin and mannose resistance is that the entire procedure from transformation to selection can be done under nonsterile conditions ***(12)***. T2 seed from putative T1 transformants can be screened on plates to isolate homozygous transformed individuals.
24. Commercial preparations of phosphinothricin (such as Finale, AgrEvo Environmental Health, Montvale, NJ) contain about 6% phosphinothricin and can be diluted 1:100 for spraying.

References

1. Feldmann, K. A. and Marks, M. D. (1987) *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach. *Molec. Gen. Genet*. **208,** 1–9.
2. Chang, S. S., Park, S. K., Kim, B. C., Kang, B. J., Kim, D. U., and Nam, H. G. (1994) Stable genetic transformation of *Arabidopsis thaliana* by *Agrobacterium* inoculation *in planta*. *Plant J.* **5,** 551–558.
3. Katavic, V., Haughn, G. W., Reed, D., Marin, M., and Kunst, L. (1994) *In planta* transformation of *Arabidopsis thaliana*. *Molec. Gen. Genet*. **245,** 363–370.
4. Bechtold, N., Ellis, J., and Pelletier, G. (1993) *In planta Agrobacterium*-mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *Comp. Rend. L'Acad. des Sci. Serie III* **316,** 1194–1199.
5. Bechtold, N. and Pelletier, G. (1998) *In planta Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration, in *Arabidopsis Protocols* (Martinez-Zapater, J. M. and Salinas, J., eds.) Humana Press, Totowa, NJ, pp. 259–266.
6. Clough, S. J. and Bent, A. F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16,** 735–743.
7. Richardson, K., Fowler, S., Pullen, C., Skelton, C., Morris, B., and Putterill, J. (1998) T-DNA tagging of a flowering-time gene and improved gene transfer by *in planta* transformation of *Arabidopsis*. *Austr. J. Plant Physiol.* **25,** 125–130.
8. Mysore, K. S., Kumar, C. T. R., and Gelvin, S. B. (2000) *Arabidopsis* ecotypes and mutants that are recalcitrant to *Agrobacterium* root transformation are susceptible to germ-line transformation. *Plant J.* **21,** 9–16.
9. Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., and Smith, J., eds. (1995) *Current Protocols in Molecular Biology, 3rd ed.* Wiley, New York.
10. Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. R. (1980) Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti. Proc. Natl. Acad. Sci. USA* **77,** 7347–7351.
11. Tzfira, T. Jensen, C. S., Wang, W., Zuker, A., Vinocur, B., Altman, A., and Vainstein, A. (1997) Transgenic *Populus tremula*: a step-by-step protocol of its *Agrobacterium*-mediated transformation. *Plant Molec. Biol. Rep*. **15,** 219–235.
12. Qing, C. M., Fan, L., Lei, Y., Bouchez, D., Tourneur, C., Yan, L., and Robaglia, C. (2000) Transformation of pakchoi (*Brassica rapa* L. ssp *chinesis*) by *Agrobacterium* infiltration. *Molec. Breed.* **6,** 67–72.

13. Curtis, I. and Nam, H. (2001) Transgenic radish (*Raphanus sativus* L. *longipinnatus* Bailey) by floral-dip method—plant development and surfactant are important in optimizing transformation efficiency. *Transgenic Res.* **10,** 363–371.
14. Tague, B. (2001) Germ-line transformation of *Arabidopsis lasiocarpa*. *Transgenic Res.* **10,** 259–267.
15. Trieu, A. T., Burleigh, S. H., Kardailsky, I. V., et al. (2000) Transformation of *Medicago truncaluta* via infiltration of seedlings or flowering plants with *Agrobacterium*. *Plant J.* **22,** 531–541.
16. Kojima, M. Shiojri, H., Nogawa, M., et al. (2004) *In planta* transformation of kenaf plants (*Hibiscus cannabinus* var. *aokawa* No. 3) by *Agrobacterium tumefaciens*. *J. Biosci Bioengr*. **98,** 136–139.
17. Chung, M. H., Chen, M. K., and Pan, S. M. (2000) Floral spray transformation can efficiently generate *Arabidopsis* transgenic plants. *Transgenic Res.* **9,** 471–476.
18. Bechtold, N., Jaudeau, B., Jolivet, S., Mabo, B., Vezon, D., Voicin, R., and Pelletier, G. (2000) The maternal chromosome set is the target of the T-DNA in the *in planta* transformation of *Arabidopsis thaliana*. *Genetics*. **155,** 1875–1887.
19. Bechtold, N., Jolivet, S., Voicin, R., and Pelletier, G. (2003) The endosperm and the embryo of *Arabidopsis thaliana* are independently transformed through infiltration by *Agrobacterium tumefaciens*. *Transgenic Res.* **12,** 509–517.
20. Bent, A. F. (2000) *Arabidopsis in planta* transformation. Uses, mechanisms, and prospects for transformation of other species. *Plant Physiol.* **124,** 1540–1547.
21. Desfeux, C., Clough, S. J., and Bent, A. F. (2000) Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the floral-dip method. *Plant Physiol.* **123,** 895–904.
22. Martinez-Trujillo, M., Limones-Briones, V., Cabrera-Ponce, L., and Herrera-Estrella, L. (2004) Improving transformation efficiency of *Arabidopsis thaliana* by modifying the floral dip method. *Plant Molec. Biol. Rep.* **22,** 63–70.
23. Haseloff, J., Siemering, K. R., Prasher, D. C., and Hodge, S. (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. *Proc. Acad. Natl. Sci. USA* **94,** 2122–2127.
24. Mylne, J. and Botella, J. R. (1998) Binary vectors for sense and antisense expression of *Arabidopsis* ESTs. *Plant Molec. Biol. Report.* **16,** 257–262.
25. Joersbo, M. (2001) Advances in the selection of transgenic plants using non-antibiotic marker genes. *Physiol. Plant.* **111,** 269–272.
26. Todd, R. and Tague, B. W. (2001) Phosphomannose isomerase: a versatile selectable marker for *Arabidopsis thaliana* germ-line transformation. *Plant Molec. Biol. Rep.* **19,** 307–319.

18

Transient Expression Assay by Agroinfiltration of Leaves

Min Woo Lee and Yinong Yang

Summary

An *Agrobacterium*-mediated transient expression assay has been described for in vivo analysis of constitutive or inducible gene expression in *Arabidopsis* plants. By simple infiltration of *Agrobacterium* cells carrying appropriate gene constructs into *Arabidopsis* leaves, transient expression assays can be performed within 3 d without using expensive instruments or complicated procedures. Two days after agroinfiltration, *Arabidopsis* plants can be treated with chemical inducers such as dexamethasone to induce transgene expression. The expression of the β-glucoronidase (GUS) reporter gene can be determined by histochemical staining. In addition, RNA and protein can be extracted from agroinfected leaves and used for reverse transcription-polymerase chain reaction (RT-PCR), Northern blot, Western blot, immunoprecipitation, and enzyme assays.

Key Words: Transient gene expression; *Agrobacterium tumefaciens*; agroinfiltration; GUS staining; dexamethasone-inducible expression.

1. Introduction

Transient expression assay has been frequently used as a convenient alternative to stable transformation because it enables a rapid analysis of gene expression. Furthermore, transient gene expression is not biased by position effect, which often occurs in stable transformation. A number of transformation methods have been used for transient expression analysis in *Arabidopsis*, which include biolistic bombardment ***(1,2)***, PEG-mediated gene transfer and electroporation of protoplasts ***(3,4)***, and *Agrobacterium tumefaciens*-mediated transformation ***(5,6)***. Biolistic bombardment requires an expensive particle gun and often has large sample variations. As a result, a control construct (e.g., a luciferase reporter gene) is often required to be cotransformed with test constructs to standardize the variation. Protoplast transformation involves a care-intensive, complicated procedure of isolating protoplasts from leaf mesophyll. Protoplasts can also respond differently from intact cells and may not be suitable for certain types of expression analysis.

In contrast to the biolistic bombardment and protoplast transformation, *Agrobacterium*-mediated transient expression can be conveniently conducted without using expensive equipment (e.g., biolistic gun or electroporation) or complicated procedures (e.g., preparation of protoplast). Using a modified β-glucuronidase (GUS) gene containing an intron, the transient expression assay was first used to detect the transfer of T-DNA from *Agrobacterium* tumefaciens to tobacco cells ***(7)***. Recently, several studies have shown that *Agrobacterium*-mediated transient expression is an efficient and versatile tool for studying foreign gene expression ***(8)***, gene silencing ***(9,10)***, host-pathogen interactions ***(11–15)***, *cis*-element/*trans*-factor interaction ***(16)***, and protein–protein interaction ***(17)***.

In this chapter, we describe a transient expression assay using simple infiltration of *Agrobacterium* cells carrying appropriate plasmid constructs into intact plant leaves. The methods

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

described are largely based on previously published procedures ***(5,16)***. The constitutive expression of the *35S::GUS* reporter gene and dexamethasone (DEX)-inducible expression of the bacterial *avrRpt2* gene were examined following the agroinfiltration of *Arabidopsis* leaves *in planta*.

2. Materials

1. Seeds of *Arabidopsis thaliana.*
2. *Agrobacterium tumefaciens* strains such as C58C1 ***(18)***.
3. Luria-Bertani (LB) media (broth and agar plate).
4. Induction medium ***(5,19)***: 10.5 g/L K_2HPO_4, 4.5 g/L KH_2PO_4, 1 g/L $(NH_4)_2SO_4$, 0.5 g/L NaCitrate, 1 g/L glucose, 1 g/L fructose, 4 g/L glycerol, 1 m*M* $MgSO_4$, 10 m*M* 2-(N-morpholino) methanesulfonic acid (MES); adjust pH to 5.6, autoclave before use (*see* **Note 1**).
5. Infiltration medium: 10 m*M* $MgSO_4$, 10 m*M* MES; adjust pH to 5.6, autoclave before use.
6. Rifampicin: Dissolve in methanol to make 25 mg/mL stock solution and store in –20°C freezer.
7. Tetracycline: Dissolve in 70% ethanol to make 12.5 mg/mL stock solution and store in –20°C freezer.
8. Kanamycin: Dissolve in water to make 50 mg/mL stock solution and store in –20°C freezer.
9. Acetosyringone (3,5-dimethoxy-4-hydroxyacetophenone, Aldrich): Dissolve 196.2 mg acetosyringone in 1 mL dimethylsulfoxide to make 1 *M* stock solution and store in –20°C freezer.
10. Dexamethasone (DEX, USB, Cleveland, OH): Dissolve 78.4 mg DEX in 1 mL ethanol to make 200 m*M* stock solution.
11. 5-bromo-4-chloro-3-indolyl β-D-glucuronide (X-Glu, Sigma, St. Louis, MO): Dissolve X-Glu in dimethylformamide to make 200 m*M* stock solution and store in –20°C freezer.
12. GUS staining solution: 50 m*M* sodium phosphate (pH 7.0), 1 m*M* EDTA, 2 m*M* X-Glu, 0.05% SDS, 0.1% sodium N-lauryl-sarcosine, 0.1% Triton X-100.
13. Vacuum pump.

3. Methods

3.1. Preparation of Agrobacterium *Suspension*

1. Streak *Agrobacterium* strain C58C1 carrying plasmid constructs (e.g., *35S::GUS* or *DEX::avrRpt2*) on LB agar plates with appropriate antibiotics (*see* **Note 2**). Incubate plates at 30°C for 2 d.
2. Inoculate agrobacteria into 5 mL LB broth with antibiotics and grow overnight at 30°C on a shaker (150 rpm).
3. Collect 2 mL overnight cultures in sterile microcentrifuge tubes by centrifugation (3000*g* for 5 min).
4. Resuspend agrobacteria in 1 mL induction medium. Transfer bacterial suspension to 15-mL sterile culture tubes. Add to a final volume of 4 mL induction medium supplemented with 100 μ*M* acetosyringone and appropriate antibiotics.
5. Grow agrobacterial suspension at 30°C for 5 to 6 h.
6. Collect bacterial cultures by centrifugation (3000*g* for 5 min). Resuspend bacterial cells in the infiltration medium (containing 200 μ*M* acetosyringone) to 0.4 OD at 600 nm.

3.2. Growing of Plants and Selection of Leaves

1. *Agrobacterium*-mediated transient transformation can be significantly influenced by *Arabidopsis thaliana* genotype, light, temperature, and humidity ***(20,21)***. Best results were obtained by growing *Arabidopsis* plants under 90 μE of light in a short-day cycle (8 h light, 16 h dark) at 23°C (*see* **Note 3**).
2. Competence for transient transformation can also be enhanced by placing plants under complete darkness for 16 h before infiltration.
3. Plant age and leaf size are very important for the efficiency of transient gene expression. Generally, 4-wk-old *Arabidopsis* plants are used for the transient expression assay. The best transient expression comes from agroinfiltration of fingernail-sized leaves. Older plants with larger leaves also work but the transformation efficiency decreases rapidly with the increase of plant age and leaf size (*see* **Note 4**).

3.3. Agroinfiltration of Arabidopsis *Leaves* In Planta

1. Agroinfiltration is conducted by infiltration of agrobacterial suspension (0.4 OD) into intercellular spaces of fingernail-sized leaves that are still attached to the intact plant. A needleless plastic syringe is used to infiltrate bacterial suspension into the underside of leaves (*see* **Note 5**).
2. Agroinfiltration is preferably conducted during late afternoon or evening so that T-DNA transfer occurs overnight.
3. After infiltration, place *Arabidopsis* plants in the 23°C growth room.

3.4. Activation of Dexamethasone-Inducible Promoter

1. Two days after agroinfiltration, plants are sprayed with 20 μ*M* DEX solution to activate the glucocorticoid-mediated transcriptional expression of target genes such as *avrRpt2* (**22**, *see* **Note 6**).
2. After DEX treatment, it is important to cover plants with transparent plastic domes. Remove the cover the next morning.
3. In addition to DEX, other treatments can also be conducted 24 to 48 h after agroinfiltration (*see* **Note 7**).

3.5. Phenotype Observation and Sample Collection

1. Phenotypic changes such as the hypersensitive reaction elicited by the AvrRpt2 protein can be observed 18 to 24 h after DEX treatment (**Fig. 1A**).
2. Leaf samples can be collected for subsequent GUS staining or RNA or protein extraction. Due to differences in protein expression and stability, it is desirable to perform time-course experiments to find out the best sampling time for extracting the maximal amount of the target protein (*see* **Note 8**).

3.6. Histochemical GUS Staining

1. Transfer leaves to 1.5-mL microcentrifuge tubes.
2. Add 1 mL or enough X-Glu staining solution to completely submerge leaves in the solution.
3. Vacuum briefly (5 min) to infiltrate X-Glu solution into leaf tissue.
4. Incubate samples overnight at 37°C.
5. Clear leaves with 75% ethanol at 70°C or with acetone:methanol solution (1:3 ratio).
6. Take photographs as needed (**Fig. 1B**).

4. Notes

1. The components of the induction medium can be mixed together and autoclaved at 120°C for 20 min. Alternatively, glucose, fructose, and glycerol can be sterilized separately and mixed with the basal induction medium before use. Add appropriate antibiotics and acetosyringone just before use.
2. The *Agrobacterium tumefaciens* C58C1 strain is resistant to rifampicin (50 μg/mL) and tetracycline (12.5 μg/mL). The plasmid constructs carry kanamycin resistance (50 μg/mL). The C58C1 strain sometimes is very sensitive to tetracycline. As a result, the final concentration of tetracycline used in selection may be reduced to as low as 2.5 μg/mL.
3. *Arabidopsis* plants are usually watered every 2 d. It is best not to water the plants for 2 d before agroinfiltration. Agroinfiltration can be conducted more easily with plants grown in relatively dry soil conditions. Wet soil conditions make infiltration of *Agrobacterium* suspension hard, as greater pressure (which could damage leaves) is needed to evenly infiltrate *Agrobacterium* suspension into entire leaves.
4. Plants of similar size should be selected for optimal comparisons of experimental controls and tests. In addition, infiltration should be performed with leaves of the same age. Usually, leaves 6 through 8 are chosen for infiltration.
5. Use 1-mL or 3-mL disposable plastic syringes (without the needle) to inject the *Agrobacterium* suspension (about 50–100 μL per leaf) into the underside of *Arabidopsis* leaves. If you have difficulty injecting the back of the leaf, try to inject the front. Generally, it is not necessary to pierce the leaf to infiltrate the bacterial suspension, although an incision can be made to facilitate infiltration. After agroinfiltration, mark the infiltrated leaves with a fine-point marker.

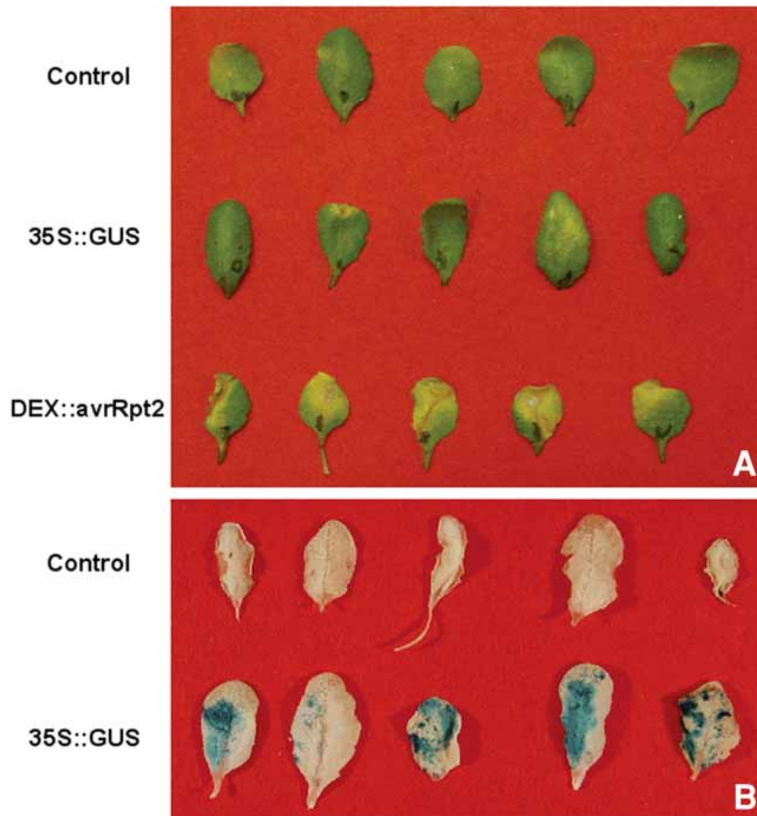


Fig. 1. Transient expression of the *GUS* and *avrRpt2* genes in *Arabidopsis* leaves. The *Agrobacterium tumefaciens* strains carrying *35S::GUS* and *DEX:avrRpt2* gene constructs were infiltrated into leaves. DEX solution (20 μ*M*) was sprayed onto leaves 2 d after agroinfiltration. **(A)** Elicitation of the hypersensitive reaction by AvrRpt2 1 d after the DEX treatment. **(B)** GUS staining of the leaves 3 d after agroinfiltration of the *35S::GUS* construct. *See* color version in insert following p. 239.

6. It is recommended to wait 2 d before spraying plants with DEX. This allows the accumulation of the chimeric receptor/transcription factor, so that a high level of target gene expression can be achieved upon induction with DEX. The 20 μ*M* concentration of DEX is routinely used but can be increased depending on specific gene constructs. Spray DEX solution until the point of imminent runoff (leaves are wet but the solution is not yet running off). The DEX concentration may be refined but 20 μ*M* is a good starting point. Following DEX treatment, it is important to cover the plant overnight with a plastic dome.
7. Depending on target gene constructs and research purposes, various treatments can be conducted at 1 or 2 d after agroinfiltration. These may include specific chemicals (e.g., salicylic acid, jasmonic acid) as well as abiotic (salt and cold) and biotic (virus or bacterial infection) treatments ***(16)***.
8. In addition to histochemical GUS staining, the GUS activity in leaves can be determined using 4-methylumbelliferyl-D-glucuronide as a substrate and quantified with a fluorometer ***(16)***. RNA and protein can also be extracted from leaves expressing a specific gene construct and used for reverse transcription-polymerase chain reaction (RT-PCR), Northern blot, Western blot, immunoprecipitation and enzyme assays, etc.

References

1. Seki, M., Komeda, Y., Iida, A., Yamada, Y., and Morikawa, H. (1991) Transient expression of β-glucuronidase in *Arabidopsis thaliana* leaves and roots and *Brassica napus* stems using a pneumatic particle gun. *Plant Mol. Biol.* **17,** 259–263.
2. Seki, M., Iida, A., and Morikawa, H. (1999) Transient expression of the beta-glucuronidase gene in tissues of *Arabidopsis thaliana* by bombardment-mediated transformation. *Mol. Biotechnol.* **11,** 251–255.
3. Damm, B., Schmidt, R., and Willmitzer, L. (1989) Efficient transformation of *Arabidopsis thaliana* using direct gene transfer to protoplasts. *Mol. Gen. Genet*. **217,** 6–12.
4. Abel, S. and Theologis, A. (1994) Transient transformation of *Arabidopsis* leaf protoplasts: a versatile experimental system to study gene expression. *Plant J*. **5,** 421–27.
5. Nimchuk, Z., Marois, E., Kjemtrup, S., Leister, T. R., Katagiri, F., and Dangl, J. L. (2000) Eukaryotic fatty acylation drives plasma membrane targeting and enhances function of several type III effector proteins from *Pseudomonas syringae*. *Cell* **101,** 353–363.
6. Johansen, L. K. and Carrington, J. C. (2001) Silencing on the spot. Induction and suppression of RNA silencing in the *Agrobacterium*-mediated transient expression system. *Plant Physiol.* **126,** 930–938.
7. Ohta, S., Mita, S., Hattori, T., and Nakamura, K. (1990) Construction and expression in tobacco of a β-glucuronidase (GUS) reporter gene containing an intron within the coding sequence. *Plant Cell Physiol*. **31,** 805–813.
8. Kapila, J., Rycke R. D., Van Montagu, M., and Agenon, G. (1997) An *Agrobacterium*-mediated transient gene expression system for intact leaves. *Plant Sci.* **122,** 101–108.
9. Schöb, H., Kunz, C., and Meins, Jr., F. (1997) Silencing of transgenes introduced into leaves by agroinfiltration: a simple, rapid method for investigating sequence requirements for gene silencing. *Mol. Gen. Genet.* **256,** 581–588.
10. Llave, C., Kasschau, K. D., and Carrington, J. C. (2000) Virus-encoded suppressor of posttranscriptional gene silencing targets a maintenance step in the silencing pathway. *Proc. Natl. Acad. Sci. USA* **97,** 13,401–13,406.
11. Scofield, S. R., Tobias, C. M., Rathjen, J. P., Chang, J. H., Lavelle, D. T., Michelmore, R. W., and Staskawicz, B. J. (1996) Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. *Science* **274,** 2063–2065.
12. Tang, X., Frederick, R. D., Zhou, J., Halterman, D. A., Jia, Y., and Martin, G. B. (1996) Initiation of plant disease resistance by physical interaction of AvrPto and the Pto kinase. *Science* **274,** 2060–2063.
13. Van den Ackerveken, G., Marois, E., and Bonas, U. (1996) Recognition of the bacterial avirulence protein AvrBS3 occurs inside the host plant cell. *Cell* **87,** 1307–1316.
14. Shao, F., Golstein, C., Ade, J., Stoutemyer, M., Dixon, J. E., and Innes, R. W. (2003) Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. *Science* **301,** 1230–1233.
15. Jamir, Y., Guo, M., Oh, H. S., et al. (2004) Identification of *Pseudomonas syringae* type III effectors that can suppress programmed cell death in plants and yeast. *Plant J.* **37,** 554–565.
16. Yang, Y., Li, R., and Qi, M. (2000) *In vivo* analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *Plant J.* **22,** 543–551.
17. Ferrando, A., Koncz-Kalman, Z., Farras, R., Tiburcio, A., Schell, J., and Koncz, C. (2001) Detection of in vivo protein interactions between Snf1-related kinase subunits with intron-tagged epitope-labelling in plants cells. *Nucleic Acids Res.* **29,** 3685–3693.
18. Koncz, C. and Schell, J. (1986) The promoter of T_L-DNA gene *5* controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* **204,** 383–396.
19. Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P. J. (1995) Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. *EMBO J.* **14,** 3206–3214.
20. Chateau, S., Sangwan, R. S., and Sangwan-Norreel, B. S. (2000) Competence of *Arabidopsis thaliana* genotypes and mutants for *Agrobacterium tumefaciens*-mediated gene transfer: role of phytohormones. *J. Exp. Bot*. **51,** 1961–1968.
21. Zambre, M., Terryn, N., De Clercq, J., et al. (2003) Light strongly promotes gene transfer from *Agrobacterium tumefaciens* to plant cells. *Planta* **216,** 580–586.
22. Aoyama, T. and Chua, N.-H. (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J.* **11,** 605–612.

19

Functional Analysis of Transcription Factors by Microparticle Bombardment

Ida Ruberti, Giovanna Sessa, and Giorgio Morelli

Summary

This chapter describes a method for rapid gene expression assays in *Arabidopsis*. Three distinct plasmids are codelivered into leaves by microparticle bombarment. The first one carries a transacting factor gene driven by the cauliflower mosaic virus (CaMV) 35S promoter. The second plasmid includes a reporter luciferase (LUC) gene transcribed from a promoter containing the TF binding site(s). The third plasmid contains a reference β-glucuronidase (GUS) gene transcribed from the 35S-CaMV promoter, which is used to normalize data derived from independent experiments.

This method could be applied to study systematically the functional properties of *Arabidopsis* transcription factors as well as to identify promoters and control region(s) regulated by specific transcription factors.

Key Words: Binding sites; GUS activity; LUC activity; particle bombardment; regulatory sequences; transacting domains; transcription factors; transient assay.

1. Introduction

Microprojectile-mediated gene transfer ***(1)*** is an established tool for delivering exogenous nucleic acids into plant cells and is a commonly employed technique in plant science. It is used for many applications, such as transformation of plant species that are otherwise impossible or very difficult to transform, transformation of organelles, and transient transformation of plant tissues ***(2,3)***. Microprojectile-mediated gene transfer has also been used to dissect and identify regulatory sequences, and to ascertain under what physiological conditions a given promoter is up- or downregulated ***(4–11)***. More recently, transient expression studies have been carried out to assay the transacting function of transcription factors in *Arabidopsis* as well as in other plant species ***(12–17)***.

In this chapter, we describe a system for rapid analysis of transcription factors in *Arabidopsis* leaves.

2. Materials

1. pGEM expression system (Promega; *see* **Note 1**).
2. *Escherichia coli* Xl1-blue strain.
3. *Arabidopsis thaliana*, Co-0 or WS ecotype.
4. Universal Soil VM type (Einheitserde-und Humuswerke Gebr. Patzer GmbH & Co. KG).
5. Plant fertilizer Mannalin-A (Wilhelm Haug Gamborg KG).
6. Plastic pots, 11 cm in diameter.
7. Agarose (IBI).
8. Sterile Petri dishes, 10 cm in diameter.
9. Murashige & Skoog (MS) salts.
10. Agar-agar (Merck).

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

11. Gold microparticles (1.6 µm in diameter; Bio-Rad).
12. Ethanol.
13. Mixer (Eppendorf).
14. Tubes (1.5 and 2 mL) and microfuge (Eppendorf).
15. Vortex.
16. 2.5 *M* $CaCl_2$.
17. 0.1 *M* Spermidine-free base.
18. Macrocarrier (Bio-Rad).
19. Biolistic PDS-1000/HE particle delivery system (Bio-Rad).
20. 1100 psi rupture discs (Bio-Rad).
21. Stopping screens (Bio-Rad).
22. Helium tank.
23. Vacuum pump.
24. Liquid nitrogen.
25. Mortar and pestle.
26. Cell culture lysis reagent (CCLR) buffer (Promega): 100 m*M* potassium phosphate (pH 7.8), 1 m*M* EDTA, 7 m*M* 2-mercaptoethanol, 1% (v/v) Triton X-100, 10% (v/v) glycerol.
27. GUS (β-glucuronidase) buffer (1X CCLR buffer containing 2 m*M* 4-methylumbelliferyl β-D-glucuronide [MUG]).
28. GUS stop buffer (0.2 *M* Na_2CO_3).
29. Thermoblock.
30. Fluorimeter (SFM25, Kontron Instruments).
31. Luciferase assay reagent (LAR) buffer (Promega): 20 m*M* Tricine (pH 7.8), 5 m*M* $MgCl_2$, 0.1 m*M* EDTA, 3.3 m*M* DTT, 270 µ*M* coenzyme A, 500 µ*M* luciferin, 500 µ*M* ATP.
32. Luminometer (TD-20e, Turner Design).

3. Methods

The methods described below outline the construction of expression plasmids, tissue preparation for particle bombardment, preparation of DNA-coated gold particles, leaf bombardment, and determination of the relative transactivation levels. In the PDS-1000/He biolistic device, high-pressure helium is used to drive a plastic disk (macrocarrier) down a barrel to accelerate gold particles (microcarrier).

3.1. Expression Vectors

Three distinct genes carried by separate plasmids are used in the transient assay (**Fig. 1**; *see* **Note 2**). **Subheadings 3.1.1.**, **3.1.2.**, and **3.1.3.** describe the characteristics of the three plasmids.

3.1.1. Transacting Gene (TF Plasmid)

The transacting plasmid contains the transcription factor (TF) coding sequence or its derivative(s) transcribed from the –343 cauliflower mosaic virus (CaMV) 35S promoter and flanked at the 3' end by a nopaline synthase (nos) polyA fragment (**Fig. 1**).

3.1.2. Reporter Gene (LUC Plasmid)

The reporter plasmid has the luciferase (LUC) reporter gene ***(18)*** transcribed from a chimeric promoter constituted by the TF binding site (*see* **Note 3**), or its mutant sequence, placed 5' of the –46 CaMV 35S promoter. The LUC reporter gene is flanked at the 3' end by a nos polyA fragment (**Fig. 1**).

3.1.3. Reference Gene (GUS Plasmid)

The reference plasmid contains the *uidA* gene (19), encoding a β-glucuronidase (GUS), transcribed from the –343 CaMV 35S promoter. The *uidA* reporter gene is flanked at the 3' end by a nos polyA fragment (**Fig. 1**).

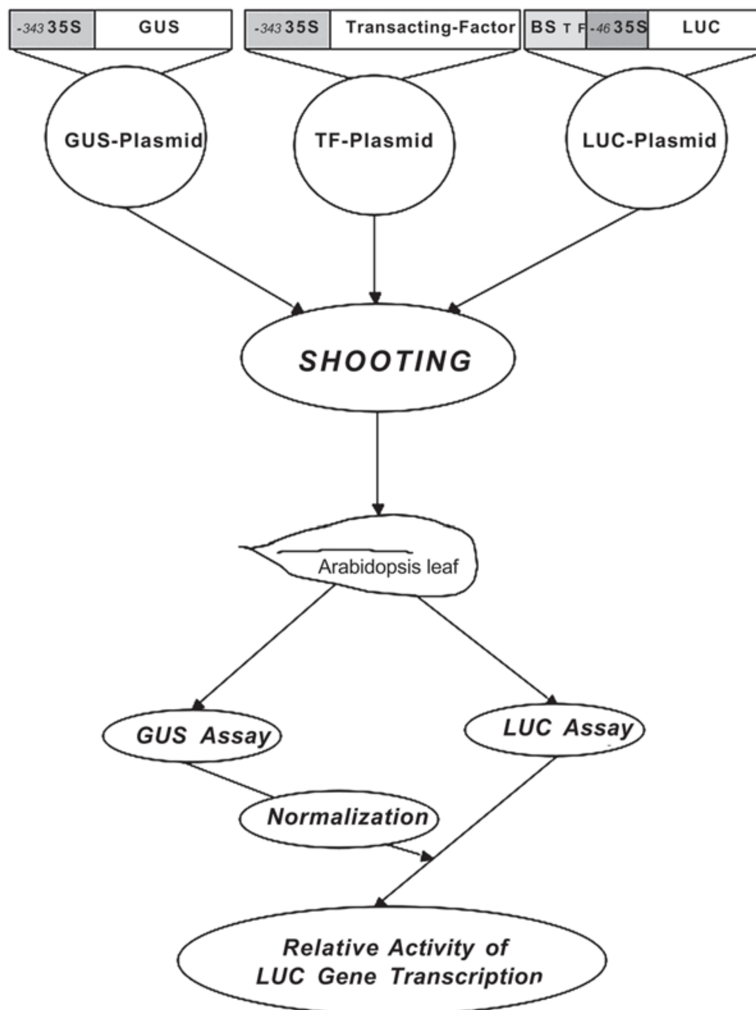


Fig. 1. Schematic representation of the functional assay for transcription factors in *Arabidopsis*. Three distinct plasmids are used in the assay. The first plasmid carries a gene coding a TF or its derivative(s) driven by the CaMV 35S promoter. The second plasmid contains a chimeric promoter recognized by the TF driving the expression of the LUC reporter gene. The third plasmid carries a GUS gene driven by the CaMV 35S promoter, which is used as an internal control. The three plasmids are codelivered into *Arabidopsis* leaves by shooting. Twenty hours after bombardment, the leaves are assayed for LUC and GUS activities. LUC activity is normalized with respect to GUS activity for each bombardment experiment. In a typical assay, the average values of three independent experiments are calculated for each combination of transacting and reporter genes.

3.2. Arabidopsis *Growth and Tissue Preparation for Particle Bombardment*

3.2.1. Arabidopsis *Growth*

1. Weight out 2 mg of *Arabidopsis* seeds, and add 10 mL of sterile, double-distilled H_2O. Keep the seeds in H_2O for 4 d at 4°C in complete darkness.
2. Pour out the H_2O and add 3 mL 0.1% (w/v) agarose solution.

3. Use a Pasteur pipet to sow the seeds in three pots containing Universal Soil VM type.
4. Grow in a light/dark cycle 16:8 h at 21°C under high fluorescent light (Osram L58 W31, Photosynthetically Active Radiation 120 μmol/s/m), and water regularly using Mannalin-A at 0.5 g/L. For further instructions on growth conditions, please *see* Chapter 1 in this book.

3.2.2. Arabidopsis *Tissue Preparation*

1. Collect 3 or 4 leaves from 3- to 4-wk-old plants (*see* **Note 4**) and place them in a 3-cm diameter circle in the center of a Petri plate containing 0.5X MS and 0.9% (w/v) agar. Prepare two Petri plates for each sample to be shot (*see* **Note 5**).
2. Incubate the plates for 30 min to 2 h in a growth chamber before bombardment.

3.3. Preparation of DNA-Coated Gold Particles for Leaf Bombardment

3.3.1. Microcarrier Preparation

1. Weight out 60 mg of gold microparticles, add 1 mL 96% ethanol, mix for 10 min, and then incubate 15 min at room temperature.
2. Pellet the microparticles by spinning for 10 s in a microfuge, and remove the supernatant (*see* **Note 6**).
3. Add 1 mL sterile, double-distilled H_2O, vortex for 1 min, spin for 10 s in a microfuge, and remove the supernatant. Repeat this step three times (*see* **Note 6**).
4. Add 1 mL sterile, double-distilled H_2O, so that the final concentration is 60 mg/mL, assuming no loss during preparation.
5. Store the microcarrier at 4°C (*see* **Note 7**).

3.3.2. Preparation of DNA-Coated Gold Particles

1. Mix the microcarrier for 10 min to suspend and disrupt the agglomerated particles.
2. Transfer 40 mL of microcarrier solution, while vortexing, into a 1.5-mL microfuge tube.
3. Place the microfuge tube on a mixer, and add in order: 4 μL DNA (1.2 μg/μL; *see* **Note 8**), 40 μL 2.5 *M* $CaCl_2$, 16 μL 0.1 *M* spermidine-free base, and continue mixing for 3 min.
4. Spin for 10 s in a microfuge, and remove the supernatant (*see* **Note 6**).
5. Add 200 μL 96% ethanol without disturbing the pellet, and remove the supernatant.
6. Add 120 μL 96% ethanol without disturbing the pellet, and remove the supernatant.
7. Add 48 mL 96% ethanol, and resuspend the pellet by pipetting 3 or 4 times and by mixing for 20 s.
8. While mixing, remove four 10-μL aliquots of DNA-coated gold particles (*see* **Note 9**), and transfer each of them to the central 1 cm of a macrocarrier (*see* **Note 10**).
9. Dry particles on macrocarrier on filter paper in air.

3.4. Leaf Bombardment

3.4.1. Bombardment Conditions

Arabidopsis leaves are bombarded using the PDS-1000/HE biolistic device (*see* **Note 11**) with 1100 psi rupture disks. The macrocarrier is placed 6 mm from the stopping screen, and the samples to be bombarded are placed 9 cm from the macrocarrier. The vacuum applied to the chamber is 25 mmHg/in.

1. Bombard each leaf sample with 0.5 mg of gold particles coated with a DNA mixture that includes 0.15 μg of transacting plasmid, 0.6 μg of reporter plasmid, and 0.25 μg of reference plasmid.
2. Rotate each sample 90 degrees, and bombard it again.

3.4.2. Post-Bombardment Treatment

1. Incubate the bombarded leaves 20 h (*see* **Note 12**) at 21°C in a growth chamber, and then harvest and freeze them in liquid nitrogen.
2. Grind the leaves from each bombardment in a mortar and pestle precooled with liquid nitrogen.
3. Transfer the powder to a microfuge tube containing 0.2 mL of 1X CCLR buffer.
4. Centrifuge 20 min in a microfuge (20,000*g*) at 4°C, and recover the supernatant.

3.5. Determination of Relative Transactivation Levels

Total cell extracts are assayed for their specific GUS and LUC activities, essentially as previously described (***19,20***, and chapters 22 and 24 in this book). The relative transactivation level is calculated by normalizing LUC activity with respect to GUS activity (*see* **Note 13**).

3.5.1. GUS Assay

1. Transfer 100 µL of extract into a microfuge tube containing 100 µL of GUS buffer, and mix by pipetting.
2. Transfer 100 µL of this mixture into a microfuge tube containing 2 mL of GUS stop buffer, and store at 4°C.
3. Incubate the remaining 100 µL of the mixture for 1 h at 37°C.
4. Stop the reaction by adding 2 mL of GUS stop buffer.
5. Read the sample emission at 455 nm when the excitation is 365 nm using a fluorimeter (*see* **Note 14**).

3.5.2. LUC Assay

1. Transfer 50 µL of extract into a luminometer cuvet, and incubate 5 min at room temperature.
2. Read the luminescence of the sample in a luminometer for 30 s after injecting 150 µL of the LAR buffer.

4. Notes

1. In our experiments we routinely used pGEM plasmids. However, any type of plasmid with similar characteristics can be utilized.
2. For competition experiments between a positive and a negative regulator binding the same target sequence, two distinct transacting plasmids can be codelivered together with reporter and reference plasmids ***(14)***. The total amount of transacting plasmid DNA has to be the same as that used in standard transient assays performed with three plasmids. The relative ratio of the two transacting DNAs must be optimized for each combination of genes to be tested.
3. In our experiments we used reporter plasmids containing six copies of the TF binding site tandemly arranged.
4. We use entire leaves, as we noticed that cutting the leaves in smaller pieces may decrease the efficiency of the assay.
5. The two leaf samples must be processed independently in all the steps.
6. We noticed that longer centrifugations may result in particle aggregation.
7. Microcarrier can be stored at 4°C for up to 2 wk.
8. DNA must be free of protein; otherwise it forms clumps with gold particles. Moreover, DNA concentration has to be carefully checked to compare wild-type and mutant constructs. We used the Qiagen kit for DNA plasmid preparations. After spectrophotometer reading, DNA quantity and quality were checked by agarose gel electrophoresis.
9. Make sure the particles are well suspended while dispensing them. Discard the last 8 µL.
10. Be sure that the gold/DNA mixture is spread evenly, because any chunk will result in cell death.
11. Of course, different biolistic systems can be used for this assay. If this is the case, helium pressure, position of the macrocarrier from the stopping screen, and distance of the sample from the macrocarrier have to be optimized. The accelerating pressure influences the velocity as well as the distribution pattern of the gold particles. Moreover, optimal accelerating pressure varies depending on the type of tissues to be bombarded.
12. After bombardment, leaves were incubated in a growth chamber set with a light/dark cycle 16:8 h; leaves were usually exposed to 8 h of light, 8 h of dark, and 4 h of light before harvesting.
13. The luciferase value expressed as light units (lu) was corrected with the GUS activity expressed as pmol methylumbelliferone (MU) as follows: lu/30 s/pM MU/min. In a typical assay, we use three independent replicates, each of which is composed by two leaf samples. The normalized LUC values are averaged and the standard deviation is calculated.
14. The GUS activity is a good indicator of the efficiency of the assay. We routinely obtain ~0.5 pM MU/min.

Acknowledgments

This research was supported partly by MIUR-FIRB Post-Genoma Programme (contract no. RBNE01CFKB to I.R and G.M.) and European Union Life Science Programme (contract no. QLK3-CT-2000-00328 to I.R.).

References

1. Klein, T. M., Wolf, E. D., Wu, R., and Sanford, J. C. (1987) High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* **327,** 70–73.
2. Sanford, J. C., Smith, F. D., and Russell, J. A. (1993) Optimising the biolistic process for different biological applications. *Methods Enzymol*. **217,** 483–509.
3. Taylor, N. J. and Fauquet, C. M. (2002) Microparticle bombardment as a tool in plant science and agricultural biotechnology. *DNA Cell Biol*. **21,** 963–977.
4. Bruce, W. B., Christensen, A. H., Klein, T., Fromm M., and Quail, P. H. (1989) Photoregulation of a phytochrome gene promoter from oat transferred into rice by particle bombardment. *Proc. Natl. Acad. Sci. USA* **86,** 9692–9696.
5. Okubara, P. A., Williams, S. A., Doxsee, R. A., and Tobin, E. M. (1993) Analysis of genes negatively regulated by phytochrome action in *Lemna gibba* and identification of a promoter region required for phytochrome responsiveness. *Plant Physiol*. **101,** 915–924.
6. Kehoe, D. M., Degenhardt, J., Winicov, I., and Tobin E. M. (1994) Two 10-bp regions are critical for phytochrome regulation of a *Lemna gibba Lhcb* gene promoter. *Plant Cell* **6,** 1123–1134.
7. Viret, J. F., Mabrouk, Y., and Bogorad, L. (1994) Transcriptional photoregulation of cell-type-preferred expression of maize rbcS-m3: 3' and 5' sequences are involved. *Proc. Natl. Acad. Sci. USA* **91,** 8577–8581.
8. Purcell, M., Mabrouk, Y. M., and Bogorad, L. (1995) Red/Far-red and blue light-responsive regions of maize rbcS-m3 are active in bundle sheath and mesophyll cells, respectively. *Proc. Natl. Acad. Sci. USA* **92,** 11.504–11.508.
9. Ezcurra, I., Wycliffe, P., Nehlin, L., Ellerstrom, M., and Rask, L. (2000) Transactivation of the *Brassica napus* napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different cis-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. *Plant J.* **24,** 57–66.
10. Guan, L. M., Zhao, J., and Scandalios, J. G. (2000) Cis-elements and trans-factors that regulate expression of the maize *Cat1* antioxidant gene in response to ABA and osmotic stress: H_2O_2 is the likely intermediary signaling molecule for the response. *Plant J.* **22,** 87–95.
11. Skadsen, R. W., Sathish, P., Federico, M. L., Abebe, T., Fu, J., and Kaeppler, H. F. (2002) Cloning of the promoter for a novel barley gene, *Lem1*, and its organ-specific promotion of Gfp expression in lemma and palea. *Plant Mol. Biol*. **49,** 545–555.
12. Aoyama, T., Dong, C. H., Wu, Y., et al. (1995) Ectopic expression of the *Arabidopsis* transcriptional activator Athb-1 alters leaf cell fate in tobacco. *Plant Cell* **7,** 1773–1785.
13. Baumann, K., De Paolis, A., Costantino, P., and Gualberti, G. (1999) The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the rolB oncogene in plants. *Plant Cell* **11,** 323–334.
14. Steindler, C., Matteucci, A., Sessa, G., et al. (1999) Shade avoidance responses are mediated by the ATHB-2 HD-Zip protein, a negative regulator of gene expression. *Development* **126,** 4235–4245.
15. Kizis, D. and Pages, M. (2002) Maize DRE-binding proteins DBF1 and DBF2 are involved in rab17 regulation through the drought-responsive element in an ABA-dependent pathway. *Plant J.* **30,** 679–689.
16. Casaretto, J. and Ho, T. H. (2003) The transcription factors HvABI5 and HvVP1 are required for the abscisic acid induction of gene expression in barley aleurone cells. *Plant Cell* **15,** 271–284.
17. Zhang, Z. L., Xie, Z., Zou, X., Casaretto, J., Ho, T. H., and Shen, Q. J. (2004) A Rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. *Plant Physiol*. **134,** 1500–1513.
18. de Wet, J. R., Wood, K. V., De Luca, M., Helinski, D. R., and Subramani, S. (1987) Firefly luciferase gene: Structure and expression in mammalian cells. *Mol. Cell. Biol.* **7,** 725–737.
19. Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987) GUS fusion: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6,** 3901–3907.
20. Millar, A. J., Short, S. R., Chua, N.-H., and Kay, S. A. (1992) A novel circadian phenotype based on firefly luciferase expression in transgenic plants. *Plant Cell* **4,** 1075–1087.

20

Transfection Assays With Protoplasts Containing Integrated Reporter Genes

Shiv Tiwari, Shucai Wang, Gretchen Hagen, and Tom J. Guilfoyle

Summary

Transient expression assays with protoplasts that utilize stably integrated reporter genes along with transfected effector genes provide several advantages over assays in which both the reporter gene and effector gene(s) are transfected into protoplasts. A protocol for carrying out transient expression assays with *Arabidopsis* leaf mesophyll protoplasts containing single-copy integrated reporter genes is described.

Key Words: Arabidopsis mesophyll protoplasts; protoplast transfection assays; stably incorporated reporter gene.

1. Introduction

Transient transfection assays with plant protoplasts are commonly used to analyze ***(1)*** gene expression in response to hormones, metabolites, environments, mutations, and other factors, ***(2)*** promoter elements involved in regulating expression of genes, ***(3)*** roles played by transcription factors or signal-transduction proteins in regulating gene expression, and ***(4)*** subcellular localization of proteins (reviewed by in ref. ***1***). Most types of transient assays in protoplasts rely on assessing the levels of gene expression for a transfected reporter gene containing a specific type of promoter (e.g., hormone-inducible, heat-inducible, etc.) fused to a gene that encodes an easily detectable gene product (e.g., β-glucuronidase [GUS], luciferase [LUC], green fluorescent protein [GFP], etc.). In many cases, one or more effector genes that consist of a constitutive or inducible promoter fused to a structural gene may be cotransfected with the reporter gene to test the response of the reporter gene to the effector gene product. The effector gene product might be a transcription factor, a protein involved in signal transduction, or some other type of protein. Although transient assays in protoplasts with transfected reporter genes have provided a wealth of information about signals and signaling pathways in regulating inducible gene expression, *cis* elements required for promoter function, and transcription factors and signaling proteins that regulate expression of genes, there are several potential limitations and pitfalls inherent in these types of assays. First, at least two or more independent transfection events must occur: transfection of the reporter gene and transfection of the effector gene or genes into the same cell. Second, if these independent events do occur, multiple copies of the effector gene and/or reporter gene may be transfected into cells. Transfection of multiple reporter genes provides many potential targets for both endogenous transcription factors and those expressed from effector gene(s). Third, the reporter genes are transfected as nonreplicating plasmids (naked DNA) and must be transported to the nucleus, unlike integrated templates which are localized in the nucleus, replicated, and bound in an ordered chromatin arrangement. Fourth, if transfected reporter genes eventually do reach the nucleus and assemble into chromatin-bound templates, these templates may not entirely resemble the ordered chromatin structure of integrated, replicated templates ***(2)***.

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

To overcome some of the potential pitfalls with transfected reporter genes, we have been conducting transfection assays in protoplasts using integrated reporter genes. The use of single-copy, stably integrated reporter genes offers several advantages for monitoring a reporter gene's response to chemical and environmental signals or an effector gene product. The expression of a transfected and an integrated auxin-responsive reporter gene ***(3,4)*** in wild-type and *nph4-2/arf7* mutant ***(5,6)*** *Arabidopsis* leaf mesophyll protoplasts will be used to illustrate the different results obtained with the two types of reporter genes in response to auxin and a transfected effector gene that encodes a transcriptional activator, namely auxin response factor 7 or ARF7 ***(7–9)***. The differences and advantages of using the integrated auxin-responsive reporter genes compared to transfected reporter genes are discussed.

2. Materials

2.1. Plant Material

Arabidopsis thaliana (ecotype Columbia) seeds were germinated in pots containing moistened Pro-Mix (Premier Horticulture Inc, Red Hill, PA), and plants were grown at 20°C under continuous white light. Plants were fertilized once per week (Peters All-Purpose Plant Food 20-20-20). Leaves from 25 to 35 d-old plants were used for protoplast isolation (*see* **Note 1**). For callus generation, seeds were surface sterilized and treated as described in **Subheading 3.3.1.**

2.2. Buffers and Solutions

1. Cellulase-macerozyme solution for *Arabidopsi*s leaf mesophyll protoplasts: 1% w/v Cellulase R10 (Crescent Chemical Co., Islandia, NY), 0.25% w/v Macerozyme R10 (Crescent Chemical Co.), 400 m*M* mannitol, 10 m*M* CaCl2, 5 m*M* MES (2-[N-morpholino]ethanesulfonic acid). Adjust to pH 5.7 with 1 *N* KOH; filter-sterilize (0.45 µm cellulose acetate). Store at 4°C for up to 1 wk.
2. Driselase solution for *Arabidopsis* cell suspension culture protoplasts: 5 m*M* MES, 400 m*M* sorbitol. Adjust to pH 5.0 with 1 *N* KOH. Add 2% w/v Driselase (Sigma, cat. no. D9515) and stir for 30 min at 4°C. Spin at 4600*g* for 1 h and pass the supernatant through a 0.45-µm cellulose acetate filter. Store at –20°C in 25 to 40-mL aliquots.
3. 200 m*M* $CaCl_2$.
4. 2,4-D: 10 mg/mL stock, pH 6.0.
5. Kinetin: 5 mg/mL stock in 10% HCl.
6. Mg-mannitol solution: 400 m*M* mannitol, 15 m*M* $MgCl_2$, 4 m*M* MES. Adjust to pH 5.7 with 1 *N* KOH; filter-sterilize (0.22 µm cellulose acetate). Store at 4°C for up to 3 mo.
7. WI solution: 500 m*M* mannitol, 20 m*M* KCl, 4 mM MES. Adjust to pH 5.7 with 1 *N* KOH. Filter-sterilize (0.22 µm cellulose acetate). Store at 4°C for up to 3 mo.
8. W5 solution: 154 m*M* NaCl, 5 m*M* KCl, 125 m*M* $CaCl_2$, 5 m*M* glucose. Filter-sterilize (0.22 µm cellulose acetate) and store at room temperature.
9. MC solution: 5 m*M* MES, 20 m*M* $CaCl_2$, 500 m*M* mannitol. Adjust to pH 5.7 with 1 *N* KOH. Filter-sterilize (0.22 µm cellulose acetate). Store at –20°C in 25–40 mL aliquots.
10. 40% PEG (polyethylene glycol) solution:
    a. Make 500 mL Ca-mannitol solution (100 m*M* $CaNO_3$, 400 m*M* mannitol).
    b. Add 160 g of PEG (avg. mol. wt. 3350) in 280 mL of Ca-mannitol solution. Warm the solution and stir continuously until the solution is clear.
    c. Adjust solution to pH 10.0 by dropwise addition of 1 *N* KOH. Make up volume to 400 mL with Ca-mannitol solution.
    d. Filter-sterilize (0.45 µm HV Durapore Membrane, Millipore Corp., Bedford, MA, Cat. no. SCHVU05RE) and store at –200°C in 25 to 40 mL aliquots.
11. Callus-inducing Medium (CIM): 4.3 g Murashige and Skoog Salt Mixture (MS salts), 1 mL Gamborg's Vitamin Solution (1000X), 200 µL 2,4-D, 10 µL kinetin, 20 g sucrose, 0.8% w/v agar (Type A; Plant and Cell Culture Tested; Sigma). Adjust to pH 5.7 with 1 *N* KOH and bring to 1 L with deionized H_2O. Autoclave for 20 min and pour into plastic Petri dishes.

12. Cell suspension medium (CSM): 6.2 g Gamborg's B-5 Basal Salt Mixture, 2 mL Gamborg's Vitamin Solution (1000X), 100 μL 2,4-D, 10 μL kinetin, 20 g sucrose, 0.6 g MES. Adjust to pH 5.7 with 1 *N* KOH and bring to 1 L with deionized H_2O. Autoclave for 20 min.
13. Protoplast suspension medium (PSM): 4.3 g MS salts, 1 mL Gamborg's Vitamin Solution (1000X), 40 g sucrose. Adjust to pH 5.7 with 1 *N* KOH and and bring to 1 L with deionized H_2O. Autoclave for 20 min.
14. MUG buffer: 50 m*M* sodium phosphate buffer (pH 7.0), 10 m*M* EDTA, 10 m*M* dithiothreitol, 0.1% (w/v) sodium-lauryl sarcosine, 0.1% (v/v) Triton X-100.
15. MUG assay solution: 1 m*M* 4-methylumbelliferyl-β-D-glucuronide (MUG) in MUG buffer.
16. MUG stop solution: 600 m*M* Na2CO3. Filter-sterilize (0.22 μm cellulose acetate) and store at room temperature.
17. MU stock solution: 100 m*M* sodium methylumbelliferone (MU) in deionized water. Store at 4°C in the dark. Dilute the MU stock solution with MUG stop solution to make different standards.
18. 50% Bleach (2.6% sodium hypochlorite).
19. 95% ethanol.

2.3. Other Media and Equipment

1. Cell culture lysis reagent (Promega Corp., Madison, WI; cat. no. 153A).
2. Luciferase assay system (Promega Corp., cat. no. E1501).
3. 15-mL disposable sterile centrifuge tubes.
4. 200 (mm mesh nylon (Spectra/Mesh Nylon, Spectrum Laboratories, Inc., Rancho Dominguez, CA).
5. Microplate fluorometer used for MUG assays. Sterile 96-well cell culture cluster plates.
6. Luminometer used for measuring luciferase activity.
7. Hemacytometer.
8. Rotary shaker.

3. Methods

3.1. Selection of Source for Protoplasts

In many cases, the choice between *Arabidopsis* leaf tissue or cell suspension cultures to prepare protoplasts for transfection assays is somewhat arbitrary. There are both advantages and disadvantages to choosing one versus the other (we limit our discussion of these two reporter gene expression assays). With transfected reporter genes, either mesophyll protoplasts or cell suspension protoplasts can generally be used to analyze a promoter for *cis* elements or transcription factors and signaling proteins that modulate the promoter. However, the choice of mesophyll or suspension cell protoplasts can depend on whether the appropriate transcription factors or signaling pathways are present in the type of protoplasts used for reporter gene analysis. With some auxin response genes (e.g., *GH3, DR5*; *see* 7) , we have found that either mesophyll or cell suspension protoplasts can be used for gene expression analysis if the reporter gene is transfected into protoplasts. With integrated reporter genes, however, cell suspension protoplasts are not an option, because the cell cultures require auxin, which activates the reporter gene and results in high levels of gene product (e.g., GUS) in the isolated protoplasts. This makes analysis of the transgene difficult, if not impossible, because of the high background reporter gene product present in protoplasts. It is possible that background reporter gene expression might be eliminated if cell suspension cultures are starved for auxin prior to protoplast isolation, but we have not attempted to do this. On the other hand, the relatively low level of auxin present in leaf tissue fails to significantly activate the reporter genes, and integrated auxin-responsive reporter genes in isolated mesophyll protoplasts are silent or nearly silent prior to auxin application, making reporter gene expression responses easy to measure. With integrated reporter genes, it is much more facile or convenient to use mesophyll protoplasts for evaluating responses of a reporter gene to effector genes that encode transcription factors or signaling proteins. Once lines are selected for homozygous, single-copy reporter genes, seeds need only be germinated and plants grown under uniform conditions for about 1

mo before leaf tissue is ready for harvest and protoplast preparation. If cell suspension protoplasts are chosen for analysis, cell suspension lines need to be established for each homozygous, single-copy reporter gene line that is selected, making the process time-consuming and cumbersome.

3.2. Isolation of Protoplasts from Arabidopsis Leaves and Transfection

The protocol described below for the isolation of protoplasts from *Arabidopsis* leaves and transfection assays is modified from protocols described previously ***(10,11)***.

3.2.1. Isolation of Leaf Mesophyll Protoplasts

1. Cut *Arabidopsis* leaves (~1 g) with a razor blade into 0.1 to 1-mm strips.
2. Transfer cut strips into 25 mL cellulase-macerozyme solution in a 9-mm Petri dish and vacuum-infiltrate at room temperature for 20 min.
3. Incubate for 90 min with gentle shaking (~40 rpm) on a rotary shaker in the dark.
4. Shake for additional 1 min at 80 rpm to release protoplasts. Filter protoplasts through a 200 μm mesh nylon into a 50-mL centrifuge tube.
5. Immediately dilute the protoplasts with 1/3 vol 200 m*M* $CaCl_2$, mix gently, and spin at 180*g* for 3 min to recover protoplasts (*see* **Note 2**).
6. Wash protoplasts once with 25 mL W5 solution and spin at 180g for 3 min to recover protoplasts.
7. Resuspend protoplasts in 25 mL W5 Solution and count protoplasts/mL using a hemacytometer (*see* **Note 3**).
8. Keep protoplasts in W5 solution for 20 to 30 min before transfection.

3.2.2. Protoplast Transfection Assays

1. Remove W5 solution by centrifugation at 180*g* for 3 min. Resuspend the protoplast pellet in Mg-mannitol solution at a final concentration of 3×10^5 protoplasts/mL.
2. Add 10 to 20 μg effector and/or reporter plasmid DNA (*see* **Note 4**) and 200 μL protoplasts solution to a 15 mL disposible sterile centrifuge tube and mix well.
3. Add an equal volume (200 μL) PEG solution and mix thoroughly, but gently, being careful to not damage or rupture the protoplasts.
4. Incubate at room temperature for 20 min.
5. Carefully add 0.8 mL W5 solution and incubate for an additional 10 min.
6. Mix thoroughly, but gently and spin at 180g for 3 min and remove the PEG solution carefully using a pipet with aspiration.
7. Resuspend protoplasts in 1 mL WI Solution (*see* **Note 5**).
8. Incubate in darkness for 18 to 24 h, keeping tubes in a horizontal position.
9. Spin at 180*g* for 3 min; remove the solution completely.
10. Add 100 μL 1X cell culture lysis reagent, mix thoroughly by vortexing, and perform reporter gene assays as described in **Subheading 3.4.**

3.3. Isolation of Protoplasts from Arabidopsis Cell Suspension Cultures and Transfection

3.3.1. Generation of Arabidopsis Cell Suspension Cultures

Arabidopsis calli were generated from sterilized seeds using protocols similar to those described previously ***(12)***. Seeds were surface-sterilized in sterile 1.5-mL microcentrifuge tubes by suspending the seeds in 1 mL 95% ethanol for 1 min, followed by centrifugation in a microcentrifuge for 1 min at 10,000*g* and removal of the ethanol with a pipet. Seeds were suspended in 1 mL 50% bleach for 10 min with intermittent vortexing, followed by centrifugation and removal of the bleach as described above. Seeds were then washed in sterile, deionized H_2O three times. Sterilized seeds were plated on callus-inducing medium (CIM) and grown at 20°C under continuous white light. Calli were transferred to fresh CIM plates every 3 wk. *Arabidopsis* cell suspension cultures were generated by transferring several calli into 25 mL of cell suspension medium (CSM). Cell suspension cultures were grown for 1 wk on a rotary shaker at 150 rpm under dim light. The friable cell masses released in the liquid medium were carefully

transferred to 25 mL of fresh CSM. Cells were subcultured every week by diluting 1:1 with fresh CSM. This was repeated for several weeks until a culture was fully established. The established cell cultures were transferred every week by diluting 1:5 with fresh CSM.

3.3.2. Isolation of Cell Suspension Culture Protoplasts

This protocol was adopted from the original protocol designed for carrot cell suspension cultures ***(8,13)*** and modified for *Arabidopsis cell* suspension cultures. In general, we use *Arabidopsis* cells 4 to 5 d after subculturing for making protoplasts and carrying out transfection.

1. Spin down 40 mL cells at 180*g* for 2 to 3 min (*see* **Note 2**).
2. Decant the supernatant, add 40 mL of Driselase solution to the cells, and mix gently.
3. Transfer the cell suspension to a 150-mm Petri dish, cover the dish with aluminum foil, and shake the dish gently at 40 to 50 rpm on a rotary shaker for 2.5 to 3 h.
4. Monitor the release of protoplasts using a light microscope, and continue gentle shaking as required.
5. Filter protoplasts through a 200-µm Spectra/Mesh nylon membrane to remove cellular debris.
6. Pellet the protoplasts at 180*g* for 2 to 3 min. Carefully remove the Driselase solution and wash protoplasts once with 40 mL W5 solution.
7. Suspend the protoplasts in 40 mL of W5 solution and count the protoplasts/mL using a hemacytometer. Keep the protoplasts in W5 solution for 20 to 30 min before carrying out transfection assays.

3.3.3. Protoplast Transfection Assays

1. Immediately before transfection, pellet the protoplasts and suspend in MC solution at a final concentration of 2×10^6 protoplasts/mL.
2. In a 15-mL sterile conical tube, add 10 to 20 µg of purified effector and/or reporter plasmid DNA (*see* **Note 4**) and 300 µL of protoplast solution. Mix gently.
3. Add an equal volume (300 µL) of 40% PEG solution and mix the contents by gentle shaking.
4. Incubate the protoplast–DNA–PEG mixture at room temperature for 15 to 20 min. After incubation, add 4 mL of PSM, mix gently and incubate for 18 to 24 h in the dark, keeping the tubes in a horizontal position (*see* **Note 5**).
5. Pellet the protoplasts at 180*g* and carefully remove the supernatant. Protoplasts are now ready for cytological analysis, protein analysis, and reporter gene assays.
6. For reporter gene assays and protein analysis, lyse the protoplasts with 200 µL of 1X cell culture lysis reagent.

3.4. Reporter Gene Assays

3.4.1. MUG Assays

1. Add 10 µL of lysed protoplasts in a sterile 96-well cell plate and add 100 µL of MUG assay solution to each well. Cover the plate and incubate for 0.5 to 1 h at 37°C.
2. After incubation, add 100 µL of MUG stop solution to each well and measure the fluorescence using a microplate fluorometer.
3. Make standard solutions of MU and measure the fluorescence as described above for the standard curve and quantitative measurement of enzyme activity in the samples.

3.4.2. Luciferase Assays

1. Add 2 µL of lysed protoplasts into a 500 µL microfuge tube without lid and mix with 100 µL luciferase assay reagent.
2. Immediately place the tube in the luminometer and measure the light produced.

3.5. Comparison of Auxin-Responsive Gene Expression With a Transfected Versus an Integrated Reporter Gene

The *DR5:GUS* reporter gene ***(3,4)*** provides an example to illustrate the differences observed for transient versus integrated reporter gene expression in response to auxin. In reporter gene

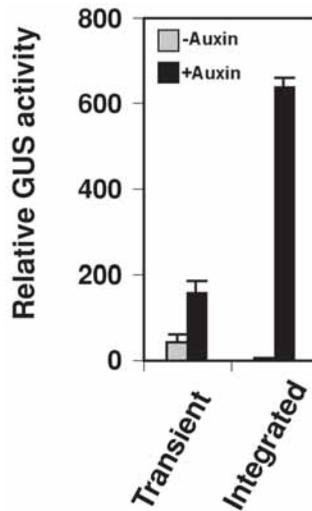


Fig. 1. *DR5:GUS* reporter gene expression in wild-type *Arabidopsis* mesophyll protoplasts. GUS activity was measured using a transfected *DR5:GUS* (10 µg) or a single-copy integrated *DR5:GUS* reporter gene.

transfection assays with wild-type *Arabidopsis* mesophyll protoplasts, *DR5:GUS* is activated three to four-fold in the presence of auxin (**Fig. 1**). This contrasts with the 50-fold increase in reporter gene expression observed with the integrated *DR5:GUS* gene in mesophyll protoplasts. The auxin response is much more robust in terms of GUS activity (i.e., units of GUS) as well as fold increase (i.e., minus auxin compared to plus auxin) with the integrated reporter gene.

The expression of a transfected or integrated *DR5:GUS* gene in *nph4-2/arf7* mutant (i.e., null for ARF7) *Arabidopsis* mesophyll protoplasts provides an example that compares the two types of reporter genes in response to a transfected effector gene. With *nph2-4* mesophyll protoplasts, both the transfected and integrated reporter genes display low levels of gene activity in the presence or absence of auxin (**Fig. 2A** and **2B**, no effector gene), presumably because ARF7 is required for high levels of expression of the *DR5:GUS* gene in *Arabidopsis* leaf mesophyll cells. Transfection of an effector gene encoding ARF7 under control of the cauliflower mosaic virus (CaMV) *35S* promoter results in strong activation of the reporter gene; however, this activation is entirely dependent on auxin with the integrated reporter gene, but less dependent on auxin with the transfected reporter gene (**Fig. 2A** and **2B**, ARF 7 effector gene). While the transfected reporter gene shows high constitutive expression in the presence of the ARF7 effector gene and shows only a modest four to five-fold increase in response to auxin, the integrated reporter gene is tightly regulated, being silent in the absence of auxin and induced 25 to 50-fold in the presence of auxin.

We believe assays like those described above for integrated reporter genes can be applied to a wide variety of other types of reporter genes, and that protoplast effector gene transfection assays with integrated reporter genes offer several advantages over assays that require transfection of both reporter genes and effector genes. First, the inducibility of integrated reporter genes by chemical or environmental agents is likely to be more tightly controlled and possibly more robust than transfected reporter genes because of the chromosomal nature of the integrated reporter genes. We have observed this with the auxin-responsive reporter genes, and similar observations have been made with the *c-jun* promoter:chloramphenicol acetyltransferase reporter gene in mouse F9 embryonal carcinoma cells ***(14)***. Second, the ordered chromatin structure associated with replicating chromosomal templates provides a more natural chromatin context for reporter genes than nonreplicating, nonintegrated reporter templates. These more

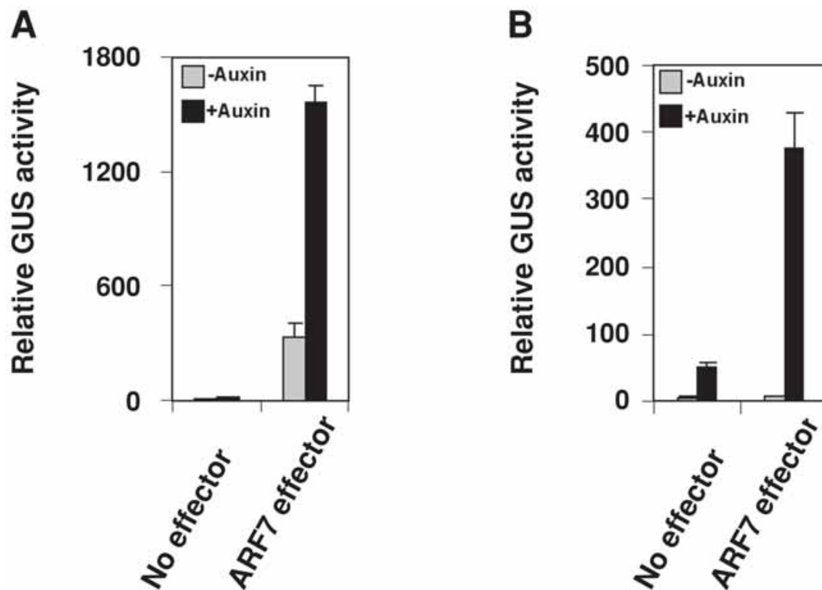


Fig. 2. *DR5:GUS* reporter gene expression in *nph4-2/arf7* mutant *Arabidopsis* mesophyll protoplasts. GUS activity was measured using a transfected *DR5:GUS* (10 μg) or a single-copy integrated *DR5:GUS* reporter gene with or without a transfected *CaMV35S:ARF7* effector gene (10 μg).

natural templates may reveal additional chromosomal factors involved in regulating specific genes, such as histone deacetylases, histone acetyltransferases, and chromatin remodeling enzymes. Third, the use of integrated reporter genes may reveal additional domains or domain functions within transcription factors that are not functional on transfected reporter genes. DNA binding of transcription factors may differ on integrated versus nonintegrated templates, or activation/ repression domains may be altered in their responses to cues that only operate on integrated genes. For example, in metazoan cells, some repression and activation domains function only on integrated templates. In human cells, a transcriptional activation domain was identified in the Epstein-Barr virus nuclear antigen 1 transcription factor that functioned on integrated, but not nonintegrated templates ***(15)***. In *Drosophila* cultured cells, a repression domain of the Engrailed repressor protein showed much stronger activity with integrated templates compared to transfected templates ***(16)***.

When combined with mutants that do not produce a specific transcription factor(s) or signaling protein(s) or produce defective transcription factor(s) or signal protein(s), the employment of integrated reporter genes can provide additional insight into regulatory processes involved in gene expression. The many mutants available in *Arabidopsis* provide fertile ground for studies with protoplasts harboring integrated reporter genes.

4. Notes

1. In our hands, the age of the *Arabidopsis* leaves is an important factor in obtaining protoplast preparations that give consistent results. In general, we use leaves from plants that are 25 to 35 d old.
2. We routinely pellet the protoplasts using a JS7.5 rotor (Beckman Coulter, Fullerton, CA) at 180g for 3 min.
3. One gram of *Arabidopsis* leaves will generally yield >10^7 protoplasts.
4. Plasmid DNAs are purified by the Endofree maxi kit from Qiagen, Valencia, CA, and stored as 1 to 5 mg/mL stocks in TE buffer (10 m*M* Tris-HCl, 1 m*M* EDTA, pH 8.0).

5. It is at this stage in the protocol that we add or withhold exogenous auxin (1-naphthaleneacetic acid or 1-NAA) to the protoplasts.

Acknowledgments

We thank Dr. Jen Sheen and Dr. Xiao-Jun Wang for helpful discussions on *Arabidopsis* mesophyll protoplast isolation and transfection protocols, and we thank Dr. Mannie Liscum for the *nph4-2* mutant line. This work was supported by National Science Foundation Grant MCB 00800096 to T. J. G. and G. H.

References

1. Sheen, J. (2001) Signal transduction in maize and *Arabidopsis* mesophyll protoplasts. *Plant Physiol.* **127,** 1466–1475.
2. Smith, C. L. and Hager, G. L. (1997) Transcriptional regulation of mammalian genes in vivo. A tale of two templates. *J. Biol. Chem.* **272,** 27,493–27,496.
3. Ulmasov, T., Murfett, J., Hagen, G., and Guilfoyle, T. J. (1997) Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *Plant Cell* **9,** 1963–1971.
4. Murfett, J., Wang, X-J., Hagen, G., and Guilfoyle, T. J. (2001) Identification of Arabidopsis histone deacetylase HDA6 mutants that affect transgene expression. *Plant Cell* **13,** 1047–1061.
5. Stowe-Evans, E. L., Harper, R. M., Motchoulski, A. V., and Liscum, E. (1998) NPH4, a conditional modulator of auxin-dependent differential growth responses in *Arabidopsis*. *Plant Physiol.* **118,** 1265–1275.
6. Harper, R. M., Stowe-Evans, E. L., Luesse, D. R., et al. (2000) The NPH4 locus encodes the auxin response factor ARF7, a conditional regulator of differential growth in aerial *Arabidopsis* tissue. *Plant Cell* **12,** 757–770.
7. Guilfoyle, T. J. and Hagen, G. (2001) Auxin response factors. *J. Plant Growth Reg.* **10,** 281–291.
8. Ulmasov, T., Hagen, G., and Guilfoyle, T. J. (1999) Activation and repression of transcription by auxin response factors. *Proc. Natl. Acad. Sci.USA* **96,** 5844–5849.
9. Tiwari, S. B., Hagen, G., and Guilfoyle, T. (2003) The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell* **15,** 533–543.
10. Abel, S. and Theologis, A. (1994) Transient transformation of *Arabidopsis* leaf protoplasts: a versatile experimental system to study gene expression. *Plant J.* **5,** 421–427.
11. Kovtun, Y., Chiu W-L., Tena, G., and Sheen, J. (2000) Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Nat. Acad. Sci. USA* **97,** 2940–2945.
12. Doelling, J. H. and Pikaard, C. S. (1993) Transient expression in *Arabidopsis thaliana* protoplasts derived from rapidly established cell suspension cultures. *Plant Cell Rep.* **12,** 241–244.
13. Liu, Z. B., Ulmasov, T., Shi, X., Hagen, G., and Guilfoyle, T. J. (1994) The soybean GH3 promoter contains multiple auxin-inducible elements. *Plant Cell* **6,** 645–657.
14. Kitabayashi, I., Kawakami, Z., Chiu, R., et al. (1992) Transcriptional regulation of the c-jun gene by retinoic acid and E1A during differentiation of F9 cells. *EMBO J.* **11,** 167–175.
15. Kennedy, G. and Sugden, B. (2003) EBNA-1, a bifunctional transcriptional activator. *Mol. Cell. Biol.* **23,** 6901–6908.
16. Tolkunova, E. N, Fujioka, M., Kobayashi, M., Deka, D., and Jaynes, J. B. (1998) Two distinct types of repression domain in Engrailed: One interacts with the Groucho corepressor and is preferentially active on integrative target genes. *Mol. Cell. Biol.* **18,** 2804–2814.

21

Chloroplast Genetic Engineering Via Organogenesis or Somatic Embryogenesis

Amit Dhingra and Henry Daniell

Summary

Chloroplast genetic engineering offers a number of unique advantages, including high-level transgene expression, multigene engineering in a single transformation event, transgene containment via maternal inheritance, lack of gene silencing, position and pleiotropic effects and undesirable foreign DNA. More than 40 transgenes have been stably integrated and expressed via the tobacco chloroplast genome to confer desired agronomic traits or express high levels of vaccine antigens and biopharmaceuticals. Despite such significant progress, this technology has not been extended to other important plant species. For example, *Arabidopsis* may be an ideal model system for chloroplast functional genomics. The employment of chloroplast transformation technology in *Arabidopsis* has been hampered by the lack of an efficient and reproducible protocol that provides fertile chloroplast transgenic plants. Transformation of the *Arabidopsis* chloroplast genome was achieved via organogenesis but the efficiency was at least a 100-fold lower than in tobacco and had the drawback of polyploidy in the leaf tissue that resulted in sterile transgenic plants. This problem can be overcome by adapting procedures that are now available to regenerate plants from both diploid and tetraploid explants via callus. In addition, it is feasible to regenerate *Arabidopsis* via somatic embryogenesis. Recent breakthroughs in highly efficient plastid transformation of recalcitrant crops such as cotton and soybean have opened the possibility of engineering *Arabidopsis* plastid genome via somatic embryogenesis. Therefore, protocols of recent improvements in tissue culture, DNA delivery, and the novel vector designs are provided here in order to achieve highly efficient plastid transformation in Arabidopsis.

Key Words: Plastid transformation; maternal inheritance; transgene containment; chloroplast-derived agronomic traits; chloroplast-derived therapeutic proteins; non-green plastids; *Arabidopsis*.

1. Introduction

Chloroplast genetic engineering is a relatively new approach for genetic modification of plants that offers high levels of foreign gene expression and containment of the transgenes in most plant species. In addition to addressing low levels of transgene expression, this approach overcomes gene silencing, position effects, pleiotropic effects, and presence of antibiotic-resistant genes or vector sequences in transformed genomes ***(1,2)***. Chloroplast transgenic approach has been successfully used in our laboratory to confer desired plant traits, including herbicide ***(3)***, insect ***(4)*** and disease ***(5)*** resistance, drought tolerance ***(6)***, salt tolerance ***(7)***, and phytoremediation ***(8)***. Employing knowledge gained through these studies, we have further demonstrated expression and assembly of several vaccine antigens ***(9–11)***.

To advance the concept of oral delivery of therapeutic proteins, we developed an antibiotic free chloroplast transformation system ***(12)*** and plastid transformation in carrot, with high levels of transgene expression in the edible part, the root ***(7)***. Several successful approaches to

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

eliminate selectable marker genes from transgenic chloroplasts *(13–15)* are also available. We have also demonstrated expression of very small (20 aa) therapeutic proteins such as magainin *(5)* or large proteins such as anthrax protective antigen (85 kDa, ref. *10*) or human blood proteins such as human serum albumin, which are highly susceptible to proteolytic degradation *(16)*. Interferon γ, which is susceptible to proteolytic degradation was stabilized in transgenic chloroplasts by fusing with β-glucuronidase *(17)*. We have also expressed and assembled interferon α or multisubunit proteins such as Guy's 13 monoclonal antibody in transgenic chloroplasts, with proper folding and disulfide bonds *(18,19)*. Furthermore, we have demonstrated integration of bacterial operons or multiple transgenes via the chloroplast genome *(8,20)*, with exceptionally large accumulation of foreign proteins (up to 46% total leaf protein, *20*). This has opened the door for engineering novel biosynthetic pathways for nutritional enhancement or several vaccine antigens (multivalent vaccines) through a single transformation event.

Daniell and McFadden *(21)* reported the very first foreign gene expression in higher plant plastids in 1987. Since then this concept has been extensively employed in the model plant tobacco. In the past few years, this technology has been extended to potato *(22)* and tomato *(23)* but the efficiency of transformation is extremely low. Most recently, highly efficient protocols for chloroplast transformation of carrot *(7)*, cotton *(24)* and soybean *(25)* via somatic embryogenesis have been developed. Given the fact that most plant species are regenerated via somatic embryogenesis, including *Arabidopsis*, this should serve as a model system for chloroplast transformation of other plant species.

The chloroplast genome consists of a circular molecule capable of self-replication. In most plant species it consists of a large single copy (LSC), small single copy (SSC), and two inverted repeat regions. Chloroplast genome size varies from 120 to 220 kb among different plant species *(26)*. Cytogenomic analysis has revealed a highly dynamic structural organization of chloroplast DNA. Contrary to an earlier held notion, it is arranged in both linear and circular conformation with 1 to 4 tandem copies of the genome *(27)*. Complete sequence information is now available for only 10 chloroplast genomes (six crops) from higher plants (**Table 1**).

Chloroplast transformation is achieved by biolistic or polyethylene glycol (PEG)-mediated delivery of the transformation cassette. Chloroplast flanking sequences on either side of the transgene facilitate homologous recombination and its site-specific integration. Initially a few genome copies are integrated but after about 15 to 20 cell divisions under selection pressure, a relatively homogenous population of transformed plastid genomes is obtained. Several transgenes have been integrated at different spacer regions into the chloroplast genome (*see* **Tables 2** and **3**). A transgene integrated into one inverted repeat region duplicates itself into the second inverted repeat region by the phenomenon of copy correction (**Fig. 1**). The presence of a chloroplast origin of replication within the chloroplast flanking sequence *(28,29)* might facilitate its replication inside the chloroplast, thereby increasing its copy number. This enhances the probability of transgene integration and expedites the process of achieving a homogenous population of transformed plastid genomes, a state called as homoplasmy *(3,30,31)*. Stable chloroplast transformation in higher plants has been achieved in tobacco *(32)*, Arabidopsis *(33)*, potato *(22)* and tomato *(23)* and most recently in carrot *(7)*, cotton *(24)* and soybean *(25)*. Detailed protocols for tobacco chloroplast transformation have been published earlier *(34–36)*. This chapter focuses on stable plastid transformation of *Arabidopsis*.

2. Materials

2.1. Chloroplast Transformation Vector Construction

1. Total cellular DNA or chloroplast DNA. Routinely, Arabidopsis thaliana ecotype RLD–mature leaf tissue is used.
2. Chloroplast DNA sequence information (*see* **Note 1**). Primers are designed to anneal with the chloroplast genome for the amplification of flanking sequences and the requisite regulatory elements like promoters and 5' and 3' untranslated regions. The size of the primers could range from 20 to 25 bp, and could be designed using any of the freely available web based primer

Table 1
List of Sequenced Higher Plant Chloroplast Genomes

Plant		Size	Accession no.
1.	*Arabidopsis thaliana*	154478 bp	NC_000932
2.	*Atropa belladonna*	156687 bp	NC_004561
3.	*Lotus corniculatus*	150519 bp	NC_002694
4.	*Medicago truncatula*	124033 bp	NC_002694
5.	*Nicotiana tabacum*	155939 bp	NC_001879
6.	*Oryza sativa*	134525 bp	NC_001320
7.	*Spinacia oleracea*	150725 bp	NC_002202
8.	*Triticum aestivum*	134545 bp	NC_002762
9.	*Zea mays*	140387 bp	NC_001666
10.	*Saccharum officinarum*	141182 bp	NC_006084

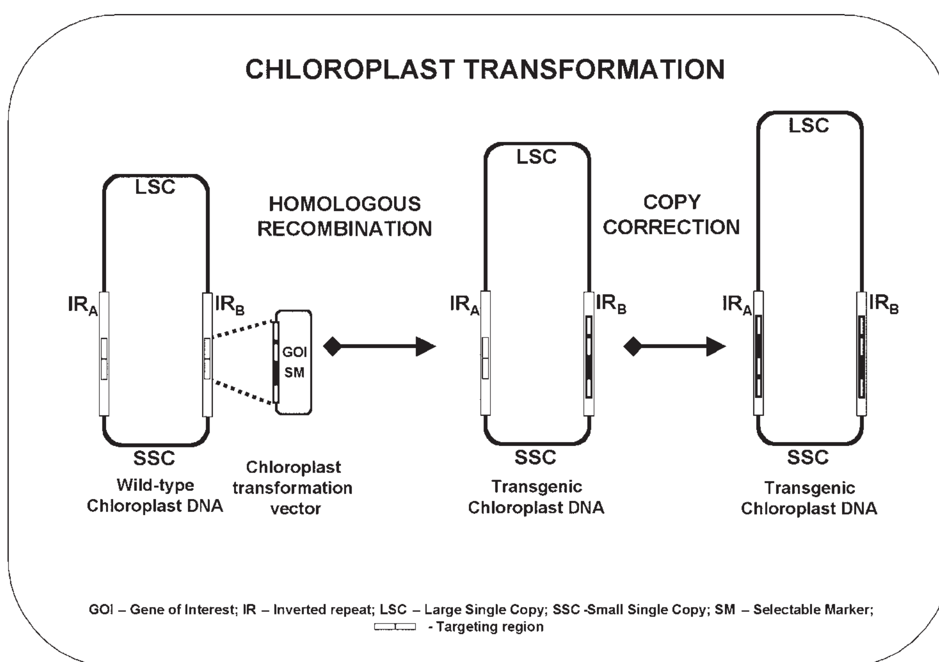


Fig. 1. Schematic representation of chloroplast transformation showing the phenomenon of homologous recombination and copy correction.

designing software. A preferred site of integration would be the intergenic region between the trnI-trnA genes of the 16S operon. Please see **Tables 2** and **3** for the sites of integration used for chloroplast transformation.

3. PCR mix: 100 to 200 ng DNA, 1X buffer (according to manufacturer's protocol), 200 to 300 μ*M* dNTPs, 15 to 20 pmole primer 1, 15 to 20 pmoles Primer 2, 1 to 2.5 units DNA polymerase. Make up the total volume to 50 μL with sterile Milli Q (Millipore) grade water.
4. Pfu-based DNA polymerase (Pfx-Platinum, Invitrogen; Pfu Turbo, Stratagene).
5. DNA modifying enzymes: *Taq* DNA polymerase, T4 DNA polymerase for producing blunt ends, calf intestinal alkaline phosphatase (CIAP) to remove 5' and 3' phosphoryl groups from nucleic acids, and T4 DNA ligase to form phosphodiester bonds.
6. PCR cloning kit: TOPO cloning kit (Invitrogen), including the pCR 2.1 vector.
7. Plasmid purification kit: Plasmid Midi-Kit (Qiagen) or similar.

Table 2
List of Genes Integrated into the trnI-trnA Intergenic Spacer Region of the rRNA Operon Within the Chloroplast Genome

Insertion site	Transcription status	Transgene integrated	% Total soluble Protein	Protein size (kDa)	Ref.
*trnI/trn*A	Active (same strand promoterless spacer region)	*aadA/aro*A^D	ND	47.623	3
		*aadA/ctx*B^D	4%	11.6	9
		*aad*A/*ltx*B^D	2.5%	11.6	43
		*aadA/ctx*B-CPV^D	31.1%	14.0	11
		aadA/gfp-CPV^D	22.6 %	29.0	11
		aadA/pag^D	18.1 %	83.0	10
		*aad*A/CaF1-Lcrv^D	14.8%	53.0	19,44
		*aad*A/EG121^D	ND		31
		*aad*A/MSI-99^D	21.5%	2.381	5
		*aad*A/IGF-1^D	33%	7.6	18,45
		*aad*A/INFa5^D	ND	23.0	18,46
		*aad*A/INF-a2b^D	19%	21.5	18,47
		*aad*A/HSA^D	11.1%	66.5	16
		*aad*A/Guy's 13^D	ND	50.5, 23.6	18
		*aad*A/*Cry2*Aa2^D	2-3%	65.0	4
		aadA/*Cry2*Aa2 operon^P	46.1%	65.0	20
		*aad*A/*tps*^{D}	ND	56.0	6
		*aad*A/*mer*A-*mer*B^P	ND	69.0, 24.0	8
		*aad*A/*badh*^{D}	ND	54.275	12
		(*Daucus carota, carrot)*			7
		*aad*A/*Rbc*S^{D,M}	ND	14.559	41
		*aad*A^M	ND	29.447	48
		*npt*II/*aph*A6^{D,M} (*Gossipium hirsutum*, cotton)	ND	ND	24
		*aad*A/gfp^D	ND	62.0	49
		*aad*A/*ubi*C^M	35%	ND	54

D, Dicistron; P, Polycistron; M, Monocistron

8. 500 m*M* EDTA.
9. 7.5 *M* ammonium acetate (pH 5.5).
10. Phenol/chloroform (50:50 v/v).

2.2. Tissue Culture Media (see Note 2*)*

1. ARM medium: 1X Murashige and Skoog (MS) salts, 3% sucrose, 0.8% TC agar, 2 mL/L of vitamin solution (100 mg myoinositol, 5 mg vitamin B1, 0.5 mg vitamin B6, 0.5 mg nicotinic acid, 1 mg glycine, 0.05 mg biotin per mL). The ARM medium is used for rooting.
2. ARM1 medium: ARM medium containing 3 mg indoleacetic acid, 0.15 mg 2,4-dichlorophenoxyacetic acid, 0.6 mg benzyladenine (BA), and 0.3 mg isopentenyladenine (2iP) per liter.
3. ARMIIr medium: ARM medium supplemented with 0.2 mg/L 1-naphthaleneacetic acid (NAA) and 0.4 mg/L indole-3-propionic acid (IPA).
4. ASI-NIBI medium: ARM medium supplemented with 1 mg/L NAA and 1 mg/L BA. This is the shoot induction medium.
5. ARM5 medium: ARM medium supplemented with 5% sucrose. This is the seed culture medium.

Table 3
List of Other Integration Sites Used for Chloroplast Transformation and Their Transcriptional Status

Insertion site	Transcription status	Ref.
*trn*H/*psb*A	Read-through, Same strand with promoter	50
*trn*G/*trn*fM	Silent, divergent genes, opposite strands	51
(L. esculentum, tomato)		23
*ycf*3/*trn*S	Silent, divergent genes, opposite strands	52,42
*rbc*L/*acc*	Read-through, same strand with promoter	53,3
(*S. tuberosum*, potato)		22
*pet*A/*psb*J	Silent, divergent genes, opposite strands	42,52
*pet*D/*rpo*A	Silent, divergent genes, opposite strands	42
3_*rps*12/*trn*V	Silent, divergent genes, opposite strands	56
(A. thaliana)		33
(*Glycine max*)		25
*trn*V/*rrn*16	Read-through, same strand with	57
(*S. tuberosum*)	promoter	22
*trn*N/*trn*R	Silent, divergent genes, opposite strands	52,58
*rp*l32/*trn*L	Read-through, same strand with promoter	59,60,61

6. 100 mg/mL spectinomycin; 100 mg/mL streptomycin; 50 mg/mL kanamycin. All stock solutions in distilled water. Antibiotic solutions are filter-sterilized under aseptic conditions in the laminar flow hood and stored at –20°C.

2.3. Particle Bombardment

1. Particle bombardment: Particle gun—PDS 1000 He, Microcarriers—gold or tungsten particles, Macrocarriers, macrocarrier holders, rupture disks—450, 650, and1100 psi (Bio-rad).
2. 2.5 *M* $CaCl_2$ (freshly prepared filter-sterilized solution).
3. 0.1 *M* spermidine-free base.
4. 50% (v/v) glycerol.
5. 70% ethanol.
6. Absorbent paper tissues, e.g., Kimwipes.

2.4. Molecular Analysis of Transformed Plants

1. DNeasy™ Plant Mini Kit (Qiagen, Valencia, CA).
2. Depurination solution: 250 m*M* HCl.
3. Denaturation solution: 1.5 *M* NaCl, 0.5 *N* NaOH.
4. Neutralization solution: 1 *M* Tris-HCl, pH 7.4, 1.5 *M* NaCl.
5. 20X SSC: 3 *M* NaCl, 0.3 *M* trisodium citrate.
6. Nylon membranes, e.g., Hybond N (Amersham Biosciences).

3. Methods

3.1 Cloning of Homologous Flanking Sequences

Homologous flanking sequences mediate homologous recombination and result in the insertion of the transgene cassette. These sequences can be obtained directly from chloroplast DNA either by restriction digestion of the desired fragment or by amplifying the region of interest by polymerase chain reaction (PCR). For the latter approach, a polymerase with proofreading properties should be used.

1. Design and synthesize primers that are 5' forward and 3' reverse complementary to the chloroplast sequence of interest.
2. Prepare the PCR mix: In a 50 µL PCR reaction add 1 µL of 100 ng/µL genomic DNA, 5 µL of 10X PCR reaction buffer (according to manufacturer's protocols), 1 µL of each 10 µ*M* primer solution, 200–250 m*M* each of dNTPs, 0.5 µL (2.5 units) Proof-reading DNA polymerase and sterile distilled water. Denature the PCR mix at 94°C for 5 min.
3. Perform 30 cycles of denaturation at 94°C for 30 s, annealing at 55–60°C for 30 s depending on the primers, and extension at 72°C or 68°C for 2–4 min depending on the size of the expected PCR product.
4. Extend the final product 8–10 minutes at 72°C or 68°C. Amplified DNA fragment is treated with Taq DNA polymerase in the presence of dATP in order to add A-overhangs that facilitate direct cloning of PCR products into pCR 2.1 cloning vector. For the addition of A-overhangs, the PCR product is incubated with 200 µ*M* dATP, *Taq* DNA polymerase at 72°C for 10 min. For cloning the DNA fragment with modified ends into pCR 2.1 vector, follow the manufacturer's protocol.

The cloned fragment representing the flanking sequences is excised from the vector by digesting with appropriate restriction enzymes and is blunt-ended using T4 DNA polymerase:

1. Resuspend the DNA in 1X T4 DNA polymerase reaction buffer supplemented with 100 m*M* dNTPs. Add one unit of T4 DNA polymerase per µg DNA.
2. Incubate for 15 min at 12°C.
3. Stop the reaction by adding EDTA to a final concentration of 10 mM and heating to 75°C for 20 min.

The blunt-ended fragment is then ligated with the help of T4 DNA ligase to PvuII digested pBluescript II KS dephosphorylated with CIAP:

1. Purify the digested DNA to be dephosphorylated by ethanol precipitation.
2. Resuspend the pellet in 40 µL of 10 m*M* Tris-HCl (pH 8.0).
3. Set up the reaction as follows: 40 µL DNA (up to 10 pmol of 5'-ends), 5 µL 10X CIAP reaction buffer (following manufacturer's protocol), up to 5 µL diluted CIAP (0.01 U/mL). Adjust the total volume to 50 µL with sterile Milli Q-grade water (Millipore) and incubate at 37°C for 30 min.
4. With another aliquot of diluted CIAP (equivalent to the amount used earlier), continue the incubation at 37°C for an additional 30 min.
5. Add 300 µL of CIAP stop buffer, extract with phenol:chloroform, and precipitate the DNA by adding 0.5 vol 7.5 *M* ammonium acetate (pH 5.5) and 2 vol of 100% ethanol to the final aqueous phase.

3.2. Construction of Chloroplast Expression Cassette

The generic expression cassette consists of a strong promoter, a selectable marker, and 5' and 3' untranslated regions to enhance the efficiency of transcription and translation of the introduced gene (**Fig. 2**). The chloroplast promoters and regulatory elements can be amplified directly from the total cellular or chloroplast DNA using specific primers designed on the basis of the sequence information available for the chloroplast genome of *Arabidopsis* (Accession no. NC_000932). The expression cassette is inserted into a unique site between the flanking sequences. Approximately 1 kb of homologous flanking regions is adequate to facilitate

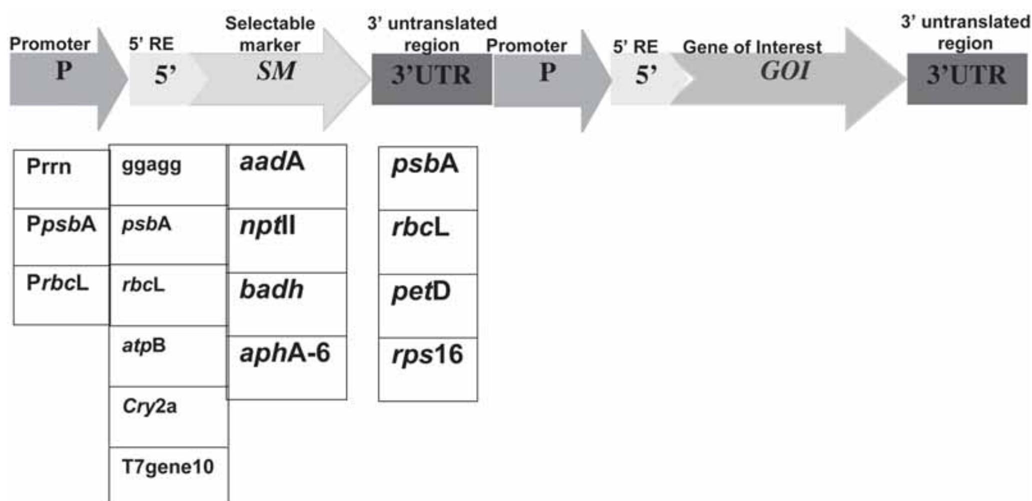


Fig. 2. Schematic representation of the chloroplast specific expression cassette. For a list of regulatory elements and genes of interest used for chloroplast transformation, refer to **Tables 2** and **3**.

recombination. Usually the genes are integrated in the intergenic spacer regions in order to avoid disrupting any chloroplast genes.

1. Clone the gene of interest under the control of the selected chloroplast promoter and 3' end regulatory region between a direct tandem duplicated chloroplast genome region (as prepared in **Subheading 3.1.**), in a colE1 replicon-based vector (pBluescript or similar).
2. Prepare a large-scale plasmid purification using a plasmid DNA purification kit, according to the manufacturer's instructions.

3.3. Particle Bombardment

3.3.1. Preparation of Gold Particle Suspension

The gold particles to be used for bombardment may be prepared in advance, at least 1 d prior to bombardment.

1. Weigh out 50 mg of gold particles (0.6 µm) into a 1.5-mL centrifuge tube.
2. Add 1 mL of 100% ethanol (molecular biology grade); vortex for 3 min.
3. Pellet the gold particles by centrifuging the tube for 2–5 min at maximum speed in a benchtop microcentrifuge and then discard the supernatant.
4. Add 1 mL of 70% ethanol and vortex for 2 min.
5. Incubate the tube for 15 min at room temperature. Mix the contents of the tube about three times during the incubation.
6. Centrifuge the tube at maximum speed for 3 min, and then discard the supernatant.
7. Add 1 mL of distilled water and vortex for 1 min or until particles are completely suspended.
8. Allow the particles to settle down for 1 min at room temperature and then centrifuge the tube for 2 min. Discard the supernatant.
9. Repeat **steps 7** and **8** two additional times.
10. Add 50% (v/v) glycerol to the gold particles at a final concentration of 50 mg/mL. Store the gold particles at –20°C until ready to use.

3.3.2. Preparation of Consumables for Bombardment

Consumables to be used for particle bombardment must be prepared prior to bombardment. The macrocarrier holders, stopping screens, filter paper, and Kimwipes need to be autoclaved.

The macrocarriers and rupture disks are sterilized by submerging in 100% ethanol for 10 min. The macrocarriers and rupture disks are placed over autoclaved Kimwipes and dried prior to use in the laminar flow hood.

3.3.3. DNA Coating of Gold Particles (see **Note 3**)

1. Vortex the previously prepared gold particles that were stored at –20°C until they are completely resuspended.
2. Pipet out 50 µL of the gold particle suspension into a 1.5-mL microcentrifuge tube.
3. Add 10 µL of plasmid DNA that has a concentration of 1 µg/µL, vortex for 5 s.
4. Add 50 µL of freshly prepared filter-sterilized 2.5 $CaCl_2$, and vortex for 5 s.
5. Add 20 µL of 0.1 *M* spermidine-free base and vortex for 5 s.
6. Vortex the mixture for 20 min at 4°C.
7. Add 200 µL of absolute ethanol at room temperature to the mixture, vortex for 5 s, and then centrifuge the mixture for 30 s at 3000 rpm. Remove the supernatant and repeat step 7 four times.
8. After the final step, resuspend the pellet in 30 µL of 100% ethanol.
9. Keep the particles on ice until use.

3.3.4. Macrocarrier Loading (see **Note 4**)

1. The macrocarrier is placed inside the macrocarrier holder with its concave side facing upward. Use the macrocarrier insertion cap to push the macrocarrier into place inside the holder.
2. Resuspend the gold particles completely by vortexing and pipetting in order to eliminate any clumps. Aliquot 5 µL of the gold particles coated with the plasmid DNA (microcarrier) and spread it out in the center of the macrocarrier.
3. Repeat **steps 1** and **2** with additional macrocarriers based on the number of leaves to be bombarded.
4. After loading all the macrocarriers, the remaining DNA-coated gold particles may be used for a second application on macrocarriers that already contain gold particles. Alternatively, the remaining gold particles coated with DNA could be run on agarose gels to test DNA binding efficiency.

3.3.5. Preparation of Plant Tissue for Bombardment

Mature leaves (5–6 wk old) used as recipients for the chloroplast transformation vector were obtained from aseptically propagated plants on ARM5 media.

1. Harvest the leaves (15–30 mm in length) and place on agar-solidified ARM1 medium.
2. Place about 12 to 18 leaves to cover an area 4 to 5 cm in diameter.
3. Repeat the steps listed above for additional samples to be bombarded.

3.3.6. Gene Gun Setup for Bombardment of Samples (see **Note 5**)

The particle bombardment procedure is per earlier reports ***(34–36)***.

1. To sterilize the particle gun, spray the inner chamber and its components with 100% ethanol. Let it air-dry or wipe with autoclaved Kimwipes.
2. Switch on the vacuum pump and open the pressure regulator valve on the helium gas tank.
3. Tighten the handle of adjustment valve to bring the helium gas approximately to 250 psi above the desired pressure.
4. Switch on the power switch on the gene gun.
5. Place the rupture disk in the rupture disk holder and screw on tightly to the gas acceleration tube.
6. Place the stopping screen in the macrocarrier holder and then place the macrocarrier coated with gold particles facing downward toward the screen. Secure the macrocarrier assembly with the macrocarrier cover lid and slide it into its slot in the gene gun chamber.
7. Place the Petri dish containing the sample to be bombarded on the target plate holder and insert it into the slot in the gene gun chamber. Close the chamber door and secure it with the entrapment attached to the gene gun.
8. The Vac switch is a three-way switch. In its first position it allows for vacuum to build up, in the center position (Vent) it bleeds the vacuum, and in its lower position it holds the vacuum.

Set the Vac switch to first position and let the vacuum reach to its maximum (up to 28 in. Hg in the vacuum gauge display). Flip the switch swiftly to the third position to hold the vacuum. Now keep the Fire switch pressed until the rupture disk bursts.

9. Toggle the switch to Vent position and release the vacuum to open the chamber and remove the bombarded sample.
10. Repeat **steps 6–9** for additional samples to be bombarded.
11. After completing the bombardment, close the pressure regulator valve on the helium gas tank. In order to release the helium in the helium tube, close the gene gun chamber and repeat step 9. Fire repeatedly until both pressure gauges show zero reading. Release the vacuum and turn off the gene gun and the vacuum pump.

For additional information on particle gun bombardment procedures, *see also* Chapter 19 in this book.

3.4. Obtaining Homoplasmic Chloroplast Transgenic Plants via Organogenesis

Chloroplast transgenic plants have been routinely obtained in tobacco via organogenesis (*see* **Note 6**). This method entails direct regeneration of putative chloroplast transgenic shoots from the bombarded leaf material. Achievement of homoplasmy is not a prerequisite before regeneration, as heteroplasmic shoot sections can be placed on selective shoot regeneration media for a second time to achieve homoplasmy. Even if homoplasmy is not achieved after the second round of selection, seeds obtained from heteroplasmic plants can be germinated on medium containing high concentration of the selection agent to obtain homoplasmic chloroplast transgenic plants.

Until recently, in all the reports of stable chloroplast transformation, chloroplast transgenic plants were obtained via organogenesis. In the model system tobacco our lab routinely obtains, on average, 15 successful chloroplast transformation events per plate ***(9)***, which is very efficient. In contrast, the efficiency of chloroplast transformation was found to be disappointingly low when chloroplast transformation was achieved via organogenesis in *Arabidopsis* (1 event per 40 or 151 bombarded plates, ***33***), potato (1 event per 25 bombarded plates, ***22***) and tomato (1 event per 10 bombarded plates or six events in 520 selection plates, ***23***; *see* **Table 4**). Besides the low efficiency, only sterile chloroplast transgenic plants were obtained in *Arabidopsis*. The authors cited extensive polyploidy of the leaf tissue as the reason for obtaining sterile plants ***(33)***. This problem could be overcome by adapting procedures that are now available to obtain regenerants from both diploid and tetraploid explants. In a study by Fras and Maluszynska ***(37)***, eight different calli lines derived from diploid and tetraploid cotyledons, hypocotyls, leaves, and root tips were analyzed. It was observed that during the period of culture all explants demonstrated a high level of polyploidization (2X–8X in diploid explant–derived callus and 4X–14X in tetraploid explant-derived callus) but only diploid, triploid, and tetraploid plants were regenerated using standard regeneration protocols. Callus formation was observed in all explants used but only the leaf explant demonstrated formation of somatic embryos. Thus, by optimizing the conditions for somatic embryogenesis, the limitations of polyploidy can be overcome. It is now feasible to achieve highly efficient regeneration via somatic embryogenesis in *Arabidopsis* ***(38)***. In addition, somatic embryogenesis could also be induced in *Arabidopsis* under various stress conditions ***(39)***. Therefore a combination of all the above improvements in tissue culture protocols and the novel vector designs that provided a breakthrough in carrot and cotton (**Note 7**) may help in overcoming the limitations of the earlier reported work on Arabidopsis chloroplast transformation ***(33)***. This approach differs from organogenesis in the fact that homoplasmy needs to be established prior to regeneration of transgenic plants. This fact is evident in carrot suspension cells that reveal heteroplasmy in the initial stages; once homoplasmy was established, plants could be regenerated that were homoplasmic (compare lanes 2 and 3 in **Fig. 3B**). Similar observation was made in case of cotton where plants were regenerated from embryos after achieving homoplasmy. However, when Southern analysis reveals heteroplasmic status of the regenerated plants, seeds obtained from heteroplasmic plants can be germinated on stringent selection medium to obtain homoplasmic plants.

Table 4
Comparison of Chloroplast Transformation Efficiency Among Different Plant Species

No. of plates	Rupture disc (psi)	Distance (cm)	Independent events	Events plate
Carrot cell cultures (1-mm; thick calli layer with 2-cm; diameter per plate) (*7*)				
30	650	6	0	0
24		9	1	1 per 24
26		12	0	0
36	900	6	0	0
34		9	1	1 per 34
36		12	0	0
38	1100	6	0	0
30		9	2,1	1 per 10
30		12	1,2,1	1 per 7.5
Cotton Cell Cultures (1-mm thick calli layer with 2-cm diameter per plate) (*24*)				
15	650	6	0	0
31		9	3,1,3,3,2,1	1 per 2
17	900	9	2,2	1 per 4
18		12	3	1 per 6
32	1100	9	2	1 per 18
46		12	1,2,1, 1	1 per 9
Tobacco leaf explant (*9*)				
5	1100	9 cm	~75	~15 per 1
Arabidopsis leaf explant (*33*)				
40	1100	12	1	1 per 40
151	450	9	1	1 per 151
Potato leaf explant (*22*)				
104	1100	Not given	3	1 per 35
Tomato leaf explant (*23*)				
60	1100	Not given	6	1 per 10

3.4.1. Tissue Regeneration and Selection

A protocol for chloroplast transformation of *Arabidopsis* via direct organogenesis is given below.

1. Incubate the bombarded leaves on ARM1 medium for 2 additional days.
2. After this period, stamp the leaves with a stack of 10 razor blades to create parallel incisions 1 mm apart.
3. Transfer the stamped leaves on solid ARM1 medium supplemented with 500 µg/mL spectinomycin for selection.
4. After 8 to 10 d transfer the leaf strips to ARMIIr medium supplemented with 500 µg/mL spectinomycin.
5. Transfer the green resistant calli obtained on ARMIIr medium to ASI-NIBI medium containing 500 µg/mL spectinomycin or 500 µg/mL each of spectinomycin and streptomycin for regeneration of resistant plants (*see* **Note 8**).
6. Transfer the plantlets to ARM medium for rooting and shoot elongation.
7. Finally, transfer the plants to soil.

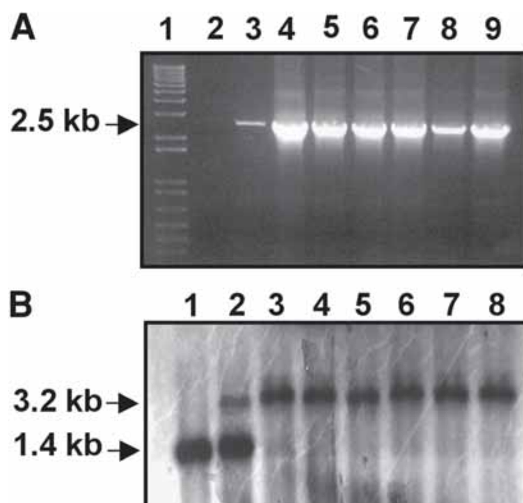


Fig. 3. Molecular analysis of chloroplast transgenic plants. Top panel **A**, PCR analysis. Lane 1-1 kb ladder; Lane 2, untransformed; Lanes 3–9, chloroplast transgenic lines. Bottom panel **B**, Southern blot analysis. Lane 1, untransformed; Lane 2, transformed but heteroplasmic; Lanes 3–8, transformed homoplasmic lines.

3.5. Molecular Analyses of Transformed Plants

3.5.1. PCR Analysis

PCR analysis is used to screen the transgenic plants and distinguish true chloroplast transgenic plants from mutants or nuclear transgenic plants. Site-specific chloroplast integration of the transgene cassette is determined by using a set of primers, one of which anneals to the native chloroplast genome and the other anneals within the transgene cassette. Mutant and nuclear transgenic plants are not expected to produce a PCR product with these primers. An example of PCR-based screening is provided (**Fig. 3A**). As is evident, the 2.5-kb PCR product is visible only in the chloroplast transgenic plants while it is missing from the wild-type sample. This initial screening is very important for eliminating the mutants and nuclear transgenic plants from the transgenic plant population.

1. Isolate total DNA from wild-type and transgenic plants using DNeasy™ Plant Mini Kit and use this DNA as a template for PCR reactions.
2. In a 50-µL PCR reaction add: 1 µL of 100 ng/µL genomic DNA, 5 µL of 10X PCR reaction buffer, 1 µL each of 10 µ*M* 5' forward and reverse primers, 200 to 250 µ*M* each of dNTPs, 0.5 µL (2.5 units) *Taq* DNA polymerase, and sterile distilled water.
3. Perform the PCR under following reaction conditions: denaturation for 5 min at 94°C, followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 60 to 65°C for 1 min, and extension at 72°C for 1 min. The time of extension can be varied depending on the size of the amplicon. Usually 1 min per kilobase is the prescribed elongation time for most DNA polymerases. This is followed by a final step of 7 to 10 min of elongation at 72°C.
4. Detect PCR amplification products in agarose gel by electrophoresis and stain with ethidium bromide (EtBr).

It is possible to determine the degree of homoplasmy by PCR using primers that anneal to flanking sequences used for homologous recombination. Unfortunately, most commercially available polymerases preferentially amplify the shorter fragment (native chloroplast genomes) as opposed to longer fragments (from transformed genomes), giving a misleading degree of

heteroplasmy. Therefore, published protocols use (with few exceptions) Southern blot analysis to determine homoplasmy.

3.5.2. Southern Analysis

Southern analysis is performed to determine the copy number and degree of homoplasmy of the introduced transgene. A single plant cell harbors about 10,000 copies of the chloroplast genome. Initially, the transgene cassette integrates into a few of these genomes and under selection pressure its copy number gradually increases. After several rounds of selection, the untransformed genome copies are replaced by the transgenic genome copies, leading to complete homoplasmy. This can be ascertained by Southern analysis of the total DNA isolated from plants generated after the third round of selection.

1. Isolate total cellular DNA from the plant material representing putative chloroplast transgenic tissue and wild-type plant tissue, using a Qiagen Plant DNA extraction kit.
2. Quantify the DNA concentration using an ultraviolet (UV) spectrophotometer and set up a restriction digestion with 1 μg of total cellular DNA in a 50 μL volume.
3. Incubate the samples at the prescribed temperature for at least 12 h before electrophoresing on agarose gel for transfer of DNA to a membrane.
4. Because the chloroplast transgenic samples harbor additional genetic material integrated via homologous recombination, the chosen restriction enzyme should be absent in the introduced foreign DNA and should restrict the chloroplast genome on either side of the integration site.
5. Depurinate the genomic DNA by incubating the gel in depurination solution for 15 min. Rinse with distilled water.
6. Denature the DNA by soaking the gel in denaturation solution for 30 min.
7. Rinse with distilled water. Soak in neutralization solution for 30 min. Repeat the neutralization step once.
8. Using capillary transfer, blot the nucleic acids onto nylon filters with 20X SSC.
9. Crosslink the DNA on to the membrane using an uv light cross-linker.

The southern blot is probed with radiolabeled flanking sequences used for homologous recombination. The transgenic plants possess a higher molecular weight plastid genome that is distinguishable on the autoradiogram from the smaller molecular weight fragment representing the untransformed plastid genome. If the transgenic plants are heteroplasmic, a wild-type fragment is visible along with the larger transgenic fragment. Absence of the wild-type fragment confirms the establishment of homoplasmy (**Note 9**). A typical Southern blot of chloroplast transgenic plants illustrates this point (**Fig. 3B**). Lane 1 represents wild type sample (1.4 kb fragment) and lane 2 represents heteroplasmic line (both 1.4-kb and 3.2-kb fragments). Lanes 3 through 8 represent homoplasmic transgenic plants (only 3.2-kb fragment) derived after repetitive subculture in liquid medium under selection.

3.6. Maternal Inheritance

Chloroplast genomes are maternally inherited in most plant species, including *Arabidopsis* ***(40)***. Maternal inheritance of chloroplast-integrated transgenes has been demonstrated for the solanaceous crops, tobacco ***(3)*** and tomato ***(23)***, cotton ***(24)*** and soybean ***(25)***. To test if the chloroplast-integrated transgenes are maternally inherited:

1. Grow the *in vitro* produced transgenic plants in the growth chamber along with wild-type.
2. Pollinate the emasculated flowers of the wild-type plant (female) with pollen grains derived from chloroplast transgenic lines (male).
3. Germinate the seeds obtained from chloroplast transgenic line and F1 crosses (wild-type female × male chloroplast transgenic line) on a suitable medium supplemented with the selection agent.
4. No germination of seeds from F1 crosses (wild-type female × male chloroplast transgenic line) on selection medium confirms strict maternal inheritance of the transgene. All seedlings derived from self-pollinated chloroplast transgenic lines or those cross-pollinated with pollen from wild-type should be resistant to the selection agent (*see* **Fig. 4**).

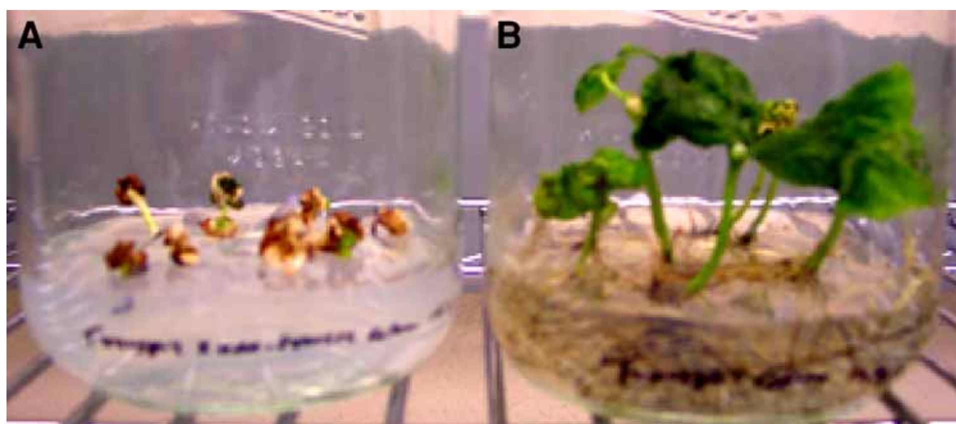


Fig. 4. Maternal inheritance of chloroplast-integrated transgenes in cotton. (**A**) Seeds obtained from F1 cross (wild-type female × male chloroplast transgenic line) and (**B**) seeds obtained from self-pollinated chloroplast transgenic lines or pollinated with pollen from wild-type germinating on the selection agent (kanamycin).

Unfortunately the chloroplast transgenic plants obtained so far in *Arabidopsis* have been sterile. Therefore maternal inheritance could not be tested in this species. Fertile plants obtained using optimized protocols described in this chapter will be required for this purpose.

4. Notes

1. The overall efficiency of chloroplast transformation of any species is dependent on several factors. The site of insertion of the transgene in the chloroplast genome seems to be one of the most important factors. In seven out of nine biopharmaceutical proteins, five out of seven vaccine antigens, and five of the six agronomic traits engineered via the chloroplast genome so far (*see* **Tables 2** and **3**), transgenes were integrated into the transcriptionally active spacer region between *trnI/trn*A genes within the *rrn* operon. The foreign gene expression levels obtained from genes integrated at this site are among the highest ever reported (46% total leaf protein; ***20***). Thus, it appears that this preferred site is unique and allows highly efficient transgene integration and expression. One of the flanking regions used at this site of integration contains the chloroplast origin of replication ***(28,29)*** and this might facilitate replication of foreign vectors within chloroplasts ***(30)***, enhance the probability of transgene integration and achieve homoplasmy even in the first round of selection ***(31)***. This is further illustrated by the first successful Rubisco engineering obtained by integrating the RbcS gene at this site ***(41)***. All other earlier attempts at Rubisco engineering at other integration sites in the chloroplast genome were unsuccessful. Various transgenes integrated into the transcriptionally active *trn*I-*trn*A spacer region or other read-through or silent intergenic spacer regions are listed in **Tables 2** and **3**.
   The urgent need to extend this technology to other crop species is impeded by the lack of chloroplast genome sequence information. Availability of DNA sequence information would allow for faster vector construction, identification of appropriate regulatory elements and analysis of chloroplast transgenic plants. So far sequence information is available for only six crop chloroplast genomes (**Table 1**).
2. Generally the tissue culture medium is autoclaved for 20 min at 121 psi and allowed to cool to 45°C before adding any antibiotics or heat-sensitive growth hormones to the medium. Sterilized growth medium is poured into sterile Petri dishes, jars, or magenta boxes as required. It is important not to autoclave tissue culture medium for more than 30 min because the phytohormones added to the medium may break down, inhibiting tissue regeneration.
3. The order of adding gold particles, DNA, CaCl2, and spermidine is important for the proper coating of the gold particles. CaCl2 should be prepared fresh and the DNA-coated gold particles must be used within 2 h.

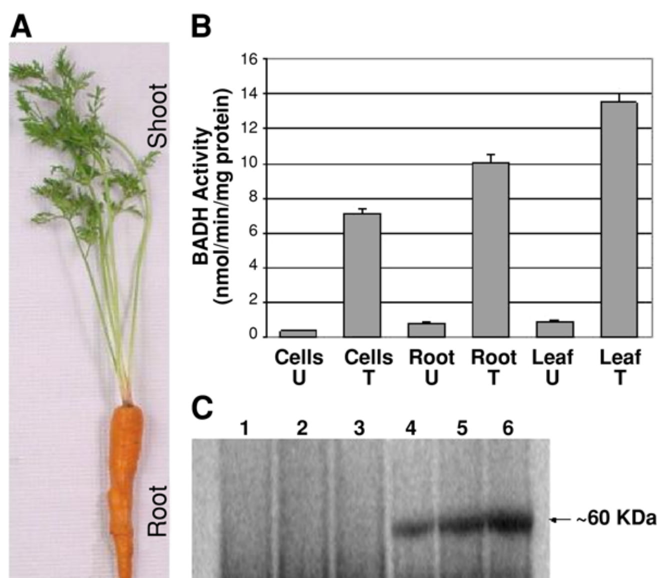


Fig. 5. BADH enzyme activity and BADH expression in control and chloroplast transgenic lines. **(A)** A chloroplast transgenic line is shown with taproot and shoot. **(B)** BADH activity in untransformed **(U)** and transformed **(T)** cells suspension, root and leaf. **(C)** Western blot using polyclonal anti-BADH serum. Antigenic peptides were detected using horseradish peroxidase-linked secondary antibody. Lanes 1–3, untransformed cell culture, root and leaf; lanes 4–6, transformed cell culture, root, and leaf.

4. The preparation of leaf tissues, macrocarriers, and bombardment must take place under aseptic conditions in a laminar flow hood. Before each addition of gold particles into the macrocarrier, make sure to vortex the particles for at least 30 s to resuspend, and use immediately. It is essential to avoid clumps of gold particles when loading the macrocarriers because this will damage the leaf tissue during bombardment and decrease transformation efficiency.
5. The microcarrier launch assembly must be placed in level one (L1 = 3 cm); meanwhile, the target plate shelf must be placed according to the requirements of individual explant type of a given plant species. The stopping screen support is to be placed between the spacer rings (one under and one above the stopping screen). This setup is essential for efficient chloroplast transformation.
6. The employment of chloroplast transformation technology in *Arabidopsis* has been hindered by lack of an efficient and reproducible protocol that provides fertile chloroplast transgenic plants. *Arabidopsis* could be used as an ideal model system for chloroplast functional genomics. So far, tobacco mutants have been employed in order to optimize chloroplast transformation *(42)* because of inefficient plastid transformation system in *Arabidopsis*. This chapter provides the particle bombardment parameters and tissue culture protocols that have been successfully applied to transform recalicitrant species such as cotton.
7. Unique chloroplast transformation vectors provided success in transforming the chloroplast genomes of carrot and cotton *(7,24)*. The antibiotic-resistance genes present in the expression cassette were engineered to express around the clock and also in green and non-green plastids. This is vital as most of the bombarded calli or suspension cultures are non-green tissues containing proplastids instead of chloroplasts. Gene expression and gene regulation systems in proplastids are quite different from green chloroplasts. Transformed proplastids should develop into mature chloroplasts and transformed cells should survive the selection process during all stages of development in the light and the dark. This is absolutely critical because only one or two chloroplasts are transformed in a plant cell that should have the ability to survive the selection pressure, multiply and establish themselves, whereas all other untransformed

Table 5
List of Reporter/Selectable Marker Genes Expressed Via the Chloroplast Genome

Reporter / selectable genes	Organism	Ref.
cat— chloramphenicol acetyl transferase	a. Cucumber etioplasts^T	a. *21*
	b. Cultured tobacco cells^T	b. *30*
*uid*A—β-glucuronidase	Wheat leaves and calli^T	*62*
*aad*A- aminoglycoside adenine transferase	a. Chlamydomonas^S	a. *63*
	b. Tobacco^S	b. *53*
*npt*II—neomycin phosphotransferase	Tobacco^S	*64*
*aph*A-6—aminoglycoside phosphotransferase	a. Chlamydomonas^S	a. *65*
	b. Tobacco^S	b. *52*
badh—Betaine aldehyde dehydrogenase	Tobacco^S	*12*
gfp—Green fluorescent protein	a. Tobacco and *Arabidopsis*^{T}	a. *66*
	b. Potato^S	b. *22*
*aad*A-*gfp*—fusion protein	Tobacco^S and rice^T	*67*
hphs—hygromycin phosphotransferase—synthetic	Tobacco^S	*55*

T, transient expression; S, stable integration

plastids are eliminated. "Double Barrel" plastid transformation vectors accomplish this by using genes coding for two different enzymes capable of detoxifying the same selectable marker (or spectrum of selectable markers), driven by regulatory signals that are functional in proplastids as well as in mature chloroplasts. The efficiency of non-green tissue derived expression is exemplified by the BADH protein expression in non-green carrot roots that was found to be 75% that of leaf expression (**Fig. 5**).

8. Besides the extension of the chloroplast transformation technology to other crop species, the focus of future research would be to eliminate antibiotic based selectable markers (*see* **Table 5** for currently used reporter and selectable markers). Several approaches like marker excision via direct repeats ***(14)***, CRE-lox recombination system ***(13)***, and transiently integrated selectable marker genes ***(15)***, should facilitate achieving that goal. The edible selectable marker *badh* also holds potential to substitute the antibiotic-based selectable markers ***(12)***. With the establishment of somatic embryogenesis-based protocols, chloroplast transformation technology has come of age and should lead to successful chloroplast engineering of most crop species including monocots.
9. The presence of the foreign gene into the nuclear or mitochondrial genome can be detected by using the foreign gene as a DNA probe and prolonged exposure of blots to x-ray films.

Acknowledgment

The results reported from investigations in the Daniell laboratory are currently supported by NIH R 01 GM 63879 and USDA 3611-21000-017-00D grants to HD.

References

1. Daniell, H. (2002) Molecular strategies for gene containment in transgenic crops. *Nat. Biotechnol.* **20,** 581–586.
2. Daniell, H., Khan, M. S., and Allison, L. (2002) Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. *Trends Plant Sci.* **7,** 84–91.
3. Daniell, H., Datta, R., Varma, S., Gray, S., and Lee, S. B. (1998) Containment of herbicide resistance through genetic engineering of the chloroplast genome. *Nat. Biotechnol.* **16,** 345–348.
4. Kota, M., Daniell, H., Varma, S., Garczynski, S. F., Gould, F., and William, M. J. (1999) Overexpression of the *Bacillus thuringiensis* (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. *Proc. Natl. Acad. Sci. USA* **96,** 1840–1845.
5. DeGray, G., Kanniah, R., Franzine, S., John, S., and Daniell, H. (2001) Expression of an antimicrobial peptide via the chloroplast genome to control phytopathogenic bacteria and fungi. *Plant Physiol.* **127,** 852–862.

6. Lee, S. B., Kwon, H. B., Kwon S. J., et al. (2003) Accumulation of trehalose within transgenic chloroplasts confers drought tolerance. *Mol. Breeding* **11,** 1–13.
7. Kumar, S., Dhingra, A., and Daniell, H. (2004) Plastid expressed betaine aldehyde dehydrogenase gene in carrot cultured cells, roots and leaves confers enhanced salt tolerance. *Plant Physiol.* **136,** 2843–2854.
8. Ruiz, O. N., Hussein, H., Terry, N., and Daniell, H. (2003) Phytoremediation of organomercurial compounds via chloroplast genetic engineering. *Plant Physiol.* **132,** 1344–1352.
9. Daniell, H., Lee, S. B., Panchal, T., and Wiebe, P. O. (2001) Expression of the native cholera toxin B subunit gene and assembly as functional oligomers in transgenic tobacco chloroplasts. *J. Mol. Biol.* **311,** 1001–1009.
10. Watson, J., Koya V., Leppla S., and Daniell, H. (2004) Expression of *Bacillus anthracis* protective antigen in transgenic chloroplasts of tobacco, a non-food/feed crop. *Vaccine* **22,** 4374–4384.
11. Molina, A., Hervás–Stubbs, S., Daniell, H., Mingo-Castel, A., and Veramendi J. (2004) High-yield expression of a viral peptide animal vaccine in transgenic tobacco chloroplasts. *Plant Biotech. J.* **2,** 141–153.
12. Daniell, H., Muthukumar, B., and Lee, S. B. (2001) Marker tree transgenic plants: engineering the chloroplast genome without the use of antibiotic selection. *Curr. Genet.* **39,** 109–116.
13. Corneille, S., Lutz, K., Svab, Z., and Maliga, P. (2001) Efficient elimination of selectable marker genes from the plastid genome by the CRE-lox site-specific recombination system. *Plant J.* **27,** 171–178.
14. Iamtham, S. and Day, A. (2000) Removal of antibiotic resistance genes from transgenic tobacco plastids. *Nat. Biotechnol*. **18,** 1172–1176.
15. Klaus, S. M. J., Huang, F. C., Golds, T. J., and Koop, H. U. (2004) Generation of marker-free plastid transformants using a transiently cointegrated selection gene. *Nat. Biotechnol.* **22,** 225–229.
16. Fernandez-San Millan, A., Mingo-Castel, A., and Daniell, H. (2003) A chloroplast transgenic approach to hyper-express and purify human serum albumin, a protein highly susceptible to proteolytic degradation. *Plant Biotechnol. J.* **1,** 71–79.
17. Leelavathi, S. and Reddy, V. S. (2003) Chloroplast expression of His-tagged GUS-fusions: a general strategy to overproduce and purify foreign proteins using transplastomic plants as bioreactors. *Mol. Breeding* **11,** 49–58.
18. Daniell, H., Carmona-Sanchez, O., and Burns, B. B. (2004) Chloroplast derived antibodies, biopharmaceuticals and edible vaccines, in *Molecular Farming* Fischer, R. and Schillberg, S. (eds.), WILEY-VCH Verlag, Weinheim, Germany, pp.113–133.
19. Chebolu, S. and Daniell, H. (2005) Chloroplast derived vaccine antigens and biopharmaceuticals: expression, folding, assembly and functionality. *Curr. Trends Microbiol. Immunol.,* in press.
20. DeCosa, B., Moar, W., Lee, S. B., Miller, M. and Daniell, H. (2001) Overexpression of the Bt Cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat. Biotechnol.* **19,** 71–74.
21. Daniell, H. and McFadden, B. A. (1987) Uptake and expression of bacterial and cyanobacterial genes by isolated cucumber etioplasts. *Proc. Natl. Acad. Sci. USA* **84,** 6349–6353.
22. Sidorov, V. A., Kasten, D., Pang, S. Z., Hajdukiewicz, P. T. J., Staub, J. M., and Nehra, N. S. (1999) Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. *Plant J.* **19,** 209–216.
23. Ruf, S., Hermann, M., Berger, I. J., Carrer, H., and Bock, R. (2001) Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat. Biotechnol.* **19,** 870–875.
24. Kumar, S., Dhingra, A., and Daniell, H. (2004) Stable transformation of the cotton plastid genome and maternal inheritance of transgenes. *Plant Mol. Biol.* **56,** 203–216.
25. Dufourmantel, N., Pelissier, B., Garçon, F., Peltier, J. M., and Tissot, G. (2004) Generation of fertile transplastomic soybean. *Plant Mol Biol.* **55,** 727–741.
26. Palmer, J. D. (1985) Comparative organization of chloroplast genomes. *Ann. Rev. Genet.* **19,** 325–354.
27. Lilly, J. W., Havey, M. J., Jackson, S. A., and Jiang, J. M. (2001) Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. *Plant Cell* **13,** 245–254.
28. Kunnimalaiyaan M. and Nielsen, B. L. (1997) Chloroplast DNA replication: mechanism, enzymes and replication origins. *J. Plant Biochem. Biotechnol.* **6,** 1–7.
29. Kunnimalaiyaan, M. and Nielsen, B. L. (1997) Fine mapping of replication origins (*ori*A and *ori*B) in Nicotiana tabacum chloroplast DNA. *Nucleic Acids Res.* **25,** 3681–3686.
30. Daniell, H., Vivekananda, J., Nielsen, B. L., Ye, G. N., Tewari, K. K., and Sanford, J. C. (1990) Transient foreign gene expression in chloroplasts of cultured tobacco cells after biolistic delivery of chloroplast vectors. *Proc. Natl. Acad. Sci. USA*. **87,** 88–92.

31. Guda, C., Lee, S. B., and Daniell, H. (2000) Stable expression of biodegradable protein based polymer in tobacco chloroplasts. *Plant Cell Rep.* **19,** 257–262.
32. Daniell, H., Cohill, P., Kumar, S., and Dufourmantel, N. (2004) Chloroplast genetic engineering, in Molecular biology and biotechnology of plant organelles (Daniell, H. and Chase, C., eds.), The Netherlands, Kluwer Academic Publishers, pp. 423–468.
33. Sikdar, S. R., Serino, G., Chaudhuri, S., and Maliga, P. (1998) Plastid transformation in *Arabidopsis*. *Plant Cell Rep.* **18,** 20–24.
34. Daniell, H. (1997) Transformation and foreign gene expression in plants mediated by microprojectile bombardment. *Methods Mol. Biol.* **62,** 453–488.
35. Kumar, S. and Daniell, H. (2004) Chloroplast genetic engineering for medical molecular farming. *Methods Mol. Biol.* **267,** 365–383.
36. Daniell, H. Ruiz, O. N., and Dhingra, A. (2004) Chloroplast genetic engineering to improve agronomic traits. *Methods Mol. Biol.* **286,** 111–137.
37. Fras, A. and Maluszynska, J. (2003) Regeneration of diploid and tetraploid plants of *Arabidopsis thaliana via callus*. *Acta Biolog. Cracoviensia Ser. Bot*. **45,** 145–152.
38. Gaj, M. D. (2004) Factors influencing somatic embryogenesis induction and plant regeneration with particular reference to Arabidopsis thaliana (L.) *Heynh. Plant Growth Reg.* **43,** 27–47.
39. Ikeda-Iwai, M., Umehara, M., Satoh, S., and Kamada, H. (2003) Stress-induced somatic embryogenesis in vegetative tissues of *Arabidopsis thaliana*. *Plant J.* **34,** 107–114.
40. Hagemann, R. (2004) The sexual inheritance of plant organelles, in Daniell, H., and Chase C., eds., Molecular Biology and Biotechnology of Plant Organelles, Kluwer Academic, The Netherlands, pp 93–113.
41. Dhingra, A., Portis, Jr., A. R., and Daniell, H. (2004) Enhanced translation of a chloroplast expressed *Rbc*S gene restores SSU levels and photosynthesis in nuclear antisense *Rbc*S plants. *Proc. Natl. Acad. Sci. USA* **101,** 6315–6320.
42. Klaus, S. M. J., Huang, F. C., Eibl, C., Koop, H. U., and Golds, T. J. (2003) Rapid and proven production of transplastomic tobacco plants by restoration of pigmentation and photosynthesis. *Plant J.* **35,** 811–821.
43. Kang, T. J., Loc, N. H., Mi-Ok, J., Jang, Y. S., Kim, Y. S., Seo, J. E., and Yang, M. S. (2003) Expression of the B-subunit of *E. coli* heat–labile enterotoxin in the chloroplasts of plants and its characterization. *Transgenic Res.* **12,** 683–691.
44. Singleton, M. L. (2003) Expression of CaF1 and LcrV as a fusion protein for a vaccine against *Yersinia pestis* via chloroplast genetic engineering. MS thesis, University of Central Florida, USA.
45. Ruiz, G. (2002) Optimization of codon composition and regulatory elements for expression of the human IGF-1 in transgenic chloroplasts. MS thesis, University of Central Florida, USA.
46. Torres, M. (2002) Expression of interferon α5 in transgenic chloroplasts of tobacco. MS thesis, University of Central Florida, USA.
47. Falconer, R. (2002) Expression of interferon α2b in transgenic chloroplasts of a low-nicotine tobacco. MS thesis, University of Central Florida, USA.
48. Muhlbauer, S. K., Lossl, A., Tzekova, L., Zou, Z. R., and Koop, H. U. (2002) Functional analysis of plastid DNA replication origins in tobacco by targeted inactivation. *Plant J.* **32,** 175–184.
49. Jeong, S. W., Jeong, W. J., Woo, J. W., Choi, D. W., Park, Y. I., and Liu, J. R. (2004) Dicistronic expression of the green fluorescent protein and antibiotic resistance genes in the plastid for selection and tracking of plastid-transformed cells in tobacco. *Plant Cell Rep.* **22,** 747–751.
50. Carrer, H. and Maliga, P. (1995) Targeted insertion of foreign genes into the tobacco plastid genome without physical linkage to the selectable marker gene. *Biotechnology* **13,** 791–794.
51. Bock, R. and Maliga, P. (1995) Correct splicing of a group II intron from a chimeric reporter gene transcript in tobacco plastids. *Nucleic Acids Res.* **23,** 2544–2547.
52. Huang, F. C., Klaus, S. M. J., Herz, S., Zou, Z., Koop, H. U., and Golds, T. J. (2002) Efficient plastid transformation in tobacco using the aphA-6 gene and kanamycin selection. *Mol. Genet. Genomics* **268,** 19–27.
53. Svab, Z. and Maliga, P. (1993) High frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *Proc. Natl. Acad. Sci. USA* **90,** 913–917.
54. Viitanen, P. V., Devine, A. L., Khan, M. S., Deuel, D. L., Dyk, D. E. V., and Daniell, H. (2005) Metabolic engineering of the chloroplast genome using the E. coli ubiC gene reveals that chorismate is a readily abundant plant precursor for p-Hydroxybenzoic acid biosynthesis. *Plant Physiol*. **136,** 4048–4060.

55. Dhingra, A. (2000) Analysis of components involved in regulation of plastid gene expression, plastome organization and its manipulation with emphasis on rice. Ph.D. thesis, University of Delhi South Campus, New Delhi, India.
56. Zoubenko, O. V., Allison, L. A., Svab, Z., and Maliga, P. (1994) Efficient targeting of foreign genes into the tobacco plastid genome. *Nucleic Acids Res.* **22,** 3819–3824.
57. Staub, J. M. and Maliga, P. (1993) Accumulation of D1 Polypeptide in tobacco plastids is regulated via the untranslated region of the *psb*A messenger RNA. *EMBO J.* **12,** 601–606.
58. Zou, Z., Eibl, C., and Koop, H. U. (2003) The stem-loop structure of the tobacco psbA 5' UTR is an important determinant of mRNA stability and translation efficiency. *Mol. Gen. Genom.* **269,** 340–349.
59. Eibl, C., Zou, Z. R., Beck, A., Kim, M., Mullet, J., and Koop, H. U. (1999) In vivo analysis of plastid *psb*A, *rbc*L, *rpl*32 UTR elements by chloroplast transformation: tobacco plastid gene expression is controlled by modulation of transcript levels and translation efficiency. *Plant J.* **19,** 333–345.
60. Koop, H. U., Steinmuller, K., Wagner, H., Rossler, C., Eibl, C., and Sacher, L. (1996) Integration of foreign sequences into the tobacco plastome via PEG mediated protoplast transformation. *Planta* **199,** 193–201.
61. Thum, K. E., Kim, M., Morishige, D. T., Eibl, C., Koop, H. U., and Mullet, J. E. (2001) Analysis of the barley chloroplast *psb*D light responsive promoter elements in transplastomic tobacco. *Plant Mol. Biol.* **47,** 353–366.
62. Daniell, H., Krishnan, M., and McFadden, B. F. (1991) Transient expression of beta-glucuronidase in different cellular compartments following biolistic delivery of foreign DNA into wheat leaves and calli. *Plant Cell Rep.* **9,** 615–619.
63. Goldschmidt Clermont, M. (1991) Transgenic expression of aminoglycoside adenine transferase in the chloroplast—a selectable marker for site-directed transformation of Chlamydomonas. *Nucleic Acids Res.* **19,** 4083–4089.
64. Carrer, H., Hockenberry, T. N., Svab, Z., and Maliga, P. (1993) Kanamycin resistance as a selectable marker for plastid transformation in tobacco. *Mol. Gen. Genet.* **241,** 49–56.
65. Bateman, J. M. and Purton, S. (2000) Tools for chloroplast transformation in *Chlamydomonas*: expression vectors and a new dominant selectable marker. *Mol. Gen. Genet.* **263,** 404–410.
66. Hibberd, J. M., Linley, P. J., Khan, M. S., and Gray, J. C. (1998) Transient expression of green fluorescent protein in various plastid types following microprojectile bombardment. *Plant J.* **16,** 627–632.
67. Khan, M. S. and Maliga, P. (1999) Fluorescent antibiotic resistance marker for tracking plastid transformation in higher plants. *Nat. Biotechnol.* **17,** 910–915.

22

β-Glucuronidase as Reporter Gene

Advantages and Limitations

Kye-Won Kim, Vincent R. Franceschi*, Laurence B. Davin, and Norman G. Lewis

Summary

The β-glucuronidase (GUS) gene is used extensively in plant biology studies; this analysis summarizes its advantages and limitations. With the advances in genomic sequencing and computational analyses (including bioinformatics), its application in the study of plant gene expression is now an integral component of modern day plant science. This chapter focuses on the detailed challenges of carrying out GUS studies for both qualitative and quantitative analyses, including the increasing employment of GUS from *Bacillus* strains, rather than *E. coli*; the *Bacillus* GUS genes encode proteins with enhanced properties, such as both increased thermostability and stability in the presence of crosslinking fixatives.

Key Words: β-glucuronidase (GUS); GUS-promoter fusions; *Arabidopsis*; Cruciferae; gene expression localization; thermostability; fixatives.

1. Introduction

Localization of gene expression is an experimental approach often used toward obtaining a comprehensive understanding of both gene function and regulation during plant growth/development. Typically, such investigations are targeted towards identification of constitutive gene expression patterns ***(1,2)***; detection of overlapping (functionally redundant) metabolic networks ***(2)***; establishment of specific gene expression responses that are induced by either biotic or abiotic stresses ***(3)***; and confirmation of transgene expression patterns in selected tissues or organs ***(4)***. Studies of this type are increasingly more feasible with the recent advances in both molecular techniques and database analyses ***(5)***, which, in turn, have permitted the relatively facile isolation of genes of interest and their corresponding promoters. Accordingly, transformation procedures are routinely applied to an increasing number of plant species, with the result being that reporter gene techniques are now widely employed in establishing patterns of gene expression.

One of the earliest reporter genes used in plants is *gus*A (or *uid*A), which encodes β-glucuronidase (GUS) (β-D-glucuronoside glucuronosohydrolase, E.C. 3.2.1.31). Initially isolated from *Escherichia coli* ***(6)***, *gus*A is part of an operon that includes *gus*B, *gus*C and *gus*R. *Gus*B encodes a glucuronide-specific permease, whereas the product of *gus*C is of unknown function, and *gus*R (upstream of *gus*A) is separately transcribed and encodes a specific repressor of the *gus* operon ***(7)***. GUS is also present in various other bacteria, including *Thermotoga maritima*, *Bacillus* sp., *Staphylococcus warneri*, *S. homini*, *Salmonella* sp., and *Enterobacter* sp. ***(8)***. GUS has a monomeric molecular weight of approximately 68,000 Da, but exists in vivo as a tetramer ***(9)***.

*Prof. V. R. Franceschi deceased (April 30, 2005).

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

The use of GUS as a reporter has several technical advantages that have led to its wide adoption in a variety of biological studies. One of its advantages in plant studies is that it is capable of tolerating large amino-terminal additions for translational fusions without loss of enzymatic activity ***(10)***. Another advantage is its quite broad substrate versatility—i.e., GUS is able to catalytically hydrolyze β-*O*-glycosidic linkages of a wide variety of glucuronides, and it is this substrate versatility that has been employed for either qualitative or quantitative assessments of patterns of gene expression. The GUS substrates that are commercially available find different applications due to the distinct chromogenic, spectrophotometric and fluorogenic properties of their aglycone forms. The most widely used of these is the colorless chromogenic substrate, 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc). Cleavage of its glycosidic linkage by the action of GUS and subsequent oxidative dimerization/dehydrogenation of the resulting aglycone (**Fig. 1**, reaction A) produce the characteristic indigo blue chromogenic precipitate ***(10)*** that is extensively utilized to delineate patterns of gene expression in intact biological tissues. Both the intensity of color of the final dichloro-dibromo-indigo (ClBr-indigo) blue dye precipitate, and the precision of its spatial resolution in identifying the location of gene expression, are enhanced when ferricyanide/ferrocyanide salts are added ***(11)***.

A second, albeit less commonly employed, GUS substrate is *p*-nitrophenyl-β-D-glucuronide (*p*NPG), which is typically used in standard spectrophotometric assays ***(7)***. In this case, GUS activities in vitro can be quantified by measuring the absorbance of the aglycone, *p*-nitrophenol (λ_{max}= 390 nm), which is released following GUS hydrolysis (**Fig. 1**, reaction B). A third substrate used for in vitro assays is 4-methylumbelliferyl-β-D-glucuronide (4-MUG). In this case, GUS hydrolysis generates the aglycone, 4-methylumbelliferone (4-MU; 7-hydroxy-4-methyl coumarin) (**Fig. 1**, reaction C), which gives a (detectable) fluorescent emission at 455 nm following excitation at 365 nm. Quantification of GUS activity using this fluorescent substrate is generally two to three orders of magnitude more sensitive than either X-Gluc or *p*NPG ***(4,7)***.

Although not yet widely employed, other microbial GUS genes have been isolated that encode GUS with enhanced characteristics. A *Bacillus* sp. GUS has, for example, superior properties compared to that of the GUS derived from *E. coli* ***(12)***, as the former is thermostable with a half-life of 10 min at 65°C and enzymatically stable in the presence of crosslinking fixatives such as formaldehyde and glutaraldehyde ***(8)***. Since 2001, it has been made commercially available as GUS*Plus*™ in pCAMBIA 1305. Other improved vectors for GUS expression have also been generated from the CAMBIA vector series (www.cambia.org.au/), and Curtis and Grossniklaus ***(13)*** engineered an *Agrobacterium* sp. binary vector system (Gateway-compatible) that reportedly facilitates fast and reliable DNA cloning.

An important prerequisite for the use of reporter genes is the absence of endogenous activities in the organisms to be transformed. While Jefferson et al. ***(12)*** reported no detectable endogenous GUS-like activity in a broad spectrum of higher plants, endogenous GUS-like activity was, however, detected in the reproductive, but not vegetative, tissues by others ***(14,15)***. Although this is a potential limitation, such effects can largely be overcome by carrying out the appropriate controls, either of GUS assays of untransformed tissues, or of tissues transformed with a gene other than GUS. Another limitation is that because GUS is also very stable, staining intensities can increase over time even when overall expression levels of the gene of interest are low ***(16)***. This can make it difficult to distinguish between strong promoter activity at discrete but short developmental stages versus that of weaker but prolonged promoter activity.

Even with the limitations of this technique, it is extremely useful for many types of studies; the GUS transcriptional or translational gene fusion system is now used extensively as a reporter in plants because of its three major advantages: simplicity, versatility, and robustness ***(7)***. The assay is extremely sensitive; with application of the appropriate method it is possible to obtain both quantitative (i.e., level of expression) and qualitative (i.e., specificity of expression in tissues and organs) data with the same reporter gene. Furthermore, GUS has been successfully used as a reporter gene in a very broad range of studies, ranging from identification of tissue specificities of gene expression to that of developmental and pathogen-induced patterns. In *Arabidopsis*, GUS has been used to explore vein development ***(17)***, photomorphogenic gene

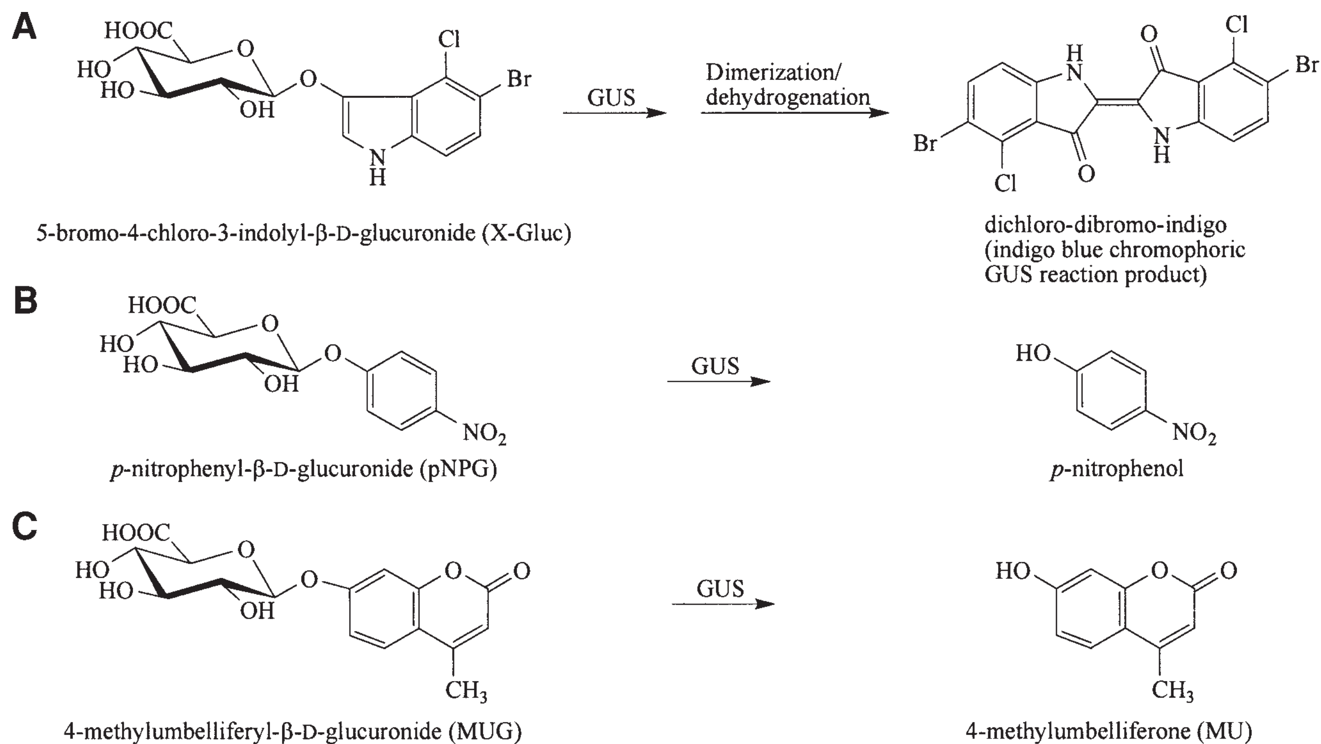


Fig. 1. β-Glucuronidase specifically hydrolyzes β-linked D-glucuronides to D-glucuronic acid and aglycone. (**A**) 5-bromo-4-chloro-3-indolyl-β-D-glucuronide, (**B**) *p*-nitrophenyl-β-D-glucuronide, and (**C**) 4-methylumbelliferyl-β-D-glucuronide.

expression *(18)*, root *(19)*, and root-cap *(20)* differentiation/organization, as well as hormone-related root tissue expression *(1,21)*. It has also been employed to investigate cell wall *(22)* and cell membrane *(23)* related proteins, stress-induced expression *(24)*, as well as, for example, differential expression patterns of the western red cedar dirigent multigene family *(2,25)*.

GUS reporter genes can also be used in mutant screening—for example, with gene/promoter *(26)*, secretion *(27)*, or enhancer *(28)* trap techniques—and this method has the advantage of not relying upon a phenotype induced by gene disruption. In addition, the screening can be performed at the heterozygous stage of the potential mutants, as recently demonstrated by the isolation of a putative β-xylosidase gene from *Arabidopsis* secondary cell wall tissue *(26)*. One limitation of using gene trap insertions was noted in secretory pathway encoding genes *(27)*, where N-linked glycosylation specific to the secretory pathway inhibited β-glucuronidase activity. This limitation was overcome by addition of tunicamycin, which inhibits glycosylation.

The anti-β-glucuronidase antibody can also be used to detect the GUS enzyme in transformed plant tissue using either Western blotting or immunohistochemical techniques *(29)*. Additionally, Jenik and Irish *()* have suggested that use of the anti-GUS antibody results in greater consistency in probing tissue sections for the determination of cell division patterns that are regulated differently at distinct stages of floral development, as compared with using GUS staining directly.

These examples thus serve to help illustrate the range of applications that the GUS reporter system has played in modern plant biology studies. They also illustrate that it is a tool that can be used for many different types of studies and as part of a broad range of rigorous experimental protocols. As with any technique, its application requires experience or access to detailed information about the protocol in order to avoid artifacts and maximize effectiveness and sensitivity. Included here are detailed descriptions on application of the GUS reporter system that should prove useful as a general reference. Notes are also given that can help optimize the system for a given usage.

2. Materials

2.1. GUS Activity Quantification: Buffers, Reagents and Equipment/Supplies

1. 500 m*M* sodium phosphate buffer (pH 7.0).
2. 500 m*M* EDTA, disodium salt (pH 8.0).
3. 10% (w/v) SDS in H_2O.
4. 10% (v/v) Triton X-100 in H_2O.
5. GUS extraction buffer (*see* **Note 1**): Prior to use, the GUS extraction buffer is prepared from the above stock solutions to obtain the following final concentrations: 50 m*M* sodium phosphate buffer (*see* **Note 2**), 10 m*M* EDTA (*see* **Note 3**), 0.1% SDS and 0.1% Triton X-100; β-mercaptoethanol is added to a final concentration of 10 m*M*.
6. 4-MUG (Fluka) is added to the GUS extraction buffer to give a final concentration of 25 m*M*. Store solution at 4°C in the dark.
7. Sodium carbonate (0.2 *M*) as a reagent to stop the reaction.
8. 4-MU solutions (200, 150, 100, 50, and 10 n*M*) are prepared in order to obtain a standardized linear relationship of 4-MU concentration versus fluorescence. Prepare a 10 m*M* 4-MU (Sigma) stock solution in dimethylsulfoxide (DMSO, 5 mL) (*see* **Note 4**) and dilute to 100 µ*M* with DMSO. Store at 4°C in the dark. Dilute the 100 µ*M* 4-MU stock solution to 500 n*M* with 0.2 *M* sodium carbonate, and prepare serial dilutions to 200, 150, 100, 50, and 10 n*M* in 0.2 *M* sodium carbonate buffer.
9. Acetone:water (9:1, v/v). Keep cold.
10. Fluorescence measurements: Using a spectrofluorometer (e.g., LS 50B luminescence spectrometer, Perkin-Elmer), obtain a standard plot of fluorescence versus 4-MU concentration.
11. Bradford reagent (Bio-Rad).
12. Liquid nitrogen.
13. Mortar and pestle.

2.2. Histological Staining

1. 500 m*M* sodium phosphate buffer (pH 7.0).
2. 10% (v/v) Triton X-100 in H_2O.
3. 50 m*M* potassium ferrocyanide (store at 4°C in the dark).
4. 50 m*M* potassium ferricyanide (store at 4°C in the dark).
5. GUS staining buffer: Prior to use, the GUS staining buffer is prepared from the above stock solutions such that the following final concentrations are obtained: 50 m*M* sodium phosphate buffer (pH 7.0; *see* **Note 5**), 0.2% Triton X-100 (*see* **Note 6**), 2 m*M* potassium ferrocyanide, and 2 m*M* potassium ferricyanide (*see* **Note 7**).
6. 20 m*M* X-Gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronide cyclohexylammonium salt; Gold BioTechnology, Inc., St Louis, MO) in N,N-dimethylformamide (DMF). Store aliquots at –20°C in the dark. Stock solution should be colorless or near colorless.
7. 70% ethanol in water.
8. 50 m*M* PIPES [L-piperazine-N-N'-*bis* (2-ethane sulfonic acid)] buffer, pH 7.2.
9. 2% glutaraldehyde in 50 m*M* PIPES, pH 7.2.
10. Dehydrating solutions of 30, 40, 50, 60, 70, 80, and 90% ethanol in water, and 100% ethanol.
11. Prepare 0.5% aq. Safranin-O. This solution should be filtered.
12. Other materials and equipment: LR White resin (medium grade); double edged-razor blades (EM Scientific); ultramicrotome for thick (1 µm) and thin (TEM) resin sections; gelatin-coated glass slides; slide warmer set to 60°C; immersion oil.

3. Methods

Quantitative GUS activity assays are first described using the fluorogenic substrate 4-MUG, this being a method whereby time-course kinetics can be employed to confirm that the detected fluorescence was derived from the fluorogenic product, 4-MU (**Fig. 1C**). This is followed by a description of the various techniques utilized for localization of GUS activity *in situ*, with application of these technologies at the organ (**Fig. 2A,B,D,E**), cellular (**Fig. 2C, 3**), and subcellular levels ***(7)*** using dissecting, light, and electron microscopy, respectively. In visualization using electron microscopy, the blue GUS reaction product creates visible deposits as a result of its electron density.

We have also developed a procedure suitable for obtaining high-resolution images at the single-cell level. This modification differs from previous protocols by De Block and van Lijsebettens ***(11)***, who performed the X-Gluc reaction on semithin sections (20 µm) of plastic embedded material, or that of Weigel and Glazebrook ***(31)***, who used paraffin embedding for cross-sections and histological localization, which gave only low-resolution images. Using this modified protocol, histochemical GUS activity in whole tissues can be detected and high-resolution images from plastic resin embedding (2 µm thick) can be obtained in a straightforward manner. Specifically, this was achieved by modification of the method to reduce the time for embedding. Moreover, the effort required for sectioning is dramatically reduced via quick visual identification and prescreening of stained samples.

Figures 2 and **3** show, for illustration purposes only, differential spatial and temporal gene expression of selected dirigent promoter::*GUS* fusions of some of the 16 members of the putative dirigent gene (homolog) family in *Arabidopsis* (***2***, and references therein; Kim et al., manuscript in finalization). In this way, the patterns of gene expression of this large gene family were systematically delineated.

3.1. Quantitation of GUS Activity

GUS activity can be accurately determined using the fluorogenic substrate 4-MUG ***(12)***, which is hydrolyzed by GUS to liberate the fluorochrome 4-MU. Using an excitation wavelength of 365 nm and measuring the emission at 455 nm, the amounts of 4-MU generated can be quantified as follows ***(12)***:

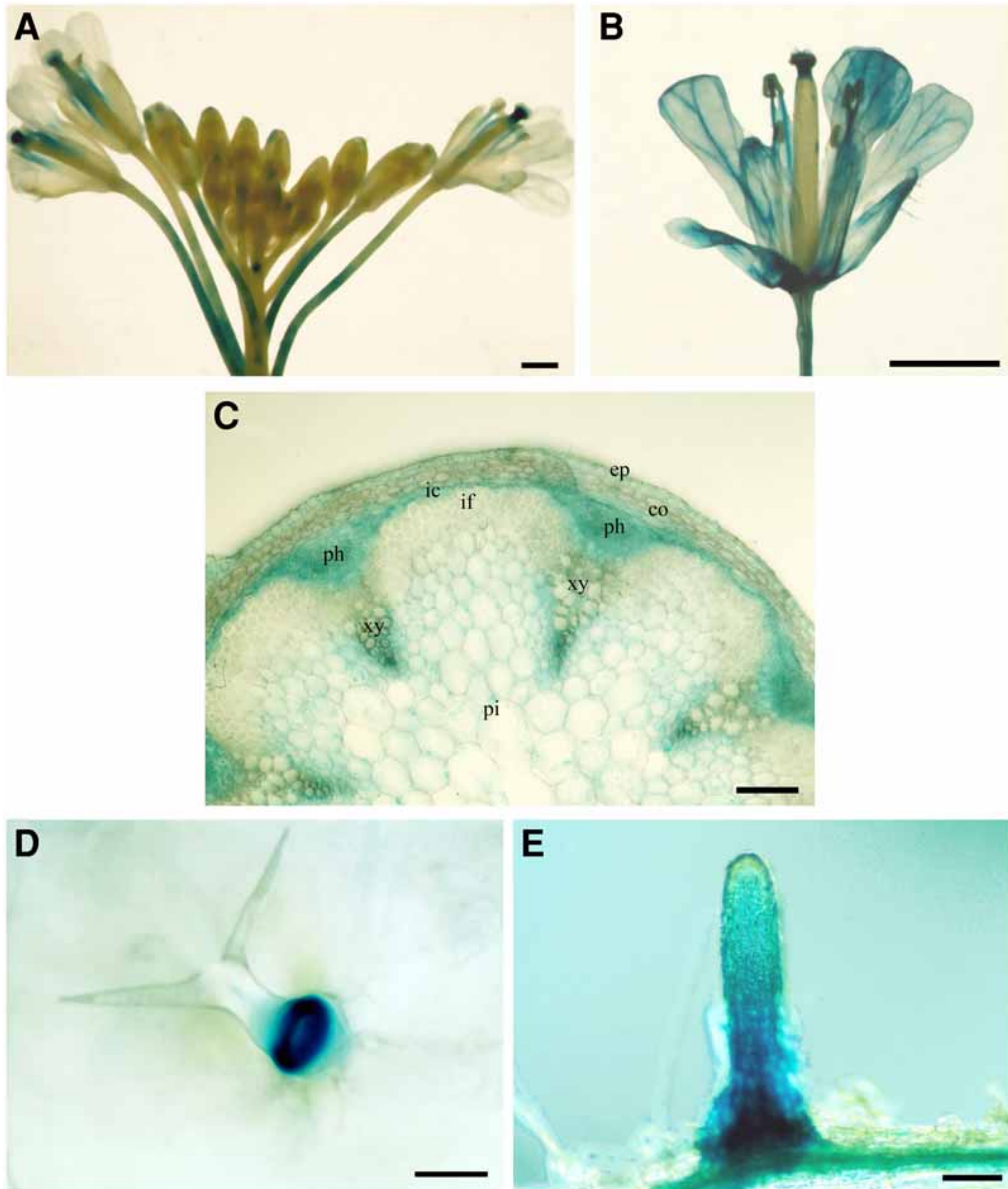


Fig. 2. Histochemical localization of GUS in *Arabidopsis* tissues. Four- **(A)** and five- **(B)** wk-old flowers and basal stem transverse section (**C**) of plants transformed with *At*Dir13 promoter::*uidA*. Figures 2A and 2B illustrate expression patterns in vascular tissue, i.e., the filament of the stamen, the peduncle, and the veins of flowers and sepals. Figure 2C depicts a cross-section of the stem, where the gene expression is also clearly noted in the interfascicular cambium, the cambium, the phloem, and the developing xylem. **(D)** Five-wk-old rosette leaf trichome of *Arabidopsis* plants transformed with the *At*Dir3 promoter::*uidA*, with specific expression observed at the base of the trichome. **(E)** Seven-d-old roots of *Arabidopsis* transformed with the *At*Dir10 promoter::*uidA*, with the *At*Dir10 promoter expressed in secondary root tissue. Abbreviations: cortex (co), epidermis (ep), interfascicular cambium (ic), interfascicular cells (if), phloem (ph), pith (pi), and xylem (xy). Size bars: **(A)** and **(B)** = 1 mm; **(C)** = 100 µm; **(D)** and **(E)** = 50 µm. [*At*Dir3: At5g 4904, *At*Dir10: At2g 28670, *At*Dir13: At4g 11190.] *See* color version in insert following p. 239.

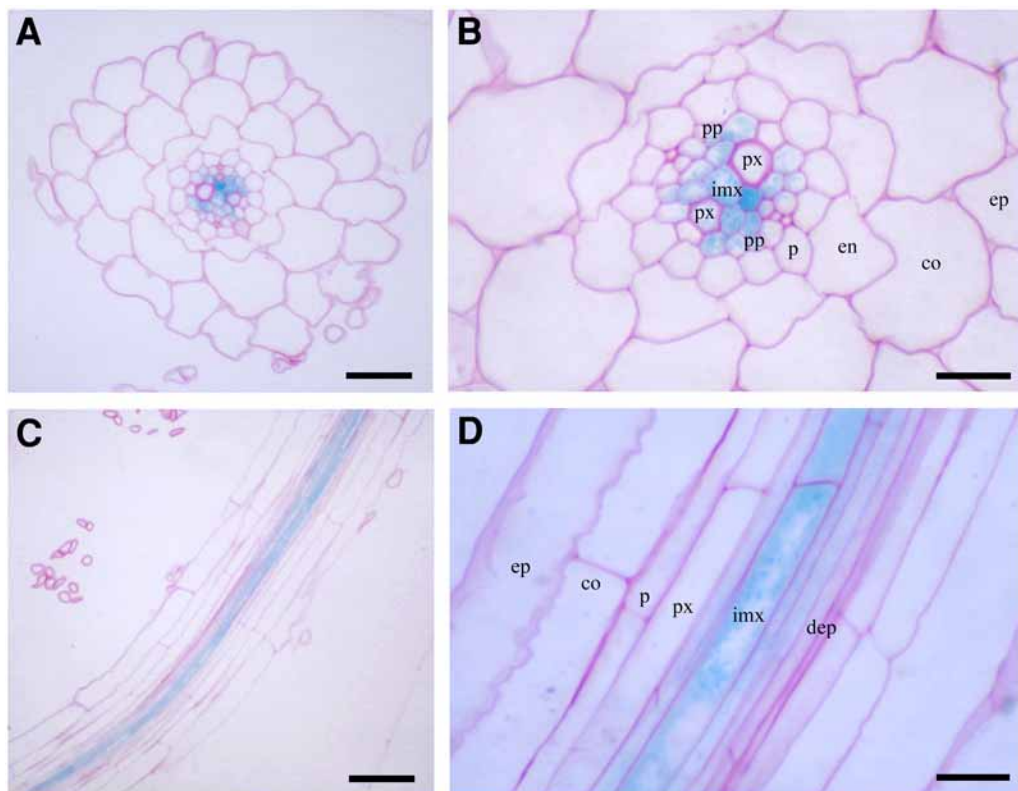


Fig. 3. Histochemical localization of GUS in *Arabidopsis* roots of 5-d-old plants transformed with *At*Dir6 promoter::*uidA*. (**A**,**B**) Transverse and (**C**,**D**) longitudinal cross-sections. Note the pattern of expression at the single cell-level in the immature xylem of the developing root tissue and to a lesser extent in the protophloem. Abbreviations: cortex (co), developing phloem (dep), endodermis (en), epidermis (ep), immature xylem (imx), pericycle (p), protophloem (pp) and protoxylem (px). Size bars **(A)** = 20 µm; **(B)** and **(D)** = 10 µm; **(C)** = 50 µm. [*At*Dir6: At4g 23690.] *See* color version in insert following p. 239.

1. Place 100 mg tissue of interest in a 1.5-mL microcentrifuge tube. Freeze immediately in liquid nitrogen; frozen tissues can be stored at –80°C if necessary until assaying.
2. Transfer frozen tissue to a mortar and grind tissue by means of a pestle to a fine powder (*see* **Note 8**).
3. Transfer powdered tissue to a 1.5-mL microcentrifuge tube; add 200 µL GUS extraction buffer and store on ice while processing remaining samples.
4. Vortex each sample and centrifuge at 20,000*g* for 20 min at 4°C.
5. Transfer the supernatant (crude extract) to a fresh 1.5-mL microcentrifuge tube and place on ice.
6. Determine total protein concentration of each sample (e.g., Bradford method *[32]*).
7. For each sample, prepare five microcentrifuge tubes containing 800 µL of 0.2 *M* sodium carbonate (stop solution).
8. Prepare a reaction mixture containing 1 m*M* 4-MUG in GUS extraction buffer (from 25 m*M* stock solution), transfer 980-µL aliquots into microcentrifuge tubes, and prewarm to 37°C.
9. Add 20 µL of each crude extract (between 10 and 50 µg of protein) to individual microcentrifuge tubes containing 1 m*M* 4-MUG (*see* **Note 9**) and vortex. Take a 200-µL aliquot and add to the microcentrifuge tube containing 800 µL stop solution; this sample will be used in the reference cell and is referred to as the time 0 min sample.
10. After 5, 15, 30, and 60 min incubation, transfer a 200-µL aliquot of each reaction mixture to the corresponding tube containing the stop solution and vortex (*see* **Note 10**).

11. Prepare the 4-MU standard solution as described in **Subheading 2.1.**
12. Measure the fluorescence of the standard 4-MU solutions (*see* **Note 11**) (excitation wavelength 365 nm; emission wavelength 455 nm; filter wavelength at 430 nm; *see* **Note 12**) and plot a standard curve of fluorescence against 4-MU concentration.
13. Measure the fluorescence of each set of samples.
14. Calculate GUS activity in pmol/min/mg protein using the standard curve.

3.2. Histological Staining

GUS histochemical staining is a powerful tool to study promoter expression at the tissue level throughout the plant and during its growth and development. The procedure described below has been employed to localize GUS expression in transformed *Arabidopsis*.

3.2.1. Whole-Mount GUS Staining

1. Place the tissue of interest (e.g., entire seedlings, individual organs [*see* **Note 13**], or dissected stems [*see* **Note 14**]) into microcentrifuge tubes or 20-mL glass scintillation vials (depending on size of sample) containing cold acetone–H_2O (9:1, v/v). Keep on ice until all samples are harvested.
2. When all samples have been harvested, individually fix in acetone-H_2O (9:1) for an additional 20 min at room temperature (*see* **Note 15**).
3. Wash each sample in staining buffer (*see* **Subheading 2.2.**) three times on ice.
4. Prepare staining solution by adding X-Gluc stock solution to staining buffer to a final concentration of 1 m*M* (*see* **Note 16**).
5. Add staining solution to the fixed samples and slowly infiltrate them *in vacuo*.
6. Incubate (typically overnight) at 37°C with gentle agitation (*see* **Note 17**).
7. Add 70% ethanol to remove solubles (e.g., chlorophyll), and repeat extraction with a fresh solution several times until the sample is either essentially colorless or very faintly colored. At this point, the tissue can be stored at 4°C.
8. Examine the tissue under a dissecting microscope. Place the tissue on a glass slide and coat with 70% ethanol (*see* **Note 18**). Place the slide on the microscope stage, and then immediately take pictures. Transfer the sample back to 70% ethanol for storage. A well-slide can be used for large bulky samples.
9. If the tissue is to be sectioned, continue with the embedding procedure described below. For best resolution, it is often desirable to examine GUS patterns in tissue sections (*see* **Fig. 2**).

3.2.2. Embedding and Sectioning

1. Fix the tissue samples that have been stained for GUS activity for 2 h with 2% glutaraldehyde in 50 m*M* PIPES (pH 7.2) at 4°C (*see* **Note 19**).
2. Rinse the samples in 50 m*M* PIPES buffer (3 times, 20 min) at room temperature.
3. Dehydrate sample through a graded series of ethanol–H_2O: 30, 40, 50, 60, 70, 80, and 95% EtOH (15 min each), and finally 100% EtOH (3 times, 15 min).
4. Infiltrate the sample with LR White acrylic resin in ethanol as follows: acrylic resin–EtOH (v/v): 25, 34, 50, 67, and 75% (12 h each) and finally infiltrate with pure LR White acrylic resin (3 times, 12 h).
5. Sample embedding in LR White resin is performed in 30-mL plastic cups (or gelatin capsules, depending on the size of the sample) with a 24-h polymerization at 60°C (*see* **Note 20**).
6. Trim the block around the specimen to the desired size and region.
7. Cut thick sections (2 μm) with an ultramicrotome equipped with a glass knife and fitted with a water boat.
8. Place a drop of distilled water onto a gelatin-coated slide.
9. Transfer sections to the water.
10. Dry the sections onto the slides on a slide warmer set at ~60°C.
11. Preexamine the sections without additional counterstaining.
12. If needed, counterstain for 30 s to 2 min with 0.5% aq. Safranin O (*see* **Note 21**).
13. Coat the sections with immersion oil and place a cover slip on them. Seal the cover slip on the slide with nail polish.

14. Visualize under the compound light microscope using bright or dark field optics: With bright field optics the blue precipitate can be seen within cell or tissue types expressing the promoter fusion (*see* **Fig. 3**). Dark field microscopy causes the GUS precipitate to appear as bright purple particles against a dark background.

4. Notes

1. In solanaceous plants, it has been demonstrated that organic solvents can both inhibit endogenous GUS-like activities and enhance the exogenous *gus*A-dependent activities ***(33)***. Addition of methanol–H_2O (1:4) to the extraction buffer mixture drastically inhibits plant-derived GUS-like activity.
2. The endogenous GUS-like activity is also pH-dependent and can be inhibited further by adjusting the pH of the extraction buffer to pH 8.0 ***(7)***.
3. GUS activity is inhibited by certain divalent cations: addition of ethylenediaminetetraacetic acid (EDTA) can reduce inhibition ***(12)***.
4. According to the Weigel and Glazebrook protocol ***(31)***, water was used to dissolve 4-MU. However, in our hands, 4-MU was not readily soluble in water.
5. Different buffers (e.g., potassium phosphate, Tris-HCl) at different concentrations in the staining solution provided no significant differences in either intensity or pattern of staining ***(7,15)***.
6. The presence of Triton X-100 gives a more uniform staining pattern in leaves ***(7)***.
7. Use of ferro- and ferricyanide enhances the oxidative dimerization/dehydrogenation of two molecules of the indoxyl aglycone intermediate and formation of the insoluble indigo-colored pigment, indigo blue (**Fig. 1A**), that is formed at the GUS reaction site. However, these reagents can also interfere with GUS activity. Thus, a balance between concentration of the reagent and enzyme activity is important, with 2 m*M* ferricyanide/ferrocyanide working well for most applications. However, if the staining is not precisely localized (because of diffusion), higher concentrations can be employed (e.g., 10 m*M*) ***(34)***.
8. When a comparison of the GUS activity is required with respect to the amount of the sample materials (and not protein concentration), weigh the powdered sample at this step.
9. Triplicate measurements of activities are required.
10. In order to confirm that the detected activity is due to the fluorogenic compound 4-MU, time-course kinetics can be performed. The GUS-derived fluorescence should increase with time, whereas the relative fluorescence due to other fluorogenic compounds should remain constant.
11. The absolute fluorescence units of standard solutions may vary depending on the type of spectrofluorometer used, but the standard curve must show a linear correlation between rise in fluorescence and increasing concentration of 4-MU in 0.2 *M* sodium carbonate stop solution.
12. The background fluorescence of the product is negligible when using an excitation at 365 nm and measuring the emission at 455 nm. It is important to note, however, that the substrate does exhibit a slight intrinsic fluorescence (emission at 375 nm) if excited at 316 nm. Therefore, the precise excitation source and/or appropriate filter sets are required ***(7)***.
13. Several independent transformants must be examined so that representative expression patterns can be evaluated.
14. Thin sections of stem samples can be prepared prior to staining. After fixation of a 1-cm stem in 0.5% (v/v) formaldehyde for exactly 1 min, wash twice with 50 m*M* sodium phosphate buffer (pH 7.0). In the same buffer, stem sections can be cut to ~0.2-mm thickness using a double edged-razor blade.
15. Fixation is necessary to minimize intrinsic/endogenous GUS activity. Instead of using acetone, 4–6% formaldehyde solution (on ice for 2–3 min) or 2–3% glutaraldehyde (on ice for 2–5 min) can be used ***(7)***.
16. Most of the previous protocols recommend 2 m*M* X-Gluc as a substrate, but we obtained comparable results at lower concentration (0.4 m*M*). Thus 1 m*M* X-Gluc is recommended for the first trial, even though staining intensity varies depending on the promoter strength.
17. Exact staining time may depend on GUS activity in selected transgenic lines. Although it is possible to let the reaction proceed for two or more days, tissue integrity begins to deteriorate after long incubation periods.

18. To use an optically favorable medium such as glycerol or immersion oil, the sample tissue must be infiltrated first with a gradual decrease of ethanol accompanied by a gradual increase of glycerol. The sample and glycerol solution can be infiltrated under vacuum to remove residual small bubbles.
19. Each tissue is subsequently cut into <2.5 mm^2 sections in the fixation solution using new double-edged razor blades. Fixation time varies depending on the size of tissue piece but is typically 2 h.
20. Spurr's resin should not be used because it requires acetone for dehydration and the indigo blue pigment is soluble in organic solvents such as acetone.
21. A simple counter-staining procedure has been developed *(25)* for comparative β-glucuronidase expression and anatomical localization in transgenic herbaceous *Arabidopsis* and tobacco.

Acknowledgments

The authors thank the US National Science Foundation Arabidopsis 2010 Project Grant MCB-0117260, the National Aeronautics and Space Administration (NAG 2-1513), the Department of Energy (DE-FG-0397ER20259) and the G. Thomas and Anita Hargrove Center for Plant Genomic Research for generous support.

References

1. Sabatini, S., Beis, D., Wolkenfelt, H., et al. (1999) An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell* **99,** 463–472.
2. Kim, M. K., Jeon, J.-H., Davin, L. B., and Lewis, N. G. (2002) Monolignol radical-radical coupling networks in western red cedar and *Arabidopsis* and their evolutionary implications. *Phytochemistry* **61,** 311–322.
3. Merkouropoulos, G. and Shirsat, A. H. (2003) The unusual *Arabidopsis* extensin gene *atExt1* is expressed throughout plant development and is induced by a variety of biotic and abiotic stresses. *Planta* **217,** 356–366.
4. James, V. A., Worland, B., Snape, J. W., and Vain, P. (2004) Strategies for precise quantification of transgene expression levels over several generations in rice. *J. Exp. Bot.* **55,** 1307–1313.
5. Fizames, C., Muños, S., Cazettes, C., et al. (2004) The *Arabidopsis* root transcriptome by serial analysis of gene expression. Gene identification using the genome sequence. *Plant Physiol.* **134,** 67–80.
6. Jefferson, R. A., Burgess, S. M., and Hirsh, D. (1986) β-Glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc. Natl. Acad. Sci. USA* **83,** 8447–8451.
7. Gallagher, S. R., ed. (1992) *GUS Protocols: Using the GUS Gene as a Reporter of Gene Expression*. Academic Press, San Diego, CA.
8. Jefferson, R. A. and Mayer, J. E. (2003) Microbial β-glucuronidase genes, gene products and uses thereof. U.S. Patent no. 6,641,996, 103 pp.
9. Tussey Bracey, L. and Paigen, K. (1987) Changes in translational yield regulate tissue-specific expression of β-glucuronidase. *Proc. Natl. Acad. Sci. USA* **84,** 9020–9024.
10. Naleway, J. J. (1992). Histochemical, spectrophotometric, and fluorometric GUS substrates, in *GUS Protocols: Using the GUS Gene as a Reporter of Gene Expression* (Gallagher, S. R., ed.), Academic Press, San Diego, CA, pp. 61–76.
11. De Block, M. and van Lijsebettens, M. (1998) β-Glucuronidase enzyme histochemistry on semithin sections of plastic-embedded *Arabidopsis* explants, in *Methods in Molecular Biology* (Martínez-Zapater, J. M. and Salinas, J., eds.), Vol. 82, Humana Press, Totowa, NJ, pp. 397–407.
12. Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6,** 3901–3907.
13. Curtis, M. D. and Grossniklaus, U. (2003) A Gateway cloning vector set for high-throughput functional analysis of genes *in planta. Plant Physiol.* **133,** 462–469.
14. Hu, C. Y., Chee, P. P., Chesney, R. H., Zhou, J. H., Miller, P. D., and O'Brien, W. T. (1990) Intrinsic GUS-like activities in seed plants. *Plant Cell Rep.* **9,** 1–5.
15. Plegt, L. and Bino, R. J. (1989) β-Glucuronidase activity during development of the male gametophyte from transgenic and non-transgenic plants. *Mol. Gen. Genet.* **216,** 321–327.
16. Mantis, J. and Tague, B. W. (2000) Comparing the utility of β-glucuronidase and green fluorescent protein for detection of weak promoter activity in *Arabidopsis thaliana. Plant Mol. Biol. Rep.* **18,** 319–330.

17. Koizumi, K., Sugiyama, M., and Fukuda, H. (2000) A series of novel mutants of *Arabidopsis thaliana* that are defective in the formation of continuous vascular network: calling the auxin signal flow canalization hypothesis into question. *Development* **127,** 3197–3204.
18. von Arnim, A. G. and Deng, X.-W. (1994) Light inactivation of *Arabidopsis* photomorphogenic repressor *COP1* involves a cell-specific regulation of its nucleocytoplasmic partitioning. *Cell* **79,** 1035–1045.
19. Malamy, J. E. and Benfey, P. N. (1997) Organization and cell differentiation in lateral roots of *Arabidopsis thaliana. Development* **124,** 33–44.
20. Tsugeki, R. and Fedoroff, N. V. (1999) Genetic ablation of root cap cells in *Arabidopsis. Proc. Natl. Acad. Sci. USA* **96,** 12,941–12,946.
21. Oono, Y., Ooura, C., Rahman, A., Aspuria, E. T., Hayashi, K.-i., Tanaka, A., and Uchimiya, H. (2003) *p*-Chlorophenoxyisobutyric acid impairs auxin response in *Arabidopsis* root. *Plant Physiol.* **133,** 1135–1147.
22. Verica, J. A., Chae, L., Tong, H., Ingmire, P., and He, Z.-H. (2003) Tissue-specific and developmentally regulated expression of a cluster of tandemly arrayed cell wall-associated kinase-like kinase genes in *Arabidopsis. Plant Physiol.* **133,** 1732–1746.
23. Javot, H., Lauvergeat, V., Santoni, V., et al. (2003) Role of a single aquaporin isoform in root water uptake. *Plant Cell* **15,** 509–522.
24. Denekamp, M. and Smeekens, S. C. (2003) Integration of wounding and osmotic stress signals determines the expression of the *AtMYB102* transcription factor gene. *Plant Physiol.* **132,** 1415–1423.
25. Kim, M. K., Choi, J.-W., Jeon, J.-H., Franceschi, V. R., Davin, L. B., and Lewis, N. G. (2002) Specimen block counter-staining for localization of *GUS* expression in transgenic *Arabidopsis* and tobacco. *Plant Cell Rep.* **21,** 35–39.
26. Goujon, T., Minic, Z., El Amrani, A., et al. (2003) *AtBXL1*, a novel higher plant (*Arabidopsis thaliana*) putative β-xylosidase gene, is involved in secondary cell wall metabolism and plant development. *Plant J.* **33,** 677–690.
27. Groover, A. T., Fontana, J. R., Arroyo, J. M., Yordan, C., McCombie, W. R., and Martienssen, R. A. (2003) Secretion trap tagging of secreted and membrane-spanning proteins using *Arabidopsis* gene traps. *Plant Physiol*. **132,** 698–708.
28. Holding, D. R. and Springer, P. S. (2002) The *Vascular Prepattern* enhancer trap marks early vascular development in *Arabidopsis. Genesis* **33,** 155–159.
29. Chan, M.-T., Chao, Y.-C., and Yu, S.-M. (1994) Novel gene expression system for plant cells based on induction of α-amylase promoter by carbohydrate starvation. *J. Biol. Chem.* **269,** 17,635–17,641.
30. Jenik, P. D. and Irish, V. F. (2000) Regulation of cell proliferation patterns by homeotic genes during *Arabidopsis* floral development. *Development* **127,** 1267–1276.
31. Weigel, D. and Glazebrook, J., eds. (2002) *Arabidopsis: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
32. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72,** 248–254.
33. Kosugi, S., Ohashi, Y., Nakajima, K., and Arai, Y. (1990) An improved assay for β-glucuronidase in transformed cells: methanol almost completely suppresses a putative endogenous β-glucuronidase activity. *Plant Sci*. **70,** 133–140.
34. Sessions, A., Weigel, D., and Yanofsky, M. F. (1999) The *Arabidopsis thaliana MERISTEM LAYER 1* promoter specifies epidermal expression in meristems and young primordia. *Plant J.* **20,** 259–263.

23

Use of Fluorescent Proteins as Reporters

Guido Jach

Summary

Fluorescent proteins (FPs) have established themselves as valuable reporter proteins in plant molecular biology. Beside general background information about spectral properties, protein structure and maturation of different commonly used FPs, this chapter provides detailed protocols about cloning of suitable expression cassettes for GFP and DsRED and detection of FPs in *Arabidopsis* protoplasts and plants transiently or stably expressing these constructs by fluorescence microscopy.

Key Words: GFP; DsRED; mRFP1; fluorescent proteins; transient gene expression; protoplasts; transgenic plants; particle bombardment.

1. Introduction

Reporter proteins and their genes are essential tools in molecular biology, frequently used to monitor developmental and spatial gene expression patterns. Since their introduction in 1994, fluorescent proteins (FPs) have proved to be highly valuable reporters, thus extending the set of popular reporter systems employing the enzymes β-glucuronidase (GUS), luciferase (LUC), and chloramphenicol-acetyl-transferase (CAT). They have also gained high interest owing to their unique ability to allow for noninvasive, nondestructive detection, abolishing the need for preparation of protein extracts, addition of suitable substrates, and performance of enzyme assays.

During the past decade numerous applications of fluorescent reporter proteins have been established, including passive applications such as the use of FPs as fluorescence tags in fusion proteins to monitor the appearance, degradation, location, or translocation of appropriate partner proteins, as well as more active applications measuring biochemical parameters such as metabolite concentrations, enzyme activity, or protein–protein interactions by their effects on the fluorescence properties of appropriately designed derivatives of fluorescent proteins (biochemical sensors/indicators) ***(1)***.

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is by far the best-understood FP. GFP was first observed in 1962 as a companion protein to aequorin (a blue-light-emitting chemiluminescent *Aequorea* protein) ***(2)*** and the first prediction of the chemical structure of the GFP fluorophore (the light emitting component) dates back to 1979 ***(3)***. Ultimate proof for the postulated 4-(*p*-hydroxy-benzylidene)imidazolidin-5-one structure was finally provided in 1992 by Cody et al. ***(4)***.

Probably the most important milestones in the history of FPs are the successful cloning of GFP cDNAs (1992) and the demonstration (1994) that GFP cDNA expression is sufficient to produce green fluorescence in organisms other than jellyfish, thus proving that the GFP gene contains all the information necessary for the posttranslational synthesis of the fluorophore ***(5–7)***.

During the past decade different new and improved GFP isoforms with enhanced or altered fluorescence phenotypes ***(9–16)*** were generated using molecular evolution techniques. In addition, severe limitations of the wild-type protein, such as its thermosensitivity, which causes

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

accumulation of improperly folded nonfluorescent, insoluble protein *(8)*, were overcome. Currently, several independent GFP mutants and variants emitting blue, cyan, and yellow light are available (BFP, CFP, and YFP, respectively). An overview is presented in **Table 1**.

Aequorea GFP remained the only cloned FP gene available until Matz et al. reported in 1999 about the cloning of six novel FP genes from nonbioluminescent anthozoan species *(17)*. One of the proteins, dsFP583, has the unique ability to emit red light peaking at 583 nm and today is commercially available as DsRED. Since then, numerous other FP genes have been cloned from a variety of marine organisms *(18,19)*. Today about 30 cloned members of the FP gene family are known. However, suitability as reporters is proven for only a small subset of these proteins.

Mature GFP-like proteins share a common structural motif described as β-can, even if their homology at amino-acid level is low (**Fig. 1**). The β-can is formed by 11 β-sheets connected by short α-helical segments, while an encapsuled central helix contains the fluorophore whose conjugated -electron resonance system accounts for the fluorescent characteristics of the protein *(20)*. The encapsulation provides the fluorophore with an environment of charged residues and water molecules/hydrogen bonds, which makes it physicochemically very stable *(20,21)*.

GFP maturation starts with the correct folding of the apoprotein, which buries the nonfluorescent tripeptide fluorophore motif deep inside the mature protein. Presumably promoted by the 3D-structure, the mature fluorescent fluorophore is sequentially formed autocatalytically by two distinct chemical processes: cyclization and oxidation, with the latter being rate-limiting *(22)*. Cyclization proceeds through a nucleophilic attack of the amide group of Gly67 to the carbonyl residue of Ser65 (**Fig. 2**). Oxidation of the hydroxybenzyl side chain of Tyr66 by molecular oxygen establishes the active fluorophore structure *(21)*. Originating from the tripeptide Gln66-Tyr67-Gly68 the DsRED fluorophore is structurally similar but possesses an extended conjugated -electron system formed by an additional, rate-limiting dehydrogenation step. Consequently, DsRED maturation proceeds via a green-light-emitting intermediate state (**Fig. 2**).

In general, amino acid exchanges within the fluorophore itself and/or its environment, thus causing fundamentally altered fluorophore resonance properties, will change the spectral properties of the mutant protein. In GFP, substitution of Ser65 by aliphatic residues such as threonine, alanine, or glycine results in red-shifted excitation spectra and higher molar extinction, probably owing to an altered hydrogen-bonding network around the fluorophore *(21)*. The well known blue- and cyan-FPs are also GFP mutants (BFP: Tyr66His; CFP: Tyr66Trp) resulting from the introduction of imidazole or indole side chains at position 66 (**Fig. 2**; **Table 1**). Likewise the introduction of the point mutation Thr203Tyr into GFP results in a yellow fluorescent protein (YFP), owing to stacking of the -electron systems of the tyrosine side chain and the fluorophore .

For DsRED nonfluorophore mutations, altering their spectral properties was also described, but the underlying mechanisms are less well understood and only one mutant, the “fluorescent timer,” is of value as a reporter *(23,24)*. This particular mutant carries the point mutations Val105Ala and Ser197Thr and changes its light emission peaks from green to red over time. The heavily mutated monomeric DsRED variant mRFP1 (33 point mutations) emits far-red light peaking at 607 nm *(25)*.

2. Materials

1. Fluorescence stereo microscope (Leica MZFL-III).
2. Fluorescence microscope (Leica DMRB) equipped with xenon and Hg lamps (XBO 75W/HBO 100W).
3. DISKUS-Imaging system (using a JVC KY-F70B video camera; Ingenieurbuero Hilgers, Koenigswinter, Germany, www.hilgers.com).
4. GFP filters:
   a. HQ-GFP-LP (for LEICA-DMRB, long-pass-filter; AF Analysentechnik, Tübingen, Germany, www.ahf.de): HQ 470/40 (exciter)/Q 495 LP (beam splitter)/HQ 500 LP (emitter).
   b. HQ-GFP (for LEICA-DMRB, band pass filter; AF Analysentechnik): HQ 470/40 (exciter)/Q 495 LP (beam splitter)/HQ 525/50 (emitter).

c. GFP1 (for LEICA-MZFL-III, longpass-filter): 425/60 (exciter)/480nm (barrier-filter)
d. GFP3 (for LEICA-MZFL-III, band pass filter): 470/40 (exciter)/ 525/50 (barrier filter).

5. DsRED filters: DsRED (for LEICA-DMRB and LEICA-MZFL-III, bandpass, AF-Analysentechnik; D546/10 (exciter)/565 LP (beam splitter)/D600/40 (emitter).

3. Methods

The methods outlined in this article describe (1) the construction of plant expression and transformation vectors for FPs (GFP and DsRED), and (2) analysis of ectopically expressed FPs by fluorescence microscopy. The underlying methods for transient gene expression in protoplast and intact leaf tissue, and generation of transgenic *Arabidopsis* plants are provided in separate chapters of this book.

3.1. Construction of Plant Expression and Transformation Vectors

All cloning steps were done according to standard procedures ***(26)*** not described here. *Escherichia coli* strains DH10 or XL1-Blue were used throughout. The expression cassettes of all FP plant expression vectors described here are based on the plasmid pCK-GFP-S65C (*see* **Note 1**) whose cloning was described elsewere ***(27)***.

3.1.1. Expression Vectors for GFP and DsRED

3.1.1.1. Cytosolic GFP and DsRED.T3 Expression

The plasmid pGJ280 (encoding EGFP) was constructed from pCK-GFP-S65T by site-directed mutagenesis using the chameleon-mutagenesis kit (Stratagene) according to the manufacturer's instructions.

Wild-type DsRED proved to be a suitable reporter for transient and stable gene expression in plant cells ***(28)***. Efficacy in transient gene expression studies, however, is limited by its slow maturation rate. In this respect, use of one of the recently published fast-maturing DsRED mutants (DsRED.T1 [= DsRED-Express], DsRED.T3 or DsRED.T4; *see* **Table 1**) is recommended. In fact, transient DsRED.T3 expression in tobacco BY2 cells proved to give rise to at least five times more protoplasts displaying well perceptible red fluorescence after overnight incubation (Jach, unpublished; *see* **Note 2**). Construction of the DsRED.T3 expression vector (pGJ2513) was done by mutagenesis of pGJ1425 (wt-DsRED).

3.1.1.2. ER-Targeted GFP Expression

The coding sequence for an ER retention signal (KDEL) was inserted at the 3' end of the EGFP gene of plasmid pGJ280 by site-directed mutagenesis, as described above. Finally, an XhoI/XbaI-fragment of this intermediate was generated by PCR and inserted into plasmid pGJ460, resulting in the expression vector pGJ1273 (35S-TL-hspEGFP-ER).

3.1.1.3. Translational Fusions with GFP or DsRED

The experimental details for the successful generation of translational fusions between a given protein of interest and either GFP or DsRED are different for each individual case. Nevertheless the following guidelines should be taken into account: (1) In general, the protein of interest might be added to either the N- or C-terminus of the FPs. For both scenarios functional fusion proteins have already been described. However, the best-suited approach can only be determined empirically. (2) Do not delete or change any N- or C-terminal amino acid residues from the FPs, as they are likely to get nonfluorescent (especially true for DsRED). (3) Usually it is not required to introduce a peptide linker between the FPs and the protein of interest, although it might help in individual cases.

To add a protein of interest to the N-terminus of GFP or DsRED it is recommended to introduce a NcoI-site at the 5' end and BspHI-site (removing the stop codon; *see* **Note 3**) at the 3' end of your coding sequence (by polymerase chain reaction [PCR]) and subclone the result-

Table 1
Overview of Commonly Used Fluorescent Proteins and Their Properties

Protein	Mutation	Peak wavelength's (nm)		EC	QY	rel. fluorescence		Ref.
		Ex.	Em.			abs	spec / app	
Green fluorescent proteins								
GFP (gfp10)		395 (471)	508	25000 –30000	0,79	1	1 / 1	*(5)*
S65T	S65T	489	511	52000–58000	0,64	1,6	6 /–[c]	*(12)*
GFPmut1 (EGFP)	F64L, S65T	488	508	56000	0,60	1,5	35 / 39	*(13)*
Cycle3GFP (smGFP)	F99S, M153T, V163A	398 (480)	508	30000	0,79	1,1	–/ 42	*(9)*
mGFP5	V163A, I167T, S175G	395 473	510	–	–	–	– / 20	*(8)*
mGFP5(S65T)	S65T, V163A, I167T, S175G	475	510	–	–	–	–/ 33	*(8)*
smRS-GFP	S65T, F99S, V163A, I167T	–	–	–	–	–	1,68 /	*(40)*
Emerald	S65T, S72A, N149K, M153T, I167T	487	509	57500	0,68	1,8	– / #	*(38)*
Blue fluorescent proteins								
EBFP	F64L, Y66H, Y145F	380 (383)	440 (447)	26300 (31000)	0,17 (0,26)	0,9 (1,6)	– / 5,6	*(41)*
Cyan fluorescent proteins								
W1B (ECFP)	F64L, S65T, Y66W, N146I, M153T, V163A	434 (452)	476 (505)	32500	0,40	1,3	–/ 1,3	*(42)*
Cerulean	F64L, S65T, Y66W, S72A, Y145A, N146I, H148D, M153T, V163A	433	475	43000	0,62	3,2	– / –	*(43)*
Yellow fluorescent proteins								
10C (EYFP)	S65G, V68L, S72A, T203Y	514	527	83400	0,61	1,1	– / 9,7	*(38)*
Topaz	S65G, S72A, K79R, T203Y	514	527	94500	0,60	1,2	– / 16,7	*(38)*
Citrine	S65G, V68L, Q69M, S72A, T203Y	514	524	77000	0,76	1,3	– /–	*(44)*
Venus (SEYFP-F46l)	F46L, S65G, V68L, S72A, M153T, V163A, S175G, T203Y	515	528	92200	0,57	1,1	– / ##	*(45)*

(continued)

Table 1 ***(Continued)***
Overview of Commonly Used Fluorescent Proteins and Their Properties

Protein	Mutation	Peak wavelength's (nm)		EC	QY	rel. fluorescence		Ref.
		Ex.	Em.			abs	spec / app	
Red fluorescent proteins								
DsRED		558	583	52–57	0,68	1[a]	1 / 1	(*17*)
DsRED2	R2A, K5E, K9T, V105A, I161T, S197A	561	587	43,8	0,55	0,65	– / 2	(*32*)
DsRED.T1 (DsRED-Express)	R2A, K5E, N6D, T21S, H41T, N42Q, V44A C117S, T217A	554	586	30,1	0,42	0,34	– / –	(*33*)
DsRED.T3	R2A, K5E, N6D, T21S, H41T, N42Q, V44A, A145P	560	587	49,5	0,59	0,79	– / 7[b]	(*33*)
Fluorescent timer	V105A, S197A	–	499/582	–	–	–	– / 7–20.	(*24*)
mRFP1 (monomeric!)	R2A, K5E, N6D, T21S, H41T, N42Q, V44A, V71A, K83L, C117E, F124L, I125R, V127T, L150M, R153E, V156A, H162K, K163M, A164R, L174D, V175A, F177V, S179T, I180T, V195T, Y192A, Y194K, S197I, T217A, H222S, L223T, F224G, L225A	584	607	44,0	0,25	0,30	– / –	(*25*)

[a]values were calculated using 54.4 as average of the extinction coefficient (EC) of wildtype DsRED

[b]After 24-32 h maturation (as deduced from graphs provided in ref. ***33***)

[c]EC = molar extinction coefficient; QY = quantum yield; (= values not described/determined

[d]All values are relative to wild-type FPs or the first described mutant FP of the individual category, abs = calculated from the individual brightness values (product EC*QY), spec. = published values measured with equal protein amounts, app. = published values measured with whole cells or equal volumes of cellular extracts
protein has only be compared to the GFP mutants GFP-S65T and EGFP. App. fluorescence was found to be 8*S65T and 5*EGFP.
protein was compared to EYFP only . App. fluorescence was found to be 30*EYFP.

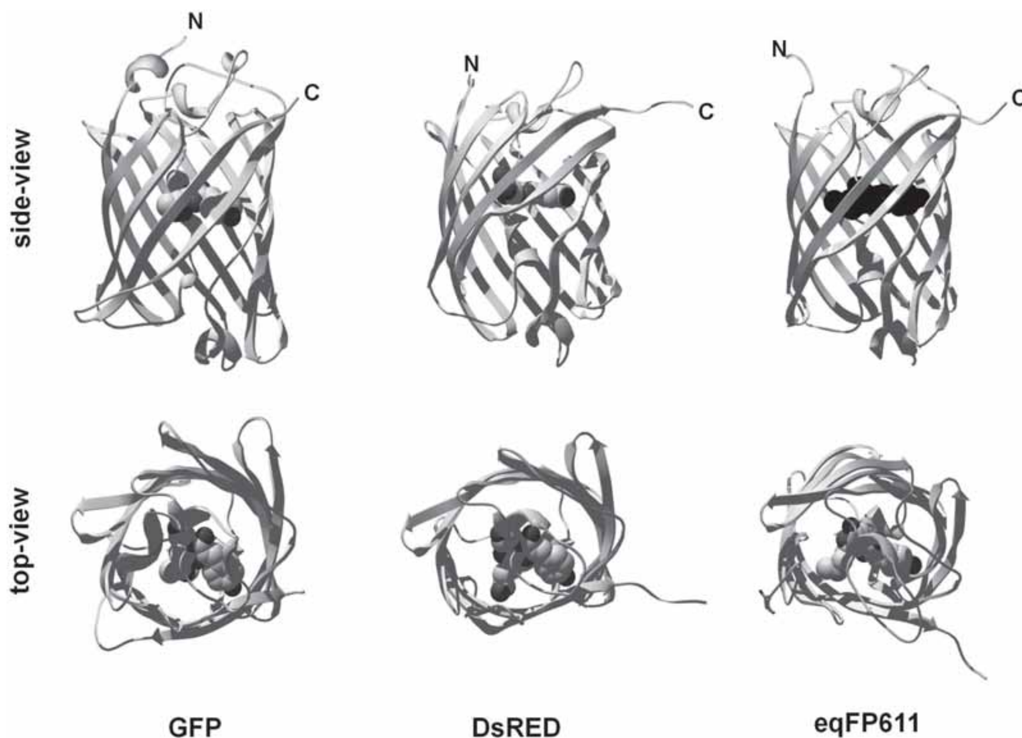


Fig. 1. Ribbon models of the 3D structure of three different FPs are presented. The names of the proteins and descriptions of the viewing angles are given. Space-filling modeling was used for the fluorophore structures. β-Sheet structures appear as flat arrows, whereas coiled, nonhelical structures are modeled as thin rods.

ing gene into the NcoI-site of the described plant expression vectors (*see* above), which marks the translational start site (ATG) of the GFP and DsRED coding sequences.

Likewise, fusions of your protein of interest to the C-terminus of the FPs can be achieved by amplifying the GFP/DsRED codon sequences and their subcloning in front of your given coding sequence. Of course, this requires the removal of the stop codons from the FP genes.

Most likely the sequences of the restriction sites can be adjusted according to your cloning needs, but functionality of the resulting fusion protein should be tested in recombinant expression systems (*E. coli*) whenever possible.

3.1.2. Plant Transformation Vectors

Plant transformation vectors were generated by inserting the GFP and DsRED expression cassettes (*see* **Subheading 3.1.1.**) as EcoRV/XbaI fragments into pGJ2164 (Jach, unpublished), a derivate of the pPZP200 ***(29)*** harboring the *bar* gene as selectable marker gene (conferring phosphino-tricine/BASTA resistance) and a streptomycin/spectinomycin-resistance gene for selection of *Agrobacteria*.

3.2. Transient Gene Expression

These methods are often used to quickly test whether or not a given gene is suitable for plant expression. Furthermore, protein localization studies and other approaches can be done. Several methods are available, among which PEG-mediated DNA uptake into protoplasts and particle bombardment are the most widely used (*see* **Note 4**). For experimental details please *see* Chapters 18, 19, and 20 in this book.

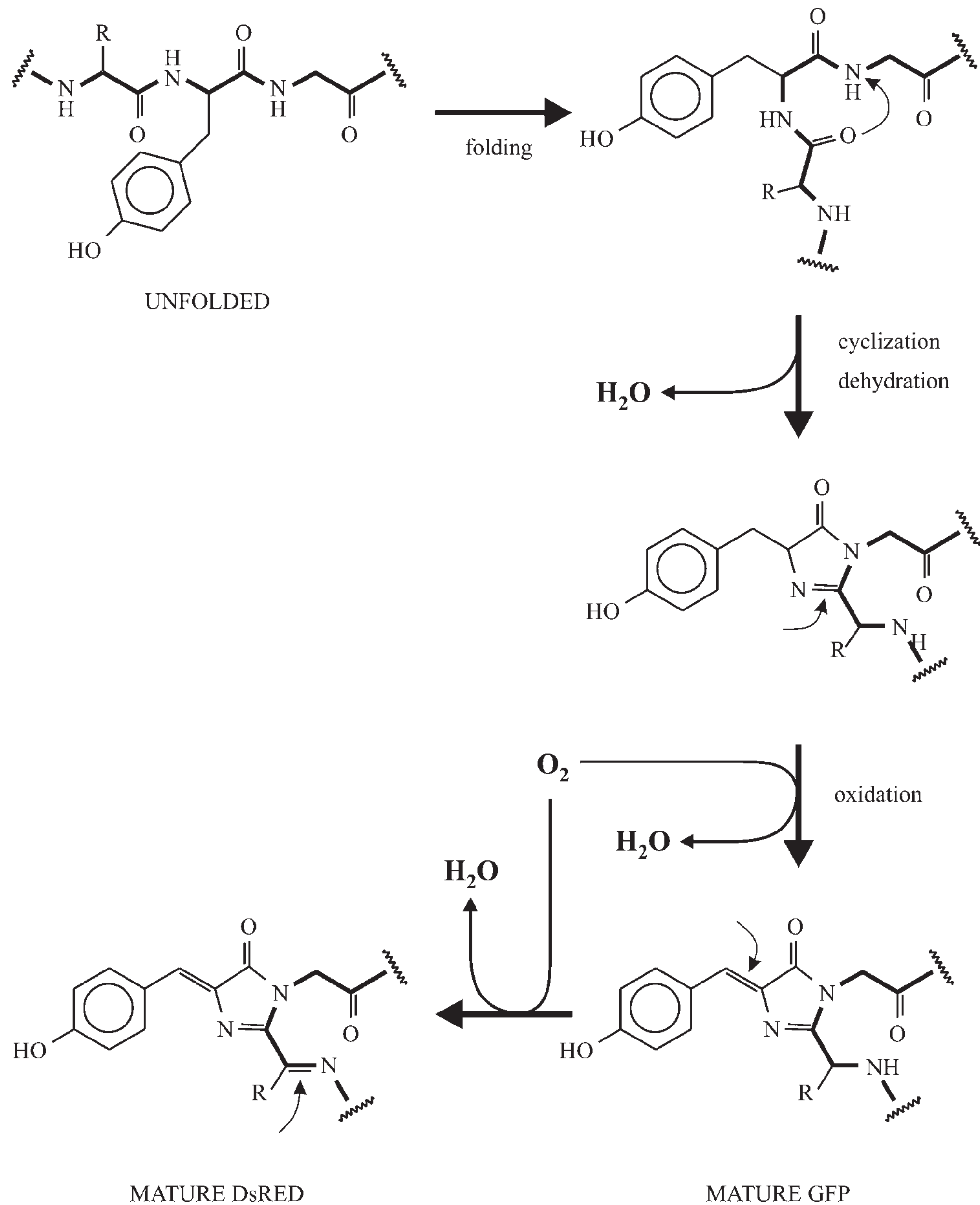


Fig. 2. The proposed mechanism of the fluorophore maturation of GFP-like fluorescent proteins is depicted. R = side-chain of the first amino-acid of the fluorophore-tripeptide (GFP: Ser65; DsRED: Gln66).

3.3. Stable Gene Expression

The easiest and least laborious way to generate transgenic *Arabidopsis* plants is provided by the so-called floral-dip transformation method (***30***; *see also* Chapter 17 in this book). When binary plant-transformation vectors containing the *bar* gene are used, the screening for transgenic plant lines can be done the greenhouse, thus avoiding any tissue-culture steps (*see* **Note 5**).

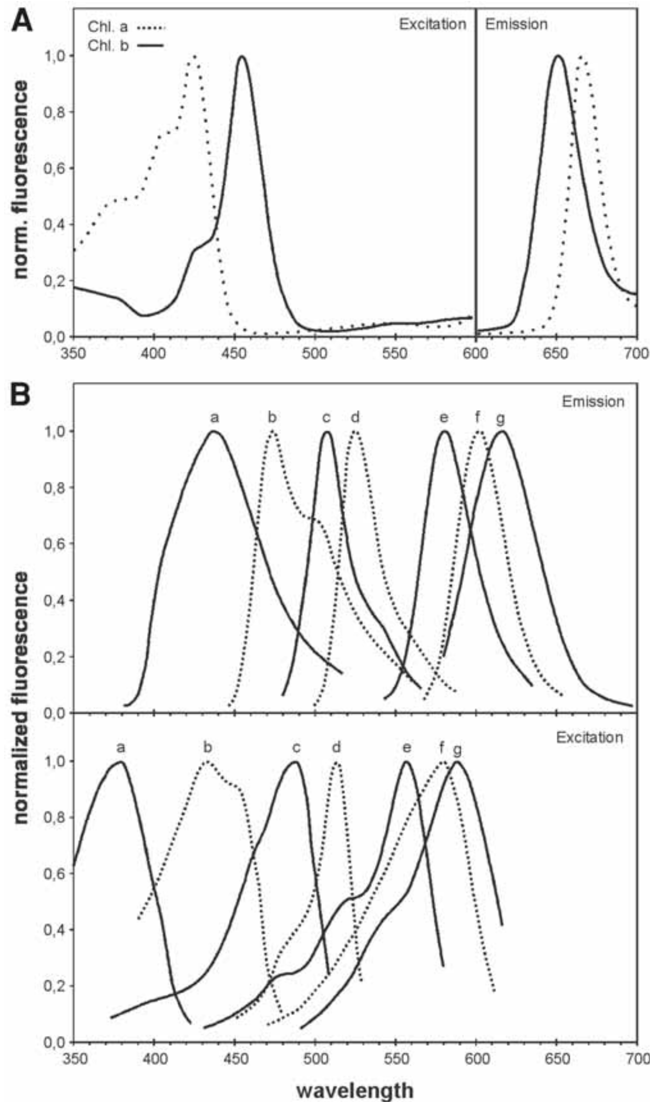


Fig. 3. A comparison of the fluorescence spectra of chlorophyll a, chlorophyll b, and commonly used FP reporters is presented. A: chlorophyll fluorescence (excitation and emission). Chl. a = Chlorophyll a, Chl. b = chlorophyll b. B: excitation and emission spectra of common FPs. a = BFP, b = CFP, c = GFP, d = YFP, e = DsRED, f = mRFP1 and g = HcRED.

3.4. Analysis of Protoplasts and Plant Tissues by Fluorescence Microscopy

Fluorescence microscopy is an easy way to detect and analyze expression of fluorescent proteins in transfected protoplasts, bombarded tissues, and/or transgenic plant lines (*see* **Note 4**). If available, the use of a confocal laser scanning microscope should be considered when working with thick samples (intact leaves). Here, its ability to generate optical sections and to avoid fluorescent blurred backgrounds comes in quite handy and can result in vastly improved image quality (*see* **Note 6**).

When working with green tissues and mesophyll protoplasts, GFP detection can be problematic due to the presence of chlorophyll, which emits strong red fluorescence, when excited with blue light (**Fig. 3A**), which is also used to excite GFP (*see* **Fig. 3B**). Specialized filter sets

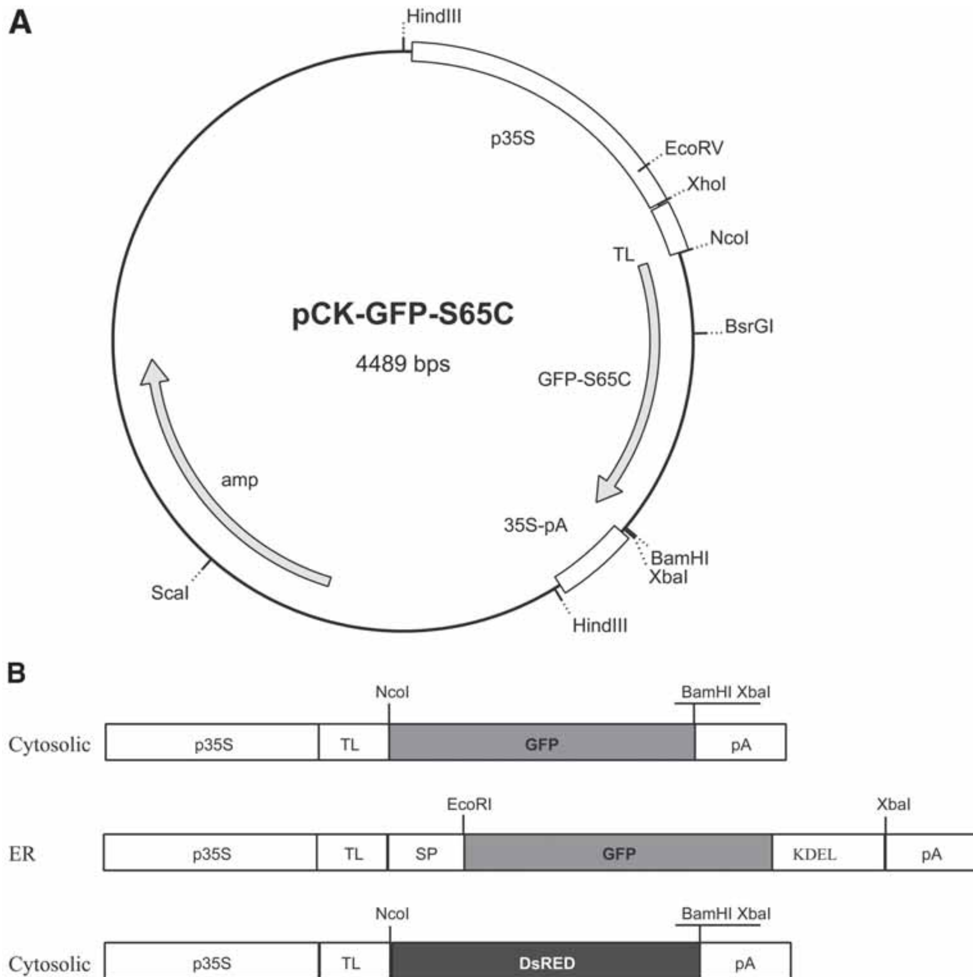


Fig. 4. **(A)** The map of the plant expression vector pCK-GFP-S65C is shown. The vector is based on pRTL2-GUS and provided the expression cassette for all GFP and DsRED constructs used in this study. **(B)** Schematic view of the GFP and DsRED expression cassettes used in this study Amp = β-lactamase gene, p35S = CaMV-35S promoter, TL = translational enhancer from tobacco etch virus (TEV), 35S-pA = polyadenylation signal of the CaMV 35S gene, SP = signal peptide, KDEL = ER retention signal. Some key restriction-enzyme sites are depicted.

(GFP-band pass filters) passing green light but blocking red signals (*see* **Fig. 5**) should be used in this situation, although the intensity of the light passing these filters is significantly reduced compared to long-pass filters (passing green to red light).

Analysis of DsRED is less problematic, as the green light used to excite the protein does not effectively excite chlorophyll (**Fig. 3B**). Moreover, properly chosen DsRED filters can very effectively block any fluorescence emitted from chlorophyll during microscopic analysis (**Fig. 5**). In consequence, the visible background when analyzing healthy *Arabidopsis* leaf tissue or mesophyll protoplasts is close to zero (*see* **Note 7**).

3.4.1. Microscopic Analysis of Protoplasts

Starting materials for this approach are protoplast suspensions derived from either transient gene expression studies or protoplastation of leaf tissues from transgenic plants stably express-

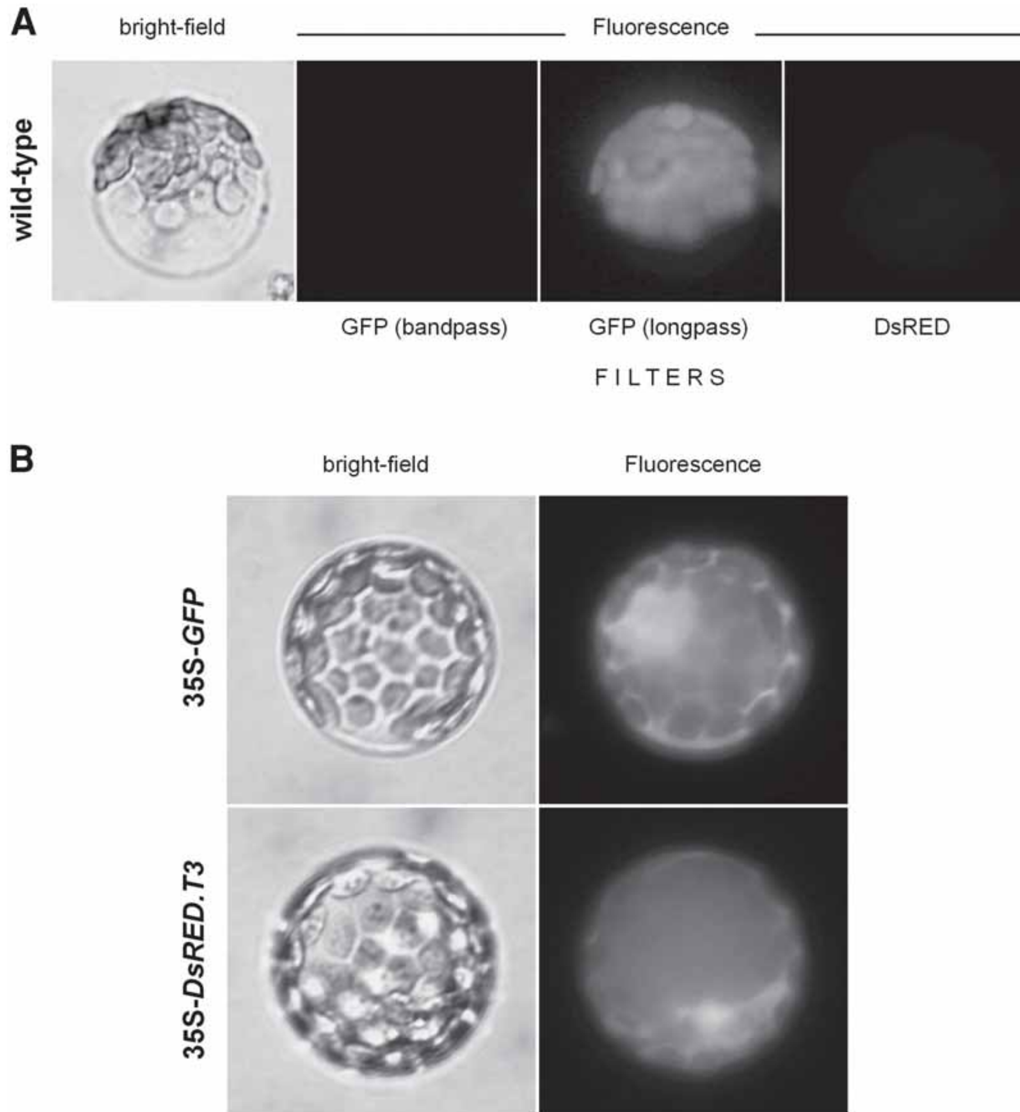


Fig. 5. Transient expression of GFP and DsRED.T3 in *Arabidopsis* mesophyll protoplasts is shown. (**A**) Background fluorescence present in non-transfected protoplasts when analyzed with different optical filter sets. GFP-LP = long-pass filters (blue-light excitation/passes green to red light emission, GFP-BP = band-pass filters (blue-light excitation/ passes green light only) RFP = band-pass filter for DsRED detection (green light excitation/ passed red light of 580 to 620 nm). (**B**) The lower panel shows *Arabidopsis* protoplasts showing cytosolic expression of GFP and DsRED.T3, respectively. Pictures were taken with GFP bandpass and DsRED filters.

ing FPs (for experimental details, refer to Chapters 19 and 20 in this book). After you have made yourself familiar with the available equipment, perform fluorescence microscopy as follows:

1. Place a drop (25–50 µL) of your protoplast suspension on a glass slide and place a cover slip on top of it. Be careful to avoid disruption of protoplasts from to mechanical shearing. Avoid excess of liquid, otherwise the cover slip will "swim" and microscopy at higher magnification using oil-immersion objectives will be very difficult.

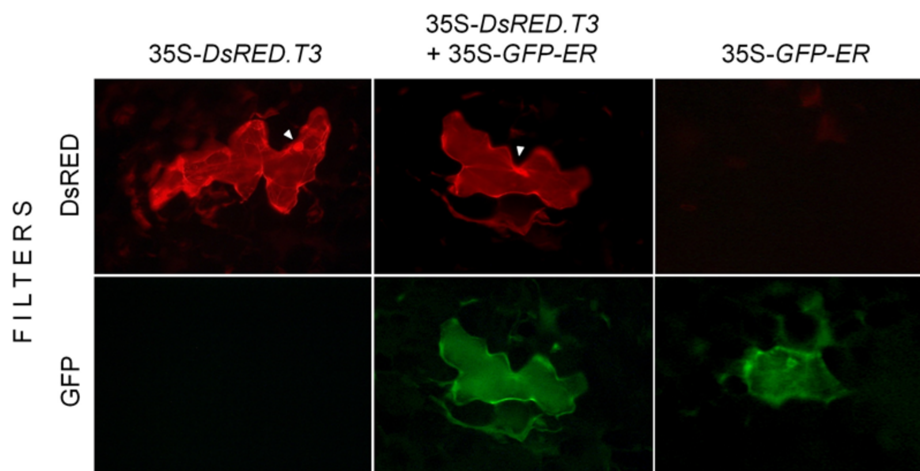


Fig. 6. Transient expression of EGFP-ER and DsRED.T3 in bombarded Arabidopsis leaves is presented. As indicated, leaves were bombarded with either the construct 35S-*EGFP*, 35S-*DsRED.T3*, or a mixture thereof. All leaves were analyzed by fluorescence microscopy using GFP bandpass and DsRED filters. Positions of stained nuclei are marked by arrowheads. *See* color version in inser following p. 239.

2. Start microscopy of your specimen. Find the focal plane in the bright field at low magnification (10×).
3. Switch to fluorescence microscopy by choosing a suitable filter set for your given FP (still using 10× magnification). Get a first overview and find suitable cells showing satisfactory reporter gene expression (*see* **Notes 6** and **8**).
4. Switch to higher magnifications (40–100-fold) to obtain high-resolution pictures of individual cells (*see* **Note 9**).

Sometimes movement of protoplasts while performing fluorescence microscopy is disturbing. To fix the cells the following protocol can be applied:

1. Add 0.3% agarose to protoplast medium and heat to dissolve.
2. Cool the solution and maintain it at 45°C until use
3. Mix (briefly and gentle) 1 vol of protoplast suspension with 0.5 vol of the warm agarose solution.
4. Immediately put a sample (25–50 µL; depending on the size of the cover slip) on a glass slide and (carefully) place the cover slip on top of it.
5. Incubate for 5 min at room temperature (until agarose becomes solid).
6. Perform fluorescence microscopy as described above.

3.4.2. Microscopic Analysis of Leaf Tissues

This analysis is usually required for the investigation of transient gene expression after particle bombardment or monitoring of reporter gene expression in transgenic plant lines:

1. Use a fluorescence stereomicroscope with 2 to 10× magnification to identify leaves showing FP expression (*see* **Note 7**).
2. For closer investigation of the leaf surface (epidermal cells, trichomes, stomata, etc.) remove the identified leaves from the Petri dish or transgenic plant and mount it on top of a glass slide with the aid of commonly used transparent double-faced adhesive strips (alternatively use a drop of water and a cover slip).
3. If required prepare cross-sections (e.g., tissue-specific gene expression) of your tissue of interest, put them with a drop of water on a glass slide and add a cover slip (*see* **Note 10**).
4. Perform microscopy as described above (*see* **Note 11**).

3.4.3. Imaging

Video image systems allow documenting your microscopic work with ease and high flexibility. However, resolution (here: 1.2 megapixels) and signal-to-noise ration of a video camera is significantly lower compared to today's CCD still cameras. Nevertheless, even with a video camera; publication-quality pictures can be obtained if a few things are taken into account in order to improve image quality:

1. Always use the picture-integration function (simulates "exposure time") of the video camera instead of the live picture in order to minimize electronic noise present in your picture.
2. Take additional pictures of empty glass slides for each exposure time used in your experiments and use suitable software (e.g., Adobe Photoshop; Corel Photopaint) to subtract this background noise (including the "dead" pixels of the CCD chip) from your sample pictures.

4. Notes

1. The expression vectors used in this study are based on the vector pCK-GFP-S65C (**Fig. 4**, ref. ***27***) a derivative of the vector pRTL2-GUS (all vectors are freely available from the author's lab). Even though expression plasmids harboring the 35S promoter will certainly work as well, the use of pCK-GFP-S65C-derived expression cassettes is recommended for two reasons: (a) the 35S-promoter used contains a tandem duplication of its own transcriptional enhancer leading to significantly increased transcription levels and (b) the presence of sequences of the 5' untranslated leader of the tobacco etch virus RNA acting as strong translational enhancer (TL) causes a 5- to 10-fold boost in protein yields (***31***; Jach, unpublished). Hence a strong transgene expression is ensured, which is an ultimate prerequisite for easy detection of any FP. In our hands the TL elements constantly and reproducibly yielded higher protein amounts than the Omega-element form tobacco mosaic virus (TMV), another viral sequence with translational enhancer activity. Since the TL element does not alter the regulatory properties of the promoter it is recommended in general to use it with any given promoter of interest.
   With respect to translation efficiency, the codon use also was shown to play an important role. However, commercially available codon-optimized FP genes are adapted to the codon-bias found in humans ("humanized" proteins), which is quite similar to that of monocot plants, but distinct from dicot plants such as tobacco, potato, and *Arabidopsis*. Moreover, comparisons of non-codon-optimized and optimized gene versions encoding the same GFP mutant proved a 30 to 50% lower performance of the "humanized" gene in dicot plants (Jach, unpublished). Consequently, one can only benefit from these genes when working with monocot plants (rice, barley wheat, etc.).
   Localized expression of FPs in the endoplasmic reticulum, vacuole, or cell wall requires the presence of an N-terminal signal peptide to enter the default secretory pathway. In our hands different signal peptides not only showed different secretion efficiencies, but also affected the obtained protein yields. Hence, the signal peptide to be used should be chosen carefully to ensure highest expression levels of the desired FP.
2. Besides its slow maturation rate and its oligomerization, wild-type DsRED protein has another property that might interfere with its usefulness as a reporter protein: It undergoes protein aggregation. In consequence, microscopic images may appear blurred, hampering protein localization studies ***(32)***. However, mutant DsRED proteins abolishing this problem are available. They contain a couple of N-terminal point mutations. In fact, the "wild-type" protein used in this study also contained a point mutation (Arg2Gly, ***28***) at its N-terminus. Problems due to aggregations have never been observed while working with this protein.
   The DsRED mutants DsRED-Express or DsRED.T4 are described to be even faster in maturation and might be used instead of DsRED.T3 ***(33)***. However, at least in our hands the (relatively small) difference in maturation speed is overcompensated by the about twofold higher light emission of the variant DsRED.T3, which reproducibly gave better results in our transient gene expression studies.
3. BspHI is compatible to NcoI, a restriction enzyme site frequently found or introduced at the translational start site of cloned genes. Note that in case of GFP the end sequence is changed and elongated (wt: —MDELYK-stop; mod: —MDEL***IM***R—), whereas in RFP the C-terminal

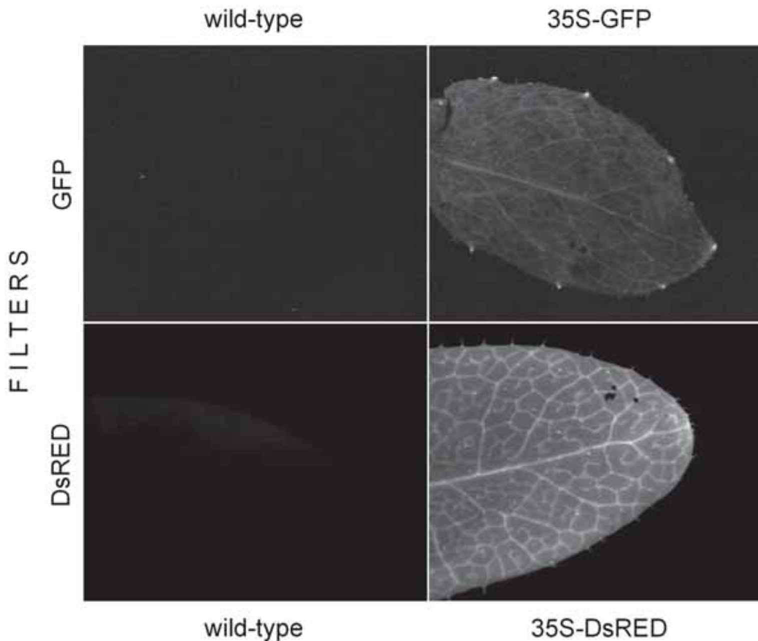


Fig. 7. Leaves of transgenic *Arabidopsis* plants expressing GFP and DsRED.T3 are shown in this figure.

amino-acid sequence is only elongated (wt: HLFL-stop; mod: HLFLEFI— and HLFLVNI—). It has been mentioned in the literature that the C-terminus of wt-DsRED probably plays a role in its oligomerization. However, functional C-terminal fusions with DsRED could be obtained, indicating that apparently addition of other proteins to its C-terminus does not interfere with this process.

4. In transient gene expression studies one will always find a broad range of different expression levels, whereas protoplasts derived from transgenic plants are much more uniform. For transient expression studies in protoplasts the standard deviation of replicates within one assay usually does not exceed a level of 5 to 10%. The absolute expression levels between independent experiments/assays vary considerably, whereas the relative expression levels are quite reproducible under standardized conditions. Avoid using old DNA preparations in these assays. Prolonged storage of plasmids (>6–9 mo) and/or numerous freeze/thaw cycles usually cause significantly reduced expression levels of the transgene. For semiquantitative comparisons of different plant expression vectors, freshly prepared plasmid DNA of the same age should be used.
5. As visible from **Fig. 7** it is, in principle, possible to generate vital transgenic *Arabidopsis* plants showing stable constitutive expression of cytosolic GFP. However, as demonstrated by Haseloff and colleagues, both the transformation efficiency and the average expression levels of the transgene are significantly increased when an ER-localized GFP version is used ***(34)***. In contrast, transformation efficiencies of cytosolic and ER-targeted GFP are about the same when used in transient gene expression studies, indicating that high-level cytosolic GFP expression somehow interferes with the plant regeneration.
6. In conventional fluorescence microscopy it is not always the best solution to choose the highest-expressing protoplast for further investigation at high magnification due to the following reasons: The fluorescence signal will increase with increasing magnification because (a) fluorescence light emission of any fluorescent molecule is directly proportional to the amount of input (light) energy and (b) for any given area the amount of input energy increases at higher magnification due to focusing of the excitation light beam. Hence, at high magnification the fluorescence signal might be overwhelming, leading to a blurred picture with little detail. By using a confocal laser scanning microscope these problems can be overcome, but

these instruments are extremely expensive, their handling needs some training, and they are thus not available in all laboratories. A nice introduction into the basics of laser scanning microscopy (LSM) is given on www.olympusmicro.com/primer/virtual/confocal/index.html.

7. Fluorescence microscopy should never be done without a suitable negative control (mock-transfected protoplasts, wild-type plants or [ideally] transgenic plants generated with an empty vector). In tissue from greenhouse plants, often single wounded or heavily stressed cells can be found that give rise to fluorescent signals quite similar in color and intensity to GFP. Likewise, necrotic tissues tend to produce strong red fluorescence mimicking DsRED. Before attempting to use FPs as reporters, make yourself familiar with the fluorescent background of the chosen target tissues/cells. In *Arabidopsis,* for example, healthy anthers and seeds already produce strong fluorescent background signals ranging from green, to red thus hampering the use of FPs as reporters.
8. As can be seen from **Fig. 3**, spectra of some FPs (e.g., GFP/YFP and CFP/GFP) show considerable overlap. Moreover, the bandwidth of the respective emission (bandpass) filter is usually at least ±20 nm beyond the peak wavelength to allow for satisfactory signal intensities. This allows for significant crosstalk (bleedthrough) of FP signals. In consequence, CFP or GFP signals can also be visualized using GFP or YFP emission filters, although to a lesser extent. Especially in multicolor studies (simultaneous expression of different FPs) bleedthrough is problematic and, therefore, differentiation between certain FPs using conventional fluorescence microscopy and optical filters is almost impossible. LSM can help overcome some of the limitations but bleedthrough still must be taken into account. However, simultaneous expression of GFP and DsRED proved to be nonproblematic in both conventional and confocal microscopy.
9. Cytosolically expressed GFP and DsRED typically stains not only the cytoplasm but also the nucleus, probably due to passage of the nuclear pore by passive means ***(28)***. In experiments intended to show nuclear localization of a protein of interest using FPs as fluorescent tags, GFP–GUS fusion proteins (or similar constructs) should be used as negative controls, as they are known to be excluded from the nucleus due to their increased size. Note that fusions of GFP with itself ("double GFP") are less well suited, as they still give rise to some fluorescence in the nucleus. The reasons are unknown but protolysis cannot be excluded. In fact, experience showed that fusions of GFP with other proteins are prone to undergo proteolytic cleavage in a way that releases intact fluorescent GFP molecules in the size of the monomer. Therefore, Western blot analyses should be done in addition to localization studies to verify the presence of the full-length fusion protein in the test system. In our analyses the monoclonal anti-GFP antibody from Roche always gives excellent results with no background, whereas the polyclonal DsRED peptide antibody from BD Biosciences (Clontech) showed in some cases crossreactions with endogenous plant proteins, but appeared to be more sensitive than the corresponding monoclonal antibody from the same company.
10. In general, simple hand-hold cross-sections should be sufficient to allow for a first approximation of the tissue specificity of gene expression. However, to produce pictures of highest possible resolution and quality, application of fixation and embedding techniques to do your sections with a microtome would be an advantage. Unfortunately, little is known about the compatibility of these techniques with the different FPs. For animal cells it has been described that GFP tolerates formalin fixation and paraffin embedding, as well as 4% paraformaldehyde 0.1% glutaraldehyde fixation and embedding in LR-white without loss of fluorescence ***(35,36)***. For DsRED and work on plant tissues no such information is available. In addition, there are no published standard protocols for these applications in plants. It is therefore recommended to use the given fixation/embedding protocols developed for *in situ* hybridization methods as guidelines ***(37)***. However, artefacts such as stress induced background fluorescence might very easily be introduced be these procedures. Therefore the use of suitable negative controls is essential and any conclusions should be drawn very carefully. However, confocal LSM might help overcome some of the problems. By acquiring series of optical sections ("Z-stacks") the microscope software is able to compute a 3D image of the fluorescence distribution in a sample without disrupting the tissue (provided the thickness of the sample is not too high).

11. It is important to note that there are also some disadvantages in the use of FPs as reporters, mainly due to the absence of signal amplification. Whereas a single copy of reporter enzymes like β-galactosidase, luciferase, or β-lactamase will catalyze the turnover of multiple substrate molecules, FPs are limited to a single fluorophore for each protein, which leads to significantly reduced sensitivity levels. Nevertheless, for certain experimental approaches such as protein localization, FPs are superior.
    In general, the sensitivity of an FPZ based reporter system is determined by numerous factors such as the protein yield, the efficiency of protein maturation/fluorophore formation, the individual properties of the fluorescent protein in use, the organism and tissue under investigation, and, last but not least, the available technical equipment ***(38)***. Moreover, each of these factors itself is influenced by several parameters. The expressed protein amounts are determined by the rate of protein synthesis vs degradation. Protein synthesis generally depends on the amount of mRNA produced (because of the given strength and temporal/spatial expression pattern of the promoter in use, the gene copy number, and the presence of transcriptional enhancer sequences) and its translation efficacy. The latter is determined by the stability (half-life) of the mRNA, its splicing efficiency, its codon use, the sequence of the translational start site, and the presence of translational enhancer sequences ***(16,34,39)***.
    The rate of protein/fluorophore maturation is an intrinsic property of each fluorescent protein and is affected by time, temperature, and oxygen availability ***(38)***. In case of fusions between host proteins and a fluorescent protein hindrance of the protein folding, thus preventing proper maturation, cannot be excluded and must be tested empirically.
    The efficacy of a certain fluorescent protein as a reporter protein and its ability to overcome/avoid possible autofluorescent background signals depends on its brightness (given by its extinction coefficient and fluorescence quantum yield) and its specific excitation and emission peak wavelengths. Among the factors limiting the usefulness of a fluorescent protein are oligomerization and susceptibility to photoisomerization/bleaching ***(14)***.

References

1. Zhang, J., et al. 2002 Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell. Biol.* **3,** 906–918.
2. Shimomura, O., Johnson, F., and Saiga, Y. (1962) Extraction, purification and properties of Aequorin, a bioluminescent protein from the luminous hydromedusa, *Aequorea. J. Cell Comp. Physiol.* **59,** 223–239.
3. Shimomura, O. (1979) Structure of the chromophore of *Aequorea* green fluorescent protein. *FEBS Lett*. **104,** 220–222.
4. Cody, C., et al. 1993 Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein. *Biochemistry* **32,** 1212–1218.
5. Prasher, D., et al. (1992) Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* **111,** 229–233.
6. Chalfie, M., et al. (1994) Green fluorescent protein as a marker for gene expression. *Science* **263,** 802–805.
7. Inouye, S. and Tsuji, F. (1994) *Aequorea* green fluorescent protein: expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Lett*. **341,** 277–280.
8. Siemering, K. R., et al. (1996), Mutations that suppress the thermosensitivity of green fluorescent protein. *Curr. Biol.* **6,** 1653–1663.
9. Crameri, A., et al. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14,** 315–319.
10. Kimata, Y., et al. (1997) A novel mutation which enhances the fluorescence of green fluorescent protein at high temperatures. *Biochem. Biophys. Res. Comm.* **232,** 69–73.
11. Heim, R. and Tsien, R. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* **6,** 178–182.
12. Heim, R., Cubitt, A., and Tsien, R. (1995) Improved green fluorescence. *Nature* **373,** 663–664.
13. Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996) Facs-optimized mutants of the green fluorescent protein (gfp). *Gene* **173,** 33–38.
14. Cubitt, A. B., et al. (1995) Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **20,** 448–455.

15. Pollok, B. A. and Heim, R. (1999) Using GFP in FRET-based applications. *Trends Cell Biol.* **9,** 57–60.
16. Yang, T. T., Cheng, L. Z., and Kain, S. R. (1996) Optimized codon usage and fluorophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* **24,** 4592–4593.
17. Matz, M. V., et al. (1999) Fluorescent proteins from nonbioluminescent *Anthozoa* species. *Nat. Biotechnol.* **17,** 969–973.
18. Labas, Y. A., et al. (2002) Diversity and evolution of the green fluorescent protein family. *Proc. Natl. Acad. Sci. USA* **99,** 4256–4261.
19. Wiedenmann, J. E. A., et al. (2002) A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor. *Proc. Natl. Acad. Sci. USA* ***99,*** 11,646–11,651.
20. Prendergast, F. G. (1999) Biophysics of the green fluorescent protein, in *Methods in Cell Biology: Green Fluorescent Proteins*, Sullivan, K. F. and Kay, S. A., eds. Academic Press, New York, p. 1–17.
21. Cubitt, A. B., Woollenweber, L. A., and Heim, R. (1999) Understanding structure-function relationships in the Aequorea victoria green fluorescent protein, in *Methods in Cell Biology: Green Fluorescent Proteins*, Sullivan, K. F. and Kay, S. A., eds. Academic Press, New York, p. 19–29.
22. Reid, B. G. and Flynn, G. C. (1997) Fluorophore formation in green fluorescent protein. *Biochemistry* **36,** 6786–6791.
23. Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. USA* **97,** 11,984–11,989.
24. Terskikh, A., et al. (2000) "Fluorescent timer": Protein that changes color with time. *Science* **290,** 1585–1588.
25. Campbell, R. E., et al. (2002), A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. USA* **99,** 7877–7882.
26. Sambrock, J., Nicklen, S., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27. Reichel, C., et al. (1996) Enhanced green fluorescence by the expression of an aequorea victoria green fluorescent protein mutant in mono- and dicotyledonous plant cells. *Proc. Natl. Acad. Sci. USA* **93,** 5888–5893.
28. Jach, G., et al. (2001) Use of red fluorescent protein from *Discosoma* sp (dsRED) as a reporter for plant gene expression. *Plant J.* **28,** 483–491.
29. Hajdukiewicz, P., Svab, Z., and Maliga, P. (1994) The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation. *Plant Mol. Biol.* **25,** 989–994.
30. Clough, S. J. and Bent, A. F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16,** 735–743.
31. Carrington, J. C. and Freed, D. D. (1990) Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. *J. Virol.* **64,** 1590–1597.
32. Yanushevich, Y. G., et al. (2002) A strategy for the generation of non-aggregating mutants of Anthozoa fluorescent proteins. *FEBS Lett.* **511,** 11–14.
33. Bevis, B. J. and Glick, B. S. (2002) Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *Nat. Biotechnol.* **20,** 83–87.
34. Haseloff, J., et al. (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic arabidopsis plants brightly. *Proc. Natl. Acad. Sci. USA* **94,** 2122–2127.
35. Walter, I., et al. (2000) Rapid and sensitive detection of enhanced green fluorescent protein expression in paraffin sections by confocal laser scanning microscopy. *Histochem. J.* **32,** 99–103.
36. Luby-Phelbs, K., et al. (2003) Visualization of identified GFP-expressing cells by light and electron microscopy. *J. Histochem. Cytochem.* **51,** 271–274.
37. Jackson, D. P. (1992) In situ hybridization in plants. In: *Molecular Plant Pathology: A Practical Approach* (Bowles, D. J., Gurr, S. J. and McPhereson, M., eds.) Oxford University Press, Oxford, pp. 163–174.
38. Tsien, R. Y. (1998) The green fluorescent protein. *Ann. Rev. Biochem.* **67,** 509–544.
39. Kozak, M. (1989) The scanning model for translation: an update. *J. Cell Biol.* **108,** 229–241.
40. Davis, S. J. and Vierstra, R. D. (1998) Soluble, highly fluorescent variants of green fluorescent protein (**Gfp**) for use in higher plants. *Plant Mol. Biol.* **36,** 521–528.
41. Patterson, G. H., et al. (1997) Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.* **73,** 2782–2790.
42. Heim, R. and Tsien, R. Y. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence energy transfer. *Curr. Biol.* **6,** 178–182.

43. Rizzo, M. A., et al. (2004) An improved cyan fluorescent protein variant useful for FRET. *Nat. Biotechnol.* **22,** 445–449.
44. Griesbeck, O., et al. (2001) Reducing the environmental sensitivity of yellow fluorescent protein—mechanism and applications. *J. Biol. Chem.* **276,** 29,188–29,194.
45. Nagai, T., et al. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20,** 87–90.

24

Luciferases as Reporter Genes

Megan M. Southern, Paul E. Brown, and Anthony Hall

Summary

Luciferase is the ideal reporter gene to provide temporal and spatial information on promoter activity in *Arabidopsis* and other eukaryotes; the noninvasive detection of luminescence and short half-life of luciferase activity allow repeated measurements of individual seedlings over several days to assay dynamic changes in gene expression. Transgenic or transiently transformed plants with a luciferase gene under the control of a promoter of interest are required. Detection of the low level of luminescence produced by the luciferase gene in *Arabidopsis* requires the use of low light detecting charge-coupled device (CCD) cameras or scintillation counters. This chapter contains protocols on assaying and imaging luciferase in vivo and the automation for high-resolution timecourses.

Key Words: Luciferase (LUC); LUC+; CBG68luc; CBG99luc; CBRluc; click beetle; luciferin; gene expression; *Arabidopsis*; genetic engineering; reporter gene; transcription.

1. Introduction

The luciferase (LUC) enzyme expressed in *Arabidopsis* emits light during the oxidation of its substrate, luciferin. *LUC* is the ideal reporter gene to provide temporal information on promoter activity, as levels can be measured noninvasively, in vivo, and repeatedly from the same individual over several days ***(1)***. In the absence of luciferin, LUC is very stable; the first readings after the exogenous application of luciferin are from LUC built up over the lifetime of the plant. In the presence of luciferin LUC loses activity, possibly by end product inhibition, with a half-life of about 2 to 3 h, allowing detection of both increases and decreases in promoter activity ***(1,2)***. In plants, LUC emits relatively low levels of light, necessitating the use of sensitive photon-detecting technology.

Gene expression can be assayed by measuring RNA levels using Northern blotting, quantitative reverse transcription-polymerase chain reaction (RT-PCR), or S1 protection assays. Tissue must be harvested for each timepoint and RNA extracted and processed further. Each timecourse is therefore laborious and technically demanding. In vivo luciferase assays avoid this, as samples can be measured nondestructively and automatically. However, a promoter:*LUC* construct must be prepared for each promoter to be studied and transformed into every plant background in which the promoter activity is to be assayed, and a luminescence detection system in a light-tight room is required. Once the investment has been made the luciferase reporter gene is the easiest method to obtain many traces of promoter activity from individual plants or parts of plants with high time resolution. An alternative in vitro luciferase technique is well established; however, this lacks many of the advantages described above. The methodology of this technique will not be discussed here but can be found at www.promega.com.

A variety of *LUC* genes are available from different organisms with different modifications for use as a reporter gene. The most used luciferase is from the firefly *Photinus pyralis*. It emits

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

a photon at 560 nm during the Mg^{2+}- and ATP-dependent catalysis of beetle luciferin and O_2 to dehydroluciferin and CO_2. Some features of the native *LUC* are not ideal for a reporter gene: a peroxisomal translocation sequence, restriction sites and glycosylation sites. A modified form has been produced, *LUC*+, in which these have been removed and the codon use improved for mammalian cells ***(3)***. In *Arabidopsis*, *LUC*+ produces five to 20-fold brighter luminescence than native *LUC*. We routinely use *LUC*+, and this will be the luciferase used in the protocol. Other luciferases are available with different properties to firefly *LUC* that allow them to be measured simultaneously. Click beetle, *Pyrophorus plagiophthalamus*, luciferase is related to firefly *LUC* and has been engineered to emit different fluences and remain in the cytoplasm ***(4)***. *CBG68luc* and *CBG99luc* are green-emitting at 537 nm and *CBRluc* is red-emitting at 613 nm. Different genes can therefore be measured simultaneously by using filters or optical splitters to distinguish among the different luminescence spectra. Like firefly *LUC*, their substrate is luciferin. Photomultiplier tube-based detection systems, such as scintillation counters, are often less sensitive in the red region, and the signal from the red-emitting reporter would appear weaker than that from the green emitting and firefly luciferases; this has not been found in charged-coupled device (CCD) camera detection systems. The sea pansy, *Renilla reniformis*, luciferase is evolutionarily distinct from the previous examples. It does not require cellular ATP or Mg^{2+} so it would be more suitable in assays where these are limited ***(5)***. Renilla *LUC* utilizes a different substrate than firefly *LUC*, coelenterazine, enabling their activities to be distinguished. Coelenterazine is unstable in light, restricting the use of *Renilla LUC* to protocols in which material is maintained in the dark. All the luciferase genes mentioned are available from Promega (Southampton, UK).

A construct should be designed with the luciferase gene downstream of the promoter fragment of interest and with selection markers. We use the plant binary vector pPCV812, which contains a bacterial ampicillin resistance cassette to allow selection in *Escherichia coli* and *Agrobacterium tumefaciens* and a hygromycin resistance cassette to allow selection of transformed plants. The *LUC*+ gene is supplied by Promega in the pSP-*luc*+ vector. Reproducing the regulation of a gene as completely as possible when the *cis*-acting regulatory elements are unknown include all the 5' sequences, including the leader but not the ATG, to the end of the next coding region. In plants, the important regulatory fragments are usually found within 2 to 1.5 kb upstream of the ATG, so this is generally sufficient. The promoter should be fused to the *LUC*+ gene ATG. The *nos* termination sequence should be attached to the 3' of the *LUC*+ gene ***(6,7)***. Alternatively, protein luciferase fusions can be used to measure dynamics in translation as demonstrated with our C-terminal PHYB protein fusion ***(6)*** (**Fig. 1**). Some *Arabidopsis* lines with luciferase transgenes are available from the stock centers; the following are a selection with the Nottingham Arabidopsis Stock Centre numbers. Chlorophyll A/B Binding Protein2 (CAB2) is available with both *LUC*+ and native *LUC*, *CAB2::LUC*+ in Ws ***(8)*** (N9352), and *CAB2::LUC* in C24 ***(1)*** (N3755). Phytochrome B red light photoreceptor, *PHYB::LUC*, is available in the Ws ecotype (N9343); the gibberellin-responsive promoter *GASA1::LUC* ***(9)*** (N57945), and the jasmonate-inducible promoter *LOX2::LUC* ***(10)*** are in Col-0 (N57953).

Luciferase levels can be detected noninvasively by measuring the luminescence emitted in the presence of luciferin. The two classes of luminescence detection systems are low-light CCD cameras and scintillation counters; this chapter contains a protocol for an example of each, an OrcaII CCD camera and a Packard TopCount Scintillation Counter that uses photomultiplier tubes. During measurement in a TopCount seedlings are grown in 96-well microtiter plates, space and nutrients are limited, and seedlings become stressed younger than in the square plates used with a camera. The light illumination levels are greater and more even in a camera than in a TopCount, where only relatively low and uneven levels of light can be reflected to the TopCount plate. The TopCount is very sensitive and has a huge dynamic range, low signal-to-noise, ratio, and a very high quantum efficiency, making it extremely effective to detect low light levels, such as those from *CAB2::LUC*. The TopCount has a higher throughput and time resolution; the precise details depend on the promoter, machine and light environment. For ~400 seedlings expressing *CAB2::LUC*+ in constant light, the time resolution can be higher in the TopCount (every 45 min) than the OrcaII (every 2 h).

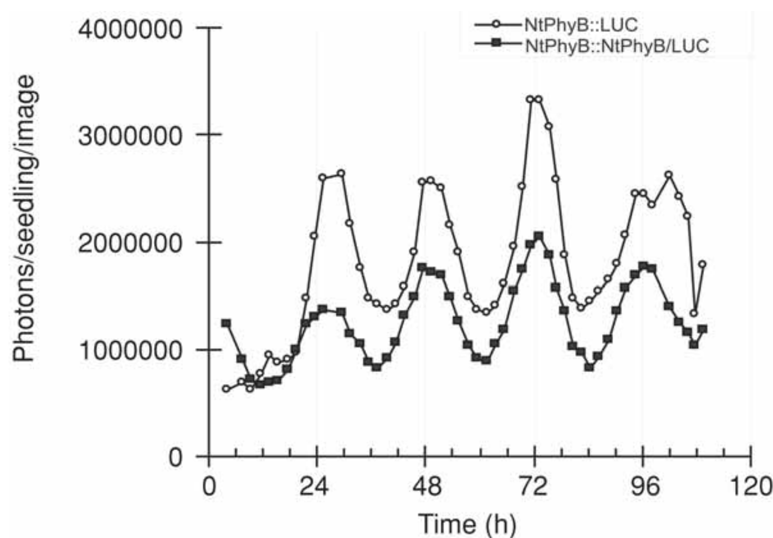


Fig. 1. Luciferase can be fused to the promoter of a gene or fused to a protein. Tobacco plants expressing either the *PHYB* promoter::*LUC* (empty circles) or *PHYB* promoter::*PHYB* protein *LUC* fusion (filled squares) constructs were entrained in light/dark cycles and then assayed under constant condition using a cooled CCD camera.

The challenges of imaging luciferase activity are multiplied when high magnification is required, partly because the aim is often to characterize the spatial distribution of luciferase activity. Useful numerical data can be obtained from images with low signal-to-noise ratios but convincing spatial information typically requires higher-quality images. Particular applications vary enormously, so we suggest guidelines that will be broadly applicable. The solution usually depends on the optics of light capture, rather than the preparation of samples or the camera.

Photographic lenses provide the best results for imaging fields of view down to one or two centimeters. "Macro" lenses tend to have low apertures, so a better approach is to shorten the minimum focal distance of a high-aperture lens. Most simply, the lens can be moved away from the body of the camera. The Perkin Elmer "Night Owl" system has a motorized device to do this under software control. For small adjustments of screw-mounted lenses, this can be done by partially unscrewing the lens from the camera. A safer method is to use extension rings or tubes, which are available for many types of lenses from photographic suppliers; a 10-mm ring was used for **Fig. 2A** and **B**. "Close-up" magnifying lenses are also available, which add only one glass element to the lens but may not maintain image quality across the whole image. Such lenses were used for the close-up images in ref. ***11***. A method with excellent image quality uses two high-aperture objectives, screwed together in a "head to head" configuration with a double-threaded filter ring (usually custom-made). When both lenses are focused at infinity, the tandem lens has a field of view the same size as the camera sensor, as far in front of the tandem lens as the sensor is behind it. Lenses with the same focal length give a one-to-one mapping but a combination of lenses with different focal lengths allows magnification or demagnification by the ratio of the focal lengths.

Fields of view of less than 1 cm could normally be imaged using a dissecting microscope but these often have poor light capture, owing to the large distance from sample to objective. Specialized microscopes for fluorescence imaging are now improving in this area. Our approach has been to use a compound microscope with low-magnification, high-transmission lenses. The key to imaging luciferase activity on the microscope is rigorously to minimize the number of elements in the light path from sample to camera: only an objective is required. The imaging camera should therefore be mounted in line above the objective in an upright microscope, which

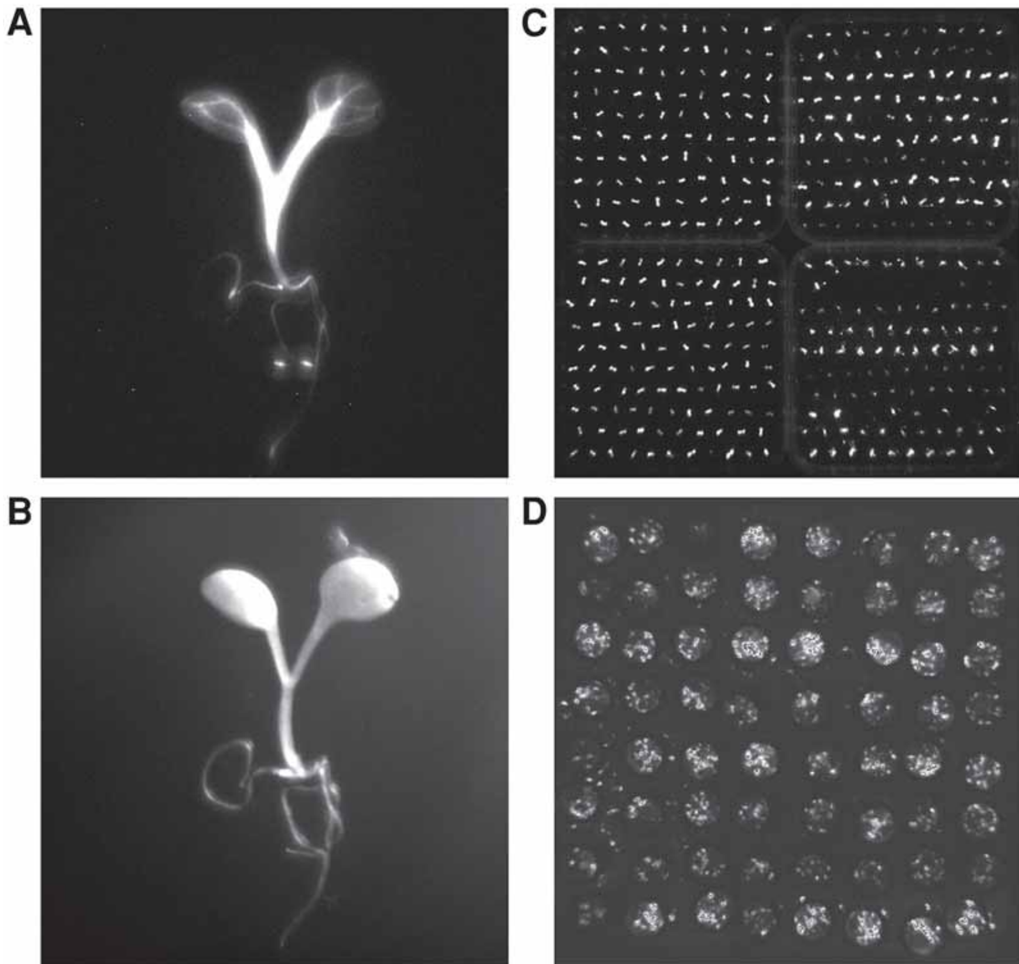


Fig. 2. Luminescence measured using the OrcaII camera. **(A)** One 10-d-old seedling expressing *CCR2::LUC*+. Extension rings held the lens 10 mm from the camera, the plate was ~ 2 cm from the lens, and the exposure time was 20 min. **(B)** Bright-field exposure of the seedling in **A** with an exposure time of 1 s. **(C)** Four 12 cm × 12 cm plates of 8-d old light-grown seedlings expressing LUC+ under the control of the *COLD*, *CIRCADIAN RHYTHM2* promoter (*CCR2::LUC*+) or *CIRCADIAN CLOCK ASSOCIATED1* promoter (*CCA1::LUC*+). Plates were ~1 m from the camera and the exposure time 20 min. Different expression levels from rows of independent transformants are clearly visible in the right-hand plates. **(D)** One 12 × 12 cm plate of 8-d old etiolated seedlings expressing *CCR2::LUC*+ in collars containing ~20 seedlings. The plate was ~ 0.5 m from the camera and the exposure time 20 min.

is the most useful configuration for imaging plants. On an inverted microscope, the camera must be mounted below the microscope body, which requires a bottom port and a table with a hole in it. Objectives should be selected for high light transmission (corresponding to a high numerical aperture) above all other criteria: the Fluar range from Zeiss are ideal (Zeiss, Welwyn Garden City, UK). The mechanical stability of the sample and of the microscope focus are also important, as exposure times may be long (tens of minutes), and eliminating background light may require additional care. Luciferase uses oxygen as a substrate, so it is essential to keep the

samples aerobic: this will not be the case under a sealed cover slip, for example. Subcellular resolution can then be achieved, both in isolated cells or protoplasts (for example, refs. *12* and *13*) and in intact tissue *(8)*.

This chapter describes how to set up an experiment to detect luminescence over a 7-d timecourse in constant light for ~400 8-d old whole seedlings, including analysis for differences in patterns of promoter activity. We cover the setup of OrcaII camera and TopCount systems, experimental setup, and analysis of results.

2. Materials

2.1. Automated Measurement of Luminescence Using CCD Camera

2.1.1. CCD Camera System Set-up

1. Low-light CCD camera, e.g., the electrically cooled OrcaII BT-1024G (Hamamatsu, UK; *see* **Note 1**) with a high-transmission lens (Xenon f/0.95, 25 mm).
2. Dark box, ~ 110 cm × 60 cm × 60 cm.
3. Two LED arrays with power supplies (*see* **Note 2**; MD Electronics, Coventry, UK).
4. Personal computer (PC) to control the camera.
5. MetaMorph® v. 5.0 (Universal Imaging Corporation; West Chester, PA)
6. Main relay switch (MD Electronics, Coventry).
7. MetaMorph Journals from www.amillar.org.

2.1.2. CCD Camera Experiment Setup

1. Sterile flow hood with a low air flow rate; a high air flow rate will cause the plants to wilt and the agar to overdry.
2. 100% ethanol.
3. 20% bleach with 0.001% Tween-20 (*see* **Note 3**).
4. Sterile distilled water.
5. Bacto-agar.
6. Murashige & Skoog Basal Salt Mixture (MS; Sigma-Aldrich, Gillingham, UK).
7. Sucrose.
8. 12 cm × 12 cm square plates.
9. 3M Micropore™ surgical tape (3M United Kingdom, Bracknell, UK).
10. 5 m*M* D-*L*uciferin, potassium salt (Biosynth AG, Switzerland). Make a 50 m*M* stock solution of luciferin in 0.1 *M* Tris-phosphate, pH 8.0. Store aliquots (e.g., 1.5 mL) in the dark at –80°C, once thawed keep at –20°C and use within 4 wk. To prepare the working solution, dilute to 5 m*M* in 0.001% Triton X-100.
11. Filter sterilizers
12. Sterilized pump spray.
13. Microtubes, tips, pipets.

2.1.3. CCD Camera Results Analysis

1. Microsoft Excel (Microsoft).
2. Fast Fourier Transform–Non Linear Least Squares (FFT-NLLS) software available from Dr. Martin Straume, ms3g@virginia.edu.
3. Biological Rhythms Analysis Software System (BRASS) from www.amillar.org.

2.2. Automated Measurement of Luminescence Using a TopCount

2.2.1. TopCount System Setup

1. Scintillation counter, e.g. Packard TopCount 12-Detector 96-Well (Perkin Elmer Life Sciences, UK).
2. Strips of 2 × 10 light-emitting diodes (LEDs) with power supply if luminescence is to be measured from seedlings growing in the light (MD Electronics).
3. Reflector plates, www.scrips.edu/cb/kay/ianda/spacer_plate.htm.

2.2.2. Setting the TopCount Assay

Assay Wizard software (included with TopCount).

2.2.3. TopCount Experiment Setup

1. Sterile flow hood with a flame and a low air flow rate, a high air flow rate; will cause the plants to wilt and the agar to overdry.
2. 100% ethanol.
3. 20% bleach with 0.001% Tween-20 (*see* **Note 3**).
4. Sterile distilled water.
5. Bacto-agar.
6. Murashige & Skoog Basal Salt Mixture (MS) (Sigma-Aldrich, UK).
7. Sucrose.
8. 12 cm × 12 cm square plates.
9. 3M Micropore surgical tape (3M United Kingdom).
10. Black 96-well plates (e.g., Fluorolux HB, DYNEX Technologies Ltd, UK; *see* **Note 4**).
11. Clingfilm.
12. Fine forceps.
13. 5 m*M* D-luciferin, potassium salt (Biosynth AG, Switzerland), storage as described under **Subheading 2.1.2.**
14. Filter sterilizers.
15. TopSeal-A (Perkin Elmer Life Sciences).
16. Sterile hypodermic needle.
17. TopCount Bar Codes (Perkin Elmer Life Sciences).
18. Microtubes, tips, pipets

2.2.4. TopCount RRsults AAalysis

1. A PC and Microsoft Excel (Microsoft).
2. Fast Fourier Transform—Non-Linear Least Squares (FFT-NLLS) software available from Dr. Martin Straume, ms3g@virginia.edu.
3. Biological Rhythms Analysis Software System (BRASS) Excel workbook from www.amillar.org.

3. Methods

3.1. Automated Measurement of Luminescence Using CCD Camera

3.1.1. CCD Camera System Setup

1. The system setup is shown in **Fig. 3**.
2. Set up the camera and dark box in a dark room to remove background light, over which the luciferase cannot be detected. The dark box should have a hole at the top for the camera lens and holes in the sides for LED wires. Mount the camera on top of the dark box and the LED arrays on the sides.
3. Place the plates of seedlings as close to the camera as possible while keeping all the seedlings in the field of view. Four 12 cm × 12 cm plates should be ~1 m from the OrcaII.
4. The controlling computer can be kept in an adjacent room to allow access to the computer while the camera remains in the dark. There needs to be a small hole in the wall between the rooms for wires to pass through, plugged with foam to prevent light leakage.
5. Install MetaMorph on the computer. Install the camera drivers and connect the camera to a parallel port.
6. Connect the LED arrays to the main relay switch, and connect this to the computer via a parallel port.
7. MetaMorph can be programmed to run a timed series of commands, "Journals," to control the connected hardware. Andrew Millar's laboratory has written a series of journals for timeseries of luciferase measurements; these can be downloaded from www.amillar.org. Unless the system is identical they are unlikely to work, but can be used as a template.

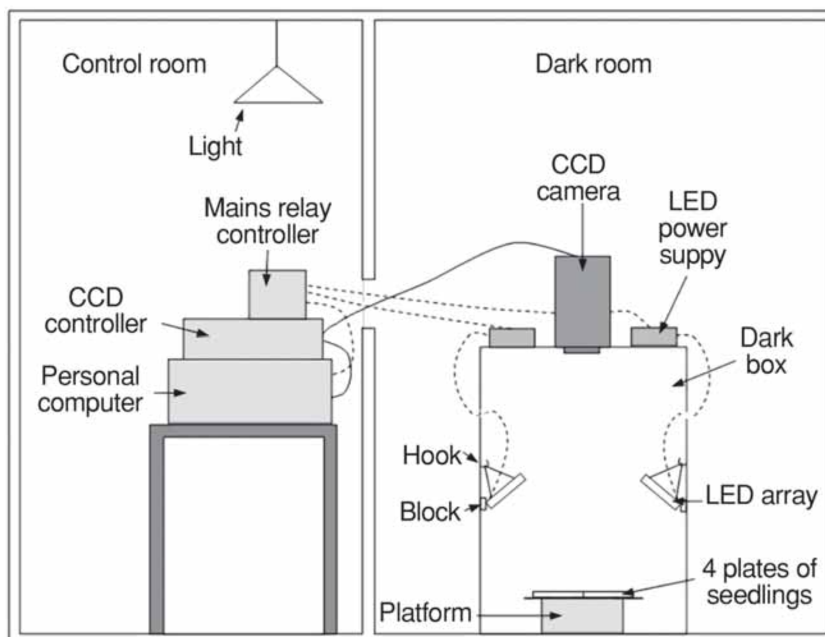


Fig. 3. CCD camera system setup. The camera is kept in a dark room; the controlling computer, CCD controller, and mains relay controller in a light room. Wires to connect them pass through a hole in the wall blocked with foam. The camera rests on a dark box of approx 110 × 60 × 60 cm, the lens drops through a hole in the top. Four 12 × 12 cm plates of seedlings are at the bottom of the box. A platform raises them as close to the camera as possible where they are all in the field of view. LED arrays provide light to the seedlings while the camera is not exposing. A grid of 15 × 12 LEDs is covered with a diffuser and cooled with fans. They are hung from hooks in the dark box and moved from the edge by blocks to position them so that the seedlings are illuminated evenly at 10–18 μM/m^2/s. Wires feed through holes in the dark box to the LED power supplies, and from the power supplies to the mains relay controller in the controller room which is controlled by the computer.

a. Constant light: this runs a timecourse in cycles of 95 min light then 25 min dark. The LEDs are switched on for 95 min then off for 5 min (see Note 5). An image is taken with an exposure time of 20 min (*see* **Note 6**) and 2 × 2 binning (*see* **Note 7**) and saved to the hard drive with a sequential file name, the LEDs are switched back on, and the cycle repeated until stopped.
b. Constant darkness: as constant light but without switching on the LEDs; images are taken every 120 min.
c. For photoperiods: cycles between the required h of "constant light" and the required h of "constant dark" (*see* **Note 8**).

8. The time resolution depends on the brightness of the construct and the timecourse growth conditions. For example, to monitor circadian rhythms of *CAB2::LUC+* in the light, 20 min exposures are taken every 2 h. A higher time resolution can be achieved with shorter exposure times if the plants are bright enough or with shorter or no light intervals.

3.1.2. CCD Camera Experiment Setup

1. Surface-sterilize the seed (*see* **Note 9**). Wash in 100% ethanol. Wash in 20% bleach with Tween-20 (*see* **Note 10**). Wash twice with sterile distilled water. Suspend in 0.1% agar.

2. Autoclave MS, 3% sucrose, 1.5% agar. In a sterile flow hood, pour into four 12 × 12 cm plates to a depth of ~0.5 cm.
3. In a sterile flow hood, sow individual seed onto the agar in grids of ~10 × 10 (**Fig. 2C**; *see* **Note 11**) and seal the lid to the plate with Micropore tape (*see* **Note 12**).
4. Cold treat for 2 to 4 days at 4°C (*see* **Note 13**).
5. Grow seedlings for 7 d e.g., at 12 h, 75 $\mu M/m^{-2}/s^{-1}$ light:12 h dark cycles at 22°C.
6. Spray the plate with filter sterilized 5 m*M* luciferin using a pump spray (*see* **Note 14**) 24 h before the start of the experiment (*see* **Note 15**). Return to the growth environment.
7. At the start of the experiment move the plants to the dark box.
8. Check that the camera is on and cooled (on the CCD controller) and the LEDs are plugged in and switched on. Load MetaMorph and run the constant light journal.

3.1.3. CCD Camera Results Analysis

Steps 1 to **5** use MetaMorph.

1. Build a stack of the individual images. Use "Select Plane" to scroll through the stack.
2. Use the ellipse region tool to drag a loop around each seedling to analyze (*see* **Note 16**). Select an empty region far from the seedlings to monitor the level of background signal (*see* **Note 17**). Save the regions (*see* **Note 18**).
3. Open a "Dynamic Data Exchange" "Data Log" to connect MetaMorph to the spreadsheet package.
4. Select the information MetaMorph will send to the spreadsheet under "Apps"/ "Graph Intensities..."; select the following options: "Measure From"—"Stack"; "Measure Regions Over"—"Time"; "Region Measurement"—"Average Intensity." In "Configure Log," under "Parameter Configuration" ensure only the following are ticked: "Image Name," "Image Plane," "Image Date and Time," "Elapsed Time" and "Integrated" and under "Logging Options" select "Log column titles."
5. Windows titled "Graph" and "Graph Intensities" appear. Select "log data" and "begin." MetaMorph will produce a spreadsheet of the data. The first row contains a title for each column. The other rows are the data for each timepoint. The "Experiment Name" column contains the name of the experiment. "Image Plane" contains the number of the image the data are from. "Image Date and Time" contains the date and time the image was taken in fractions of a day (*see* **Note 19**). Columns titled "ROI 'number' Ellipse (Integrated)" are the data columns. The region of interest (ROI) "number" is the number of the region. The data values are the luminescence value from that region at the time for that row.
6. The data are ready for analysis; **Fig. 4** graphs promoter activity over time. Automated mathematical analysis to determine rhythms in gene expression can be done very simply using FFT-NLLS, which adds a series of sine waves to mathematically define the closest fitting curve to each data trace, and BRASS, which acts as a front end. The type of information wanted will vary depending on the promoter and experiment, we will cover the possibilities, and a comprehensive instruction manual is available to download with BRASS. BRASS contains functions for normalizing the data to the average of the trace; sorting data series into experimental groups, such as genotype; and creating averages. It will run FFT-NLLS and read the results back into Excel; values of period and amplitude can be quickly compared and waveforms are quantified using an implementation of the statistical functions skewness and kurtosis. BRASS can also measure the phase of a peak of expression in each trace by selecting the highest point in an interval or allowing the user to hand-pick the phase of interest. BRASS can produce graphs of traces from individual seedlings or averages of genotypes, statistical analysis of all variables measured, and graphs of the statistics.

3.2. Automated Measurement of Luminescence Using a TopCount

3.2.1. TopCount System Setup

1. The TopCount scintillation counter is a complete system that uses photomultiplier tubes to measure radiation or luminescence from samples in 96-well microtiter plates. The TopCount has two stackers in which up to 20 plates can be stacked; the TopCount moves a plate from the front

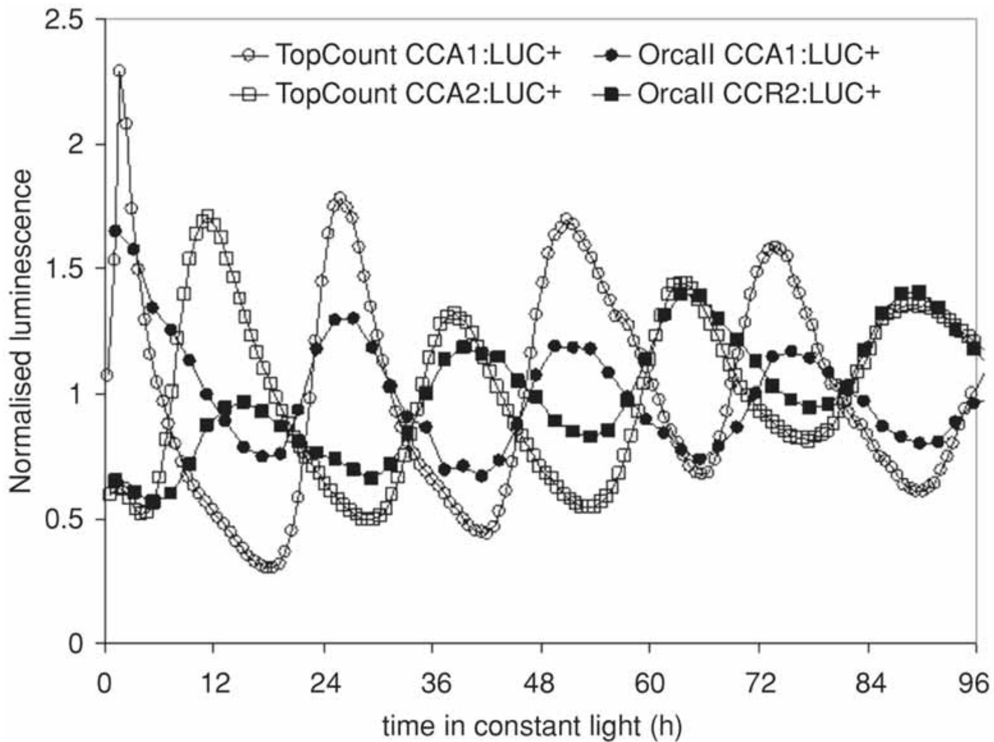


Fig. 4. Luminescence timecourses measured using the OrcaII and TopCount. Luminescence levels from whole *Arabidopsis* seedlings expressing *CCA1::LUC*+ or *CCR2::LUC*+ in 96 h constant light measured using the OrcaII camera or TopCount. Luciferin was applied once, 24 h before the start of the assay. Each trace is the average of 4 to 13 individual seedlings. The data have been normalized by dividing by the average of the data series, to allow comparison between different instruments and reporter genes.

stacker to the machine for reading then to the back stacker. Once all the plates have been read the plates are returned to the front stacker and cycle through again (*see* **Note 20**). The time resolution depends on the number of heads of the machine and the number of plates in the stack.

2. To keep the plants in light during the experiment attach LED arrays or fluorescent bulbs to the sides of stackers (**Fig. 5**). To allow the light to reach the seedlings alternate the experimental plates with reflector plates (clear plates with a reflective sheet that reflects light to the seedlings). This provides 1.5 µM m^{-2} s light to the seedlings. The LEDs do not need to be switched off during readings as the plate is inside the machine. A commercial timer can be used to make a photoperiod.
3. For experiments in a controlled light environment keep the TopCount in a light-tight room, cover the monitor with a green filter (*see* **Note 21**), and block the lights on the TopCount.

3.2.2. Setting the TopCount Assay

The TopCount must be programmed to read the correct plates at the correct times. When doing many similar experiments, set up a template assay to use for each experiment.

1. The TopCount comes with operating software. To program an assay use the "Assay Wizard."
2. The "Assay Type" is "Luminescence" and the "Plate Type" a "96 Well — Microplate."
3. The "Available samples" should contain "BLK" and "TOT", and "Selected samples" "unknowns."

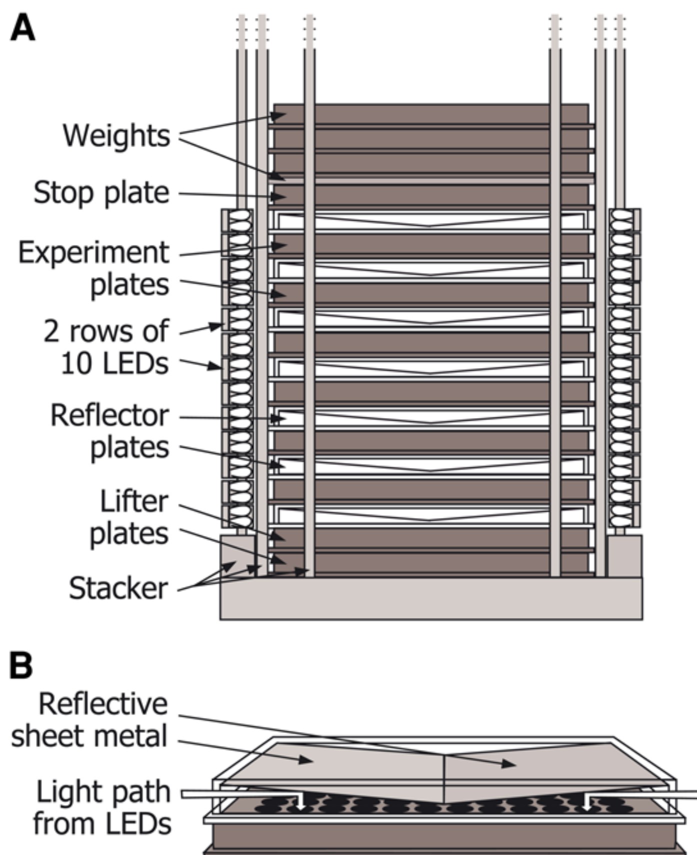


Fig. 5. TopCount modification to allow timecourses in the light. **(A)** TopCount Stackers with LEDs. **(B)** A reflector plate and an experimental plate. Strips of two rows of 10 LEDs are attached to the poles on the side of the stackers. Each plate beneath the stop plate should have a row of LEDs next to each of the shorter sides. The bottoms of the stackers are solid, preventing the use of LEDs to light the bottom plates of a stack. Lifter plates (empty plates) lift the experimental plates to the level of the LEDs. Reflector plates are clear plates with the wells removed and a folded sheet of reflective metal glued to the top. They reflect light from the sides of the stack to the seedlings in the middle of the plates; there must be one below and above each experimental plate as when in the back stacker the plate order is reversed.

4. The "Count Options" depend on the type of assay. A 1 s count of each well gives accurate results. If more seconds are selected, the TopCount counts each well for this time then divides by the number of seconds. In "Advanced Count Options" a delay can be set before each plate is counted to reduce the time resolution. For an assay that continues to run until stopped, the "Number of times to count assay" should be set to 0. The "Number of times to count each plate" and "Number of times to count each well" are the numbers per cycle, usually 1.
5. Generate reports and files only at the end of the assay.
6. The TopCount must be told which samples to count under "Map Samples." To count all wells in all plates select all the wells and under "Edit," "Fill Wells..." with the number of replicates set to 1. To read different wells on different plates, insert plates for each plate there will be and fill the wells to be read. For an assay in light conditions there will be reflector plates and lifter plates in the stack. It will substantially reduce the time resolution if these are read, so insert the number of plates beneath the stop plate and leave the wells in the plates without samples blank.

3.2.3. TopCount Experiment Setup

Steps 2 to **7** should be carried out in a sterile flow hood.

1. As in **Subheading 3.1.2.**, surface-sterilize seeds and plate them on sterile MS, 3% sucrose, 1.5% agar. The layout is less important, as individual seedlings will be picked off the agar. Cold-treat. Grow for 7 d.
2. Autoclave MS, 3% sucrose, 1.5% agar. In a sterile flow hood use a repeater pipet to add 0.3 mL to each well of the black 96-well microtiter plates. When the agar is set, wrap the plates in clingfilm so they do not dry out; they can be stored at 4°C for a few days.
3. To label the plates attach stickers to the front of the plate as the short sides are read by the TopCount.
4. Sterilize the tips of the forceps by dipping into 100% ethanol and flaming.
5. When the forceps are cool gently pick a seedling and replant in a well of the 96-well plate (*see* **Note 22**). Transfer all the seedlings and wrap each plate in clingfilm (*see* **Note 23**). Individual organs can be transferred to assay gene expression in, for example, just the roots or cotyledons
6. Filter sterilize 5 m*M* luciferin and add 15 µL to each well using a multichannel pipet.
7. Cover each plate with a TopSeal (*see* **Note 24**) and pierce above each well with a sterile hypodermic needle (*see* **Note 25**)
8. Place a bar code on the right-hand side of each plate and return to the growth condition for another day (*see* **Note 26**).
9. Place the plates in the front stacker of the TopCount as in **Fig. 5** (*see* **Notes 27** and **28**).
10. Select and start the assay. Throughout the experiment regularly check the TopCount has not jammed or stopped. At the end of the experiment stop the assay (*see* **Note 29**).

3.2.4. TopCount Results Analysis

Steps 1 to **5** use the TopCount; **step 6** uses a PC.

1. To download the data open the “Results Analysis” program, make a new file, and select the assay to analyze.
2. Under “Go To” select “Last Plate.”
3. Under “Define Report” and “ASCII File Output” select the following options: all the sample types, “Repeat Measurement” and “Replicate Average,” “Sample CPS” and “Read Time,” “Matrix (Single Line),” “Excel Import (tab del...,” “Use plate barcode,” “Increment file extension after each plate,” and “Do NOT label columns,” then generate an ASCII report.
4. Use DOS to merge the files for each plate at each timepoint into a single file for each plate by running the command, e.g., for bar code 0123 “copy 0123.* plate0123” in the directory containing the plate files.
5. The TopCount has only a floppy drive so use WinZip® to compress the files to fit on a floppy disk and transfer to a PC for analysis.
6. Import the files into BRASS and analyze as in **Subheading 3.1.3.**, **step 6**.

4. Notes

1. Many suitable camera systems are available, the selection of which can be complex. We recommend on-site demonstrations of a few potential instruments with the biological materials of interest and manufacturer’s local support staff. Important features are a low signal-to-noise ratio, high quantum efficiency, and large dynamic range. The best choice differs depending on the type of experiment; electron bombardment and intensified cameras are very sensitive but can be destroyed by high light levels, making them unstable for automation. In our experience the electrically cooled (–50°C to –90°C) back-thinned 1K CCD cameras are the most robust and versatile cameras currently on the market. Manufacturers of suitable instruments include Roper (www.roperscientific.com), Hamamatsu (http://sales.hamamatsu.com), Photek (www.photek.com), Perkin Elmer (http://las. perkinel mer.com/), and Andor (www.andor-tech.com).
2. Two arrays of 15 × 12 LEDs with a diffuser ~40 cm from the plates can provide 10–18 µ*M*/m^2/ s light to the seedlings.

3. Tween-20 is a detergent that facilitates wetting of the seed and prevents seeds from sticking to each other or floating.
4. The microtiter plates must have solid edges to take a bar code label and be black to avoid signal crosstalk between wells and reduce the autofluorescence associated with white plates; some makes frequently jam the TopCount.
5. Chlorophyll autofluoresces after exposure to light; this would interfere with the luciferase luminescence if the image was taken immediately after the lights were switched off.
6. The exposure time may need to be altered for the promoter::*LUC*+ construct so the signal is detectable and not saturating.
7. The luminescence is measured as the average of 4 pixels to reduce noise.
8. Alternatively use a "constant light" protocol with the LEDs controlled by a commercial timer.
9. The growth media will also support bacteria and fungi; seeds are sterilized to avoid measuring the response of the plant to infection.
10. The seeds may float and need to be spun down in a microfuge.
11. Seed can also be grown on selective media and selected seedlings transferred to fresh plates before imaging. The density of plants is determined by the area they will cover once the experiment is finished, i.e., if seedlings overlap at the end of the experiment this segment cannot be measured from either seedling. Different genotypes and growth conditions will affect this. For wild types grown in 12 h 75 μ*M*/m^2/s light: 12 h dark cycles at 22°C, 100 seedlings arranged in a 10 × 10 grid per plate, as in **Fig. 2C**, will have relatively little overlap at 14 d. Etiolated seedlings grown like this fall over and overlap; to image etiolated seedlings we grow groups of ~20 seedlings in collars to separate different genotypes, as in **Fig. 2D**. To make the collars cut the pointed bottom and cap off microtubes, autoclave, and push (cut side down) into the agar; black collars reduce crosstalk between groups. To increase the plate height use a second plate base as a lid.
12. Micropore tape prevents infection while allowing gas exchange.
13. Cold treatment synchronizes germination; the duration is not critical but if too short germination is uneven, and if too long germination occurs in the refrigerator.
14. Luciferin diffuses into the seedlings. This luciferin should be sufficient for at least 10 d, but levels will decrease over time and cause the appearance of a decrease in expression; to avoid this spray the plate with more luciferin at regular intervals, e.g., twice a day.
15. During 24 h the luciferase built up during the life of the plant loses activity.
16. The luminescence from each region will be measured over the timecourse. The region should include as much of the seedling as possible without any other seedlings. Scroll through the stack to ensure that the regions are fine for all images. The number of regions that can be transferred to an Excel spreadsheet is limited to 253 by the number of columns in Excel.
17. The background level should be subtracted from each experimental level to show the correct amplitude.
18. MetaMorph sometimes shuts down during the next step, losing unsaved regions.
19. To show the time in hours format the cells to a custom format of "[hh]:mm."
20. Plates sometimes fail to drop evenly under gravity alone and jam the machine; this is reduced with motorized stackers.
21. Plants are relatively insensitive to green light; a green filter allows operation of the monitor with minimal effect to the seedlings.
22. Be gentle; it is easy to crush the seedling. A relatively strong part to hold is the collet, the small junction between the hypocotyl and the root. The roots should all be touching or submerged in the agar.
23. This is a time-consuming process with the rate-limiting step of a TopCount screen. It can take a day to transfer seedlings to 10 to 20 plates depending on experience.
24. Seedlings often stick to the TopSeal. If they are to be rescued after the experiment and a slight reduction in luminescence readings is not important, cut 0.5 cm from each side of a TopSeal and stick it (sticky side to sticky side) to the middle of a complete TopSeal and seal the plate.
25. This is done to provide oxygen for the plants and luciferase.
26. An option to reduce jams is to wipe the sides of the plates with WD-40.
27. The stop plate is a plate with two small bar codes on the right hand side; it tells the TopCount that the end of the stack has been reached and to repeat the assay.

28. Weights reduce the chances of plates dropping unevenly and causing a jam.
29. Remember to delete the assays; a full memory is a common cause of crashes as the TopCount stops mid-assay.

Acknowledgments

The imaging systems and techniques were developed at the University of Warwick supported by grants from the Gatsby Charitable Foundation, BBSRC, and the Royal Society to Andrew Millar. We wish to thank Andrew Millar and László Kozma-Bognár for helpful input to the manuscript, Kieron Edwards for critical comments on the manuscript, and Nazir Shariff for maintenance of the imaging systems.

References

1. Millar, A. J., Short, S. R., Chua, N. H., and Kay, S. A. (1992) A novel circadian phenotype based on firefly luciferase expression in transgenic plants. *Plant Cell* **4,** 1075–1087.
2. van Leeuwen, W., Hagendoorn, M. J. M., Ruttink, T., van Poecke, R., van der Plas, L. H. W., and van der Krol, A. R. (2000) The use of the luciferase reporter system for in planta gene expression studies. *Plant Mol. Biol. Rep.* **18,** 143a–143t.
3. Sherf, B. A. and Wood, K. V. (1994) Firefly luciferase engineered for improved genetic reporting. *Promega Notes* **49,** 14–21.
4. Almond, B., Hawkins, E., Stecha, P., et al. (2003) A new luminescence: not your average click beetle. *Promega Notes* **85,** 11–14.
5. Zhuang, Y. O., Butler, B., Hawkins, E., Paguio, A., Orr, L., Wood, M. G., and Wood, K. V. (2001) New synthetic renilla gene and assay system increase expression, reliability and sensitivity. *Promega Notes* **79,** 6–11.
6. Bognar, L. K., Hall, A., Adam, E., Thain, S. C., Nagy, F., and Millar, A. J. (1999) The circadian clock controls the expression pattern of the circadian input photoreceptor, phytochrome B. *Proc. Nat. Acad. Sci. USA* **96,** 14,652–14,657.
7. Hall, A., Kozma-Bognar, L., Bastow, R. M., Nagy, F., and Millar, A. J. (2002) Distinct regulation of CAB and PHYB gene expression by similar circadian clocks. *Plant J.* **32,** 529–537.
8. Hall, A., Kozma-Bognar, L., Toth, R., Nagy, F., and Millar, A. J. (2001) Conditional circadian regulation of PHYTOCHROME A gene expression. *Plant Physiol.* **127,** 1808–1818.
9. Raventos, D., Meier, C., Mattsson, O., Jensen, A. B., and Mundy, J. (2000) Fusion genetic analysis of gibberellin signaling mutants. *Plant J* **22,** 427–438.
10. Jensen, A. B., Raventos, D., and Mundy, J. (2002) Fusion genetic analysis of jasmonate-signalling mutants in *Arabidopsis*. *Plant J.* **29,** 595–606.
11. Toth, R., Kevei, E., Hall, A., Millar, A. J., Nagy, F., and Kozma-Bognar, L. (2001) Circadian clock-regulated expression of phytochrome and cryptochrome genes in *Arabidopsis. Plant Physiol.* **127,** 1607–1616.
12. Gallie, D. R., Lucas, W. J., and Walbot, V. (1989) Visualizing mRNA expression in plant protoplasts: factors influencing efficient mRNA uptake and translation. *Plant Cell* **1,** 301–311.
13. Kost, B., Schnorf, M., Potrykus, I., and Neuhaus, G. (1995) Non-destructive detection of firefly luciferase (LUC) activity in single plant cells using a cooled, slow-scan CCD camera and an optimized assay. *Plant J.* **8,** 155–166.

25

Aequorin as a Reporter Gene

Christoph Plieth

Summary

Reporter proteins allow one to monitor cellular parameters that are involved in signal transduction, development, metabolic processes, and transport. There are targeting strategies available to direct the indicator protein exactly to the locale inside the organism from which information is desired. This circumvents experimental reductionism and allows experimentation with whole intact and undisturbed organisms. The outstanding advantages of self-reporting organisms make it worth to shoulder cost- and time-consuming molecular work. Here, the luminescent Ca^{2+} indicator aequorin is introduced and a rough guideline is given from early planning the molecular work and assembling an experimental setup to experimentation with luminescent *Arabidopsis*, data processing, and control experiments.

Key Words: *Arabidopsis*; aequorin; coelenterazine; calcium; signal transduction; reporter genes; self-reporting plants; protein targeting; calcium indicator; self-reporting organisms; genetically encoded indicator; luminescence; luciferase; reactive oxygen species.

1. Introduction

Genetically encoded markers and indicators such as green fluorescent protein (GFP) and aequorin enable monitoring of subsystems that still reside in the intact organism. There are actually two types of recombinant indicators, namely fluorescent and luminescent proteins. Fluorescent proteins glow when excited by light with a suitable wavelength (*see* Chapter 23). In contrast, luminescent proteins, so-called luciferases, catalyse a redox reaction, by which light is produced (**Fig. 1A**). Some of these luciferases can be used to probe free cellular Ca^{2+} ion concentrations. The best characterized Ca^{2+} indicator protein is the photoprotein aequorin ***(1)***. It has been expressed in a broad range of plant species, as well as in fungi and many other nonanimal organisms (listed in ref. ***2***). Further efforts are being made to develop more indicator proteins for a broader range of relevant cellular parameters that can then be visualized in vivo ***(3,4)***.

Many animals use bioluminescence for different purposes (*see* **Note 1**). Luciferases catalyzing light-generating reactions have evolved in the animal kingdom separately more than 30 times. The different origins make luciferases a very heterogenic group of enzymes, with light generation being the only property they have in common. Coelenterate luciferases, in particular, are oxidoreductases attacking the CH-OH group of an organic electron donor (luciferin) while consuming oxygen (**Fig. 1A**).

Aequorin is a unique light emitting monomeric protein of about 21.4 kDa. It was originally characterized in the luminescent medusa *Aequorea victoria*, a hydrozoan jellyfish of the Pacific, also named *Aequorea aequorea*, *A. forskålea*, or *A. coerulescens* (*see* **Note 1**). Aequorin has two properties that in other luminescent organisms are usually separated: First, it is a luciferase, able to catalyze oxidization of special low-molecular-weight luciferins called coelenterazines (CTZ). Second, it is a luciferin-binding protein and is hence called a

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

A
O_2
CTZ
reconstitution
Apoaequorin
Aequorin
Ca^{2+}
Ca^{2+} binding
discharge
CO_2
CTA
Light
B
I
II
III
Ca^{2+}
1
10
20
40
50
60
80
100
120
150
189
NH_2^+
COO^-
D119A
low affinity aequorin

"photoprotein." A third and most relevant property is the Ca^{2+} dependency of its luciferase activity. CTZ-charged aequorin has three Ca^{2+} binding domains (**Fig. 1B**). When calcium is bound, CTZ is oxidized to coelenteramide (CTA) and the protein undergoes a conformational change accompanied by the release of carbon dioxide and emission of blue light (approx 470 nm). Hence, aequorin bioluminescence is correlated with the actual free Ca^{2+} ion concentration ***(5)***. Other cations, such as Mg^{2+} and K^+, do not trigger luminescence, although these ions may depress Ca^{2+} sensitivity ***(6)***. Aequorin can potentially detect free calcium levels of up to 100 μ*M*, although in practice, most measurements are made in the range of 10 n*M* to 10 μ*M* ***(7)***. This makes aequorin an excellent and widely used free calcium ion concentration [Ca^{2+}] reporter protein ***(8)***.

Many synthetic Ca^{2+} probes ***(9,10)*** can quickly be introduced into plants ***(2,7,11)***. Genetically encoded probes, in contrast, need laborious molecular work for being stably introduced into plants. Ca^{2+} protein indicators can be targeted to many different plant tissues, cell structures, organelles, and compartments by fusion with specific promoters, signal sequences, or by trapped enhancers (***12,13***; targeting strategies listed in ref. ***2***). The experimental approach is to express the genetically encoded [Ca^{2+}] indicator (here: luciferase) specifically at the site(s) of interest (targeting) and monitor [Ca^{2+}] changes by recording the optical signal (here: luminescence).

Aequorin was first cloned about 20 yr ago ***(14,15)***. The cDNA coding for apoaequorin from *Aequorea victoria* is readily accepted by many plants (*see* ref. ***2*** for a table with plant species that have been transformed to express aequorin). There are no cryptic introns detected as was the case for the native GFP, which was also derived from *Aequorea victoria* ***(16)***. Codon usage adjustments of aequorin cDNA are not necessary for good expression. Usually the sequence from Prasher et al. (accession no. M16103; identification no. AVAEQA) is used for expression in *Arabidopsis* (*see* **Note 2**). Aequorin can be fused with other reporter proteins such as GFP ***(13,17–20)***. This eases the location of expressed recombinant aequorin by fluorescence microscopy.

Aequorin has been targeted to many different organelles, compartments, and tissues (listed in ref. ***2***). In this way, many new aspects of Ca^{2+}-mediated signal transduction have been discovered. Different plant cell compartments harbor different free calcium concentrations at rest ranging from some 10 n*M* in the cytoplasm to several hundred μ*M* in the vacuole. It is necessary to choose a recombinant Ca^{2+} probe with an indicator sensitivity that meets the concentration range in which [Ca^{2+}] is to be measured. Wild type aequorin (i.e. apoaequorin reconstituted with normal coelenterazine; CTZ) has a Ca^{2+} binding affinity in the n*M* range and is thus perfect for monitoring cytoplasmic [Ca^{2+}] in the range between 100 n*M* and 10 μ*M*. Sometimes, however, different calcium affinities are desired in order to monitor either tiny [Ca^{2+}] changes (e.g., refs. ***21*** and ***22***) or to measure [Ca^{2+}] in a higher concentration range. There are three ways to choose a luciferase that is able to report [Ca^{2+}] in the desired concentration range:

1. Apart from its native "coelenterate" luciferin (CTZ), apoaequorin accepts more than 30 other CTZ derivatives ***(5,23,24)***, all of which produce Ca^{2+} probes with different Ca^{2+} binding affinities, stabilities, and emission spectra ***(25)***. In this way, semisynthetic aequorins (aequorins made up of apoaequorin coupled with a chemically synthesized analog of coelenterazine; ***5***) can be produced with a very wide range of sensitivities to Ca^{2+}, ranging from approx 0.01 to 200 times ***(7)*** that of natural aequorin (**Fig. 2**; *see* **Note 3**).

Fig. 1. *(opposite page)* The photoprotein aequorin. (**A**) Active aequorin is reconstituted *in vivo* from apoaequorin and a low-molecular-weight luminophore called coelenterazine (CTZ) in the presence of oxygen (O_2). When calcium is bound, CTZ is oxidized to coelenteramide (CTA) and the protein undergoes a conformational change accompanied by the release of carbon dioxide (CO_2) and emission of blue (469 nm) light. ***(2)***. (**B**) Primary structure ("snake" diagram) of the coelenterate luciferase aequorin ***(100)***. Aequorin has three Ca^{2+} binding domains, so-called "EF-hands." The luciferase is active only when all three are occupied by a calcium ion. The D119A mutation (green) gives an aequorin with lower Ca^{2+} binding affinity ***(34)***. The amino end proline (P189) residue is essential for luciferase activity ***(64)***, whereas the first nine base triplets coding for the carboxy end residues (gray) are dispensable.

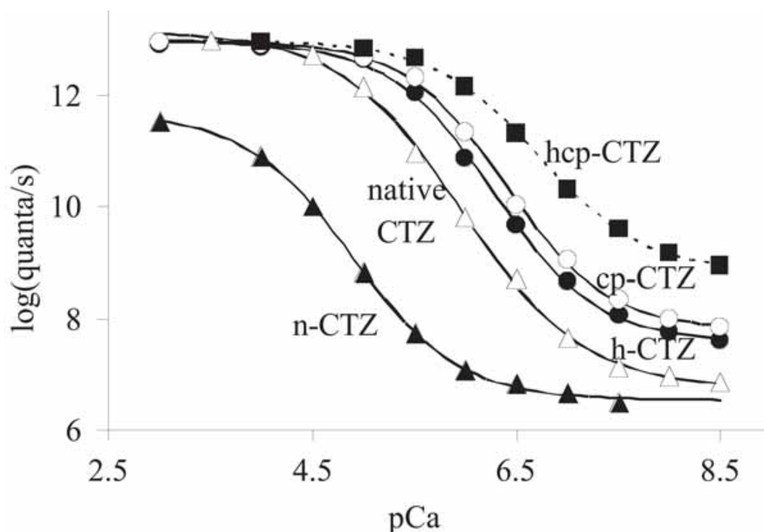


Fig. 2. Ca^{2+} dependence of semisynthetic aequorins. Relationship between $[Ca^{2+}]$ and luminescence (data taken from ref. *23* and fitted with sigmoidal Bolzmann fit) of semisynthetic native aequorin and variants reconstituted with other CTZ derivatives.

2. Many different isoforms of apoaequorin exist naturally; their sensitivities to calcium vary in such way that the most sensitive is about 10 times more sensitive than the least *(24)*. Further, there are more and more Ca^{2+}-dependent luciferases other than aequorin being characterized *(8,26)*. For instance, obelin, mitrocomin, and clytin have already been cloned *(27–31)* and found to have sequence homologies with aequorin. These new Ca^{2+} probes have $[Ca^{2+}]$ binding and spectral emission properties different from aequorin *(32)*. They also accept many different synthetic CTZ derivatives as luciferin and will probably become more important in the future for studying Ca^{2+} signaling in plants.
3. The Ca^{2+} indicator apoprotein can be mutated in order to develop new properties *(33)*. Kendall et al. *(34)*, for instance, exchanged an amino acid residue (D119A; **Fig. 1B**) in an EF-hand domain of aequorin and this way produced a variant with lower calcium binding affinity. This type of aequorin—reconstituted with normal CTZ—is used for making measurements in cellular locations where Ca^{2+} levels are expected to be orders of magnitude higher than in the cytosol—e.g., in the mitochondria *(19)*, in the nucleus *(35)*, and in the apoplast *(2,20)*. It is likely that other mutants will be produced and characterized in due course that have even lower or higher affinities than wild-type aequorin.

Depending on the desired kind of data there are two ways to obtain information about Ca^{2+} signaling from whole plants expressing aequorin: luminometry and luminescence imaging. The first can be performed with a simple luminometer setup (**Fig. 3**; *36*). The latter requires a more expensive camera setup *(37,38)*. Luminometry provides high sensitivity and fast (subsecond) time resolution. So, if kinetic analysis of time series is desired (i.e., good temporal resolution) then luminometry is the best way to go (*see* **Note 4**). However, if spatial characteristics of propagating Ca^{2+} waves and visualization of communication between cells and tissues is desired (i.e., good spatial resolution) then a sensitive camera is needed. Whatever kind of setup (camera or luminometer) has been chosen, it is good to keep in mind that spatial resolution is always obtained at the price of time resolution and vice versa (*see* **Note 5**). An important requirement is a light-tight chamber or cabinet that allows one to hold the specimen and to precisely change environmental conditions accordingly to the problem studied (**Figs. 4, 5**).

The use of aequorin in higher plants has already been reviewed several of times and many protocols are available on how to handle and use this Ca^{2+} probe in genetically engineered

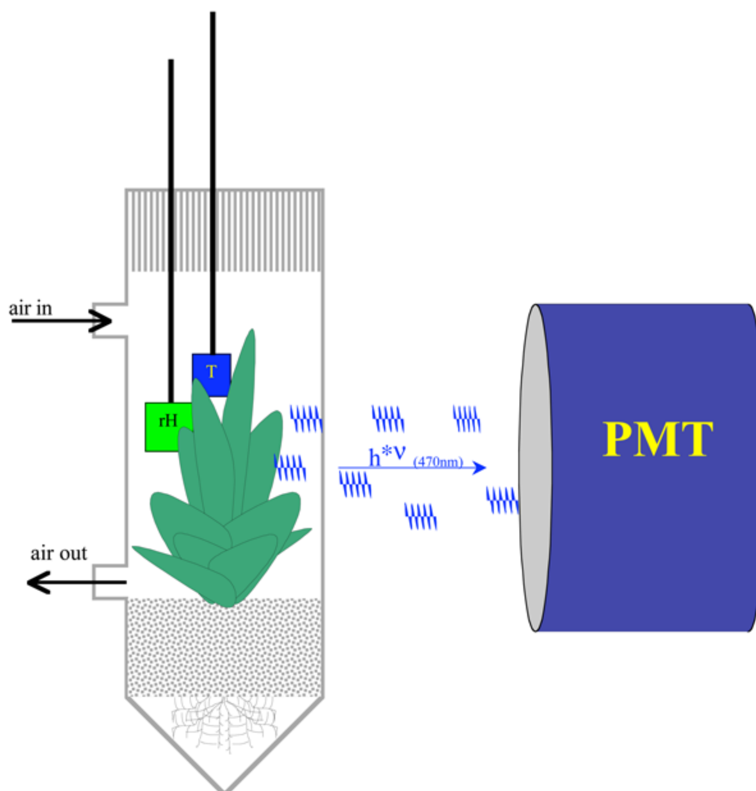


Fig. 3. A simple setup. Scheme of a simple experimental setup to study $[Ca^{2+}]_{cyt}$ in transgenic reporter plants under influence of different changing environmental factors. A transgenic *Arabidopsis* expressing aequorin and reconstituted in vivo with CTZ is encapsulated in a cuvet, or a cabinet which allows precise control of environmental factors. The plant gives off light by which in vivo free $[Ca^{2+}]$ is monitored. PMT = Photomultiplier tube, T = Thermosensor; rH = relative humidity sensor.

plants (*2,7,39–41*). This chapter is intended to update aequorin-based experimental procedures for studying Ca^{2+} signaling in *Arabidopsis* and to span a coarse guideline from early planning of experiments to in vivo luminescence measurements and data processing.

2. Materials

For molecular work a fully equipped laboratory (polymerase chain reaction thermocycler, incubators/shakers, centrifuges, vortexer, spectrometer, autoclave, laminar flow hood, fume hood, refrigerators, freezers, growth facilities, etc.) is necessary that complies with all national safety rules and regulations for handling and containment of genetically modified organisms. The standard consumables for molecular work (e.g., chemicals, DNA-modifying enzymes, disposable plastic ware, etc.) are not listed here.

2.1. Buffers, Reagents, and Molecular Tools

1. cDNAs for aequorin or obelin from Molecular Probes or NanoLight Ltd.
2. Bacterial expression vectors from Invitrogen, Qiagen, or Clontech.
3. Empty plant expression vectors like pGreen (www.pgreen.ac.uk), pART7, pDH51, or GFP containing plant cloning vectors (*42,43*).

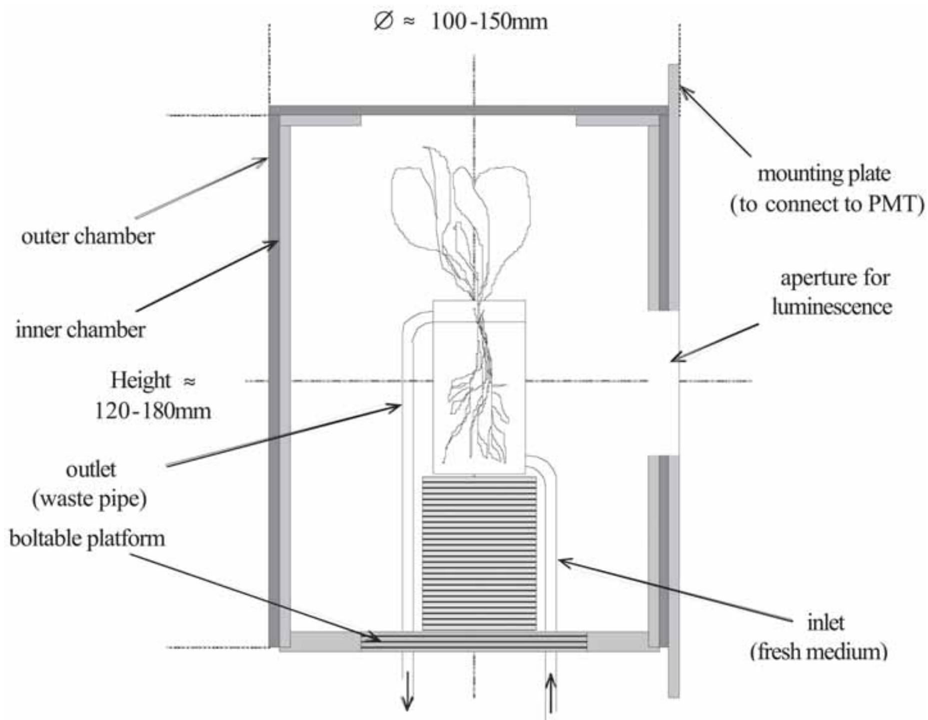


Fig. 4. Sample housing for experiments with *Arabidopsis.* Crosscut side view of an experimental set up to monitor changes of $[Ca^{2+}]_{cyt}$ by aequorin luminescence in the root system of an *Arabidopsis* plant while the nutrition medium is exchanged.

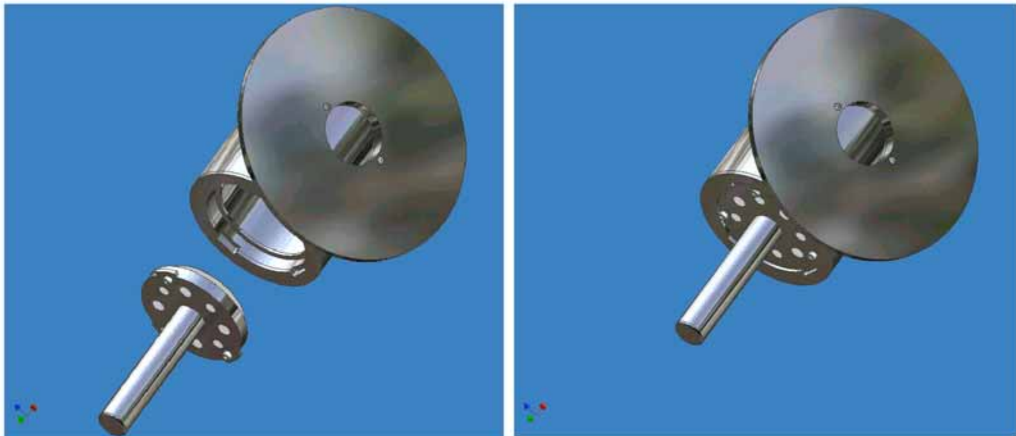


Fig. 5. Light-tight sample housing. Example for the design of a light-tight sample housing *(51)*. The housing is made of two aluminium cylinders, both having a window that fits in size the entry window of the PMT. The inner one can be turned and in this way deblock the optical path from the specimen to the PMT only when the sample holder is in place and the housing is closed. This setup allows one to mount plants in close proximity to a PMT, to precisely adjust environmental factors such as temperature, relative humidity, atmospheric composition, and composition of the nutrient solution and to make alterations to give stimuli and apply effectors to leaves or the root system.

4. Empty binary plant expression vectors such as pGreen, pART27, pBI121 ***(42)***.
5. *Escherichia coli* strains for protein expression (e.g., BL21(pLysS), Clontech; SG13009(pREP4), Qiagen).
6. NT buffer: 500 m*M* NaCl, 0.5 m*M* EGTA, 10 m*M* Tris-HCl, pH 7.5.
7. NET-buffer (NT-buffer with higher EGTA): 500 m*M* NaCl, 5 m*M* EGTA, 10 m*M* Tris-NaOH, pH 7.5.
8. MO-medium (NET with reducing agent): 500 m*M* NaCl, 10 m*M* Tris-NaOH, pH 7.5, 5 m*M* EGTA, 5 m*M* DTT or β-mercaptoethanol (freshly prepared).
9. 50 m*M* $CaCl_2$.
10. $CaCl_2$-ethanol solution: 1 *M* $CaCl_2$ in 10% (v/v) ethanol.
11. Coelenterazine(s) (NanoLight Technologies, Biosynth, Biotium, Molecular Probes, PJK). Choose the appropriate CTZ moiety that gives an aequorin to record $[Ca^{2+}]$ in the expected concentration range (*see* **Note 3**). All the following steps should be conducted under dim light.
    a. For aliquoting dissolve 1 mg (i.e., 2360 nmol) native CTZ (MW = 423.5) in, e.g., 1180 µL ice-cold (–20°C) methanol (never use DMSO; *see* **Note 6**).
    b. Keep tube with methanolic CTZ solution (e.g., 2 nmol/µL) on –20°C cold metal block to minimize evaporation.
    c. Transfer portions of 5, 12.5, 25, or 50 µL in 1.5-mL microcentrifuge tubes by using a precision microsyringe and thereby aliquot the purchased lot into 10, 25, 50, or 100 nmol portions, respectively.
    d. Put open caps in a speed-vac and vacuum-dry down at RT in the dark.
    e. Close caps with dried CTZ and store aliquots in a freezer.

2.2. Dedicated Equipment and Consumables

The light yield from aequorin-expressing plants is low; therefore an extremely sensitive light detector device is needed ***(44–46)*** (*see* **Note 7**). The spectral sensitivity of the detector must be in accordance with the spectral properties of the bioluminescent material and photon counting mode should be used in order to maximize the signal-to-noise ratio ***(47,48)***.

1. Luminometer (Berthold; EMI Electron Tubes; PJK).
2. Luminescence camera and imaging system (Photek; Hamamatsu; TILL; PJK).
3. Custom tailored cuvet holder for luminometry assay (Berthold; or lab-made, e.g., **Fig. 6**, **7**)
4. Light-tight chamber/box for experiments, custom-tailored and lab-made, e.g., **Fig. 4**, **5**.
5. Dark box for imaging (Photek; can also be lab-made).
6. Fluorescence microscope and GFP filter set.
7. Vacuum speed dryer.
8. 3.5-mL luminometer tubes or cuvet, e.g., Sarstedt #55.484.
9. 50-µL or 100-µL precision microsyringe (Hamilton).
10. 1-mL (insulin) syringe.
11. 0.8 × 100 mm spinal injection needle (Unimed).

3. Methods

3.1. Assembling DNA and Transformation of Arabidopsis

The experimental design may demand targeting of the Ca^{2+} indicator to a specific compartment in the transgenic plant (*see* **Note 8**). It is advisable to early choose a procedure to verify successful targeting in the plant (e.g., by GFP or GUS fusions, or by immunolocalization; *see* **Note 9**). An appropriate Ca^{2+}-dependent luciferase should be able to report $[Ca^{2+}]$ in the concentration range estimated at the compartment where the indicator is planned to report from (*see* **Note 3**). An estimation should also be made of the kinetics and/or spatial characteristics of the expected $[Ca^{2+}]$ signal and the time window needed for $[Ca^{2+}]$ kinetics to be studied. In this way, an appropriate experimental setup (i.e., with highest light sensitivity and optimal temporal and/or spatial resolution; *see* **Notes 4** and **5**) can be designed. For assembling/cloning the cDNA cassettes only basic molecular techniques as described ***(49,50)*** are needed.

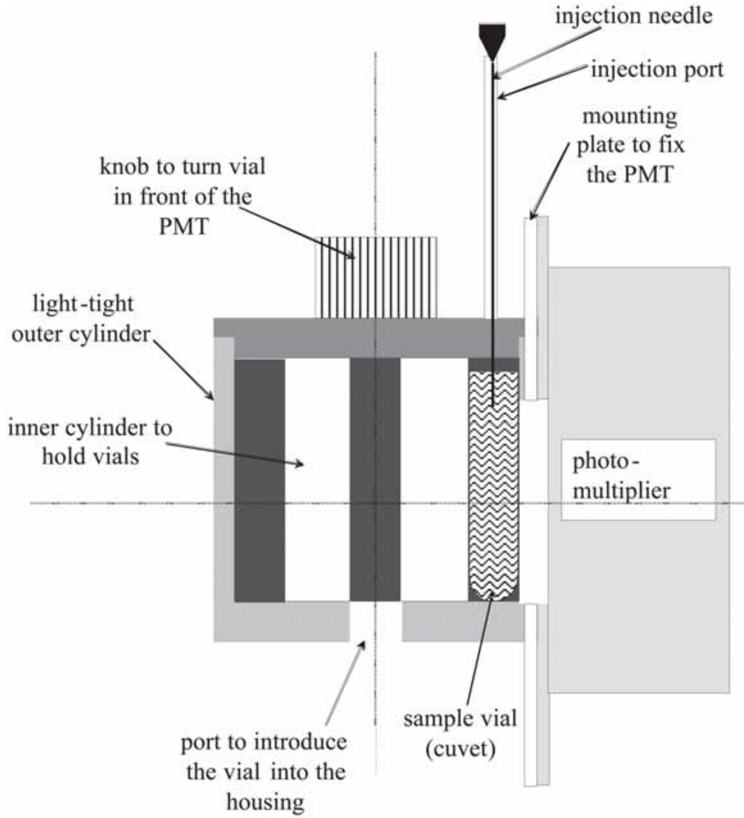


Fig. 6. Light-tight housing to hold in vitro cuvet in front of a luminometer. Crosscut side view of a simple revolver housing for holding vials with luminescent specimen in front of a PMT for in vitro assays.

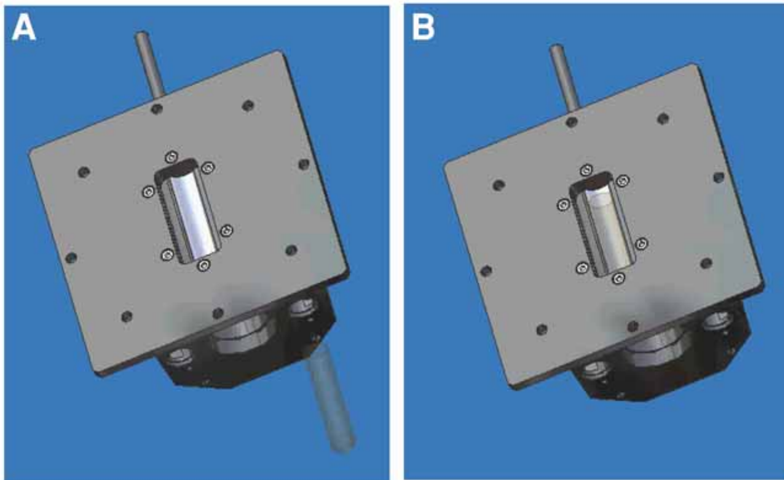


Fig. 7. Light-tight cuvet housing. Aspect of a simple revolver housing for holding vials with luminescent specimen in front of a PMT for in vitro assays. **(A)** The vial is introduced from the bottom into the revolver. **(B)** By turning the inner revolver cylinder, the vial is exposed to the PMT and light produced by the sample can be recorded.

1. Assemble your cDNA cassette (e.g., promoter, reporter gene, Ca^{2+}-indicator gene, targeting signal peptides) in appropriate plasmid(s) for plant and bacterial expression.
2. Express the reporter/indicator fusion in bacteria (*see* **Note 10**). Extract and purify the recombinant protein (*see* **Note 11**).
3. Run in vitro tests with extracted/purified protein (**Subheading 3.2.1.**; *see* **Note 12**).
4. Use the plant vector for transient transformation of plant tissue (*see* **Note 13**). Depending on the fused reporter gene, check plant expression by light or fluorescence microscopy.
5. Also check aequorin expression by luminometry assay of homogenized tissue (**Subheading 3.2.**; ***40***).
6. If transient transformation is successful, transfer the cDNA cassette from the plant expression cloning vector to a binary vector. Use the binary vector for transient expression as in **step 4**. Transfer the binary vector into an appropriate disarmed strain of *Agrobacterium tumefaciens*. Use the agrobacteria for stable transformation of *Arabidopsis* (*see* Chapter 17 in this book).
7. Screen and select adequate transgenic lines. Inspect leaves of several-week-old plants for gene expression by microscopy and luminescence assay, as has been done with transient transformation (*see* **Note 14**).

3.2. Screening of Aequorin Transgenic Lines

Once successfully targeted and stably expressed in *Arabidopsis*, experimentation with aequorin becomes easy. The usual way is to mount the plant in front of the luminescence detector (PMT or camera) and to adjust defined environmental conditions. Changes of environmental parameters are then produced in an exactly defined manner by using a dark cabinet or chamber tailored according to requirements (***51***; **Figs. 4** and **5**), and the $[Ca^{2+}]$ response in the plant is monitored by the luminescence signal coming from recombinant aequorin. Thereby all relevant environmental parameters are monitored in parallel (e.g., ***52–54***). Any experimental strategy with a newly produced Ca^{2+}-reporter *Arabidopsis* line should start with well-explored stimuli ***(1,55,56)***. When setting new kinds of stimuli or conditions, hard treatment should be applied before doing experiments that are physiologically more relevant (*see* **Note 15**).

3.2.1. Identify Aequorin Overexpressing Lines: In vitro Reconstitution of Aequorin and Luminometry Assay

The following protocol is adapted from Knight et al. ***(40)***, and Fricker et al. ***(7)***. Starting from **step 6**, this protocol can also be used to check aequorin expression in bacteria.

1. Grow transgenic plants on vertical plates to the desired age and size under appropriate conditions (*see* Chapter 1 in this book).
2. From each transgenic line to be tested, harvest a few seedlings or an equivalent amount of tissue/cells (approx 5 to 50 mg fresh weight) and transfer into a 1.5-mL microcentrifuge tube. Do not forget to also prepare samples for negative and positive controls (*see* **Note 15**). Snap-freeze the tubes in liquid nitrogen.
3. Remove tubes one at a time from liquid N_2, open with caution, and grind tissue with a precooled micropestle designed to fit the tube. Use a clean pestle for each sample.
4. Add 0.1 mL ice-cold MO medium and continue grinding. Add a further 400 µL MO, mix well, and leave on ice until all samples have been processed (**Steps 2** to **4**).
5. Spin down the debris in a microcentrifuge at maximum speed for 2 min.
6. Perform this and the following steps under dim light: Prepare a solution of 10 µ*M* CTZ in MO buffer and aliquot into 100-µL portions in luminometer tubes.
7. To each tube add 100 µL of the supernatant from step 5 or of suspension from **step 4**, mix well (vortex 2(s) and let stand in the dark at RT for 3 hr.
8. Add 0.5 mL NT buffer and place the luminometer tube in the cuvet holder of the luminometer (**Figs. 6** and **7**).
9. Take up 0.5 mL of 50 m*M* $CaCl_2$ in a 1-mL syringe with a long spinal needle and draw up a little air to avoid leakage from the syringe. Wipe the needle with a paper towel and insert the syringe needle into the injection port. Prevent entry of ambient light through the port.
10. Set the count period to 1 s and record sample background for approx 24 s.

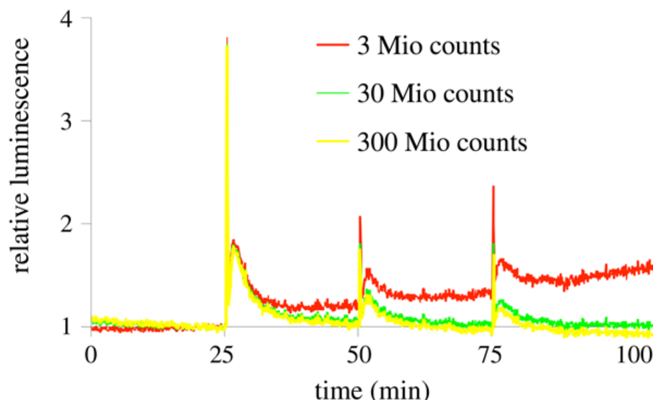


Fig. 8. Estimation of remaining aequorin. Baseline fit of an experiment to roughly estimate the amount of active aequorin still residing in the specimen after the experiment. A baseline is recorded before the experiment (prebaseline $t = 0$ to 25 min) and stimuli are set as planned ($t = 25$, 50, and 75 min). Finally, a baseline recording is performed after the experiment (postbaseline $t = 75$ to 100 min). It has been experienced from previous experiments that [Ca^{2+}] settles after stimuli within 25 min. This means that the pre-baseline and post-baseline should be at same height. An estimation of 3 Mio counts discharge gives an increasing post-baseline (red line). Hence, this estimation is too low. Estimated 300 Mio counts discharge give a post-baseline going below the pre-baseline level (yellow line). This estimation is too high. The best estimation that brings the levels before and after the experiment to same level is 30 Mio counts (green line). (*L* has been calculated as in **Subheading 3.3.** and 10,000 *L* has been plotted as relative luminescence).

11. While recording inject the $CaCl_2$ solution and count photons for 1 to 2 min until reading is down to less than two times the background.

3.2.2. Use of Aequorin-Overexpressing Lines: In vivo Reconstitution of Aequorin and Monitoring Ca^{2+} Signaling

1. Estimate the minimum volume you need for reconstitution. Usually 5 mL are enough for 1 g fresh weight of plant material.
2. For 5 mL take a 50-nmol CTZ aliquot from the freezer and redissolve in 50-µL methanol or propylene glycol. Dispense this solution in 5 mL water or appropriate aqueous buffer (*see* **Note 6**) to have a final concentration of 10 µ*M* CTZ.
3. Incubate plant material herein for a few hours or overnight in the dark (*see* **Note 16**). Let plants rest before further use (*see* **Note 15**).
4. Fix plant in an appropriate transparent vessel (e.g., perifusion cuvet or gas exchange chamber (**Fig. 4**).
5. Fix vessel in the light tight sample housing (**Fig. 5**), dark chamber, or cabinet in front of the PMT or CCD camera (*see* **Note 17**)
6. Adjust conditions that fit the requirements of your experiment. Let plants recover from transfer and let luminescence settle to a constant low level.
7. Start photon counting or image grabbing. Let the experiment run until a stable baseline is reached before giving a stimulus or changing any condition.
8. Perform experiment and set stimuli as planned. Continue photon counting for a reasonable while after the experiment (again: baseline monitoring).
9. Discharge the specimen by soaking in a $CaCl_2$/ethanol solution (*see* **Note 18**). Continue photon counting after discharge until signal is well below baseline level (*see* **Note 19**; **Fig. 8**).
10. Do not forget to perform control experiments with wild-type plants also incubated in CTZ. This is essential to verify that Ca^{2+}-triggered aequorin bioluminescence has been recorded during the experiment, and not ROS-triggered CTZ chemiluminescence (*see* **Note 15**).

3.3. Converting Light into Calcium Concentration

Unlike fluorescent indicators, the amount of aequorin decays during the experiment. Aequorin molecules are consumed in the course of their light-yielding reaction and cannot report Ca^{2+} binding twice. This is due to the irreversible oxidation of the luminophor (coelenterazine → coelenteramide + CO_2; **Figs. 2, 12**) and the lack of free coelenterazine during the experiment. Nevertheless, this feature is not as serious as commonly supposed ***(27,57,58)*** and presents the opportunity of calculating the [Ca^{2+}] from relative luminescence yield (*see* **Note 20**) by discharging remaining aequorin at the end of the experiment and integrating the light ***(40,59***; *see* **Note 18**).

Several procedures have been published to convert photon counts into [Ca^{2+}] ***(60–63)***. They all give rough approximations rather exact figures. In most cases it is sufficient to present data as relative luminescence yield, *L* (e.g., **Fig. 8**). This still enables one to compare magnitudes and kinetics of [Ca^{2+}] from different experiments.

1. Record the luminescence during the experiment. At the end of each experiment discharge the specimen, if appropriate, as indicated above.
2. Measure average background luminescence from a corresponding wild-type line. Subtract the obtained background value from all the experimental data points to have the aequorin derived counts per time interval left (c) (*see* **Note 21**).
3. For all experimental data points calculate the remaining counts, *R*. These are the sum of background corrected counts recorded after this timepoint, including discharge period.
4. For all datapoints calculate the luminescence yield ($L = c/R$).
5. Calculate pCa from L for all data points using calibration-appropriate parameters (*see* **Note 22**). Calculate [Ca^{2+}] from pCa for all data points (*see* **Note 23**).

4. Notes

1. There are many Web sites giving plenty of information about bioluminescence in general and luminescent organisms in particular, such as:

   http://archive.uwcm.ac.uk/uwcm/mb/ISCCG/living.html
   www.wsg.washington.edu/story/storyarchives/blinks.html
   http://lifesci.ucsb.edu/~biolum/
   www.mblab.gla.ac.uk/tubules/aequorin.html
   www.biochemtech.uni-halle.de/PPS2/projects/jonda/intro.htm
   http://sdb.bio.purdue.edu/dbcinema/jaffe/jaffe.html
   http://golgi.harvard.edu/hastings/images/bioluminiscence.html
   http://faculty.washington.edu/cemills/aequorea.html
   http://aesop.rutgers.edu/~crebb/bioluminiscentimages.html.
2. The first seven amino acids of the gene bank sequence (i.e., MTSEQYS; **Fig. 1**, gray aas) are not found with the mature protein in the jellyfish and can be omitted. The following two (i.e., VK) have also no impact on luciferase activity, Ca^{2+} sensitivity, or quantity and quality of light emission. The proline residue P189 at the carboxy end of the protein, however, is of great importance ***(64)***. This also means that proteins should preferably be fused to the N-terminus of aequorin, as fusions to the C-terminus have been found to reduce light production of the enzyme.
3. For instance, h-CTZ, which gives a very Ca^{2+}-sensitive aequorin (h-aequorin; **Fig. 2**), has been used to demonstrate very small changes in $[Ca^{2+}]_{cyt}$ which occur during the wound response in plants ***(38)***. Also, cp-CTZ massively increases the Ca^{2+}-sensitivity of aequorin. By using cp-aequorin it was even possible to demonstrate that $[Ca^{2+}]_{cyt}$ is also altered during the gravitropic response in *Arabidopsis* ***(22)*** and blue light perception ***(www.amillar.org)***. Other CTZ derivatives form an aequorin with lower affinity (e.g., n-CTZ; ***65***). Most semisynthetic aequorins have significantly lower stability, a marked reduction in half-life, and higher Ca^{2+}-independent luminescence yield compared with natural aequorins (**Fig. 2**).

   The synthesis of CTZ is time-consuming and laborious ***(66)***. Hence, some "exotic" CTZ derivatives are expensive, if at all commercially available. Several companies offer different CTZ derivatives apart from native CTZ (e.g., Molecular Probes, www.probes.com/servlets/

datatable?id=32122&item2944; NanoLight Technologies, www.nanolight.com/nanofuel.htm; BioSynth, www.biosynth.com/asp/index_produkte2.asp; Biotium, www.biotium.com/; PJK, www.pjk-gmbh.com/). Prices sometimes differ by more than an order of magnitude (e.g., 1 mg native CTZ is offered on the market in the range between 65 † and 2200 † with no relevant differences in quality or purity). So, it is advisable to look for a cost-efficient supplier or manufacturer.

4. Some sort of spatial resolution can be obtained with a simple luminometer setup by shielding off parts of the specimen and leaving only those unmasked from which light is desired ***(67)***. For example, when using a simple luminometer setup (**Fig. 3**) and information is required only from the roots, the green tissue can be masked in order to get only light emitted from the organ of interest. Widely used PMTs for aequorin luminescence measurements are of bialkali type, as their quantum efficiency is greater in the blue (e.g., Electron Tubes, UK; #9829A). Targeting the indicator to desired tissues and comparison of Ca^{2+} signals from differently targeted aequorin lines also provides spatial information with luminometry.
5. Optimal temporal resolution can be obtained with an imaging device by adjusting and counterbalancing the integration time with an optimum of signal-to-noise ratio. Imaging provides excellent spatial information, however, at the price of an expensive camera setup. Series of time-lapse images of luminescent plants can be found in refs. ***37*** and ***68***.
6. Dry CTZ is quite stable and must not necessarily be stored under oxygen free atmosphere, but it is advantageous to keep it cold (–20°C) and in the dark. However, when dissolved in aequeous solution, CTZ is very sensitive to light and oxidizing environment. CTZ dissolved in aqueous medium is quickly oxidized by ambient oxygen and decays with a half-life between 2 and 4 h at RT in the dark. Preparation of CTZ-containing solutions is best performed under dim light. CTZ in aequeous solution may be stored at –20°C and reused the next day, but it does not stand many freeze-thaw cycles. The better way is to always prepare freshly CTZ solutions, just as much as needed for the day or for one experiment.
7. Useful addresses for sensitive light detector devices are:

   www.electron-tubes.co.uk/pmts/pmtchem.html
   www.sciencewares.com/ipd/ipdbroch1.htm
   www.photek.com/
   www.hamamatsu.com/
   www.visatron.de/
   www.berthold-ds.com/

8. The activity of aequorin might suffer if the C-terminus is altered ***(64)*** or if other proteins are fused to this end ***(7)***. For some strategies a targeting sequence consisting of either a signal peptide leader sequence or a whole polypeptide encoding a protein that exists naturally in the chosen locale is used. Thereby in some cases the addition of a C-terminal peptide is inevitable. The C-terminal “K/H D E L/F” is necessary as retention sequence for targeting to the ER ***(69,70)***; “S R/K L,” the C-terminal signal PTS1, is needed for peroxisomal targeting ***(71–73)***, and different C-terminal sorting determinants are used for vacuolar targeting ***(74–76)***. Hence, it is always useful to express the fusion protein in bacteria and to verify full functionality in vitro (**Subheading 3.1.**) before introducing it into plants.
9. An important point to consider is the way to verify proper targeting in the plant. Immunocytochemical techniques can be used to evidence gene product location ***(35)***. Another option is to express reporter genes such as GFP or GUS in fusion with the Ca^{2+} indicator (aequorin preferably fused to the C-terminus of the reporter; *see* **Notes 2** and **8**). The cloning effort at the outset of the practical work is paid back by an easy detection of the indicator protein by fluorescence or light microscopy, and by a rough estimation of the expression level. Also here, the functionality of the Ca^{2+} indicator can be changed when a reporter is fused; this necessitates inspection of the biophysical properties in vitro after bacterial expression ***(77)***.
10. Bacterial expression of luminescent or fluorescent proteins with either pET-vectors (e.g., pRSET(a,b,c); Invitrogen) in *E. coli* BL21(DE3) or vectors such as pQE(30,31,32) (Qiagen; Hilden) in SG13009(pREP4) cells requires good oxygen supply during induction. Sometimes difficulties have been encountered, but these can be circumvented ***(78,79)***. Both systems (pRSET, pQE) append His-tags to the protein and thus allow purification via Ni-column and gelfiltration.

11. Purified aequorin has to be handled with care and requires mild reducing conditions (e.g., DTT or β-mercaptoethanol), high salt condition (i.e., 500 m*M* NaCl), and zero Ca^{2+} (e.g., EGTA) when reconstituted ***(80)*** (**Subheading 3.2.1.**). If cysteine residues are replaced by serine ***(81)***, no reducing agents are required.
12. For in vitro assays (**Subheading 3.1.**) a simple housing for holding the test tube is useful (**Figs. 6**, **7**).
13. Transient expression can be achieved by the biolistic technique (*see* Chapter 19) or by PEG transformation of isolated protoplasts (*see* Chapter 20).
14. Transgenic seedlings showing antibiotic resistance do not necessarily express the inserted gene; therefore, careful screening is essential. When aequorin has been expressed, the plant material should be assayed for aequorin activity as described in **Subheading 3.2.** When aequorin has been expressed in fusion with GFP then fluorescence microscopy will also give information about expression level. However, screening for GFP fluorescence is less sensitive because with low expression level it becomes hard to distinguish genuine GFP fluorescence from autofluorescence. A minimum of 50 transgenic antibiotic-resistant lines should be screened in order to choose lines with the strongest indicator protein expression.
15. There are pitfalls and artifacts using luciferases that can be avoided if some caution is taken. The luminescence from luciferases such as the aequorin/CTZ-system is dependent on environmental factors other than just [Ca^{2+}]. Careful control experiments are needed when [Ca^{2+}] signaling is investigated with aequorin-expressing organisms under abiotic stress situations. Control experiments designed to rule out possible artifacts can easily be performed with isolated indicator protein derived from bacterial expression (**Note 10**) or purchased from a company (e.g., NanoLight Technologies., Pinetop, AZ; Molecular Probes, Sigma, St. Louis, MO PJK, Kleinblittersdorf, Germany). Here are a few examples of environmental factors influencing luciferase activity:

    a. Temperature: The luminescence yield of luciferases generally increases with temperature ***(82)***. During cold treatment the luminescence decreases. Therefore, if a transgenic reporter organism displays an increase in luminescence during cooling ***(53)***, this is truly related to an increase in [Ca^{2+}]. Heat is thought to also produce a [Ca^{2+}] increase in plants ***(83)***. However, since luciferase is also activated by heat without [Ca^{2+}] increase, careful inspection of luminescence kinetics and control experiments with dead plant material or isolated protein are needed.
    b. Oxygen: the light-producing step of aequorin is independent from oxygen. However, oxygen is imperatively required for reconstitution of aequorin (**Fig. 1A**; ***84***). If an experiment is performed with freshly reconstituted plants, the free CTZ that has not been absorbed by apoaequorin may produce two effects that will spoil the [Ca^{2+}] recording: first, reconstitution will go on during the experiment and this may have an effect on both the luminescence per integration interval and the “remaining counts” (**Subheading 3.3.**); second, chemiluminescence from free CTZ will contribute to the luminescence signal, and this contribution is highly O_2-dependent (**Fig. 9**). Due to the ongoing turnover of aequorin, if free CTZ is present, luminescence will be decreased during O_2 deficiency (anoxia) and increased drastically when O_2 is resupplied (normoxia). A way to get around this kind of artifact is given below.
    c. Reactive oxygen species (ROS): ancient marine organisms have possibly used a monooxigenase/CTZ system to detoxify ROS. It is speculated that modern bioluminescent luciferase systems have evolved from these early antioxidant systems by a functional shift of CTZ from a general antioxidant to a special luciferin ***(85–87)***. Today CTZ is still an effective antioxidant and is actually offered as a chemiluminescent indicator for ROS ***(9)***. Moreover there are still luciferase systems existing that are triggered by ROS (e.g., Pholasin and Polynoïdin; ***88–91***). ROS-triggered luminescence of CTZ is shown in **Fig. 10**. Light release from aequorin-bound CTZ (compare **Fig. 1A**) is triggered not by ROS but solely by Ca^{2+}. However, any free CTZ still present in the system can produce light when treated with ROS (closed circles). Therefore, if luminescent plants are exposed to H_2O_2 ***(92)*** or to stress situations in which ROS are released ***(93,94)***, free CTZ might adulterate the [Ca^{2+}] signal from aequorin and lead to false interpretations. To circumvent the interference of

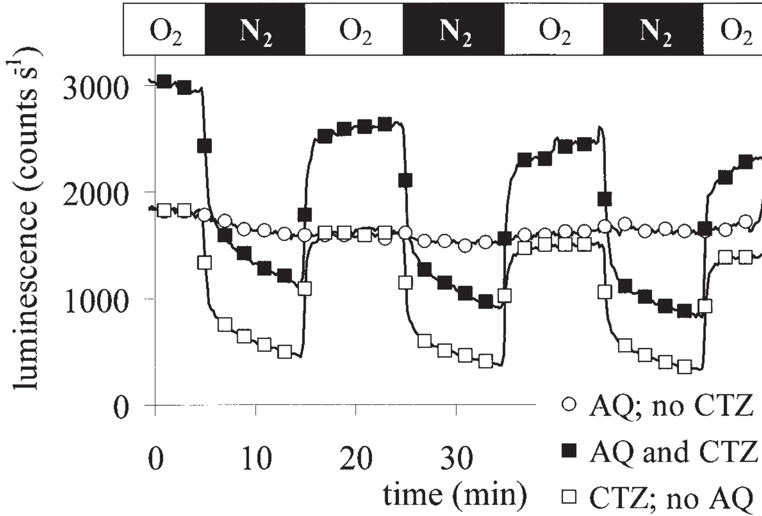


Fig. 9. Chemiluminescence of coelenterazine is O_2-dependent. Coelenterazine (2.5 µ*M* in NT buffer) and purified aequorin (1 µg/mL in NT buffer) have been aerated with ambient air (O_2) and nitrogen gas (N_2) and light output has been recorded. Chemiluminescence of CTZ (open squares) is strongly dependent on dissolved oxygen, whereas bioluminescence from aequorin is independent of O_2 (open circles). Aequorin has also been supplemented with 2.5 µ*M* CTZ to allow recycling of discharged protein (closed squares). However, there is no significant difference in O_2 dependency compared with the CTZ-only solution (open squares), so, O_2-consuming aequorin recycling (**Fig. 1A**) does not have a significant impact on this effect.

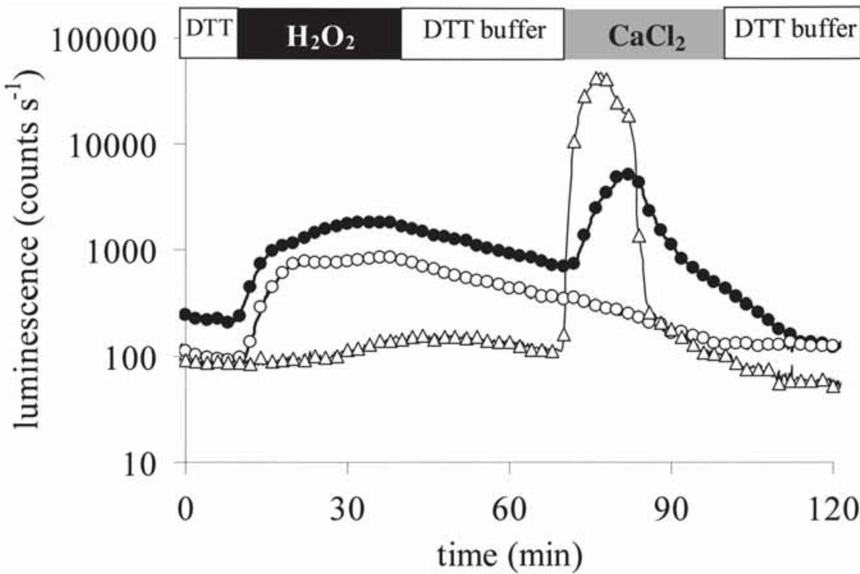


Fig. 10. CTZ chemiluminescence is dependent on reactive oxygen species (ROS). Different protein solutions (2.5 mL) have been dialyzed in front of a luminometer against the buffers indicated on top of the graph and light output has been recorded. Open triangles: purified reconstituted aequorin (0.1 µg/mL) dissolved in NET buffer; closed circles: recombinant apoaequorin from *E. coli* in NET and reconstituted with 2.5 n*M* CTZ for 6 h. Open circles: BSA (0.1 µg/mL) dissolved in NET and incubated with 2.5 n*M* CTZ for 6 h.

O_2- and/or ROS-triggered luminescence, it is advisable to let the plants rest for a reasonable while (several hours or even days) after reconstitution under usual growth conditions ***(67)***. Additionally, control experiments should always be performed with wild-type plants and with plants constitutively expressing aequorin (i.e., under the control of a promoter such as the CaMV-35S or the ubiquitin-1 promoter), in parallel and under identical conditions (i.e., identical age of *wt* plants; same CTZ treatment and recording luminescence under same experimental conditions).

d. Other conditions: Light emission of aequorin is basically unaffected by pH > 6.5 ***(82)***. Nevertheless, Ca^{2+} binding properties (dissociation and/or rate constants) may be shifted with high $[H^+]$. So, when pH falls drastically below 6.5 or if aequorin is expressed in compartments with acidic environment, it should be checked in vitro for pH dependence of luminescence. It has further been shown that the luminescence of aequorin and obelin is also affected by Mg^{2+} concentration and by monovalent ions ***(31,82,95)***. "Exotic" ions such as strontium (Sr^{2+}) and silver (Ag^+) can also have remarkable effects on aequorin luminescence ***(96,97)***.

16. CTZ is readily membrane-permeable and does not need any byproducts to facilitate loading into tissue. Also, vacuum infiltration of CTZ solution is not needed. Best results are obtained with 5 to 10 µ*M* CTZ final concentration. Always run the reconstitution reaction in the dark. It is essential to always leave the specimen in an open vessel to have sufficient oxygen for the reconstitution reaction. Avoid a final methanol concentration above 1% in the reconstitution medium. The optimal incubation time for reconstitution is best obtained empirically ***(20)***. It depends on the CTZ derivative, the luciferase, the location of the indicator protein inside the plant, and the $[Ca^{2+}]$ there. CTZ derivatives forming a high-Ca^{2+}-affinity aequorin have shorter optimal incubation times than native CTZ.
17. When working in a small closed compartment or with a perfusion system always take care to have sufficient oxygen supply. It is critical to ensure permanent aeration of perfusion solutions in order to avoid hypoxic stress ***(11)***.
18. An effective method to obtain a total discharge of the plant material is to lavish a $CaCl_2$/ethanol solution 1 *M*/10% [v/v] final concentration) on it. The ethanol destroys the cell membranes and lets the Ca^{2+} invade the cells and discharge the remaining aequorin. However, the ethanol also changes the optical properties of the specimen and makes it more transparent. This makes more light reach the detector than would have reached it without increased transparency. This leads to an overestimation of remaining aequorin and hence to an underestimation of $[Ca^{2+}]$. Another method to discharge the aequorin is by cold shocks (i.e., repetitive treatment with ice-cold water). This is effective in particular with the highly sensitive aequorin derivatives such as cp-aequorin ***(21)*** and avoids the drawbacks of the ethanol/$CaCl_2$ discharge. An alternative to the discharge procedure has been suggested ***(63)***: Different $[Ca^{2+}]_{cyt}$ transients can reproducibly be induced in plants by treatment with different NaCl concentrations ***(11)*** or $[Ca^{2+}]_{cyt}$ is clamped to different values by pH-clamping with weak acids ***(7,52)***. A standard experiment in which various $[Ca^{2+}]_{cyt}$ transients have been induced or various $[Ca^{2+}]$ been clamped stepwise can be taken as comparison, when calibrated as in **Subheading 3.3.** This requires NaCl or weak-acid treatment after each experiment exactly as in the standard experiment. The relative luminescence yield $L = c/R$ is then calculated and compared with the yield in the standard experiment.
19. There is also a method called "baseline fit" to estimate the amount of charged aequorin still available in the specimen without killing or injuring the plant. This, however, requires that the resting in vivo $[Ca^{2+}]$ (baseline) after the experiment (postbaseline) is identical to the baseline taken before setting the stimulus (prebaseline). For estimating the discharge, counting is stopped after the experiment without discharge procedure, however, not until a reasonable post-baseline recording has been appended. Data are then supplemented with a number—according to experience from ethanol-mediated discharge—that roughly matches the discharge counts and the relative luminescence yield is calculated (**Fig. 8**). If the recorded end baseline is increasing above prebaseline level (**Fig. 8**, black curve,) then the estimated discharge was too low and a higher number must be inserted. If the end-baseline is below prebaseline level (gray curve), then the approximation was too high and a smaller number must be chosen. This way a rough approximation can be achieved iteratively without killing the plant.

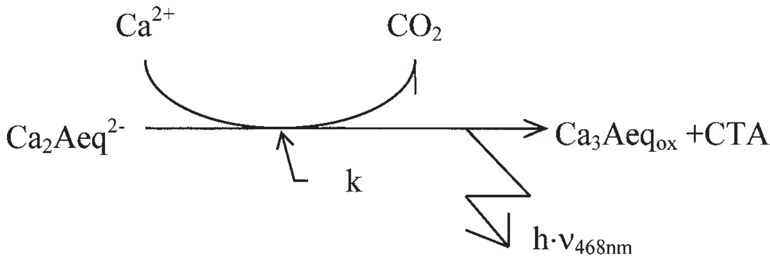


Fig. 11. Scheme of luminescence triggered by Ca^{2+} ions. The enzyme is activated when Ca^{2+} binds to the photoprotein. Bound coelenterazine is oxidised to coelenteramide (CTA) by the release of CO_2 and blue light ($h \cdot \nu_{468nm}$).

20. The *absolute* luminescence yield (in contrast to the *relative* luminescence yield) is the light produced during a particular integration interval divided by the whole light ever produced by the specimen (i.e., the quantum of aequorin discharged during the integration interval divided by the amount of aequorin present at the beginning of the experiment). This is not correlated with the $[Ca^{2+}]$.
21. Subtraction of the background is necessary to calculate the aequorin derived counts per time interval (c). The subtraction can be neglected if the background is less than one-tenth of the aequorin luminescence baseline values measured prior to the experiment.
22. A basic model to derive an equation for converting aequorin luminescence into $[Ca^{2+}]$ may be given here: Aequorin has three Ca^{2+} binding domains (**Fig. 1B**). However, light is released only when all three domains are occupied by Ca^{2+} ions (**Fig. 11**)

    The number of discharged aequorin molecules per time $\left(d/dt\left[Aeq^{2-} \right] = Aeq \right)$ is proportional to the existing concentration of undischarged Aeq^2—molecules and proportional to the free calcium ion concentration $[Ca^{2+}]$. This proportionality is defined by the rate constant k:

    $$d/dt\left[Aeq^{2-} \right] = k \cdot \left[Aeq^{2-} \right] \cdot \left[Ca^{2+} \right]$$

    or in shorter writing: $Aeq = k \cdot Aeq \cdot Ca$ (eq. 2)

    The counted number of photons per second (c) is proportional to the number of discharged aequorin molecules per s (Aeq):

    $$c = f \cdot Aeq$$

    whereas the factor f takes all optical and electronic properties of the measuring device into account. Thus, together with **eq. 2**, this can be rearranged as:

    $$c = f \cdot k \cdot Aeq \cdot Ca$$

    The number of remaining counts R at time t is proportional to the total number of available un-discharged aequorin molecules (Aeq):

    $$R = f \cdot Aeq$$

    Together with this, equation can be written as:

    $$c = k \cdot R \cdot Ca \qquad \text{(eq. 4)}$$

    By introducing the luminescence yield $L = \frac{c}{R}$ which gives the ratio of counts recorded per time interval (seconds) and the amount of remaining counts, **eq. 4** gives:

$$Ca = \frac{L}{k}$$

or in terms of pCa: $pCa = \log(k) - \log(L)$

By setting $a_0 = \log(k)$ and introducing a factor a_1, which takes care of the effects of the other two Ca^{2+} binding sites, equation can be rearranged as:

$$pCa = a_0 - a_1 \cdot \log(L) \qquad \text{(eq. 5)}$$

The factors a_0 and a_1 depend on the luciferase and on the used CTZ-derivative and must be empirically derived. For the isoform of aequorin usually used in *Arabidopsis* (ID# = AVAEQA; ***98***) together with native CTZ, the free parameters a_0 and a_1 have been empirically derived ***(99)*** and the specific calibration equation is:

$$pCa = 5.5593 - 0.332588 \cdot \log(L)$$

23. The relative luminescence yield (L) is the light measured during an integration interval (c) divided by the remaining counts in the specimen (R; i.e., the quantum of aequorin discharged during the time interval divided by the amount of charged aequorin still remaining in the specimen; *see* **Note 20**). The latter requires total discharge of all aequorin in the specimen or at least a rough estimation of the amount of charged aequorin still residing in the specimen after the experiment (*see* **Note 18**).

    A standard experiment using aequorin luminescence must not exceed two important limits:

    a. The baseline of aequorin luminescence from the transgenic plant is well above the twofold of the background luminescence (ideally five- to 10-fold the background) derived from the corresponding nontransgenic *Arabidopsis* wt-line.

    b. The counts per integration interval (c) are well below 10^{-6} times the remaining counts (R) in the specimen throughout the experiment.

Within these limits luminescence data of an experiment are almost linearily correlated with the logarithm of the free calcium ion concentration (pCa), according to **eq. 5** (*see* **Note 2**).

In wider limits as marked above the dependency of pCa on log(L) follows a curve of sigmoidal shape rather than a straight line ***(82,95)***. It is possible to take this into account by fitting luminescence data to an appropriate sigmoidal calibration curve ***(21)***. Finally, each luminescent Ca^{2+} indicator has its peculiar calibration parameters, which need to be discovered before converting relative luminescence L into [Ca^{2+}].

Acknowledgments

I thank Jürgen Nagel-Volkmann (University Bielefeld, Bielefeld, Germany) for critical reading of the manuscript and for valuable supplemental data.

References

1. Knight, M. R., Campbell, A. K., Smith, S. M., and Trewavas, A. J. (1991) Transgenic plant aequorin reports the effects of touch and cold-shock and elicitors on cytoplasmic calcium. *Nature* **352,** 524–526.
2. Plieth, C. (2001) Plant calcium signaling and monitoring – pros and cons and recent experimental approaches. *Protoplasma* **218,** 1–23.
3. Miyawaki, A. (2003) Visualization of the spatial and temporal dynamics of intracellular signaling. *Developmental Cell* **4,** 295–305.
4. Zhang, J., Campbell, R. E., Ting, A. Y., and Tsien, R. Y. (2002) Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* **3,** 906–918.
5. Shimomura, O., Musicki, B., and Kishi, Y. (1989) Semi-synthetic aequorins with improved sensitivity to Ca^{2+} ions. *Biochem. J.* **261,** 913–920.
6. Thomas, M. V. (1982) *Techniques in calcium research*. Academic Press, London, New York, Paris, San Diego, San Francisco, Sao Paulo, Sydney, Tokyo, Toronto.

7. Fricker, M. D., Plieth, C., Knight, H., Blancaflor, E., Knight, M. R., White, N., and Gilroy, S. (1999) Fluorescence and luminescence techniques to probe ion activities in living plant cells in: *Fluorescent and luminescent probes for biological activity*. Academic Press, San Diego, (Mason, B. ed.) pp. 569–596.
8. Campbell, A. K., Patel, A., Houston, W. A. J., Scolding, N. J., Frith, S., Morgan, B. P., and Compston, D. A. S. (1989) Photoproteins as indicators of intracellular free Ca^{2+}. *J. Bioluminecence Chemiluminescence* **4,** 463–474.
9. Haugland, R. P. (2003) Handbook of fluorescent probes and research chemicals. Molecular Probes Inc., Eugene OR.
10. www.probes.com/handbook/sections/2000.html.
11. Fricker, M. D., Parsons, A., Tlalka, M., Blancaflor, E., Gilroy, S., Meyer, A., and Plieth, C. (2001) Fluorescent probes for living plant cells. in: Hawes, C. and Satiai-Jeunemaitre, B. (ed.) *Plant Cell Biology: A Practical Approach*, 2nd ed., Oxford University Press, Oxford, pp. 35–84.
12. Brandizzi, F., Fricker, M., and Hawes, C. (2002) A greener world: revolution in plant bioimaging. *Nat. Rev. Mol. Cell Biol*. **3,** 520–530.
13. Kiegle, E., Moore, C. A., Haseloff, J., Tester, J., and Knight, M. R. (2000) Cell-type-specific calcium responses to drought salt and cold in *Arabidopsis roots*. *Plant J.* **23,** 267–278.
14. Inouye, S., Noguchi, M., Sakaki, Y., et al. (1985) Cloning and sequence analysis of cDNA for the luminescent protein aequorin. *Proc. Nat. Acad. Sci. USA* **82,** 3154–3158.
15. Prasher, D., McCann, R. O., and Cormier, M. J. (1985) Cloning and expression of the cDNA coding for aequorin a bioluminescent calcium-binding protein. *Biochem. Biophys. Res. Comm*. **126,** 1259–1268.
16. Haseloff, J., Siemering, K. R., Prasher, D. C., and Hodge, S. (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proc. Natl. Acad. Sci. USA* **94,** 2122–2127.
17. Baubet, V., Mouellic, H. L., Campbell, A. K., Lucas-Meunier, E., Fossier, P., and Brulet, P. (2000) Chimeric green fluorescent protein-aequorin as bioluminescent Ca^{2+} reporters at the single cell level. *Proc. Nat. Acad. Sci. USA* **97,** 7260–7265.
18. Moore, C. A. (2000) An investigation into cellular and molecular aspects of calcium-based specificity of signaling in *Arabidopsis thaliana.* D. Phil. thesis, University of Oxford.
19. Logan, D. C. and Knight, M. R. (2003) Mitochondrial and cytosolic calcium dynamics are differentially regulated in plants. *Plant Physiol.* **133,** 21–24.
20. Gao, D., Trewavas, A. J. T., Knight, M. R., Sattelmacher, B., and Plieth, C. (2004) Self-reporting *Arabidopsis thaliana* expressing pH- and [Ca^{2+}]- indicators unveil ion dynamics in the cytoplasm and in the apoplast under abiotic stress. *Plant Physiol.* **134,** 898–908.
21. Baum, G., Long, J. C., Jenkins, G. I., and Trewavas, A. J. (1999) Stimulation of the blue light phototropic receptor NPH1 causes a transient increase in cytosolic Ca^{2+}. *Proc. Nat. Acad. Sci. USA* **96,** 13,554–13,559.
22. Plieth, C. and Trewavas, A. J. (2002) Reorientation of seedlings in the earth's gravitational field induces cytosolic calcium transients. *Plant Physiol.* **129,** 786–796.
23. Shimomura, O., Kishi, Y., and Inouye, S. (1993) The relative rate of aequorin regeneration from apoaequorin and coelenterazine analogues. *Biochem. J.* **296,** 549–551.
24. Shimomura, O. (1991) Preparation and handling of aequorin solutions for the measurement of cellular calcium. *Cell Calcium* **12,** 635–643.
25. Shimomura O. (1995) Cause of spectral variation in the luminescence of semisynthetic aequorins. *Biochem. J*. **306,** 537–543.
26. Haddock, S. H. D. and Case, J. F. (1999) Bioluminescence spectra of shallow and deap-sea gelatinous zooplankton: ctenophores, medusae and siphonophores. *Marine Biol.* **133,** 571–582.
27. Blinks, J. R., Wier, W. G., Hess, P., and Prendergast, F. G. (1982) Measurement of Ca^{2+} concentrations in living cells. *Prog. Biophys. Mol. Biol.* **40,** 1–114.
28. Fagan, T. F., Ohmiya, Y., Blinks, J. R., Inouye, S., and Tsuji, F. I . (1993) Cloning, expression and sequence analysis of cDNA for the Ca^{2+}-binding photoprotein, mitrocomin. *FEBS Lett*. **333,** 301–305.
29. Inouye, S. and Tsuji, F. L. (1993) Cloning and sequence analysis of cDNA for the $Ca^{(2+)}$-activated photoprotein, clytin. *FEBS Lett.* **315,** 343–346.
30. Liu, C-J., Vysotski, E. S., Chen, C-J., Rose, J. P., Lee, J., and Wang, B-C. (2000) Structure of the Ca^{2+}-regulated photoprotein obelin at 1.7 Å resolution determined directly from its sulfur substructure. *Protein Sci*. **9,** 2085–2093.

31. Illarionov, B. A., Frank, L. A., Illaronova, V. A., Bondar, V. S., Vysotski, E. S., and Blinks, J. R. (2000) Recombinant obelin: cloning and expression of cDNA, purification, and characterization as a calcium indicator, in *Methods in Enzymology—Bioluminescence and Chemiluminescence Part C*. Ziegler, M. M., and Baldwin, T. O. (ed.), Academic Press London, pp. 223–249.
32. Markova, S. V., Vysotski, E. S., Blinks, J. R., Burakova, L. P., Wang, B-C., and Lee, J. (2002) Obelin from the bioluminescent marine hydroid Obelia geniculata: cloning, expression, and comparison of some properties with those of other Ca^{2+}-regulated photoproteins. *Biochemistry* **41,** 2227–2236.
33. Deng, L., Vysotski, E. S., Liu, Z-J., et al. (2001) Structural basis of the emission of violet bioluminescence from a W92F obelin mutant. *FEBS Lett.* **506,** 281–285.
34. Kendall, J. M., Sala-Neby, G., Ghalaut, V., Dormer, R. L., and Campbell, A. K. (1992) Engineering the Ca^{2+}-activated photoprotein aequorin with reduced affinity for calcium. *Biochem. Biophys. Res. Comm.* **187,** 1091–1097.
35. van der Luit, A., Olivari, C., Haley, A., Knight, M. R., and Trewavas, A. J. (1999) Distinct calcium signaling pathways regulate calmodulin gene expression in tobacco. *Plant Physiol.* **121,** 705–714.
36. Chiesa, A., Rapizzi, E., Tosello, V., Pinton, P., de Virgilio, M., Fogarty, K. E., and Rizzuto, R. (2001) Recombinant aequorin and green fluorescent protein as valuable tools in the study of cell signaling. *Biochem. J.* **355,** 1–12.
37. Campbell, A. K., Trewavas, A. J., and Knight, M. R. (1996) Calcium imaging shows differential sensitivity to cooling and communication in luminous transgenic plants. *Cell Calcium* **19,** 211–218.
38. Knight, M. R., Read, N. D., Campbell, A. K., and Trewavas, A. J. (1993) Imaging calcium dynamics in living plants using semi-synthetic recombinant aequorins. *J. Cell Biol.* **121,** 83–90.
39. Knight, H. and Knight, M. R. (1995) Recombinant Aequorin methods for intracellular calcium measurement in plants. *Meth. Cell Biol*. **49,** 201–216.
40. Knight, H., Trewavas, A. J., and Knight, M. R. (1997) Recombinant aequorin methods for measurement of intracellular calcium in plants. *Plant Mol. Biol. Man.* **C4,** 1–22.
41. Mithöfer, A. and Mazars, C. (2002) Aequorin-based measurements of intracellular Ca^{2+}-signatures in plant cells. in: www.biologicalprocedures.com (ed.) *Biological Procedures Online* **4,** 105–118.
42. Gleave, A. P. (1992) A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Mol. Biol.* **20,** 1203–1207.
43. Pietrzak, M., Shillito, R. D., Hohn, T., and Potrykus, I. (1986) Expression in plants of two bacterial antibiotic genes after protoplast transformation with a new plant expression vector. *Nucleic Acids Res.* **14,** 5857–5868.
44. Aikens, R. (1999) Properties of low-light-level slow-scan detectors, in *Fluorescent and Luminescent Probes for Biological Activity*. Academic Press, San Diego, (Mason, B ed.) pp. 507–516.
45. Creton, R., Kreiling, J. A., and Jaffe, L. F. (1999) Calcium imaging with chemiluminescence. *Microscopy Res. Technique* **46,** 390–397.
46. Stanley, P. E. (2000) Commercially available luminometers and low-level light imaging device in *Methods in Enzymology—Bioluminescence and Chemiluminescence Part C*. (Ziegler, M. M., and Baldwin, T. O., ed.) Academic Press, London, pp. 96–103.
47. Ellis, R. J. and Wright, A. G. (1999) Optimal use of photomultipliers for chemiluminescence and bioluminescence applications. *Luminescence* **14,** 11–18.
48. Tomkins, P. and Lyons, A. (1999) Properties of low-light-level intensified cameras, in *Fluorescent and Luminescent Probes for Biological Activity.* Academic Press, San Diego, (Mason, B., ed.) 491–506.
49. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1999) *Short Protocols in Molecular Biology*. Wiley, New York.
50. Sambrook, J. and Russell, D. (2001) *Molecular cloning — a laboratoy manual*. Vol. 1– 3; 3rd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
51. Plieth, C. (2004) Luminescence measuring chamber, in Christian-Albrechts-Universität, Kiel, Patent No. 103 18 994.7 B3 2004.03.11. Deutsches Patent- und Markenamt, München.
52. Plieth, C., Sattelmacher, B., and Hansen, U-P. (1997) Cytoplasmic Ca^{2+}-H^{+}-exchange buffers in green algae. *Protoplasma* **198,** 107–124 and **199,** 223 (correction).
53. Plieth, C., Hansen, U-P., Knight, H., and Knight, M. R. (1999) Temperature sensing by plants: the primary mechanisms of signal perception and calcium response. *Plant J*. **18,** 491–497.
54. Plieth, C., Sattelmacher, B., and Knight, M. R. (2000) Ammonium uptake and cellular alkalization in roots of *Arabidopsis thaliana*: the involvement of cytoplasmic calcium. *Physiologia Plantarum* **110,** 518–523.

55. Knight, M. R., Smith, S. M., and Trewavas, A. J. (1992) Wind-induced plant motion immediately increases cytosolic calcium. *Proc. Nat. Acad. Sci. USA* **89,** 4967–4971.
56. Russell, A. J., Knight, M. R., Cove, D. J., Knight, C. D., Trewavas, A. J., and Wang, T. L. (1996) The moss Physcomitrella patens transformed with apoaequorin cDNA responds to cold shock mechanical perturbation and pH with transient increases in cytoplasmic calcium. *Transgen. Res.* **5,** 167–170.
57. Johnson, C. H., Knight, M. R., Kondo, T., Masson, P., Sedbrook, J., Haley, A., and Trewawas, A. (1995) Circadian oscillations of cytosolic and chloroplastic free calcium in plants. *Science* **269,** 1863–1865.
58. Wood, N. T., Haley, A., Viry-Moussaid, Johnson, C. H., van der Luit, A., and Trewavas, A. J. (2001) The calcium rhythms of different cell types oscillate with different circadian phases. *Plant Phys.* **125,** 787–796.
59. Brini, M., Pinton, P., Pozzan, T., and Rizzuto, R. (1999) Targeted recombinant aequorins: Tools for monitoring [Ca^{2+}] in the various compartments of a living cell. *Micro. Res. Technique* **46,** 380–389.
60. Campbell, A. K. (1988) *Chemiluminescence — principles and applications in biology and medicine*. VCH and Ellis Horwood Ltd., New York.
61. Torrecilla, I., Leganes, F., Bonilla, I., and Fernandes-Pinas, F. (2000) Use of recombinant aequorin to study calcium homeostasis and monitor calcium transients in response to heat and cold shock in cyanobacteria. *Plant Physiol.* **123,** 161–176.
62. Allen, G. J., Chu, S. P., Schumacher, K., et al. (2000) Alteration of stimulus-specific guard cell calcium oscillations and stomatal closing in *Arabidopsis* det3 mutant. *Science* 289, 2338–2342.
63. Love, J., Dodd, A. N., and Webb, A. A. R. (2004) Circadian and diurnal calcium oscillations encode photoperiodic information in *Arabidopsis*. *Plant Cell* **16,** 956–966.
64. Watkins, A. J. and Campbell, A. K. (1993) Requirement of the C-terminal proline residue for stability of the Ca^{2+}-activated photoprotein aequorin. *Biochem. J.* **293,** 181–185.
65. Llinás, R., Sugimori, M., and Silver, R. B. (1992) Microdomains of high calcium concentration in a presynaptic terminal. *Science* **256,** 677–679.
66. Gonzalez-Trueba, G., Paradisi, C., and Zoratti, M. (1996) Synthesis of coelenterazine. *Anal. Biochem.* **240,** 308–310.
67. Sedbrook, J. C., Kronebusch, P. J., Borisy, G. G., Trewavas, A. J., and Masson, P. H. (1996) Transgenic Aequorin reveals organ-specific cytosolic Ca^{2+} responses to anoxia in *Arabidopsis thaliana* seedlings. *Plant Physiol.* **111,** 243–257.
68. Mason, B. (1999) *Fluorescent and luminescent probes for biological activity*. Academic Press, San Diego.
69. Hawes, C., Saint-Jore, C., Martin, B., and Zheng, H-Q. (2001) ER confirmed as the location of mystery organelles in *Arabidopsis* plants expressing GFP. *Trends Plant Sci.* **6,** 245–246.
70. Kaletta, K., Kunze, I., Kunze, G., and Köck, M. (1998) The peptide HDEF as a new retention signal is necessary and sufficient to direct proteins to the endoplasmic reticulum. *FEBS Lett.* **434,** 377–381.
71. Jedd, G. and Chua, N-H. (2002) Visualization of peroxisomes in living plant cells reveals acto-myosin dependent cytoplasmic streaming and peroxisome budding. *Plant Cell Physiol.* **43,** 384–392.
72. Mano, S., Nakamori, C., Hayashi, M., Kato, A., Kondo, M., and Nishimura, M. (2002) Distribution and characterization of peroxisomes in *Arabidopsis* by visualization with GFP: Dynamic morphology and actin-dependent movement. *Plant Cell Physiol.* **43,** 331–341.
73. Hayashi, M., Nito, K., Toriyama-Kato, K., Kondo, M., Yamaya, T., and Nishimura, M. (2000) AtPex14p maintains peroxisomal functions by determining protein targeting to three kinds of plant peroxisomes. *EMBO J.* **19,** 5701–5710.
74. Frigerio, L., Ombretta, F., Felipe, D. H., Neuhaus, J-M., and Vitale, A. (2001) The C-terminal tetrapeptide of phaseolin is sufficient to target green fluorescent protein to the vacuole. *J. Plant Physiol.* **158,** 400–503.
75. Neuhaus, J-M. (2000) GFP as a marker for vacuoles in plants, in *Annual Plant Reviews: Vacuolar Compartments*. Vol 5, (Robinson, DG ed.) Sheffield Academic Press, Sheffield, UK. pp. 254–269.
76. Di Sansebastiano, G-P., Paris, N., Marc-Martin, S., and Neuhaus, J-M. (1998) Specific accumulation of GFP in a non-acidic vacuolar compartment via a C-terminal propeptide-mediated sorting pathway. *Plant J.* **15,** 449–457.
77. Brini, M., Marsault, R., Bastianutto, C., Alvarez, J., Pozzan, J., and Rizzuto, R. (1995) Transfected aequorin in the measurement of cytosolic Ca^{2+} concentration $[Ca^{2+}]_c$. A critical evaluation. *J. Biolog. Chem.* **270,** 9896–9903.

78. Gonzalez, D. G. and Ward, W. W. (2000) Large Scale Purification of recombinant green fluorescent protein from Escherichia coli. *Methods in Enzymol.* **305,** 212–223.
79. Ward, W. W., Swiatek, G. C., and Gonzalez, D. G. (2000) Green fluorescent protein in biotechnology education. *Meth. Enzymol.* **305,** 672–680.
80. Shimomura, O. and Inouye, S. (1999) The in situ regenration and extraction of recombinant aequorin from *Escherichia coli* cells and the purification of extracted aequorin. *Protein Expression Purification* **16,** 91–95.
81. Shrestha, S., Paeng, I. R., Deo, S. K., and Daunert, S. (2002) Cystein-free mutant of aequorin as a photolabel in immunoassay development. *Bioconjugate Chem.* **13,** 269–275.
82. Blinks, J. R., Mattingly, P. H., Jewell, B. R., van-Leeuwen, M., Harrer, G. C., and Allen, D. G. (1997) Practical aspects of the use of aequorin as a calcium indicator: assay preparation microinjection and interpretation of signals. *Meth. Enzymol.* **57,** 292–328.
83. Gong, M., van der Luit, A. H., Knight, M. R., and Trewavas, A. J. (1998) Heat-shock induced changes in intracellular Ca^{2+} level in tobacco seedlings in relation to thermotolerance. *Plant Physiol.* **116,** 429–437.
84. Cormier, M. J., Prasher, D. C., Longiaru, M., and McCann, R. O. (1989) The enzymology and molecular biology of the Ca^{2+}-activated photoprotein aequorin. *Photochem. Photobiol.* **49,** 509–512.
85. Rees, J. F., de Wergifosse, B., Noiset, O., Dubuisson, M., Janssens, B., and Thompson, E. M. (1998) The origins of marine bioluminescence: turning oxygen defence mechanisms into deep-sea communication tools. *J. Exper. Biol.* **201,** 1211–1221.
86. Haddock, S. H., Rivers, T. J., and Robinson, B. H. (2001) Can coelenterates make coelenterazine? Dietary requirements for luciferin in cnidarian bioluminescence. *Proc. Nat. Acad. Sci. USA* **98,** 11,148–11,151.
87. de Wergifosse, B., Noidet, O., Dubuisson, M., Marchand-Brynaert, J., Baguet, F., and Rees, J. F. (1999) A new function for coelenterazine oxidation product supports the non-bioluminescent evolutionary origin of the luciferin, in *Bioluminescence and Chemiluminescence—Perspectives for the 21st Century*. Wiley, New York, (Roda, A., Pazzagli, M., Kricka, L. J., and Stanley, P. E. (ed.) pp. 396–399.
88. Müller, T., Davies, E. V., and Campbell, A. K. (1989) Pholasin chemiluminecence detects mostly superoxide anion released from human neutrophils. *J. Bioluminescence Chemiluminescence* **3,** 105–113.
89. Roberts, P. A., Knight, J., and Campbell, A. K. (1987) Pholasin—a bioluminescent indicator for detecting activation of single neutrophils. *Analy. Biochem.* **160,** 139–148.
90. Bassot, J. M. and Nicolas, M. T. (1995) Bioluminescence in scale-worm photosomes: the photoprotein Polynoïdin is specific for the detection of superoxide radicals. *Histochem. Cell Biol.* **104,** 199–210.
91. Colepicolo, P., Camarero, VCPC., Nicolas, M. T., Bassot, J. M., Karnovsky, M. L., and Hastings, J. W. (1990) A sensitive and specific assay for superoxide anion released by neutrophils or macrophages based on bioluminescence of polynoïdin. *Analyt. Biochem.* **184,** 369–374.
92. Price, A. H., Taylor, A., Ripley, S. J., Griffiths, A., Trewavas, A. J., and Knight, M. R. (1994) Oxidative signals in tobacco increase cytosolic calcium. *Plant Cell* **6,** 1301–1310.
93. Clayton, H., Knight, M. R., Knight, H., McAinsh, M. R., and Hetherington, A. M. (1999) Dissection of the ozone-induced calcium signature. *Plant J.* **17,** 575–579.
94. Chandra, S. and Low, P. S. (1997) Measurement of Ca^{2+} fluxes during elicitation of the oxidative burst in aequorin-transformed tobacco cells. *J. Biol. Chem.* **272**, 28,274–28,280.
95. Blinks, J. R. (1990) Use of photoproteins as intracellular calcium indicators. *Environ. Health Perspect.* **84,** 75–81.
96. Gilroy, S., Hughes, W. A., and Trewavas, A. J. (1989) A comparision between Quin-2 and Aequorin as indicators of cytoplasmic calcium levels in higher plant cell protoplasts. *Plant Physiol.* **90,** 482–491.
97. Shimomura, O. and Johnson, F. H. (1970) Calcium binding, quantum yield and emitting molecule in aequorin bioluminiscence. *Nature* **227,** 1356–1357.
98. Badminton, M. N., Sala-Newby, G. B., Kendall, J. M., and Campbell, A. K. (1995) Differences instability of recombinant apoaequorin with subcellular compartments. *Biochem. and Biophys. Res. Comm.* **217,** 950–957.
99. Cobbold, P. H. and Rink, T. J. (1987) Fluorescence and bioluminescence measurement of cytoplasmic free calcium. *Biochem. J.* **248,** 313–328.
100. Tsuji, F. I., Inouye, S., Goto, T., and Sakaki, Y. (1986) Site-specific mutagenesis of the calcium-inding photoprotein aequorin. *Proc. Nat. Acad. Sci. USA* **83,** 8107–8111.

26

Applications of Chemical-Inducible Expression Systems in Functional Genomics and Biotechnology

Jianru Zuo, Peter D. Hare, and Nam-Hai Chua

Summary

The LexA-VP16-ER (XVE) system is an estrogen-receptor-based chemical-inducible system for use in plants. Protocols are outlined for the inducible overexpression of transgenes, inducible transgene excision, inducible gene silencing, regulated in vivo cDNA generation, inducible activation tagging, and the generation of conditional knockouts.

Key Words: *Arabidopsis*, chemically regulated gene expression; RNAi; activation tagging.

1. Introduction

Functional annotation of plant genes has emerged as a major task since the complete sequencing of the *Arabidopsis* and rice genomes. In addition to the more traditional forward genetic approach, reverse genetics has also become a powerful tool in these endeavors. Two major strategies employed in reverse genetics are overexpression and knockdown (underexpression) of genes of interest. Constitutive promoters or enhancers have routinely been used in such efforts, although in many cases, constitutive overexpression or underexpression of a target gene may cause detrimental effects or even result in lethality of the host plant. This frequently makes it impossible to address the question of interest. To overcome these potential problems, several inducible expression systems have been developed and used successfully (reviewed in ***1***,***2***). Here we outline one of the most commonly used systems, the LexA-VP16-ER (XVE) system ***(3)***, and present several protocols for its use in plant functional genomics.

The XVE system employs the regulatory moiety of the human estrogen receptor (ER or E) to confer hormone inducibility on a chimeric transcription factor, XVE. The DNA binding domain of XVE, which is derived from bacterial repressor *LexA* (X), specifically recognizes a target promoter containing the *LexA* operator sequence. Note that the LexA operator sequence is simply used as positive *cis*-element upstream of the -46 cauliflower mosaic virus (CaMV) 35S minimal promoter. The acidic transcriptional activation domain from VP16 (V) confers competence for transcription activation on XVE ***(3)***.

Although selection of a chemically regulated gene induction system depends on the particular application, the XVE system possesses several key characteristics for effective chemical induction of trangene expression ***(2)***. The inability of endogenous *Arabidopsis* steroids to activate XVE renders 17-β-estradiol specific to the target promoter, with undetectable basal transgene expression levels in the absence of inducer. There is no evidence that the inducer disturbs endogenous gene expression or growth and development in *Arabidopsis*. The dose-dependent induction of XVE-regulated genes is equally suited to applications that require only low levels of induction as well as those demanding several thousand-fold increases in expression levels. 17-β-estradiol is readily taken up by aerial tissues as well as roots, but is not volatile. This

From: *Methods in Molecular Biology, vol. 323: Arabidopsis Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

prevents inadvertent gene activation. Owing to the difficulty in grafting *Arabidopsis*, the use of the XVE system to produce genetic chimera provides a facile approach to discriminating between local responses and those involving long-distance signaling (e.g., *5*). Finally, although dampening of transgene expression depends on stability of the transcript in question, the activity of XVE-regulated promoters routinely returns to non-detectable levels within 5 to 7 d after removal of the inducer. Expression can be reactivated repeatedly without loss of sensitivity to the inducer (*see* e.g., *7*).

The XVE system was initially used in overexpression studies for a number of genes. Subsequently, the system was modified to facilitate applications in other studies, including inducible DNA excision or marker removal *(4)*, inducible RNA interference (RNAi; *5*), regulated in vivo cDNA generation (RIDE; *6*), and inducible activation tagging *(7,8)* as well as conditional knockout studies (P.D. Hare, L.-F. Huang, and N.-H. Chua, unpublished data). Recently, the XVE system was adapted for high-throughput functional analysis of plant genes *(9)*.

2. Materials

1. XVE expression vectors pER8 and pER10.
2. Inducible DNA excision vectors pX6 and pX7.
3. Inducible RNAi vectors pSK-int, pX6, and pX7.
4. Inducible activation tagging vector pER16.
5. nducible Gateway-compatible vector pMDC7.
6. Conditional knockout vector pX8-GFP (all the above vectors are available from Nam-Hai Chua to academic users for noncommercial use).
7. Cloning vectors pBlueScript and pCRscript (Stratagene, La Jolla, CA).
8. The Gateway system (Invitrogen, Carlsbad, CA).
9. *Arabidopsis thaliana* strains Columbia-0 (Col-0) and Wassilewskija (WS).
10. *E. coli* strains DH5α (Invitrogen; formerly known as Life Technologies/Gibco-BRL; Rockville, MD) and SURE-2 (Stratagene).
11. *Agrobacterium tumefaciens* strains ABI and GV3101.
12. Lennox L Broth (LB) medium.
13. Ampicillin, spectinomycin, carbenicillin, kanamycin, and hygromycin.
14. 17-β-estradiol (Sigma, St. Louis, MO): 10 m*M* stock solution in dimethyl sulfoxide. Aliquots should be stored at –20°C. No sterilization is needed.
15. Inoculation medium: 50 g/L sucrose, 10 m*M* MgCl2, 200 µL/L Silwet L-77 (Lehle Seeds; Round Rock, TX).
16. MS (Murashige & Skoog) medium: 1X MS salts, 3% sucrose, 0.8% agarose, pH 5.7 (all MS media for protocols used in this chapter contain 3% sucrose unless indicated otherwise).
17. Gamborg's B5 medium: 1X B5 salts, 1X B5 vitamins, 3% sucrose, 0.5 g/L 2-(N-morpholino) ethanesulfonic acid (MES), pH 5.7.
18. F1 medium: 1X B5 salts, 2% glucose, 0.5 g/L MES, 0.5 mg/L 2,4-D, 0.05 mg/L kinetin, 0.2% Phytagel™, pH 5.7.
19. F2 medium: F1 medium supplemented with 20 mg/L acetosyringone (3',5'-dimethoxy-4'-hydroxy-acetophenone).
20. SR medium: 1X MS salts, 1% sucrose, 0.5 g/L MES, 1 µg/mL 2-isopentenyladenine (2-iP), 0.15 µg/mL IAA, 100 µg/mL carbenicillin, 50 µg/mL kanamycin (or 50 µg/mL hygromycin, depending on vectors used in the transformation experiments), 0.2% PhytagelTM, pH 5.7.
21. SR' medium: identical to SR medium except Phytagel is substituted with 0.6% low melting point agarose.
22. SCM' medium: identical to SR medium except without 2-ip and supplemented with 10 µ*M* 17-β-estradiol.
23. SCM' medium: identical to SCM medium except Phytagel is substituted with 0.6% low-melting agarose.
24. RI medium: identical to the SR medium except without 2-ip.
25. M medium: 1X MS salts, 1X B5 vitamins, 3% sucrose, 100 µg/mL carbenicillin, 0.8% agarose, pH 5.7.
26. Seed sterilization solution: diluted bleach (3.0% NaOCl) containing 0.01% Triton X-100.

27. Cheesecloth.
28. Oligonucleotide primers.
29. Restriction enzymes, Taq DNA polymerase, Pfu or PWO DNA polymerase, and T4 DNA ligase.
30. Expand PCR system (Roche; Penzberg, German).
31. SuperScript™ II reverse transcriptase (Invitrogen).
32. ^{32}P-dCTP.
33. Agarose; DNA and RNA gel electrophoresis equipment.
34. Duralon-UV™ membranes (Stratagene).
35. PhosphorImage (Molecular Dynamics, Sunnyvale, CA).
36. Fluorescence microscope.

3. Methods

The methods described below outline the use of the XVE system and its derivatives for (1) overexpression of transgenes, (2) inducible excision of a transgene from the host genome, (3) gene silencing by inducible RNAi, (4) regulated in vivo cDNA generation, (5) inducible activation tagging and functional cloning, (6) generation of conditional knockouts, and (7) high-throughput functional analysis of plant genes combined with the Gateway system. Although the experiments described here were performed using *Arabidopsis thaliana*, most of them have also been demonstrated in tobacco and several other dicotyledonous species. With appropriate modifications of the vectors, the entire system can be easily adapted to the use in monocotyledonous species such as rice (*see* **Note 1**). In the following subsections, we will describe related expression vectors and cloning strategies followed by data analysis and interpretation. Standard molecular methods are described elsewhere ***(10)*** as well as in instructions from manufacturers of the various reagents.

3.1. Overexpression Studies

3.1.1. XVE Expression Vectors pER8 and pER10

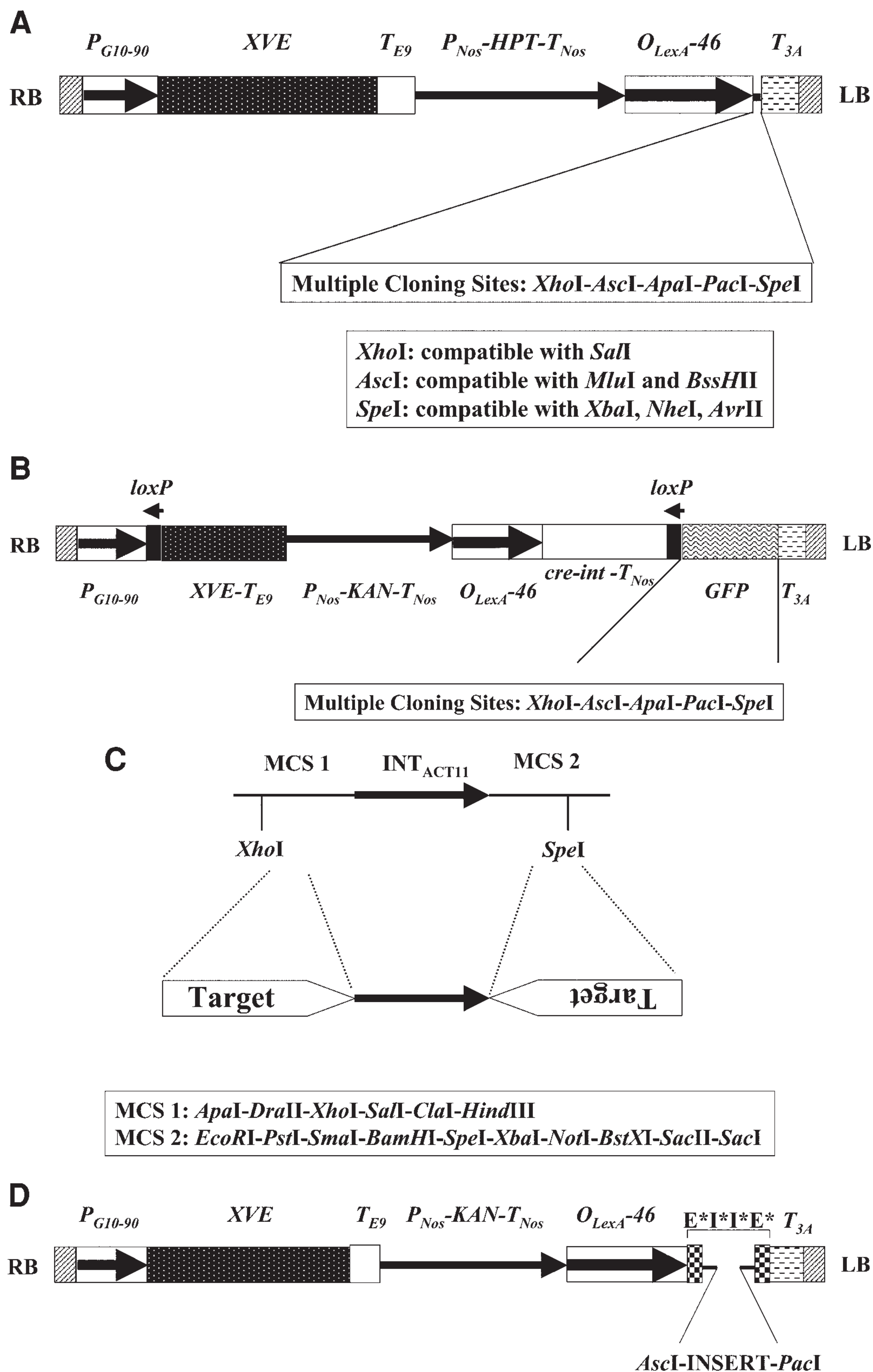
The original XVE vectors were designed to express a transgene in stably transformed plants ***(3)*** or cells ***(11)***. The XVE chimeric transcription factor, expression of which is controlled by the strong constitutive G10-90 promoter ***(12)***, is transcriptionally inactive in the absence of an inducer. When activated by the chemical inducer 17-β-estradiol, a mammalian hormone that has no detectable effects on plant growth and development, XVE specifically binds to the *LexA* operator sequence, thereby initiating transcription from the O^{LexA}*-46* promoter. A DNA fragment, either a cDNA or a genomic clone, can be cloned into the multiple cloning sites (MCS) downstream from the O^{LexA}*-46* promoter, thus allowing expression of the target gene in a 17-β-estradiol-dependent manner ***(3)***.

Using current plant transformation techniques, at least two events are associated with the expression of a transgene. Besides an elevation in the level of transgene transcript, transgene integration potentially disrupts another locus that might contribute to an altered phenotype in the transgenic line. An important advantage of inducible gene expression is that transgenic plants that have not been exposed to the inducer provide an internal control to rule out the possibility that any phenotypic consequences of transgene induction might result from positional effects arising from the site of transgene insertion in the genome.

The XVE vector pER8 (**Fig. 1A**) carries a hygromycin-selectable marker. pER10 is identical to pER8 except that the selectable marker confers resistance to kanamycin rather than hygromycin. These markers are used to identify transgenic plants or cells from non-transformed wild-type (WT) populations. All XVE vectors described in this chapter contain a spectinomycin-resistance marker for selection in *E. coli* and *Agrobacterium*.

3.1.2. Cloning

The *Xho*I and *SpeI* restriction sites in XVE vectors are most commonly used for cloning. On average, *Xho*I and *Spe*I cut *Arabidopsis* genomic sequences 6 and 5 Kb, respectively. The DNA ends left after digestion with these two enzymes are compatible with those resulting from digestion with many commonly used restriction enzymes (**Fig. 1A**), making it convenient to


A
P_{G10-90}
XVE
T_{E9}
P_{Nos}-HPT-T_{Nos}
O_{LexA}-46
T_{3A}
RB
LB
Multiple Cloning Sites: *Xho*I-*Asc*I-*Apa*I-*Pac*I-*Spe*I
*Xho*I: compatible with *Sal*I
*Asc*I: compatible with *Mlu*I and *BssH*II
*Spe*I: compatible with *Xba*I, *Nhe*I, *Avr*II
B
loxP
loxP
RB
LB
P_{G10-90}
XVE-T_{E9}
P_{Nos}-KAN-T_{Nos}
O_{LexA}-46
cre-int -T_{Nos}
GFP
T_{3A}
Multiple Cloning Sites: *Xho*I-*Asc*I-*Apa*I-*Pac*I-*Spe*I
C
MCS 1
INT_{ACT11}
MCS 2
*Xho*I
*Spe*I
Target
Target
MCS 1: *Apa*I-*Dra*II-*Xho*I-*Sal*I-*Cla*I-*Hind*III
MCS 2: *EcoR*I-*Pst*I-*Sma*I-*BamH*I-*Spe*I-*Xba*I-*Not*I-*BstX*I-*Sac*II-*Sac*I
D
P_{G10-90}
XVE
T_{E9}
P_{Nos}-KAN-T_{Nos}
O_{LexA}-46
E*I*I*E*
T_{3A}
RB
LB
*Asc*I-INSERT-*Pac*I

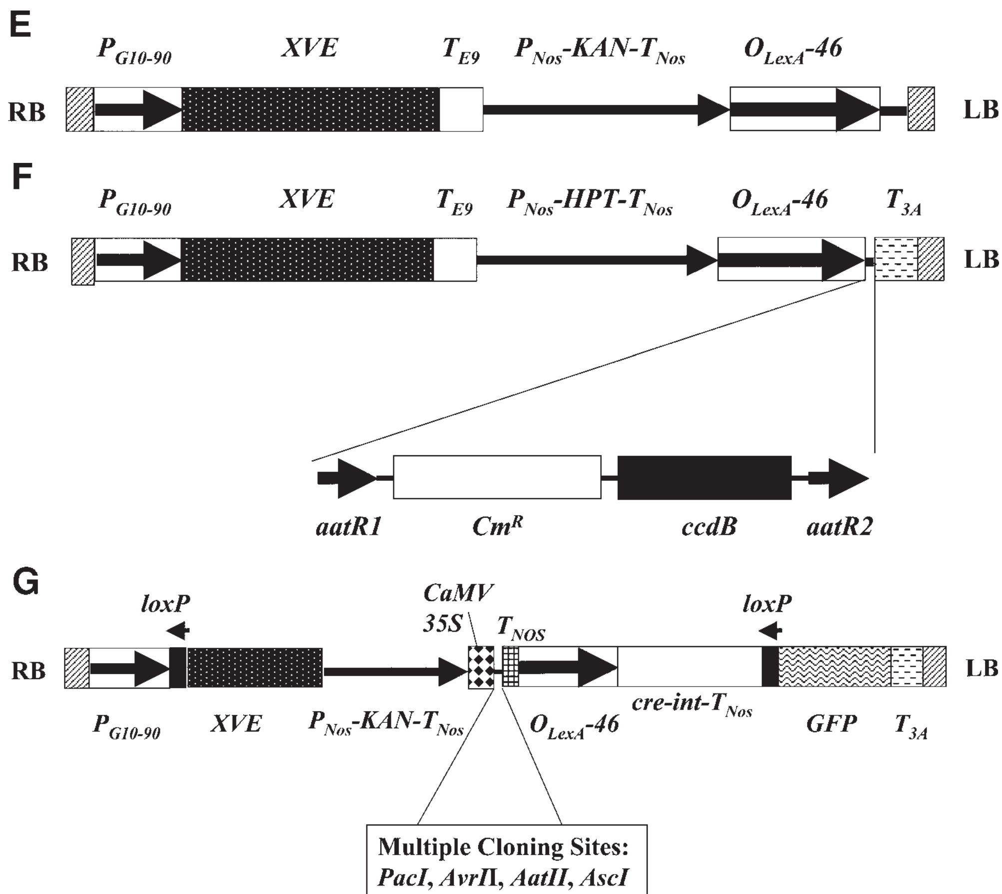


Fig. 1. Schematic diagrams of representative vectors used for applications of the XVE system. pER8, inducible expression (pER10 is a variant that confers kanamycin resistance in plants) *(3)*; pX6-GFP, inducible DNA excision (pX7-GFP is a variant that confers hygromycin resistance in plants *[5]*) *(4)*; pSK-int (used in conjunction with pX6-GFP or pX7-GFP), intermediate to prepare inverted repeats for inducible RNAi-mediated silencing *(5)*; pER10AD, regulated in vivo generation of full-length transcripts followed by cDNA generation in vitro *and* mapping of nitron/exon junctions *(6)*; pER16, inducible activation tagging *(7)*; pMDC7, gateway-compatible inducible overexpression *(9)*; pX8-GFP, conditional knockout of genes essential for gametophyte, embryo, or seedling viability. Abbreviations used: *aatR1*, *aatR2*, recombination sites for LR Clonase (Invitrogen); *ccdB*, encodes control of cell death; Cm^R, coding region of chloramphenicol resistance gene; cre-int, coding sequence of CRE recombinase interrupted by a plant intron; E*I*I*E*, the intron donor and acceptor sites from *AtABC1* separated by *Asc*I and *Pac*I sites for cloning the genomic sequence of interest; GFP, encodes green fluorescent protein optimized for expression in plants, *HPT*, encodes hygromycin phosphotransferase; INT_{ACT11}, third intron of *Arabidopsis ACTIN 11* gene; *KAN*, encodes neomycin phosphotransferase II; LB, left border of *Agrobacterium* T-DNA; *loxP*, recombination sites of CRE recombinase; O^{LexA}-*46*, eight copies of the LexA operator sequence fused to the -46 CaMV35S promoter; P_{G10-90}, G10-90 promoter; P_{NOS}, nopaline synthase promoter; RB, right border of *Agrobacterium* T-DNA; T_{3A}, *RBCS 3A* poly(A) addition sequence; T_{E9}, RBCS E9 poly(A) addition sequence; T_{NOS}, nopaline synthase poly(A) addition sequence; XVE, hybrid transactivator comprising the DNA binding domain of the bacterial repressor LexA, the acidic transcriptional activation domain of VP16, and the regulatory moiety of the human estrogen receptor.

choose alternative restriction enzymes if the DNA sequence to be cloned contains internal *Xho*I and/or *Spe*I sites. Appropriate restriction sites can be derived from commercial vectors (e.g., pBlueScript from Stratagene) or be introduced by adding linkers/adaptors through polymerase chain reaction (PCR)-mediated amplification. We provide two specific examples.

A *GFP* cDNA fragment released by *Xho*I/*Spe*I digestion of a vector described elsewhere ***(13)*** was gel-purified and then ligated with pER8 that had been digested with the same enzymes and gel-purified. The ligation mixture was transformed into *E. coli* DH5α cells by standard methods (*see* **Note 2**). The *E. coli* DH5α cells were then plated onto LB medium supplemented with spectinomycin (50–100 μg/mL), and incubated overnight at 37°C. Recombinant clones with correctly oriented *GFP* cDNA (pER8-GFP; ***3***) were identified by appropriate restriction digests and, if desired, by DNA sequencing following standard protocols. Routinely, we used XhoI/SpeI double digestion for the initial screening of candidate constructs, followed by digestion using an insert-specific internal site (*Nco*I recognition site in the *GFP* coding region) to check for the orientation of the insert.

To clone the *Arabidopsis AtIPT8/PGA22* genomic sequence into an XVE vector, we used two oligonucleotide primers to amplify the gene by PCR using genomic DNA as a template ***(8)***. The primers (*see* below for sequences), in addition to base pairing with 5'- (pga22F2) or 3'-end (pga22B2) of *AtIPT8/PGA22*, contained *Xho*I (pga22F2) and SpeI (pga22B2) sites, respectively. The PCR product was purified by phenol extraction/ethanol precipitation or by commercial kits (e.g., PCR Cleanup™, Qiagen, Hilden, Germany), and then double digested with *Xho*I and *Spe*I. After gel purification, the DNA fragment was ligated with *Xho*I/*Spe*I-digested, gel-purified pER10 vector. Identification of the recombinant plasmid pER10-PGA22 was carried out as described before.

pga22F2: 5' ggactagtgtcgactcgagCCGTATGAAATGTCCTTTGACATATCA
*Xho*I

pga22B2: 5' tctagagtcgacgactagtAAATCGAGGTGCAAAAATCTTAAACATC
*Spe*I

3.1.3. *Transformation of* Arabidopsis *by Floral Dipping and Identification of Transgenic Plants*

Agrobacterium-mediated transformation by floral dipping is the most convenient method for high efficiency transformation of *Arabidopsis*. Here is an outline of the protocol. For more details, *see* Chapter 17 in this book.

1. Dilute an overnight culture of the Agrobacterium strain of interest 1:1000 and grow the culture to stationary phase in liquid LB containing appropriate antibiotic selection (approx 20 h growth at 25–28°C, around 300 rpm).
2. Harvest the cells by centrifugation at room temperature (6000g) and resuspend in inoculation medium to a final OD_{600} of approx 1.0.
3. Invert *Arabidopsis* plants with many immature floral buds (bolts approximately 10 cm long) into the *Agrobacterium* suspension.
4. Remove the plants after approx 5 min incubation and shake off excess liquid.
5. Gently wrap the inoculated plates with transparent plastic film and incubate the plants in this humid microenvironment under low light conditions for 24 h before transfer to normal growth conditions.
6. Harvest seed collected from the inoculated plants and select transformants by germination on an appropriate selective medium.

3.1.4. *Transformation of* Arabidopsis *Root Explants and Regeneration of Transgenic Plants*

For certain applications (*see* **Subheading 3.5.**), transformation of root explants is more appropriate than transformation by floral dipping.

1. Surface-sterilize approximately 200 *Arabidopsis* seeds by vigorous agitation for 10 min in sterilization solution.
2. Remove the sterilization solution and wash the seeds with three changes of sterile distilled water.
3. Transfer 25 to 50 seeds to a 250-mL Erlenmeyer flask containing 100 mL of Gamborg's B5 medium. Incubate at 22°C on an orbital shaker (125 rpm) under a 16 h/8 h light/dark cycle.
4. Transfer 14-d-old seedlings to a sterile dish and separate the roots from the aerial tissue. Cut the roots into segments approx 10 mm in length. Wounded surfaces increase the efficiency of transformation.
5. Transfer clusters of root explants to solidified F1 medium. Ensure that all root segments are in direct contact with the medium. Incubate for 2 to 3 d at 22°C under a 16 h/ 8 h light/dark cycle. Hereafter, root explants can be stored on F1 medium at 4°C for up to 7 d.
6. Grow a culture of the *Agrobacterium* strain of interest to stationary phase in liquid LB containing appropriate antibiotic selection (approx 20 h growth at 25–28°C, around 300 rpm).
7. Harvest the cells by centrifugation at room temperature (6000*g*) and resuspend in LB medium to a final OD_{600} of approx 1.0. Repeat this step once to remove any residual antibiotics.
8. Resuspend the cells in 30 mL of Gamborg's B5 medium.
9. Mix the root explants on the F1 medium with the *Agrobacterium* suspension and incubate at room temperature for 5 to 10 min.
10. Filter the mixture of explants and *Agrobacterium* through a single layer of cheesecloth at the base of a sterile basket and drain off excess medium by transferring the explants to several layers of sterile filter paper.
11. Transfer the clusters of root explants to F2 medium and incubate for 2 to 3 d at 22°C under a 16 h/8 h light/dark cycle. By this stage the explants should be surrounded by a thin layer of *Agrobacterium*, but should not be covered by bacteria.
12. Rinse the infected explants in sterile water containing 100 mg/L carbenicillin. Collect the washed explants in a sterile basket filter and repeat the washes until the filtrate is no longer cloudy.
13. Warm SR' medium containing the appropriate selective agent(s) to 40 to 45°C, add the infected and washed explants and quickly pour onto SR medium containing the appropriate selective agent(s).
14. Incubate at 22°C under a 16 h/8 h light/dark cycle. Transformed shoots should be regenerated after 2 to 3 wk and can be transferred to RI medium to promote root growth.

3.1.5. Induction of Transgene Expression and Phenotypic Characterization

When a transgenic locus is heterozygous, identification of transgenic lines with high-level transgene expression and subsequent functional characterization can be done in the T1 generation. Induction of transgene expression can be performed by either direct germination of transgenic seeds on the 17-β-estradiol-containing medium or application of the inducer after germination. For initial screening of transgenic lines with high levels of transgene expression, we suggest germinating the transgenic seeds directly on MS media containing 10 μ*M* 17-β-estradiol (*see* **Note 3**). After germination on 17-β-estradiol-containing medium, the GFP expression in pER8-GFP transgenic plants is easily detectable by examination using a fluorescence microscope (*see* **Fig. 2** in ref. ***3***).

Analogously, the expression levels of *AtIPT8/PGA22* can be conveniently assessed by the growth phenotype ***(8)***. This study also exemplifies the importance of always testing a series of inducer concentrations to assess the sensitivity of the system to changes in transcript levels. It is strongly recommended that after initial experiments using 10 μ*M* 17-β-estradiol, researchers assess the effects of both higher and lower concentrations of inducer. Not only should preliminary data from phenotypic analyses of selected lines be confirmed by Northern blot analysis or reverse-transcription PCR (RT-PCR), but a significant advantage of the XVE system is that transgene expression levels and phenotypic effects should correlate with increasing concentrations of inducer. This enables researchers to generate the equivalent of an allelic series of mutations of different severity and thereby infer a tight dependence of the observed phenotype on the transgene in question.

In many cases, however, no growth phenotype is apparent after induced overexpression of a target gene. In such cases, Northern blot analysis or RT-PCR should be used to identify transgenic lines with high levels of transgene expression before further studies are initiated. The same confirmation of dose-dependent responsiveness outlined above is recommended.

1. Germinate transgenic T1 seeds on MS medium containing the appropriate antibiotic for selection of transformants (50 μg/mL hygromycin or kanamycin for pER8 or pER10 derivatives, respectively).
2. Transfer transgenic plants onto MS medium without antibiotics for better growth. The genes that confer hygromycin and kanamycin resistance in pER8 and pER10 respectively are under the regulation of the nopaline synthase (NOS) promoter, which is weaker than the CaMV35S promoter. Consequently, maintenance on selective media for an extended time (e.g., more than 2-3 wk), especially on kanamycin-containing medium, retards growth of the transgenic plants and may cause the appearance of bleached spots on the leaves.
3. Spray 10 μM 17-β-estradiol solution onto transgenic plants in a laminar flow biological containment hood (5 mL for a 90-mm Petri dish). Alternatively, add the inducer directly onto the plate and then gently agitate the plate to make sure the inducer solution is evenly distributed over the surface of the medium. A single application of inducer is normally sufficient to induce robust induction of transgene expression, although repeated applications at 3 to 5 d intervals may be preferable in other cases. For certain purposes, e.g., those investigating stress responses, spraying or spreading the inducer on plants may be undesirable if this elicits a stress reaction. In such cases, transfer of the plants to freshly-prepared medium containing 10 μ*M* 17-β-estradiol (*see* **Note 3**) is recommended.
4. Culture for an additional 12 to 16 h.
5. Harvest the materials for RNA or SDS/PAGE sample preparation. The material can also be frozen in liquid nitrogen and then stored at –80°C. Although transgene expression can usually be easily detected by Northern analysis or RT-PCR after 2 to 3 h of induction, it is best to perform the induction for 12 to 16 h in initial experiments.
6. Perform Northern blot or RT-PCR analysis using total RNA prepared from the 17-β-estradiol-treated plants and untreated control plants. If antibodies for the target gene product are available, a Western blot analysis will be more convenient to screen the transgenic lines. Routinely, inducer-dependent transgene expression should be seen in more than 50% transgenic lines with a range of expression levels.
7. Select representative independent lines with different transgene expression levels (low, medium, and high) for phenotypic analysis under the appropriate experimental conditions.
8. To analyze phenotypes associated with late growth and development stages (e.g., the flowering phenotype), induce the transgene of interest by directly spraying 10 μ*M* 17-β-estradiol solution onto the transgenic plants (10–15 mL per plant). A single application of inducer is normally sufficient to induce robust induction of transgene expression, although repeated application at 3 to 5 d intervals may be necessary for certain applications.
9. In all analyses, use noninduced plants as internal controls. This eliminates the possibility that any phenotypes arise from positional effects. We also recommend including lines transformed with an empty vector in induction experiments. Although in our experience we have never observed any physiological effects of 17-β-estradiol on plants, this control confirms that any phenotypic effects do not arise from inducer treatment.

3.2. Inducible DNA Excision

Inducible DNA excision was originally developed for the removal of a selectable marker gene from a transgenic genome ***(4)***. Subsequently, the system was extended to other applications (*see* **Subheadings 3.3.** and **3.6.**).

3.2.1. DNA Excision Vectors pX6 and pX7

These are modified to incorporate the CRE-*lox* site-specific recombination system of bacteriophage P1. Expression of the CRE protein causes recombination between *loxP* sites. If the

loxP sites are arranged in the same orientation as a direct repeat, recombination results in the deletion of the intervening DNA. As shown in **Fig. 1B**, a DNA segment encoding the CRE DNA recombinase was placed downstream from the O^{LexA}-*46* promoter, allowing expression of the gene under the control of the XVE system. Along with the *CRE* transcription cassette, an XVE chimeric gene and an antibiotic selectable marker (a kanamycin resistance gene in the case of pX6) were flanked by two *loxP* sites, which can be specifically recognized and cleaved by CRE. Upon induction by 17-β-estradiol, the DNA segment between the two *loxP* sites is excised from the host genome, leading to a direct fusion of the *G10-90* promoter and a downstream promoter less gene (*GFP* in this case), which in turn activates the reporter gene. Further details of the CLX system (*Cre/loxP* DNA excision system controlled by XVE) are provided elsewhere ***(4)***.

An inducible expression system can overcome problems associated with lethality during the generation of transgenic plants, but the transient expression nature also limits its use in certain cases, such as when sustained transgene expression is required. Therefore, the CLX system is particularly useful for inducible, permanent activation of a transgene. For this purpose, a MCS identical to that in XVE vectors (**Fig. 1A**) was inserted outside of a loxP site in an orientation such that the excision will lead to a direction fusion of the strong G10-90 promoter with a target gene. In **Fig. 1B,** a *GFP* cDNA was inserted into the *Xho*I/*Spe*I sites of the MCS ***(4)***. The pX7 vector is identical to pX6 except that the kanamycin resistance marker gene has been replaced with the gene conferring resistance to hygromycin ***(5)***.

3.2.2. Cloning and Plant Transformation

The same protocols are used as for the XVE vectors (*see* **Subheadings 3.1.2.** and **3.1.3.**).

3.2.3. Phenotypic Characterization and Genetic Analysis

1. Germinate pX6-GFP T1 transgenic seeds on an MS medium containing 10 μ*M* 17-β-estradiol. No antibiotics should be added in the medium, as successful removal of marker genes will result in the loss of resistance to the selection, thus killing target (marker-free) transgenic plants.
2. In parallel, germinate approx 100 to 200 T1 seeds on kanamycin-containing medium to estimate the number of transgene insertions.
3. Randomly pick 5 to 7-d-old seedlings and assess GFP expression using a fluorescence microscope. A uniform GFP signal in the roots indicates successful DNA excision from the somatic tissue. Plants grown on an inducer-free medium (i.e., from the kanamycin medium in **Step 2**) should yield no GFP signal.
4. Estimate the somatic excision efficiency. Based on T-DNA insertion number(s) (**step 2**), calculate expected GFP positive T1 plants from the test population, and estimate somatic excision efficiency using the formula:

   Observed GPF-positive seedlings/expected GFP-positive seedlings × 100%

   In our hands, the somatic excision efficiency is 100%, but may vary depending on whether multiple tandem insertions of the transgene cassette cause more complicated rearrangements between *loxP* sites.
5. Transfer GFP-positive seedlings to soil and harvest T2 seeds from individual plants. Also transfer kanamycin-resistant seedlings from **step 2** to soil and collect T2 seeds as a control in later experiments.
6. Plate T2 seeds on an MS medium with or without kanamycin. For T2 seeds collected from uninduced T1 plants, typical Mendelian segregation for the kanamycin maker gene should be observed, but no GFP signal should be expected. For T2 seeds harvested from GFP positive T1 plants, irregular Mendelian segregation for both kanamycin marker and GFP signal should be observed.
7. Calculate germline cell excision efficiency according to the formula:

   Observed GFP-positives/Expected GFP positives × 100%

   The germline excision efficiency may vary from 29 to 66% ***(4)***.

3.3. Gene Silencing by Inducible RNAi

Functional annotation of a gene by reverse genetic approach routinely requires knocking out or knocking down the expression of candidate genes. Unfortunately, specific disruption of a candidate gene by homologous recombination in higher plants is technically challenging. Recently, RNAi-mediated gene silencing has become a powerful tool in plant functional genetics. Nonetheless, as in overexpression studies, regulation of an RNAi transcript by a constitutive promoter may retard growth and development or even be lethal. An approach to ensure inducible RNAi-mediated gene silencing is thus highly desirable in functional studies of genes where loss-of-function mutations confer gametophyte or embryo lethality. The protocol described below has been used to efficiently silence both a transgene and an endogenous gene (*5*).

3.3.1. pSK-int and CLX Vectors pX6 and pX7

Efficient RNAi requires the formation of an intron-containing self-complementary "hairpin" RNA molecule. Therefore, an RNAi construct should contain a single transcription unit consisting of two inverted copies of a target gene separated by an intron. The intron provides the loop structure of the hairpin. At least two cloning steps are required to prepare an RNAi construct for stable integration into the plant genome. Owing to the larger size of most binary vectors for plant transformation, we routinely first clone the RNAi fragment in both orientations into a derivative of the pBlueScript (Stratagene) vector that contains a plant intron in the center of the MCS. The tripartite insert (intron flanked by both orientations of RNAi fragment) is then excised and cloned into a vector of the CLX series (i.e., a three-step procedure).

The pSK-int vector was made by inserting a PCR-generated DNA fragment containing intron 3 of the *Arabidopsis ACTIN11* gene into the EcoRV/EcoRI sites of pBlueScript-SK (**Fig. 1C; 5**). Consensus sequences for the donor and acceptor sites of plant gene splicing were added at the two borders of the intron.

DNA excision vectors pX6 or pX7 have been described previously (*see* **Subheading 3.2.1.**).

3.3.2. Cloning

Inverted copies of a target gene can be inserted into multiple restriction sites flanking the intron via a two-step cloning approach. The inverted-repeat-intron DNA fragment is then released and cloned into pX6 or pX7.

1. PCR-amplify a 300 to 500 bp-long DNA fragment containing the transcribed sequence of the target gene. When attempting specific silencing of a single gene, it is advisable to ensure that the amplified fragment bears minimal homology to any other transcribed sequence in the genome. Add appropriate restriction sites in the primers that can be used for cloning the fragment into the 5' arm of pSK-int (MCS 1 in **Fig. 1C**).
2. Digest with the appropriate restriction enzymes and clone into MCS 1 in an orientation as shown in **Fig. 1C**.
3. Using a similar approach, clone the 3' arm into MCS 2 in an inverted orientation relative to the 5' arm. Note that inverted repeats may be rearranged or deleted by *E. coli* DNA repair systems. Depending on the construct in question, it may be prudent to conduct the subsequent two steps in an *E. coli* strain defective in the UV repair (*uvrC*) and SOS repair (*umuC*) pathways. We routinely use the SURE-2 strain (Stratagene) to circumvent potential difficulties in cloning inverted repeats. Alternatively, should difficulties arise, it might be necessary to attempt using a different target from the transcribed sequence.
4. Release the inverted-repeat-intron sequence from pSK-int by *Xho*I/*Spe*I digestion (or digestion with compatible enzymes), and clone into the same sites of pX6 or pX7. Confirm the insertion and orientation by restriction digestion. Note that the secondary structure of the insert may prevent its sequencing.

3.3.3. Plant Transformation and Identification of Transgenic Plants

The same protocols are used as for the XVE vectors (*see* **Subheading 3.1.3.**).

3.3.4. Phenotypic Characterization and Genetic Analysis

As the same vector backbone is used in inducible DNA excision vectors to allow the permanent activation of any target gene, genetic analysis of the pX6- or pX7-RNAi transgenic plants is basically identical to that for pX6-GFP plants (*see* **Subheading 3.2.3.**). We emphasize again that all antibiotics should be omitted from the induction medium, as a successful DNA excision leads to the loss of the marker genes.

Phenotypic analysis of the resulting RNAi transgenic plants will vary depending on the gene under investigation. Examples for detailed phenotypic characterization of inducible RNAi transgenic plants can be found elsewhere ***(5)***.

3.4. Regulated In Vivo cDNA Generation

While isolation of full-length cDNAs is critical for many aspects of functional analysis of a gene of interest, this may be technically demanding for many low-abundance transcripts. The XVE inducible expression system has been successfully used for regulated in vivo cDNA generation (RIDE; ***6***). The greater capacity of the XVE system for transcriptional activation in comparison with the CaMV35S promoter ***(3)*** makes it the promoter system of choice when isolating cDNAs of extremely labile mRNAs or transcripts that cause embryo lethality when expressed above an extremely low threshold.

In its simplest form, a genomic DNA fragment encompassing the full-length or partial coding sequence of the target gene (including introns) is cloned into an XVE vector (either pER8 or pER10) as described before. The resulting construct is then transformed into *Arabidopsis* or tobacco plants. Following induction by 17-β-estradiol, total RNA is prepared and the target cDNA is then obtained by RT-PCR with the use of appropriate primer pairs ***(6)***. The XVE vectors, molecular cloning and generation of transgenic plants are essentially the same as described before (*see* **Subheadings 3.1.1–3.1.3**). The following protocol outlines an approach to generate cDNA clones from 17-β-estradiol-treated transgenic plants carrying an XVE-genomic DNA construct (*see* **Note 4**). However, pER10 has also been adapted to enable less directed post-transcriptional processing of any genomic insert. Two plant intron/exon junctions separated by *Asc*I and *Pac*I sites were introduced downstream of the O^{LexA}*-46* promoter to generate the vector pER10AD (**Fig. 1D**). Genomic libraries prepared in pER10AD should not only facilitate less time-consuming isolation of rare transcripts, but also complement conventional expressed sequence tag approaches to characterization of the entire plant transcriptome.

1. Germinate transgenic T1 seeds on MS medium supplemented with an appropriate antibiotic.
2. Treat 2 to 3 wk-old antibiotic-resistant seedlings with 10 m*M* 17-β-estradiol for 12 to 16 h.
3. Prepare total RNA from the 17-β-estradiol-treated seedlings by standard methods ***(10)*** or by using a commercial kit (e.g., Plant RNAeasy™ Kit, Qiagen, Hilden, Germany).
4. Synthesize the first strand cDNA primed with oligodT using SuperScript II according to the manufacturer's instructions.
5. PCR-amplify the first strand cDNA product with PWO DNA polymerase (Roche).
6. Run a small aliquot of PCR products on a mini agarose gel along with appropriate molecular markers to verify the size of the amplified cDNA fragment.
7. Clone the RT-PCR product into a pCRScript vector (Stratagene).
8. Identify and characterize the target cDNA clone by restriction digestion and DNA sequencing.

3.5. Inducible Activation Tagging and Functional Cloning

The identification and characterization of loss-of-function mutations is the classic approach to determination of gene function. However, the success of this method is severely challenged by the considerable functional redundancy within many plant gene families. In such cases, characterization of gain-of-function phenotypes is a powerful alternative and overexpression of candidate genes is the most common approach. As discussed above, constitutive overexpression of a gene may cause detrimental or even lethal effects on plant growth and develop-

ment or plantlet regeneration. To circumvent these concerns, the following protocols employ an inducible expression system for activation tagging and functional cloning.

3.5.1. Vectors and Cloning

Both pER16 *(7)* and pMDC7 *(9)* are XVE vector variants. The pER16 vector (**Fig. 1E**), in which the T_{3A} transcriptional terminator sequence in the O^{LexA}-*46*- T_{3A} cassette was deleted, has been used for the functional cloning of *PGA* (Plant Growth Activator) genes *(7,8)* and the generation of activation tagging T-DNA lines in Arabidopsis (J. Zhang, J. Xu, and J. Zuo, unpublished data). When inserted into a host genome, the O^{LexA}-*46* promoter is able to activate expression of a downstream gene in a 17-β-estradiol-dependent fashion. In addition, the inserted OLexA sequence can also act as an inducible enhancer to activate a target gene as far as 8 Kb from the site of integration (*7,8*; Q.-W. Niu, J. Zuo and N.-H. Chua, unpublished data).

Laborious conventional cloning technology places a considerable limitation on progress in plant functional genetics. The Gateway system enables the rapid creation of a range of constructs (fusion to various promoters, reporter genes or peptide tags for subcellular targeting or immunodetection), starting with a single donor construct. The pMDC7 vector *(9)* was made by inserting an *att* cassette (Invitrogen) into the MCS of pER8 (**Fig. 1F**), yielding a destination vector. In the presence of LR Clonase™ (Invitrogen), the *aatR1* and *aatR2* sites in the cassette mediate site-specific DNA recombination between the destination vector and an entry vector, which carries the target DNA fragment flanked by two *aat* sites. The lethal *ccdB* (*control of cell death B*) gene in the destination vector prevents growth of most *E. coli* strains, thereby allowing counterselection of successful recombinant plasmids. The vector was designed, in the combination with the Gateway technology (Invitrogen), for high-throughput functional analysis of plant genes. Technical details of using the Gateway Entry vectors and LR Clonase™/*att*-mediated site-specific DNA recombination can be found in the user's manual provided by the manufacturer. An excellent example is also provided elsewhere *(9)*.

3.5.2. Functional Cloning of PGA Genes

The pER16 vector has been successfully used to identify genes that are involved in signaling pathways that mediate somatic embryogenesis *(7)* and responses to the phytohormone cytokinin *(8)*. Rationales and principles of these screens have been discussed in detail elsewhere *(14,15)*. The following protocol was originally designed to screen for *pga* mutants *(7,8)*, but can be modified for other screens.

1. Transform *Agrobacterium* strain *ABI* with pER16. Certain strains (e.g., *GV3010*) yield a high background in the screen described below (B. Zheng and J. Zuo, unpublished data). Presumably, this results from higher levels of cytokinin production by these *Agrobacterium* strains.
2. Transform root or hypocotyl explants derived from wild type seedlings with an Agrobacterium culture carrying pER16 using standard protocols (*see* **Subheading 3.1.4.**).
3. Suspend *Agrobacterium*-infected explants in SCM' medium prewarmed to 40°C or less, and then pour the mixture onto the SCM medium (*see* **Note 5**). Also place a small portion of explants onto SR medium (full medium) to estimate the efficiency of transformation.
4. Culture the explants on SCM for 2 to 3 wk. Pick putative positive clones (green or green-yellowish calli or green shoots) and transfer onto SR (calli) or RI medium (green shoots) to allow shoot regeneration or rooting (2 to 3 wk).
5. Culture calli-derived shoots on RI medium for 2 to 3 wk for rooting.
6. Transfer well-developed shoots into soil until maturation and setting T1 seeds.
7. Transfer shoots with poor rooting or no root formation into phytacons containing M medium to set seeds in vitro.
8. Use explants derived from individual T1 plants to verify the phenotype on the SCM medium.

3.6. Inducible Conditional Knockout

Despite the usefulness of inducible RNAi-mediated silencing in activating sequence-specific degradation of transcripts that are essential for gametophyte, embryo, or seedling viability

(*see* **Subheading 3.3.**), there may be certain instances in which an alternative approach is needed to assess the roles of these transcripts later in the life cycle. First, RNAi-mediated silencing is seldom likely to be absolute. This may complicate studies involving assessment of whether certain responses are cell-autonomous or dependent on intercellular signal transmission. Second, the ability of RNAi to interfere with the silencing of transcripts homologous to the target complicates the interpretation of experiments to assess the function of members of highly conserved gene families.

The pX8-GFP vector (**Fig. 1G**) is a derivative of pX6-GFP in which a 1.8-kb fragment from the binary vector pBI121 (Clontech, Palo Alto, CA) that includes the nopaline synthase terminator (NosT) and CaMV35S promoter was cloned into the unique *Apa*I site between the coding region of *NPTII* and the NosT sequence in pX6-GFP. The pBI121-derived fragment was modified to contain *Apa*I, *Afe*I, and *Swa*I sites on the NosT end and *Pac*I, *Avr*II, *AatII*, *AscI,* and *ApaI* sites at the CaMV35S end of the insert. *AfeI*, *SwaI*, *PacI*, *Avr*II, *AatII,* and *AscI* are all unique sites within pX8-GFP.

Null mutants of *cop* (constitutively photomorphogenic) mutants germinate but are seedling lethal. We have successfully complemented the *cop10-1* and *cop1-5* phenotypes by Agrobacterium-mediated transformation of root explants from homozygous seedlings with *COP10* and *COP1-6xmyc* transgenes cloned into the unique PacI and AscI sites in pX8-GFP (P. Hare, L.-F. Huang and Nam-Hai Chua, unpublished data). Selected tissues of the adult progeny of these fertile T1 plants will be treated with 17-β-estradiol and characterization of these decomplemented mutants will enable assessment of COP10 and COP1 functions during growth and development after the seedling stage (P. Hare, L.-F. Huang, and Nam-Hai Chua, work in progress).

An advantage of the conditional "decomplementation" of a null mutant over conditional induction of RNAi-mediated silencing is that the former approach permits assessment of the extent to which excision has occurred by virtue of GFP fluorescence in mutant cells that are no longer complemented.

3.6.1. Plant Transformation, Genetic Analysis, and Phenotypic Characterization

Homozygous mutants that cause seedling lethality may be transformed by cocultivation of root or hypocotyl explants with transformed *Agrobacterium* strains (*see* **Subheading 3.1.4.**). In instances in which a null mutation in a gene of interest causes gametophyte or embryo lethality, transformation of tissues that are homozygous for the mutation is not an option. The more time- and labor-intensive alternative involves floral dip-mediated transformation of heterozygous lines (*see* **Subheading 3.1.3.**) with the pX8-GFP derivative containing the transgene of interest, selection of kanamycin-resistant transformants, and screening for their complemented homozygous progeny by PCR analysis.

4. Notes

1. The original versions of XVE vectors do not work well in rice because of poor inducible expression of transgenes, presumably resulting from poor expression of the synthetic *G10-90* promoter in rice. The problem can be addressed by replacing *G10-90* with the rice actin1 promoter.
2. Many commonly used *E. coli* strains show partial or strong resistance to spectinomycin (e.g., DH10B). We routinely use DH5α in our cloning experiments with XVE vectors that all carry a spectinomycin-resistant marker gene.
3. One-thousand fold dilutions of the 17-β-estradiol solution suitable for application to plant tissues are prepared in distilled water. Freshly prepared media (less than 3 d old and stored in darkness at 4°C) ensures the best induction of gene expression.
4. If no target cDNA clones are obtained from an experiment, given that positive controls for RT-PCR work well, Northern blot analysis should be used to check expression of the transgene induced by 17-β-estradiol. In such cases, 5 to 10 independent transgenic lines should be screened for a high level of expression of the target.
5. *Agrobacterium*-infected root explants can be placed directly on SCM medium. However, mixing the explants with the SCM' matrix normally results in more efficient transformation.

Acknowledgments

We would thank the *Arabidopsis* Biological Resources Center (Columbus, OH) for DNA and seeds, and Qi-Wen Niu for excellent technical assistance. Colleagues in the Zuo and Chua labs who contributed their unpublished data are greatly appreciated. Work in J. Z.'s laboratory was supported by grants from the Chinese Academy of Sciences (CAS; grant number: KSCX2-SW-308), National Natural Science Foundation of China (NSFC; grant numbers: 0270142, 30221002 and 30330360), and the Ministry of Science and Technology of China (grant number: 2001AA225021). J. Z. is a Bairen Jihua (Young Starter) fellow of CAS and a recipient of the NSFC Outstanding Young Investigator Award.

References

1. Gatz, C. and Lenk, I. (1998) Promoters that respond to chemical inducers. *Trends Plant Sci.* **3,** 352–358.
2. Zuo, J. and Chua, N.-H. (2000) Chemical-inducible systems for regulated expression of plant genes. *Curr. Opin. Biotechnol.* **11,** 146–151.
3. Zuo, J., Niu, Q. W., and Chua, N.-H. (2000) An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J.* **24,** 265–273.
4. Zuo, J., Niu, Q. W, Moller, S. G., and Chua, N.-H. (2001) Chemical-regulated, site-specific DNA excision in transgenic plants. *Nat. Biotechnol.* **19,** 157–161.
5. Guo, H. S., Fei, J. F., Xie, Q., and Chua, N.-H. (2003) A chemical-regulated inducible RNAi system in plants. *Plant J.* **34,** 383–392.
6. Møller, S. G. and Chua, N.-H. (2002) Chemical regulated production of cDNAs from genomic DNA fragments in plants. *Plant J.* **32,** 615–622.
7. Zuo, J., Niu Q. W., Frugis G., and Chua N.-H. (2002) The *WUSCHEL* gene promotes vegetative-to-embryonic transition in *Arabidopsis. Plant J.* **30,** 349–359.
8. Sun, J., Niu, Q. W., Tarkowski, P., et al. (2003) The *Arabidopsis AtIPT8/PGA22* gene encodes an isopentenyl transferase that is involved in de novo cytokinin biosynthesis. *Plant Physiol.* **131,** 167–176.
9. Curtis, M. D. and Grossniklaus, U. (2003) A Gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133,** 462–469.
10. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
11. Zuo, J., Niu, Q.-W., Nishizawa, N., Wu, Y., Kost, B., and Chua, N.-H. (2000) KORRIGAN, an *Arabidopsis* endo-1,4-β-glucanase, localizes to the cell plate by polarized targeting and is essential for cytokinesis. *Plant Cell* **12,** 1137–1152.
12. Ishige, F., Takaichi, M., Foster, R., Chua, N.-H., and Oeda, K. (1999) A G-box motif (GCCACGTGCC) tetramer confers high-level constitutive expression in dicot and monocot plants. *Plant J.* **18,** 443–448.
13. Kost, B., Spielhofer, P., and Chua, N.-H. (1998) A GFP-mouse talin fusion protein labels plant actin filaments in vivo and visualizes the actin cytoskeleton in growing pollen tubes. *Plant J.* **16,** 393–401.
14. Kakimoto, T. (1996) CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. *Science* **274,** 982–985.
15. Inoue, T., Higuchi, M., Hashimoto, Y., et al. (2001) Identification of CRE1 as a cytokinin receptor from *Arabidopsis. Nature* **409,** 1060–1063.

27

RNA Extraction

Huazhong Shi and Ray Bressan

Summary

RNA extraction is a routine technique in a molecular biology lab. High quality of RNA extracted from plants is a prerequisite to succeeding in subsequent experiments such as reverse transcription-polymerase chain reaction, Northern hybridization, cDNA library construction, and microarray analysis. Two methods for RNA extraction in *Arabidopsis*, small-scale and large-scale, are described here. RNA isolated by using the small-scale method can be applied in most downstream applications. The large-scale RNA preparation is recommended to extract RNA for Northern hybridization only.

Key Words: RNA extraction; gene expression; RNase inhibitor; *Arabidopsis*.

1. Introduction

Genetic information flows from the sequences of DNA to RNA and then into protein sequence. RNA is an information bridge connecting DNA to protein in gene expression. Analysis of RNA level provides important information on gene expression and regulation. RNA isolation and purification have been routine molecular techniques for such manipulations as Northern hybridization, reverse trasncription-polymerase chain reaction (RT-PCR), rapid amplifcation of cDNA ends (RACE), in vitro translation and cDNA cloning. In the past several years, considerable progress has been made regarding the ease and success of RNA isolation from plant tissues. However, beginning researchers still face a serious problem of RNA degradation during RNA extraction due to the presence of RNase activity that is difficult to inhibit. A number of methods have been developed and numerous kits are available from biotech companies for both small- and large-scale isolation of RNA ***(1–3)***. Here we describe two methods for total RNA extraction. These methods work well with *Arabidopsis* according to our experience and should be also suitable for RNA isolation in most other plant species. However, caution must be taken when applying this)methodology to different plant species and tissues, in particular those that have high levels of endogenous RNase activity, polysaccharide content, or various secondary metabolites. The first method is recommended for small-scale projects but can be scaled up as needed ***(3)***. This method is simple and allows the simultaneous processing of a large number of samples. The isolated total RNA obtained by this method can be used for most RNA analyses. The second method is designed for large-scale RNA extraction. This method is relatively time-consuming, but is economic for processing large numbers of samples. Through the use of effective RNase inhibitors in the extraction buffer, RNA degradation can be minimized. Because of the inhibition of PCR reactions by RNase inhibitors, the isolated RNA obtained from the second method is strongly recommended for use only with Northern blotting, or perhaps for mRNA purification. Almost universally, mRNA purification to isolate polyA RNA is performed by using commercially available kits. Therefore, researchers can refer to the protocols recommended in any of the excellent available kits (e.g., Promega, PolyATtract® System 1000; Stratagene, PolyA Quik® mRNA isolation kit).

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

2. Materials

2.1. Small-Scale RNA Isolation

1. Disposable plastic micropestles.
2. Microcentrifuge tubes.
3. Tissue samples in liquid nitrogen.
4. TRIzol (Invitrogen, cat. no. 15596-026).
5. Chloroform.
6. Isopropyl alcohol.
7. Diethylpyrocarbonate (DEPC).
8. RNase-free water (*see* **Note 1**).
9. 75% ethanol in RNase-free water.

2.2. Large-Scale RNA Preparation

1. Mortar and pestle.
2. Microspoon.
3. 15-mL screw-cap polypropylene centrifuge tubes.
4. Liquid nitrogen.
5. 100 m*M* stock solution of aurintricarboxylic acid (ATA) (*see* **Note 2**).
6. Extraction buffer: 50 m*M* Tris-HCl, pH 8.0, 300 m*M* NaCl, 5 m*M* EDTA, pH 8.0, 2% sodium docecyl sulfate (SDS), 2 m*M* ATA, 10 m*M* β-mercaptoethanol (*see* **Note 3**).
7. 1 *M* KCl, 8 *M* LiCl.
8. Diethylpyrocarbonate (DEPC).
9. RNase-free water (*see* **Note 1**).
10. TE-saturated phenol (*see* **Note 4**).
11. Chloroform.
12. 5 *M* NaCl.
13. 75% ethanol.

3. Methods

3.1. Small-Scale RNA Isolation

1. Harvest approx 100 mg plant tissue, wrap in aluminum foil, and immediately place in liquid nitrogen. The samples can be stored at –70°C for future use (*see* **Note 5**).
2. Prechill 1.5-mL microcentrifuge tubes and plastic micropestles in liquid nitrogen. Place 100 mg frozen tissue into a prechilled microcentrifuge tube and grind briefly by using a prechilled micropestle (*see* **Note 6**).
3. Add 600 μL of TRIzol reagent and grind to form a fine homogenate. Homogenization can be assisted by using a motor-driven pestle. Add another 400 μL of TRIzol and grind briefly (*see* **Note 7**).
4. Put the tube on ice if processing multiple samples.
5. Incubate the homogenized sample for 5 min at room temperature.
6. Add 0.2 mL chloroform, shake tubes vigorously by hand for 15 s, incubate at room temperature for 3 min and centrifuge at 12,000*g* for 15 min at 4°C.
7. Transfer the aqueous phase to a fresh tube. Do not transfer any material from the middle layer. Add 0.5 mL of isopropyl alcohol and mix. Incubate at room temperature for 10 min.
8. Centrifuge at 12,000*g* for 10 min at 4°C. The RNA forms a gel-like pellet on the side and bottom of the tube.
9. Pour off the supernatant. Add 1 mL of 75% ethanol and mix with vortex. Centrifuge at 7500*g* for 5 min at 4°C.
10. Discard ethanol. Dry the RNA pellet at room temperature in a SpeedVac for 5 to 10 min.
11. Dissolve the RNA in 30 to 50 μL of RNase-free water and store at –70°C (*see* **Note 8**).

3.2. Large-Scale RNA Preparation (*see* **Note 9**)

1. Prechill mortar and pestle with liquid nitrogen.
2. Place 2 g of frozen tissues into mortar and grind into a fine powder (*see* **Note 10**).

3. Transfer the powder into a 15-mL tube with 4.5 mL extraction buffer by using a prechilled microspoon (*see* **Note 11**).
4. Mix immediately by hand and vortex briefly. Place the tube on ice if multiple samples will be processed.
5. Add 700 μL cold 3 *M* KCl. Mix by shaking and incubate them on ice for 15 min.
6. Centrifuge at 8000*g* for 20 min at 4°C.
7. Transfer the supernatant into a new 15-mL tube. Add 2 mL of 8 *M* LiCl and mix. Incubate at 4°C for more than 4 h (*see* **Note 12**).
8. Centrifuge at 8000*g* for 20 min at 4°C.
9. Discard the supernatant and add 2 mL of RNase-free water. Resuspend the pellet by vortexing.
10. Add 2 mL of phenol:chloroform (1:1) and mix by vortexing. Centrifuge at 8000*g* for 10 min at 4°C (*see* **Note 13**).
11. Transfer approx 2 mL of the aqueous phase into a new tube. If the aqueous phase is less than 2 mL, add RNase-free water to bring to about 2 mL. Add 200 μL of 5 *M* NaCl and 5 mL absolute ethanol. Mix and incubate at –20°C for 4 h to overnight.
12. Centrifuge at 8000*g* for 20 min at 4°C.
13. Discard the supernatant. Add 2 mL of 75% ethanol and vortex briefly. Centrifuge at 8000*g* at 4°C for 5 min.
14. Discard ethanol. Dry the RNA pellet by vacuum dryer for 10 min.
15. Dissolve the pellet in 200 μL RNase-free water by brief vortex. After centrifuging briefly, transfer the RNA solution into a microcentrifuge tube and store at –70°C (*see* **Note 14**).

4. Notes

1. To prepare RNase-free water, add 1 mL of diethylpyrocarbonate (DEPC) into 1 L of ddH_2O. Stir overnight in a hood, and then autoclave.
2. To prepare 100 m*M* stock solution of aurintricarboxylic acid (ATA, Sigma A1895): weigh 4.22 g ATA and place into a beaker. Add 80 mL ddH2O and stir. Adjust pH to 8.0 with NaOH. The ATA gradually dissolves when adding NaOH. Add an additional 20 mL ddH2O. Store at room temperature in a bottle wrapped with aluminum foil to avoid light.
3. Extraction buffer should be made fresh and placed at room temperature. Placing on ice or storing for a long time may cause salt precipitation. Using each stock solution to make the extraction buffer is recommended. Calculate the volume of each component needed, including ddH_2O. Add ddH_2O first, then Tris-HCl buffer, and then others.
4. TE-saturated phenol is commercially available (e.g., Sigma P4557). Phenol is highly corrosive and can cause severe skin burns. Wear gloves, protective clothing, and safety goggles when handling phenol.
5. The tissues can be stored at –70°C for a long time without RNA degradation. However, one should pay attention to fluctuations of temperature of the freezer. Increased temperature may cause RNA degradation in tissues. Caution should be taken if tissues have been stored in –70°C freezer for an excessive time period.
6. RNase contamination is a major factor leading to RNA degradation during RNA extraction process. Exogenous RNase can be avoided by using clean materials, RNase-free solutions, and sterile and disposable plasticware, and wearing disposable gloves during RNA extraction process. Nondisposable glassware or plasticware should be RNase-free. Glassware can be baked at 150°C for 4 h, and plastic items can be soaked for 10 min in 0.5 *M* NaOH, rinsed thoroughly with water, and autoclaved. Endogenous RNase from plant tissue can be inactivated by grinding materials under frozen condition followed by rapid disruption of cells in TRIzol reagent. Do not let the frozen tissue thaw during grinding. Quickly mix the grinded tissues with TRIzol reagent and homogenize. RNase is inactivated in the present of TRIzol.
7. When using of motor-driven pestle for homogenization, 600 mL of TRIzol reagent is recommended to add to the ground tissue at first. Too much solution in a microcentrifuge tube will cause flooding when using a motor-driven pestle. After homogenization, an additional 400 mL of TRIzol is added and mixed.
8. Generally, RNA can be easily dissolved in water when using very young seedlings as material. However, when extracting RNA from mature leaves, one would experience difficulty to dis-

solve RNA because of polysaccharides in RNA pellet. Dissolving RNA can be assisted by incubating for 10 min at 55°C and passing the solution a few times through a pipet.

9. We have extensively used this method to extract RNA for Northern blotting. Comparing with other methods, the advantage of this method is that much less care is needed to protect RNA from degradation because of addition of RNase inhibitor in the extraction buffer. We just use autoclaved glass and plastic wares and solutions.
10. Do not let the frozen tissues thaw during grinding. Add liquid nitrogen into mortar during grinding if needed.
11. Prechilling microspoon in liquid nitrogen will protect the ground frozen tissue from thawing during transfer. Do not place the 15-mL tube with extraction buffer on ice before transferring tissue. Cold temperature causes salt precipitation in the extraction buffer. About 2 to 3 mL volume of ground tissue is recommended to add to 4.5 mL extraction buffer. Excessive tissue will cause incomplete homogenization and RNA degradation.
12. The supernatant can be simply poured into a new tube. It does not matter that the supernatant contains some debris, which will be removed by phenol:chloroform extraction in **step 10**. You can incubate the tubes at 4°C overnight if needed.
13. After centrifuging, a middle layer will be formed. A thick middle layer can be formed when tissues contain high polysaccharide or too much tissue is used. In these cases, aqueous phase is less than 2 mL. Just transfer aqueous phase as you can, but do not transfer anything from the middle layer.
14. The RNA pellet is red-colored. After dissolving in water, the RNA solution is reddish because of the presence of ATA. This is normal and the ATA in solution can protect RNA from degradation.

References

1. Davis, L. G., Kuehl, W. M., and Battey, J. F. (1994) Basic *Methods in molecular biology* (2nd ed). Appleton & Lange, Norwalk, CT. Chapter 11, sections 1–5.
2. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156–159.
3. Chomcyznski, P. (1993) A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* **15,** 532–535.

28

Target Preparation for DNA Microarray Hybridization

Tong Zhu, Sherman H. Chang, and Pedro Gil

Summary

DNA microarrays are widely used to analyze genome-wide gene expression patterns and to study genotypic variations. They are miniaturized collections of thousands of DNA fragments arrayed on a surface. Based on nucleic acid complementary binding, they serve as a tool to interrogate complex populations of nucleic acids for abundance or binding affinity of particular sequences. Before a nucleic acid (target) can be used for hybridization to the probes of a microarray, it needs to be extracted from the tissue and labeled. Frequently, it also needs to be amplified to increase detection sensitivity. During a hybridization process, labeled target molecules with sequences complementary to the probes are captured quantitatively. Subsequently, a reader measures the amount of label on each probe. To generate accurate and informative data, one of the most critical aspects of these experiments is the quality of both the isolated and the labeled nucleic acid samples. This chapter describes detailed procedures for the preparation of labeled RNA samples for DNA microarray analysis.

Key Words: Microarray; cDNA synthesis; cRNA synthesis; fluorescence labeling.

1. Introduction

Microarray technology is a major technical breakthrough for genomic approaches in molecular biology. DNA microarrays consist of a matrix of DNA probes ranging in number from hundreds to millions arrayed on a solid surface to measure the abundance and/or binding ability of DNA or RNA target molecules. For quantification of mRNA levels, mRNA molecules extracted from biological samples are copied into labeled nucleic acids, using enzymatic reactions that maintain in proportion the relative abundance of each particular species. Labeled targets are then hybridized to the probes under binding conditions that discriminate against noncomplementary sequences. The amount of labeled molecules captured by the DNA probes reflects the abundance of the target molecules.

The probes used in microarrays can be polymerase chain reaction (PCR)-generated cDNA fragments or DNA oligonucleotides with sequences complementary to the target sequences. Therefore, they are commonly known as cDNA microarrays or oligonucleotide microarrays. In the case of spotted cDNA or oligonucleotide microarrays, the DNA fragments are deposited onto the surface mechanically and immobilized to the surface of a chemically coated glass slide *(**1**)*. Oligonucleotide microarrays can also be directly synthesized on a surface, as in the case of GeneChip® microarrays *(**2**)*. For transcriptome profiling using spotted microarrays (cDNA or oligonucleotide), targets are usually generated by oligo-dT-primed reverse transcription, including nucleotides modified with fluorescent moieties. Because of intrinsic variations in probe amounts in each slide due to the printing procedures, in most cases the relative abundance of mRNAs are compared between two biological samples (test and control samples). Typically, two fluorescent labels with nonoverlapping emission spectra are used. This strategy allows simultaneous hybridization and detection of both targets using the same microarray. A

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

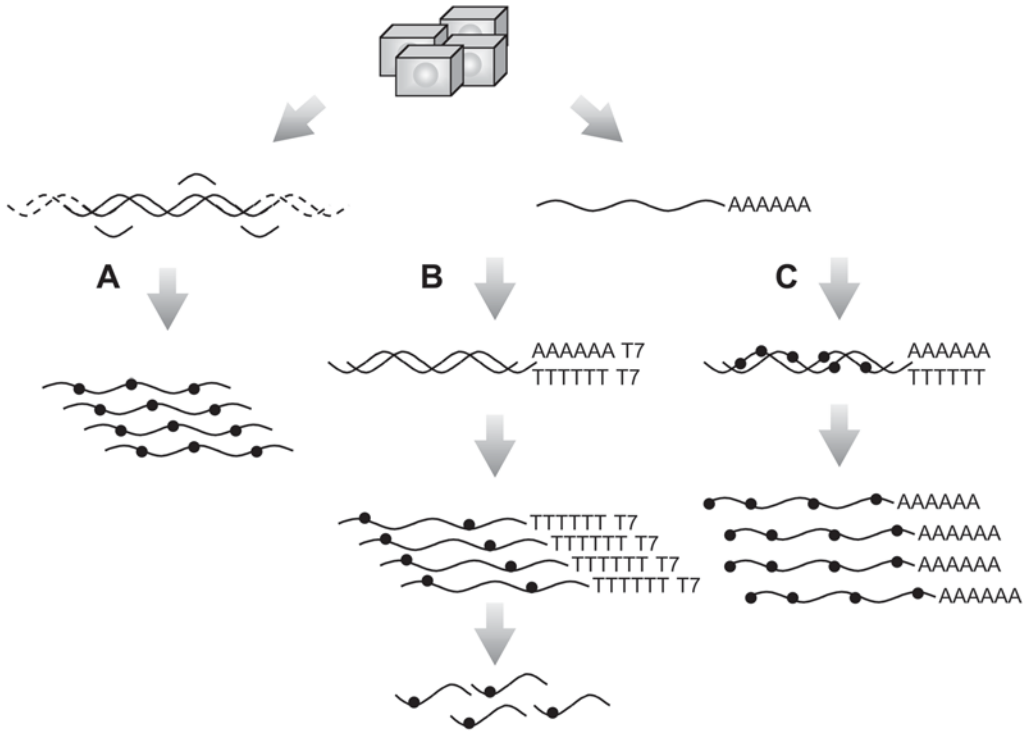


Fig. 1. A schematic overview of target preparation processes for microarray analyses. Process **A** illustrates the labeling of genomic DNA by random priming. Random hexamer primers anneal to the genomic DNA, and fragments are extended by Klenow DNA polymerase. This process is used for genomic DNA analyses using microarrays. Process **B** shows the labeling of total RNA by first synthesizing double-stranded cDNA, followed by amplifying the biotin-labeled cRNA using the double-stranded cDNA as a template. This process is used for sample labeling in transcriptome analyses using GeneChip microarrays. Process **C** depicts the direct incorporation of fluorescent-tagged nucleotides during first-strand cDNA synthesis. This process is frequently used in transcriptome analyses using spotted cDNA or DNA oligonucleotide microarrays.

charged coupled device (CCD) camera or a laser scanner is commonly used to measure fluorescent signals. In the case of transcriptome profiling using GeneChip® microarrays, unlabeled cDNA templates are generated, followed by an in vitro transcription reaction to produce biotin-tagged cRNA targets (**Fig. 1**) that can be detected by coupling with labeled antibodies ***(2)***. If the targets are genomic DNA, the fluorescent-tagged nucleotides could be incorporated to the target molecules by end-labeling or random priming method ***(3)***.

DNA microarrays are widely used in *Arabidopsis* research ***(4–7)***. Among the factors contributing to the success of a microarray experiment, nucleic acid sample preparation greatly affects the quality of the final data and, consequently, its biological interpretation. Optimization of the substrate materials, the enzymatic steps involved in target preparation, the hybridization conditions, and the method of detection also leads to enhanced performance. The parallel sample quality assays using capillary electrophoresis ***(8)*** and the parallel sample preparation methods ***(9–10)*** significantly improve the throughput of the screening without sacrificing performance. These improvements also make it possible for time-consuming and labor-intensive sample preparation processes to be automated using robots. The use of thermally stable reverse transcriptase for cDNA synthesis also enhances sensitivity and specificity (**Fig. 2**). Laser-capture microdissection (LCM) and various amplification techniques enable gene expression

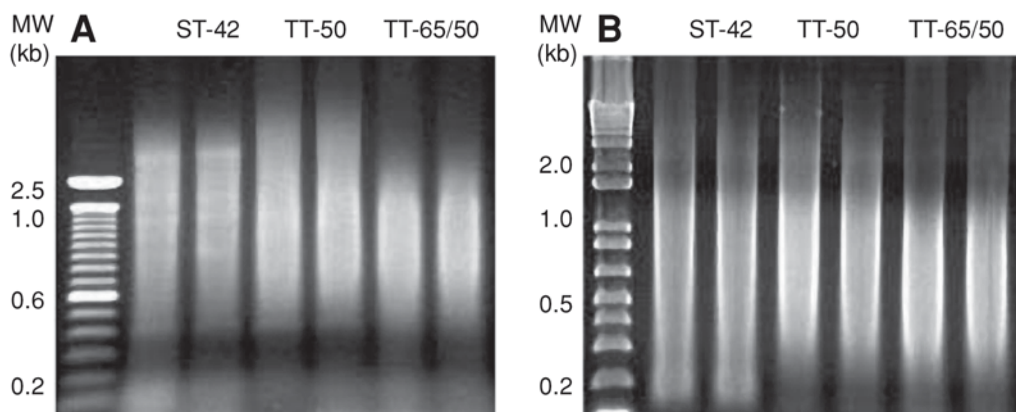


Fig. 2. Comparison of size distribution and yield of cDNA (panel **A**) and biotin-labeled cRNA (panel **B**) using regular and thermally stable reverse transcriptases (RT). The activities of SuperScript II RT, a modified Moloney murine leukemia virus RT that lacks RNase H activity at 42°C (ST-42), and ThermoScript RT, a modified avian myeloblastosis virus RT with greatly reduced RNase H activity at 50°C (TT-50) and 65°C (TT-65/50, annealing at 65°C and synthesis at 50°C) were evaluated. Newly synthesized cDNAs were purified from total RNA by RNase H digestion. Greater yield of longer cDNA fragments were obtained using thermo stable reverse transcriptase (TT) with 50°C denaturation. Subsequently, longer cRNA fragments were amplified.

profiling of very small populations of cells, enhancing enormously the resolution of the microarray analysis ***(11–12)***.

This chapter describes in detail two basic procedures of RNA sample preparation for microarray analyses. The first procedure is a single fluorescent dye labeling method coupled with complimentary RNA (cRNA) amplification ***(13)***. It has been widely used in gene expression analysis in *Arabidopsis* using GeneChip microarrays ***(14–15)***. The second procedure is a two-fluorescent-dye labeling method, and is designed for gene expression analysis using spotted microarrays ***(16)***.

2. Materials

2.1. Preparing Samples Used for GeneChip Microarrays

1. 5 µg of total RNA at a concentration of 1 µg/µL (*see* **Note 1**).
2. T7-$(dT)_{24}$ DNA oligonucleotide primer (5'-GGCCAGTGAATTGTAATACGACTCACTATA GGGAGGCGG-$(dT)_{24}$-3'). High-performance liquid chromatography (HPLC)-purified. Its concentration needs to be adjusted to 100 pmol/µL in DEPC-treated water (*see* **Note 2**).
3. SuperScript™ Double-Stranded cDNA Synthesis Kit (Invitrogen, Carlsbad, CA), 5X first strand reaction buffer: 250 m*M* Tris-HCl (pH 8.3), 375 m*M* KCl, 15 m*M* MgCl2; 0.1 *M* dithiothreitol (DTT); 10 m*M* dNTP mix: 10 m*M* each dATP, dCTP, dGTP, dTTP; SuperScript II RT (200 U/µL); 5X second strand reaction buffer: 100 m*M* Tris-HCl (pH 6.9), 450 m*M* KCl, 23 m*M* $MgCl_2$, 0.75 m*M* β-NAD+, 50 m*M* $(NH4)_2SO_4$; *Escherichia coli* DNA ligase (10 U/µL); *E. coli* DNA Polymerase I (10 U/µL); *E. coli* RNase H (2 U/µL); T4 DNA polymerase (5 U/µL) (*see* **Note 3**)
4. Enzo® BioArray™ HighYield™ RNA Transcript Labeling Kit (Enzo Biochem, Farmingdale, NY). 10X HY reaction buffer: Tris-buffered NaCl, $MgCl_2$ and spermidine, with stabilizer; 10X biotin labeled ribonucleotides; 10X concentrated mixture of ATP, CTP, GTP, UTP, Bio-UTP, and Bio-CTP; 10X DTT: 100 m*M* DTT; 10X RNase inhibitor mix: 20 U/µL in storage buffer; 20X T7 RNA polymerase.
5. Phenol:chloroform:isoamyl alcohol (PCIM, 25:24:1, v/v/v), stored at 4°C.

6. Phase lock gel (PLG), heavy, 2 mL (Brinkmann Instruments, Westbury, NY).
7. 100% and 70% ethanol, at –20°C.
8. Pellet paint (Novagen, Madison, WI).
9. 7.5 *M* ammonium acetate (NH_4OAc).
10. 5X RNA fragmentation buffer: 200 m*M* Tris-acetate, pH 8.1, 500 m*M* KOAc, 150 m*M* MgOAc (*see* **Note 4**).
11. RNeasy Mini Kit (Qiagen, Valencia, CA): RNeasy Mini Spin Columns, Collection Tubes (1.5 mL and 2 mL), RNase-free reagents and buffers.
12. Pipet tips, sterile-barrier, RNase-free (1 mL, 0.2 mL, 20 µL), and sterile-barrier and nonbarrier (0.2 mL).
13. Centrifuge tubes, sterile, RNase-free, 1.5 mL and 0.5 mL.
14. Diethlypyrocarbonate (DEPC)-treated water.

2.2. Preparing Samples Used for Spotted Microarrays

1. 25 µg of *Arabidopsis* total RNA (1.1 µg/µL) (*see* **Note 1**)
2. SuperScript™ II reverse transcriptase (200 U/µL) with 5X first-strand buffer: 250 m*M* Tris-HCl (pH 8.3), 375 m*M* KCl, 15 m*M* $MgCl_2$ (Invitrogen; *see* **Note 5**).
3. 200 µ*M* anchored oligo $(dT)_{21}$ (*see* **Note 6**)
4. 0.1 *M* DTT.
5. Deoxinucleotides: d(ACG)TP mix: 25 m*M* dATP, 25 m*M* dGTP, and 25 m*M* dCTP; 1 m*M* dTTP.
6. 10 m*M* Cy3-dUTP (Amersham Biosciences, Piscataway, NJ).
7. 10 m*M* Cy5-dUTP (Amersham Biosciences; *see* **Note 7**).
8. 3 *M* NaOH.
9. 1 *M* HCl.
10. QIAquick PCR Purification Kit (Qiagen).
11. Pipet tips, sterile-barrier, RNase-free (1, 0.2, 20 mL), and sterile-barrier and nonbarrier (0.2 mL).
12. Centrifuge tubes, sterile, RNase-free, 1.5 and 0.5 mL.
13. DEPC-treated water.

3. Methods

As a general precaution, these procedures should be conducted in nuclease-free environments. DEPC-treated water should be used for the preparation of aqueous solutions.

3.1. Preparing Samples Used for GeneChip Microarrays

3.1.1. Double-Strand cDNA Synthesis (*see* **Note 8**)

3.1.1.1. First Strand cDNA Synthesis

1. Mix 5 µL of total RNA, 1 µL of T7-$(dT)_{24}$ primer, and 6 µL of DEPC-treated water in a 0.5-µL DNase/RNase-free centrifuge tube.
2. Incubate the reaction at 70°C for 10 min.
3. Centrifuge the reaction briefly to bring down the condensation in the tube; place on ice to cool down to 4oC.
4. From the Superscript II cDNA synthesis system, add 4 µL of 5X first-strand cDNA buffer, 2 µL of 0.1 *M* DTT, 1 µL of 10 mM dNTP mix, and 1 µL of Superscript II reverse transcriptase into the reaction and mix well. Centrifuge briefly to bring down the condensation and incubate the reaction at 42°C for 1 h.

3.1.1.2. Second-Strand cDNA Synthesis

1. Place first-strand cDNA synthesis reaction mix on ice.
2. Add 91 µL of DEPC-treated water, 30 µL of 5X second strand reaction buffer, 3 µL of 10 m*M* dNTP mix, 1 µL of *E. coli* DNA ligase (10 U/µL), 4 µL of *E. coli* DNA polymerase I (10 U/µL) and 1 µL of *E. coli* RNase H (2 U/µL) to the first strand cDNA synthesis reaction tube.
3. Mix the reaction gently and centrifuge briefly to bring down the condensation. Incubate at 16°C for 2 h.

3.1.2. Purification of Double-Strand cDNA (*see* **Note 9**)

1. Before adding any reagents, centrifuge phase lock gel (PLG) tubes at ≥12,000g for 1 min.
2. Add 162 µL of PCIM into second-strand cDNA synthesis reaction, vortex briefly, and transfer to a PLG tube.
3. Centrifuge at ≥12,000*g* at room temperature for 5 min.
4. Transfer the aqueous upper phase to a 1.5-mL centrifuge tube.
5. Add 1 µL of pellet paint (*see* **Note 10**), 120 µL of 7.5 *M* NH4OAc, and 400 µL of absolute ethanol at –20°C. Mix well.
6. Precipitate the cDNA on dry ice for 20 min.
7. Centrifuge the mixture at ≥12,000*g* at 4°C for 20 min.
8. Remove supernatant and carefully wash pellet with 0.5 mL 70% ethanol at –20°C. Repeat this step at least once.
9. Centrifuge at ≥12,000*g* at 4°C for 10 min, remove supernatant, and air-dry the pellet.
10. Resuspend the dried pellet in 22 µL of DEPC-treated water. Store at –20°C if cRNA synthesis is not to be performed immediately.

3.1.3. Synthesis of Biotin-Labeled cRNA (see **Note 11**)

1. From the Enzo BioArray™ HighYield™ RNA Transcript Labeling Kit, add 4 µL of 10X T7 HY buffer, 4 µL of 10X biotin ribonucleotides mix, 4 µL of 10X DTT, 4 µL of 10X RNase inhibitor mix, and 2 µL of 20X T7 RNA polymerase to the tube containing the purified double-strand cDNA. The final volume of the reaction should be 40 µL.
2. Mix well and centrifuge briefly. Incubate at 37°C for 4 h (*see* **Note 12**).

3.1.4. Purification and Quantification of cRNA (see **Note 13**)

1. Transfer the 40 µL of synthesized cRNA in the reaction mix to a 1.5-mL centrifuge tube and add 60 µL DEPC-treated water.
2. Add 350 µL RLT buffer from the RNeasy kit and vortex briefly. Add 250 µL of 100% ethanol.
3. Place the purification column into a 2-mL collection tube.
4. Mix the solutions by pipetting up and down and transfer to an assembled Qiagen RNeasy column.
5. Wait 5 min to allow RNA to bind to the membrane inside the column.
6. Centrifuge the Qiagen RNeasy column at 5000*g* for 5 min (*see* **Note 14**).
7. Transfer the column to a new collection tube.
8. Add 500 µL of RPE buffer (ethanol-added) to wash the membrane in the column. Centrifuge at 10,000*g* for 1 min..
9. Repeat this wash step with another 500 µL of RPE buffer (ethanol-added).
10. Centrifuge the column at ≥12,000*g* for 1 min to dry the column completely.
11. Add 30 µL DEPC-treated water. Let it stand for 5 min and elute the cRNA from the column by centrifugation at 5000*g* for 5 min.
12. Repeat the eluting step with an additional 30 µL of DEPC-treated water. The total volume of the purified cRNA solution should be between 55 and 60 µL (*see* **Note 15**).
13. Take 1 µL of cRNA and mix with 4 µL of DEPC-treated water. Check the OD with a spectrophotometer to determine the cRNA yield and quality.
14. Take 1 µL of cRNA and mix with 4 µL RNA loading buffer with loading dye, heat to 70°C for 5 min, and chill on ice. Load it in a 1% agarose gel. After electrophoresis, the cRNA should be visualized either by ethidium bromide or SYBR Green II staining in 1X TBE buffer, to determine yield and distribution of sizes (*see* **Note 16**).
15. Adjust the remaining cRNA to a concentration of 1 µg/µL using DEPC-treated water. If cRNA concentration is less than 1 µg/µL, the cRNA needs to be precipitated with ethanol and resuspended in a smaller volume. Store at –80°C if a hybridization is not going to be performed immediately.

3.1.5. Fragmentation of cRNA (*see* **Note 17**)

1. Transfer 15 µL of cRNA into a 0.5-mL centrifuge tube; add 6 µL of 5X fragmentation buffer and 9 µL of DEPC-treated water. Mix well and centrifuge briefly.
2. Incubate at 94°C for 35 min. Chill on ice.

3. Check the distribution of sizes of the fragmented cRNAs by electrophoresis on 1 to 1.5% agarose gel (use 2 µL of the fragmented cRNA) (*see* **Note 18**).

3.2. Preparing Samples Used for Spotted Microarrays

3.2.1. First-Strand cDNA Synthesis (see **Note 19**)

1. In a 0.2 or 0.5-mL tube, combine 23 µL of total RNA (1.1 µg/µL) with 5 µL of 200 µ*M* anchored oligo dT_{21}.
2. Heat to 70°C for 10 min and place on ice for at least 1 min.
3. Add the following into the tube on ice in sequential order: 10 µL of 5X first-strand buffer; 5 µL of 0.1 M DTT; 1 µL of d(ACG)TP Mix; 2 µL of 1 m*M* dTTP, 1 µL of RNase inhibitor, such as RNaseOUT™ (40 U/µL); and 2 µL of 1 m*M* Cy3-dUTP (or Cy5-dUTP) (*see* **Note 20**).
4. Add 1 µL of SuperScript™ II RT (400 U/µL).
5. Mix the reaction and centrifuge briefly.
6. Incubate the reaction at 42°C for 90 min in the dark (*see* **Note 21**).

3.2.2. Purification of cDNA Targets

1. Add 5 µL of 3 *M* NaOH to the 50 µL of cDNA synthesis reaction and mix well.
2. Incubate the reaction at 65°C for 10 min to hydrolyze the RNA.
3. Add 15 µL of 1 *M* HCl and mix gently to neutralize the pH.
4. To remove the unincorporated dNTPs from the reaction, transfer to a 1.5-mL tube, add 400 µL of PB buffer, mix, and transfer to a QIAquick spin column (*see* **Note 22**).
5. Centrifuge at full speed for 1 min and discard the flow-through.
6. Add 700 µL of PE buffer to the column.
7. Centrifuge at full speed for 1 min and discard the flow-through.
8. Add 700 µL of PE buffer to the column.
9. Centrifuge at full speed for 1 min and discard the flow-through.
10. Centrifuge at full speed again for 1 min to remove any residual ethanol.
11. In the final elution, add 35 µL of water, wait for 1 min and centrifuge for 1 min at room temperature to collect the purified product. Repeat the process with another 35 µL of water. Combine the fractions.
12. Transfer the purified labeled cDNAs to a cuvet and evaluate the target yield and quality using a spectrophotometer (*see* **Note 23**).
13. Transfer the purified cDNA products to a 1.5-mL tube and proceed to the array hybridization (*see* **Note 24**). Labeled cDNAs can also be stored at –80°C.

4. Notes

1. Total RNA, rather than prepurified mRNA, is recommended for microarray analysis, because it represents a more accurate estimate of the transcript population. Several methods for RNA isolation using commercial products have been tested and validated, including RNeasy Plant Mini kit (Qiagen) and RNAwiz (Ambion). Qiagen RNeasy Column or the LiCl method ***(17)*** is recommended for preparing RNA samples from seeds or young seedlings ***(18)***. These methods effectively remove carbohydrates, storage proteins and second metabolites, which may affect the cDNA synthesis. If samples need to be stored or transported before the RNA isolation, they can be harvested, frozen in liquid nitrogen and stored at –80°C. Alternatively, tissue samples can be harvested and quickly submerged in 5 volumes of RNAlater storage buffer (Ambion) for later processing. If RNAlater is used, it is recommended to store the samples at 4°C to minimize RNA degradation. In general, 100 mg of *Arabidopsis* tissues should yield between 30 and 40 µg of total RNA. The isolated total RNA should be examined for RNA integrity by gel electrophoresis and for purity by spectrophotometric analysis (1 OD_{260nm} equivalent to 40 µg/mL of RNA). The ratio of A260/A280 and A260/230 should range between 1.9 and 2.1. RNA samples with slightly lower A260/280 ratio could be cleaned further using RNeasy columns (Qiagen).
2. The quality of the T7-$(dT)_{24}$ primer is critical not only to the success of cDNA synthesis but also for the in vitro transcription reaction. The synthesized oligonucleotide primers should be purified by high-performance liquid chromatography (HPLC). Polyacrylamide gel electrophoresis (PAGE)-purified primers do not work well with this protocol.

3. SuperScript Double-Stranded cDNA Synthesis Kit is recommended. It contains all of the components used in the GeneChip cDNA synthesis protocol.
4. To prepare 5X RNA fragmentation buffer, combine 4.0 mL of 1 *M* Tris-acetate, pH 8.1, 0.64 g of MgOAc, and 0.98 g of KOAc. Use DEPC-treated water to adjust the final volume to 20 mL. Mix thoroughly and filter through a 0.2 µm vacuum filter unit. Store at room temperature.
5. SuperScript™ Direct cDNA Labeling System (Invitrogen) is an alternative to purchasing the components for the first strand cDNA synthesis separately. It uses SuperScript III, a reverse transcriptase with higher thermal stability, for reverse transcription. In addition to the components for cDNA synthesis, the system also includes a control HeLa RNA (1 µg/µL in HE buffer) and a module for purification of cDNA products (purification columns, loading buffer, wash buffer, and amber collection tubes).
6. Anchored oligo (dT)21 primer is a mixture of 12 primers, each consisting of a string of 21 deoxythymidylic acid (dT) residues followed by two additional nucleotides represented by VN, where: V is dA, dC, or dG and N is dA, dC, dG, or dT. The VN "anchor" allows the primer to anneal only at the 5' end of the poly(A) tail of mRNA, providing a more efficient cDNA synthesis for labeling applications.
7. With a modified nucleotide composition, the labeling reaction can be used to incorporate Cy3-dCTP or Cy5-dCTP into the cDNA, instead of Cy3-dUTPs or Cy3-dUTPs.
8. For GeneChip microarray analysis, unlabeled double-stranded cDNA needs to be synthesized to provide a nonbiased representation of the transcript population in the biological sample. The double-stranded cDNA is used as a template for cRNA amplification and labeling. There are a variety of methods that can be used for double-stranded cDNA synthesis. Thermal stable reverse transcription enzymes produced from various organisms have been optimized and developed into commercial products, such as SuperScript III RT and ThermoScript RT. These thermal stable enzymes are designed to copy RNA with secondary structure and to improve specificity during priming. These enzymes could be used instead of SuperScript II RT in these methods to improve the detection of the rare messages or to quantify the amount of specific mRNA from small numbers of cells (**Fig. 2**).
9. Phase lock gels (PLG)-phenol/chloroform tubes are recommended for purifying synthesized cDNAs. They are based on a modified guanidinium isothiocyanate/acid phenol method ***(19)***. The solid barrier between the aqueous and organic phases of phenol–chloroform in the PLG allows for a more complete recovery of the cDNA in the aqueous phase and minimizes phenol-chloroform carryover during the extraction. An alternative method to purify cDNAs from the reaction mix is to use the MinElute Reaction Cleanup Kit following the manufacturer's instructions (Qiagen).
10. Adding pellet paint and other sources of glycogen helps to visualize the very small pellets of cDNA.
11. An optimal cRNA synthesis is an essential step to achieve high sensitivity in GeneChip microarray detection. The cRNA targets that represent the transcript population were synthesized and amplified by T7 RNA polymerase using cDNAs as templates during in vitro transcription (IVT). This step incorporates the biotinylated nucleotides to the targets and linearly amplifies the targets 500 to 1000 times.
12. During cRNA synthesis, it is recommended to mix gently and centrifuge briefly every 30 to 45 min to bring down the condensation in the side of the tube. Do not incubate the reaction overnight when using this labeling kit. Overnight incubation tends to produce shorter products. If overnight incubation is necessary during the cRNA synthesis, the 3'-Amplification Reagents for IVT Labeling (Affymetrix, Santa Clara) containing MEGAscript RNA polymerase (Ambion) is recommended. For details, see product instructions from Affymetrix.
13. It is essential to remove unincorporated NTPs in order to quantify the cRNA yield. It is also important to remove free biotin-labeled NTPs to reduce background from non-specific binding. To maximize recovery, it is recommended to use column-based methods.
14. To prevent the loss of the cRNA due to defective columns, it is recommended to save the flow-through until purification steps are completed.
15. Eluting cRNA twice and prolonging the eluting time could improve cRNA recovery. Verify the volume of eluted cRNA and make sure that this volume is close to 60 mL. If the volume is more than 60 mL, it is possible that there is residual Buffer RPE/ethanol contamination from the washing step.

16. The adjusted cRNA yield from 5 μg of total RNA should range from 25 to 100 μg. The ratios of A260/230 and A260/280 should be between 1.9 and 2.1. The size distribution of the purified cRNA should be from 200 to 4000 bases, with an average of around 1000 bases in length. If the cRNA yield is not sufficient due to the limited amount of total RNA, an additional cycle of target amplification can be included. In this case, the unlabeled ribonucleotides will be used in the first cycle of in vitro transcription, followed by a second cycle of double stranded cDNA synthesis, and a second cycle of in vitro transcription using biotin-labeled ribonucleotides. This can be achieved using Affymetrix's GeneChip® Two-Cycle Target Labeling and Control Reagents. See its product guide for additional information.
17. This step ensures that the labeled targets hybridized to the probes on the GeneChip microarray are in the appropriate size range (35 to 200 bases), to maximize the capability of the hybridization between the probes and targets. In order to reduce the magnesium used in the fragmentation buffer, it is recommended that the cRNA is sufficiently concentrated (1 μg/μL).
18. The fragmented cRNA is ready to hybridize to the GeneChip microarray. Please refer to the Affymetrix technical manual ***(13)*** for details of target hybridization, washing, staining, scanning, and data analysis.
19. There are various methods for Cy-labeling of total RNA targets. A fast and consistent method is the one-step labeling reaction described here. It consists of a reverse transcription that incorporates fluorescent Cy3 or Cy5 nucleotides into cDNA molecules and uses as a primer anchored oligo dT molecules annealed to the polyA tail of the *Arabidopsis* mRNAs. After the reverse transcription reaction, the newly synthesized Cy-labeled cDNA is isolated from the template RNA, unincorporated nucleotides, and other reaction components. This protocol and the subsequent hybridization (*see* **Note 24**) were optimized for hybridizations under cover slips of 20 × 20 mm on CMT-GAPS amino-saline-coated glass slides including cDNA probes printed in a 50% DMSO solution at a concentration of 250 ng/mL.
20. It is recommended to make a master mix excluding labeled dNTPs and reverse transcriptase to improve the consistence of the labeling, especially when dealing with a large set of samples. The master mix can then be divided into the individual reactions.
21. Cy3 and Cy5 dyes are very sensitive to light, so they should be kept in the dark and should be handled in a dimly lit area as much as possible. These precautions should be followed during the whole labeling procedure.
22. The unincorporated dNTPs and cDNA synthesis reaction could be removed by other methods. If a SuperScript™ Direct cDNA Labeling System is used, the cDNA purification could be performed using the enclosed purification module, including purification columns, loading buffer, wash buffer, and amber collection tubes.
23. To evaluate the quality of the labeled targets, it is necessary to monitor the cDNA yield and the incorporation of Cy-dUTP into the targets. To calculate the yield of cDNA the following formula is used: Labeled cDNA (ng) = A260 OD × 37 ng/μL × 70 μL. A yield of more than 500 ng of cDNA should be expected. The incorporation of Cy nucleotides can be monitored by calculating the number of pmoles of Cy3 or Cy5 in the labeled cDNA using the following formulas: Cy3 (pmole) = (A550-A650)/0.15 × total elution volume (μL); Cy5 (pmol) = (A650–A750)/0.25 × total elution volume (μL). A good target should incorporate between 30 and 60 pmol of Cy3 or Cy5. This measurement can be used to normalize the amount of Cy3 and Cy5 added to the hybridization. A useful measurement for the specific activity of the targets is the number unlabeled nucleotides incorporated per labeled nucleotide incorporated. It can be calculated using the following formulas: Cy3 (nt/Cy3nt) = (A260 × 1000) / (58.5 × A550); Cy5 (nt/Cy5nt) = (A260 × 1000) / (35.1 × A550). For good-quality targets, these values should be less than 80 nucleotides.
24. To perform hybridization: Prehybridize slides for 30 min at 42°C in prehybridization solution: 5X SSC, 0.1% SDS, 1% bovine serum albumin. Dry Cy3 and Cy5 targets in a Speedvac. Add 10 μL of Hybridization solution (50% formamide, 5X SSC, 0.1% SDS, 500 μg/mL yeast tRNA, 500 μg/mL poly A) to the dry Cy3-labeled target and 10 μL of Hybridization solution to the dry Cy5-labelled target. Vortex to dissolve probes. Combine Cy3 and Cy5 labeled probes to a total volume of 20 μL. Pipet the denatured targets on the DNA printed area of the slide and cover with a cover slip. Hybridize the array for at least 18 h at 42°C in a hybridization cham-

ber. Perform low stringency washes (wash 1: 10 min in 1X SSC, 0.1% SDS at room temperature; wash 2: 10 min in 0.2X SSC, 0.1% SDS at room temperature). For best resolution between related cDNA probes, it is critical to include a high-stringency 10 min wash in 0.1X SSC at 65°C. Scan slides in a microarray laser scanner.

Acknowledgments

The authors thank Wenying Xu, Makoto Ono, George Aux, Charles Chilcott, John McElver, Betsy Read, Yen Tran, Anne-Laure Wizman, and Bin Han for great help in development and optimization of sample preparation methods for microarray analyses, and helpful discussions during preparation of this chapter.

References

1. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270,** 467–470.
2. Lipshultz, R. J., Fodor, S. P. A., Gingeras, T. R., and Lockhart, D. J. High density synthetic oligonucleotide arrays. *Nat. Genet.* **21s,** 20–24.
3. Borevitz, J. O., Liang, D., Plouffe, D., et al. (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res*. **13,** 513–523.
4. Aharoni, A. and Vorst, O. (2002) DNA microarrays for functional plant genomics. *Plant Mol. Biol.* **48,** 99–118.
5. Donson, J., Fang, Y., Espiritu-Santo, G., et al. (2002) Comprehensive gene expression analysis by transcript profiling. *Plant Mol. Biol*. **48,** 75–97.
6. Mele, G. and Hake, S. (2003) Expression profiling of plant development. *Genome Biol*. **4,** 215.
7. Zhu, T. (2003) Global analysis of gene expression using GeneChip microarrays. *Curr. Opin. Plant Biol.* **6,** 418–425.
8. Khandurina, J., Chang, H.-S., Wanders, B., and Guttman, A. (2002) Automated high-throughput RNA analysis by capillary electrophoresis. *BioTechniques* **32,** 1226–1230.
9. Zarrinkar, P. P., Mainquist, J. K., Zamora, M., et al. (2001) Arrays of arrays for high-throughput gene expression profiling. *Genome Res.* **11,** 1256–1261.
10. Zhu, T., Chang, H.-S., Schmeits, J., et al. (2001) Gene expression microarrays: improvements and applications towards agricultural gene discovery. *J. Assoc. Lab. Automat*. **6,** 95–98.
11. Hertzberg, M., Sievertzon, M., Aspeborg, H., Nilsson, P., Sandberg, G., and Lundeberg, J. (2001) cDNA microarray analysis of small plant tissue samples using a cDNA tag target amplification protocol. *Plant J*. **25,** 585–591.
12. Nakazono, M., Qiu, F., Borsuk, L. A., and Schnable, P. S. (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. *Plant Cell* **15,** 583–596.
13. Affymetrix. (2003) GeneChip® Expression Analysis Technical Manual. http://www.affymetrix.com/support/technical/manuals.affx.
14. Zhu, T. and Wang, X. (2000) Large-scale profiling of the *Arabidopsis* transcriptome. *Plant Physiol.* **124,** 1472–1476.
15. Zhu, T., Budworth, P., Han, B., Brown, D., Chang, H.-S., Zou, Z., and Wang, X. (2001) Towards elucidating global gene expression in developing *Arabidopsis*: parallel analysis of 8300 genes. *Plant Physiol. Biochem.* **39,** 221–242.
16. Eisen, M. B. and Brown, P. O. (1999) DNA arrays for analysis of gene expression. *Methods Enzymol.* **303,** 179–205.
17. Chang, S., Puryear, J., and Cairney, J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **11,** 113–116.
18. Tepperman, J. M., Zhu, T., Chang, H. S., Wang, X., and Quail, P. H. (2001) Multiple transcription-factor genes are early targets of phytochrome A signaling. *Proc. Natl. Acad. Sci. USA* **98,** 9437–9442.
19. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156–159.

29

Statistical Issues in Microarray Data Analysis

Willem A. Rensink and Samuel P. Hazen

Summary

Microarrays provide the ability to quantitatively measure the abundance of specific RNA transcripts through sample hybridization to a solid-state grid of oligonucleotides or amplicons. The prospect of measuring the entire transcriptome is extremely alluring, but as with any experiment, it should be met with caution and great consideration. The level of confidence we can assign to the results depends on the skill at which the experiment is conducted, the quality of the experimental design and subsequent analysis, and, most important, the power in the study. Any microarray experiment consists of several components: (1) carrying out an appropriately designed (replicated) plant experiment; (2) array processing, which includes several steps of data acquisition and normalization; and (3) analysis of expression data to identify differentially expressed genes and overall patterns of expression. Numerous software packages are available to assist in performing these steps and it is not our intent to provide a software users manual or a statistical review. It is our intent to provide a brief user's explanation of these various components and present the commonly used methods.

Key Words: Microarray data analysis; experimental design; normalization; differential expression, cluster analysis.

1. Introduction

1.1. Gene Expression Platforms

There are essentially two types of array platforms: spotted and *in situ* synthesized (ISS) arrays. Spotted arrays are glass slides with a grid of probes applied by placing a droplet of probe solution on the surface, a solution of either an amplicon or a synthesized oligonucleotide ranging from 50 to 80 bases. Due to difficulty in precisely controlling the amount of liquid spotted on a slide, there is significant variation in probe quantity across slides. In the case of ISS arrays, multiple features for each gene are synthesized on the array itself one base at a time; therefore, there is less variation for each feature across arrays than with spotted arrays. Only a single sample is hybridized to an ISS array, whereas two samples are hybridized simultaneously to a spotted array. Several companies, including Affymetrix, NimbleGen, Febit, and Oxford Gene Technology, manufacture ISS arrays. One exception is Agilent Technologies arrays; 60 base oligonucleotides are synthesized directly on the glass slide to be used for dual-label hybridizations.

1.2. Experimental Design

Experimental design is paramount to obtaining the final goal, an accurate estimate of *relative* gene expression. Replication of the microarray experiment is needed to estimate the levels of technical variation and biological variation in order to detect differential expression with a high level of confidence. The inherent variability in the biological system and the desired accuracy will in part determine the number of replicates. Regardless, it is highly recommended to

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

use at least three biological replicates. If the total number of microarrays is limited, they should be divided equally over the different treatments. More replicates will give a better estimate of the variance and thus more confidence in calling a gene differentially expressed. Technical variation is caused by differences in efficiency of the labeling reaction, i.e., a difference between two reactions from the same RNA sample, slide-to-slide variation, and artifacts that may arise during washing, hybridization, and quantification. Technical replicates provide an estimate of the reproducibility of the microarray assay and should be considered repeated measures in contrast to biological replicates. Only by measuring the differences among independent plant experiments do we gain the power to understand the significance of the differences measured between treatments. Plant growth conditions should be kept as constant as possible, i.e., sample the same type of tissue at the same reproductive age from plants cultivated together (same growth media, nutrients, light quality and quantity, etc.). Deviating from stringent conditions will add noise and reduce power. Great consideration should be given to determining the time of day at which the tissue is sampled. More than 10% of the transcriptome is under diurnal or circadian clock regulation ***(1)***. Consider a time course experiment in which leaf tissue is sampled every 3 h following a cold treatment. Due to circadian clock regulation, changes in gene expression may be measured that occur regardless of treatment and the researcher wrongly considers them directly related to the treatment. Untreated plants grown in parallel are required to control for circadian effects. There are several excellent reviews on microarray experimental design ***(2–4)***.

1.3. Normalization

The intensity measure from each probe or probe-set on an array serves as an estimate of the abundance of the corresponding sequence in the experimental sample. There is a great deal of inherent bias in those values. Prior to labeling, the samples may differ in absolute quality of RNA and there may be subsequent differences in labeling and detection. There may also be random spatial variation across the array. The purpose of normalization is to adjust for these biases as they are caused by the microarray technology rather than the biological variation in the experiment. Normalization can be applied within an array or between arrays and is necessary prior to any subsequent analysis of the data.

1.4. Statistical Issues in Microarray Data Analysis

Extracting biologically relevant information from a microarray experiment requires appropriate statistical methods. In the early days of microarrays, a simple fold change cutoff was used to determine differential expression. Today, this practice is generally unacceptable. Various statistical methods have been applied to microarray data using estimates of experimental error that include probability statistics. In the case of microarray data, there are extremely high numbers of observations for each sample and the number of replicates is generally low. Due to the large number of genes (observations) on an array, a traditional significance threshold of $P = 0.01$ will identify 1% of the genes by chance alone. In the context of 10 to 50 thousand genes, that value is unacceptable. Modifications to traditional analysis of variance (ANOVA) and *t*-test have been made in order to adjust for multiple measures and small sample sizes ***(5)***. Adjusting for multiple measures is accomplished by selecting a significance threshold with an acceptable false discovery rate (FDR), i.e., the rate at which genes are incorrectly declared differentially expressed ***(6)***.

1.5. Cluster Analysis

Microarray experiments often include large numbers of arrays as well as multiple treatments, tissues, genotypes, or time points. It is beyond our ability to visualize or even imagine the patterns within the data set simply by looking at the expression level of several thousand genes in multiple samples. The goal of multivariate analysis is to reveal underlying structure, i.e., patterns of gene expression in an otherwise overwhelming data set. When clustering genes, we seek to find which genes have similar expression patterns among conditions. When clustering treatment, we seek to find which genes have similar expression patterns and as a result

reveal coregulation and regulatory networking. For this purpose, hierarchical clustering, nonhierarchical clustering, and ordination have successfully been applied to microarray data. Many software packages are capable of performing several types of analysis. Several forms of cluster analysis should be attempted. The most informative and appropriate depend on the needs of the researchers and the design of the experiment.

1.6. Two-Color Experiments—Specific Issues

Due to variation in probe quantity across slides, two samples are competitively hybridized to a spotted array. The expression value derived from a spotted array is the ratio of hybridization intensity from the two samples, thus internally controlling for slide-to-slide variation. In a dual-labeling experiment, the researcher must select a query sample and hybridize it together with a reference sample. There are several experimental designs for choosing the appropriate reference, the most common types are direct comparison (with or without dye-swap), reference design, and loop design ***(2).*** Biological replicates must always be included regardless of design type. The advantage of the direct comparison is that a specific control can be used for every condition. In the reference design a common reference is used on all arrays in the experiment, allowing all slides to be indirectly compared through this control, including array assays across multiple laboratories. The disadvantage of a reference design is that only a single biological condition is assayed per slide. In the loop design all samples are interrelated, enabling indirect comparisons across arrays and direct comparisons on arrays. Dye swaps are recommended to control for bias from the different dyes.

1.7. ISS Array Experiments Specific Issues

On an ISS array every gene is represented by multiple features. After scanning and quantification the multiple values for the gene have to be summarized into a single value. The area of ISS array probe set summary is very active and competitive ***(7)***. We recommend using either DNA chip analyzer (dChip; ***8***) or robust multi array analysis (RMA; ***9***). For every oligonucleotide on many Affymetrix array designs, including the AtGenome1 and ATH1 *Arabidopsis* arrays, there is a corresponding mismatch (MM) feature. A MM feature contains a single nucleotide change in the middle of the sequence. Hybridization to that feature is predicted to represent background and nonspecific hybridization. The Affymetrix MAS 4.0 and 5.0 software considers the values from these features for probe set summarization. The most efficient approaches to normalizing ISS arrays today do not include the use of MM features ***(7)***. Therefore, we recommend against using MAS 4.0 or 5.0 for probe set summarization. Array data sets derived from the use of MAS 4.0 or 5.0 should be resummarized and normalized using other methods.

2. Methods

2.1. Image Acquisition for Two-Color Arrays

The first step in data acquisition is the physical scanning of the microarray slide. In a dual labeling experiment the array is scanned at two wavelengths: 532 nm and 635 nm for Cy3 and Cy5, the most commonly used labeling dyes. All scanners are accompanied by image acquisition software. This software allows for adjustments in laser power and photo multiplier tube (PMT) voltage values. By manipulating these values, the array can be scanned with a balance of the two wavelengths. Different scanners and their software have different options to achieve this goal. It is common practice to try to balance the two channels as much as possible after which the data are further normalized.

2.1.1. Spot Intensity Quantification for Two Color Arrays

A number of commercial and open-source software packages are available for spot intensity quantification. The software will first have to recognize all the features in the image and then quantify the pixels within the feature to calculate the relative fluorescent intensities for each probe. This can be challenging, as spots may have irregular shapes and the array can have areas with high background. Reviewing the differences between the available software packages is

outside the scope of these protocols and progress is still being made. Regardless of choice, the end products are signal and background intensity values for both channels. The most commonly used expression measure is the mean or median pixel intensity of the feature, although the total intensity can also be used. As this is the product of average intensity and the total number of pixels in the feature, there is no principal difference between the two methods.

2.1.2. Background Subtraction for Spotted Arrays

It is common practice to subtract a local background value from the spot intensity. Although the background subtraction should be done with great care and the experimenter should check the effect on the data, as it may affect reproducibility. As with acquisition and quantification, the options and methods for computing background values can be software-specific. The background values enable the calculation of a signal-to-noise ratio for determining the confidence with which one can quantify a signal peak of a given value, especially a signal near background. The confidence increases as the variation in background decreases, regardless of the absolute value of the average background. The signal-to-noise ratio is calculated as the difference between median signal and median background divided by the standard deviation of the background. For every array a signal-to-noise cutoff can be used to determine whether to use a probe or not.

2.1.3. Flagging Bad Features

The final step in data acquisition for both dual- and single-labeled systems is to "flag" bad features. Flagging allows the user to discard data points or spots that do not meet certain criteria, such as diameter or number of saturated pixels. Areas with high background of scratches can also be excluded. The disadvantage will be that the data set contains a number of missing values. Because of technical issues such as mechanical and washing artifacts and detection limits, it is virtually impossible to obtain expression ratios for every spot on the array. Special consideration should be given to these situations in which a signal can be detected in only one of the channels. In these cases a ratio cannot be calculated but the spot may still provide valuable information.

2.2. Normalization of Two-Color Arrays

2.2.1. Global Intensity Normalization

Global methods assume that the intensities in the two channels are related by a constant factor. The distribution of the intensity log ratios is forced to have a median of zero for each slide. The disadvantage of this method is that it does not take into account intensity or spatial biases.

2.2.2. Intensity-Dependent Normalization

2.2.2.1. Global

With this normalization method, the two channels are adjusted by an additional factor. A locally weighted scatter-plot smoothing (lowess) curve is constructed by performing a series of local regressions and the corresponding value can be used for normalization. The lowess scatter-plot smoother performs robust locally linear fits whereas with the global intensity normalization a single adjustment factor is used. The advantage of the local linear regression is that it is not affected by outliers and corrects for intensity-based trends.

2.2.2.2. Block (Print-Tip)

A refinement of the intensity-dependent normalization is to calculate the lowess fit on a per-grid basis. This is highly recommended as every grid is printed by a different print-tip, if one print-tip deposits more DNA than other tips the intensities in that particular block tend to be higher. The block-lowess normalization will compensate for this print-tip spatial effect. Print-tip lowess normalization provides a well-tested general-purpose normalization method that has given good results on a wide range of arrays.

2.2.2.3. Composite Lowess Normalization

The normalization methods mentioned thus far use all of the genes on the array. An alternative approach is to use a subset of genes on the array using the same algorithms. For example, a set of control spots can be used that are known not to be differentially expressed. These spots should span a range of intensities as wide as possible.

2.2.2.4. Scale Normalization

The global intensity and intensity-dependent normalization methods adjust the mean of the log ratios, but they do not affect the variance of the log ratios. To adjust for these differences, an additional normalization step, known as scale normalization or standard deviation regularization, can be applied to the data. Starting with the lowess location normalized data, a regularization factor can be calculated to adjust the variance of the log ratios. A potential disadvantage of this method is that it can limit the detection of differentially expressed genes, but may reduce false positives. This scale normalization can also be applied between slides. For more detailed information on these normalization methods *see* refs. ***10*** and ***11***.

2.3. Normalization of ISS Arrays

These methods are discussed by Bolstad et al. ***(12)*** and are part of the Bioconductor package, http://www.bioconductor.org.

2.3.1. Lowess

This is similar to lowess normalization as described for two-color arrays; however, multiple arrays are normalized by looking at all pairwise combinations of arrays. Normalization of a large number of arrays may be computationally time consuming.

2.3.2. Contrasts

Normalization is done by transforming the data to a set of contrasts, then normalizing a scale by the geometrical mean across the arrays ***(13)***.

2.3.3. Quantile

The goal of this quick and favored method is to force the distribution of each array to be equal ***(12)***. The values for each array are ordered and then averaged across the array before being returned to the original order. A potential drawback of this approach is losing signal in the tail of the distribution, but empirical evidence suggests that this is not a problem.

2.3.4. Nonlinear Method

A distinct method to using all of the array information for normalization is to fit the arrays to a baseline array. A model-based approach executed by DNA-chip analyzer (dChip) uses nonlinear smooth curves and a baseline array selected from the experiment ***(8,14)***. The main drawback of this approach is that the results rely on a reference array.

2.3.5. Scaling

Each array is adjusted so that the average intensity across arrays is the same. This approach erroneously assumes a linear relationship among arrays and across intensities and should be avoided.

2.4. Detection of Differential Expression

2.4.1. t*-Test*

A *t*-test is a statistical test used to determine if two groups of numbers are different. In the case of microarray data it can be used to determine whether the mean expression of a gene is different between treatments. A *t*-test approach that is modified to account for the difficulty in estimating error variance due to small sample size such as the statistical analysis of microarrays

(SAM), regularized *t*-test or the B statistic ***(5)*** should be used. The SAM ***(15)*** adds a small positive constant to the denominator of the *t*-test. With this modification, genes with small fold changes will not be selected as differentially expressed. This results in a compromise between fold change and relative change and guards against artificially small errors estimated from very few replicates.

2.4.2. False Discovery Rate

A false discovery occurs when a gene is actually not differentially expressed, but is called as such. A false discovery rate is the percentage of truly nondifferentially expressed genes among those that are called differentially expressed. First create a "nonsense" data set by calculating *t*-statistics for each possible array pair. The FDR is the average number of genes found to be differentially expressed (i.e., exceed the threshold) divided by the number of genes exceeding the threshold in the genuine array comparisons. This approach is highly adaptable. Which FDR will be acceptable depends on the follow-up study; if the follow-up study is to conduct extensive research on each gene identified as differentially expressed, very few false positives are acceptable. If the study is to survey changes in gene expression in response to a particular treatment, then the researcher may be willing to consider more false discoveries for the sake of avoiding elimination of relevant changes.

2.4.3. Analysis of Variance

Analysis of variance (ANOVA) is another technique for detecting statistical significance. This model accounts for multiple sources of variation including interaction effects such as that between treatments and genotypes that cannot be accounted for jointly in a *t*-test ***(16)***. An *F*-statistic, calculated as the ratio of the factor variance and the error variance, is used to estimate the *P*-value. An FDR can also be used to determine the appropriate significance threshold. These methods, as well as a mixed-model approach ***(17)***, are part of many software analysis packages, including the R-package.

2.4.4. Linear Models for Microarray Data

For assessing statistically significant differential expression, linear models for microarray data (LIMMA) uses an empirical Bayes method to moderate the standard errors of the estimated log-fold changes. This results in more stable inference and improved power, especially for experiments with small numbers of arrays. The moderated *t*-statistics use sample standard deviations that have been shrunk toward a pooled standard deviation value. For arrays with within-array replicate spots, LIMMA uses a pooled correlation method to make full use of the duplicate spots ***(18)***. The LIMMA package is part of Bioconductor.

2.5. Multivariate Methods

2.5.1. Distance and Similarity Matrix

The first step in clustering methods is to create a matrix of distance, such as Euclidean, maximum, Mahalonobis, and Manhattan; or similarity, such as Pearson's correlation (noncentered, and centered). This can be done for either samples or genes.

2.5.2. Hierarchical Clustering

There are several algorithms for hierarchical cluster analysis: single, complete, centroid, average (aka UPGMA), Ward's, and flexible beta. The UPGMA clustering is very simple to carry out and has proven to be very helpful ***(19,20)***. Eisen et al. ***(20)*** provide an early example of this technique. There are several drawbacks to hierarchical clustering: (1) The dendrogram may be visually difficult to interpret. (2) Be aware that a dendrogram will be produced regardless of the level of structure in the data as the analysis imposes a hierarchical structure whether or not such a structure exists. (3) An object (a gene or a sample) can only be part of a single

cluster. (4) The method is bottom-down and does not evaluate multiple possibilities. It is recommended to asses the stability of the clusters by bootstrapping and building support trees ***(21)***.

2.5.3. Nonhierarchical Clustering

A method of partitioning, *k*-means clustering, assigns genes to a predetermined number of groups so that the variation within group is minimized and the variation among groups maximized. A disadvantage of this analysis, unlike hierarchical clustering, is that the number of clusters must be predetermined. Also, as with hierarchical clustering, genes can be part of only one group. One approach is to determine the number of groups from hierarchical clustering and principal component analysis. A modified version of this analysis, fuzzy *k*-means clustering, allows for genes to be part of more than one group, thus reducing the effect of having to a *priori* identify a number of clusters ***(22)***.

2.5.4. Principal Component Analysis

This approach reduces multidimensional data to two and three dimensions. This allows us to visualize it in two- or three-dimensional figures. The new components are uncorrelated and axis or dimensions of the new space are ordered with respect to the amount of variance they explain.

2.5.5. Self-Organizing Maps

Self-organizing maps (SOMs) assign objects to discrete groups and are less restrictive than hierarchical methods ***(23)***. One needs only to impose a partial structure, thus avoiding erroneous estimation of group number. We recommend that the user employ a number of multivariate methods. For example, first create a dendrogram using UPGMA clustering of a correlation matrix ***(20)***. Determine the number of clusters by simply looking for what appear to be natural groupings by focusing on the upper levels. Analyze the data using a nonhierarchical method, possibly setting the group number to that determined from the hierarchical clustering. Ordination analysis should also provide similar results. A consistent result across methods implies that the structure in the data set is real and not an artifact of the analysis. See ref. ***24*** for further discussion of the aforementioned clustering methods as well as supervised methods.

Acknowledgments

We would like to thank Justin Borevitz, Yunda Huang, and Ivan Baxter for helpful comments and discussion.

References

1. Harmer, S. L., Hogenesch, J. B., Straume, M., et al. (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290,** 2110–2113.
2. Yang, Y. H. and Speed, T. (2002) Design issues for cDNA microarray experiments. *Nat. Rev. Genet.* **3,** 579–588.
3. Churchill, G. A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.* **32 Suppl**, 490–495.
4. Kerr, M. K. and Churchill, G. A. (2001) Statistical design and the analysis of gene expression microarray data. *Genet. Res.* **77,** 123–128.
5. Cui, X. and Churchill, G. A. (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* **4,** 210.
6. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57,** 289–300.
7. Cope, L. M., Irizarry, R. A., Jaffee, H. A., Wu, Z., and Speed, T. P. (2004) A benchmark for Affymetrix GeneChip expression measures. *Bioinformatics*. **20,** 323–331.
8. Li, C. and Wong, W. H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* **98,** 31–36.
9. Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31,** e15.

10. Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30,** e15.
11. Quackenbush, J. (2002) Microarray data normalization and transformation. *Nat. Genet.* **32 Suppl**, 496–501.
12. Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19,** 185–193.
13. Astrand, M. (2003) Contrast normalization of oligonucleotide arrays. *J. Comput. Biol.* **10,** 95–102.
14. Schadt, E. E., Li, C., Ellis, B. and Wong, W. H. (2001) Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J. Cell Biochem. Suppl.* **Suppl 37,** 120–125.
15. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98,** 5116–5121.
16. Kerr, M. K., Martin, M., and Churchill, G. A. (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.* **7,** 819–837.
17. Wolfinger, R. D., Gibson, G., Wolfinger, E. D., et al. (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *J. Comput. Biol*. **8,** 625–637.
18. Smyth, G. K., Yang, Y. H., and Speed, T. (2003) Statistical issues in cDNA microarray data analysis. *Methods Mol. Biol*. **224,** 111–136.
19. Yeung, K. Y., Haynor, D. R., and Ruzzo, W. L. (2001) Validating clustering for gene expression data. *Bioinformatics* **17,** 309–318.
20. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95,** 14,863–14,868.
21. Kerr, M. K. and Churchill, G. A. (2001) Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc. Natl. Acad. Sci. USA* **98,** 8961–8965.
22. Gasch, A. P. and Eisen, M. B. (2002) Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biol.* **3,** RESEARCH0059.
23. Tamayo, P., Slonim, D., Mesirov, J., et al. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* **96,** 2907–2912.
24. Brazma, A. and Vilo, J. (2000) Gene expression data analysis. *FEBS Lett.* **480,** 17–24.

30

Preparation and Quality Assessment of RNA From Cell-Specific Samples Obtained by Laser Microdissection

Regine Kleber and Julia Kehr

Summary

Laser microdissection is a powerful tool to obtain cell-specific isolates from complex tissue samples. This chapter outlines how to prepare plant material for microdissection and methods to extract and measure high-quality RNA suitable for a variety of different downstream applications.

Key Words: Single cell; RNA; extraction; laser microdissection; laser pressure catapulting; tissue embedding.

1. Introduction

The measurement of RNA abundance of a given gene, or even a set of different genes, is one of the major goals of many experiments in the era of functional genomics. In multicellular organisms these types of experiments are complicated by the fact that tissues are composed of a range of specialized cell types with different complex transcription profiles. Most standard experiments are based on the analysis of samples containing a mixture of heterogeneous tissues, thus averaging gene expression of a vast number of different tissue and cell types. This often prevents a meaningful interpretation of results, especially because abundant cell types and transcripts can mask rare ones. Attempts to overcome this averaging effect require refining the spatial resolution of analyses to a tissue- or even cell-specific level *(1)*. Fundamental for the successful analysis of tissue-specific transcripts are the isolation of sufficient cell-specific material, the possibility to reproducibly isolate RNA, and the assessment of RNA quality to allow comparison of samples from different origins.

The following protocol describes the fixation and embedding of *Arabidopsis* material for obtaining morphologically well-preserved tissue sections, the isolation of cells by laser microdissection coupled to laser pressure catapulting (LMPC), and the extraction and quality assessment of RNA from microdissected material.

2. Materials

1. Diethylpyrocarbonate (DEPC, min. 97%, Sigma). Store under moisture-free conditions at 2 to 8°C.
2. RNaseZAP® (Sigma).
3. RNase-free water and solutions (DEPC-treated or purchased, for example, from Ambion, USA).
4. Fixative: 75% ethanol, 25% acetic acid; prepare fresh as required.
5. Alcohol series: 100, 95, 90, 85, 80, and 75% ethanol.
6. Eosin Y disodium salt (Sigma).
7. Xylene.
8. Paraffin wax (Paraplast, Sigma).
9. Embedding molds or small Petri dishes.
10. Polyethylene naphthalate(PEN)-slides (PALM Microlaser GmbH, Germany).

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

11. UV lamp (254 nm).
12. Rotary microtome with equipment.
13. PALM® MicroBeam system (PALM Microlaser GmbH).
14. LPC solution: 0.5% Igepal CA-630 sterile filtrated (Sigma), 1 U/μL RNasin (Promega). Prepare fresh every time before use.
15. QIAshredder® columns (Qiagen).
16. RNeasy® Micro Kit (Qiagen). Store the RNeasy® MiniElute Spin Column and the RNase-free DNase set at 4°C.
17. Agilent 2100 bioanalyzer with equipment (Agilent Technologies, USA).
18. RNA 6000 Pico LabChip® Kit (Agilent Technologies).
19. RNA 6000 ladder (Ambion).

3. Methods

The methods described below outline (1) the preparation of specimens (2) the process of laser microbeam microdissection and laser pressure catapulting (LMPC) using the PALM® MicroBeam system, (3) the isolation of high quality RNA out of the minute amounts of microdissected tissue, and (4) the examination of the achieved RNA yield and integrity using the capillary electrophoresis technology of the Agilent's 2100 bioanalyzer in combination with Agilent's "Lab on a Chip" system.

As a general precaution the entire procedure requires lab work under RNase-free conditions.

1. Wear gloves at all times and change them frequently.
2. Work on decontaminated RNase-free surfaces and use only RNase-free equipment (if necessary decontaminate surfaces and equipment by rinsing or wiping with an RNase-destructive agent such as RNaseZAP®.
3. To prevent contaminations (e.g., from previous experiments or from lab work where RNase steps are included, such as isolation of DNA or plasmids), it is recommended to define a special working space in the lab for RNA work only and to use aerosol-filtered tips and sterile, disposable plasticware.
4. All water and solutions should be RNase-free by treatment with 0.1% DEPC to completely inactivate RNases; alternatively, solutions should be prepared with DEPC-treated water (*see* **Note 1**).
5. To perform DEPC treatment, add 1 mL DEPC solution to 1 L of water or aqueous solution and stir the solution until the DEPC is completely dissolved.
6. Incubate the solution for a few hours, usually overnight. DEPC is sensitive to moisture and therefore, in aqueous solutions, decomposes to ethanol and carbon dioxide.
7. Because DEPC may disturb downstream applications, autoclave the solution twice to destroy any residual DEPC.

3.1. Specimen Preparation

The preparation of *Arabidopsis thaliana* tissue for subsequent laser microdissection and RNA isolation is described in **Subheadings 3.1.1.–3.1.3.** This includes (1) tissue fixation and dehydration, (2) embedding in paraffin wax, and (3) sectioning followed by slide preparation.

3.1.1. Tissue Fixation and Dehydration

The first aim in specimen preparation is to prevent breakdown of RNA by endogenous RNases, and therefore preserve the expression pattern of RNA species within the sample (*see* **Note 2**).

1. Prepare a fresh mixture of fixative.
2. Dissect *Arabidopsis* organ of interest and cut it in smaller pieces (*see* **Note 2**).
3. Submerge the dissected tissue pieces immediately in the fixative. The fixative:specimen volume ratio should not be less than 10:1.
4. To facilitate infiltration of the fixative apply a smooth vacuum at the submerged tissue by means of a desiccator and a vacuum pump for a few seconds and release the vacuum slowly. Repeat this step three times if necessary (*see* **Note 3**).

5. Incubate the tissue at least for 30 min in the fixative (*see* **Note 4**).
6. Transfer the fixed tissue into 75% ethanol, facilitate infiltration as described in **step 4**, and subsequently incubate the tissue for at least 30 min.
7. Transfer the tissue into 85% ethanol and add just a few crystals of eosin (*see* **Note 5**) to stain the tissue. Facilitate infiltration as described in **step 4**, and subsequently incubate the tissue for at least 30 min.
8. Transfer the tissue into 90% ethanol, facilitate infiltration as described in **step 4**, and subsequently incubate the tissue for at least 30 min.
9. Transfer the tissue into 95% ethanol at room temperature, facilitate infiltration as described in **step 4**, and subsequently incubate the tissue for at least 30 min.
10. Transfer the tissue into 100% ethanol at room temperature, facilitate infiltration as described in **step 4**, and subsequently incubate the tissue for at least 30 min.
11. To ensure that the tissue is completely dehydrated, **step 10** must be repeated.

3.1.2. Paraffin Wax Embedding

1. Prepare a fresh mixture of 50% ethanol and 50% xylene (*see* **Note 6**).
2. Melt an appropriate amount of paraffin wax at 60°C in a water bath (*see* **Note 7**).
3. Prewarm an appropriate amount of 100% xylene at 60°C.
4. Remove air bubbles of the melted paraffin wax by applying a smooth vacuum if necessary and keep it at 60°C (*see* **Note 8**).
5. Place the fixed and dehydrated tissue into the mixture of 50% ethanol and 50% xylene and incubate for 30 min at room temperature.
6. Transfer the tissue into 100% xylene and incubate it for 1 h at room temperature while replacing the xylene once.
7. Transfer the tissue into prewarmed 100% xylene and incubate it for 30 min at 60°C.
8. Place the tissue into melted, air-free paraffin wax at 60°C and incubate the tissue for 1.5 h at 60°C while replacing the paraffin wax three times. There should be no remaining xylene odor recognizable.
9. Pour paraffin wax into a mold or Petri dish and place the specimen in the desired position at 60°C by means of a prewarmed needle and tweezers (*see* **Note 9**).
10. Let the paraffin wax cool down for hardening at room temperature.
11. Store the embedded tissue at 4°C with silica gel as desiccant.

3.1.3. Sectioning and Slide Preparation

1. Prepare RNase-free PEN slides (*see* **Note 10**):
   a. Dip the PEN slides for 30 s in RNaseZAP. Handle the slides with RNase-free tweezers and rock them gently in the solutions. Work carefully to avoid any injuries to the membrane.
   b. Wash the slides three times for 30 s with RNase-free water.
   c. Let the slides air-dry at 37°C in an oven (dried slides can be stored in a dustproof box for several weeks).
2. Irradiate RNase-free PEN-slides with ultraviolet (UV) light (253 nm) for 30 min to decrease hydrophobicity of the PEN membrane and therefore facilitate the adhesion of sections.
3. Clean every surface of the microtome as well as everything that will get in contact with the sections (e.g. blade, object table, tweezers, etc.) with RNaseZAP.
4. Mount the paraffin wax block on a carrier and clamp it in the microtome.
5. Cut sections of 4 to 40 µm thickness (*see* **Note 11**).
6. Wet a glass bar with RNase-free water, pick the sections while cutting directly from the microtome blade, and transfer them onto a waterdrop placed on the top of a PEN-slide.
7. Place the PEN slide on a heating plate and incubate at 42°C until the sections get stretched.
8. Let the sections air-dry at room temperature. If necessary absorb excess water carefully with a clean tissue to shorten the drying time.
9. Attach the sections onto the PEN slide by placing them for a few seconds on a heating plate at 60°C just until the sections get melted.

Store the PEN slides with the sections at 4°C with silica gel as a desiccant until microdissection (*see* **Note 12**).

3.2. Laser Microdissection

The process of laser microdissection of single cells, which is described in **Subheadings 3.2.1. –3.2.2.** starts with the removal of the paraffin wax from the sections and is completed by laser cutting and the collection of the cells by catapulting with the PALM MicroBeam system.

3.2.1. Removal of Paraffin Wax

1. Place the slide with the mounted sections in fresh xylene and allow the paraffin wax to dissolve for 1 to 3 min.
2. Let the slide dry under a fume hood for a few minutes and process immediately with laser microdissection (*see* **Note 13**).

3.2.2. Laser Microdissection and Pressure Catapulting

One alternative to get single-cell samples is to make use of LMPC using the PALM MicroBeam system (*see* **Note 14**).

1. Start up and calibrate the PALM MicroBeam system following the manufacturer's instructions.
2. Prepare fresh LPC solution and keep it on ice.
3. Prepare fresh lysis buffer by adding 10 µL β-mercaptoethanol (β-ME) per 1 mL RLT buffer (Qiagen) (*see* **Notes 15**, **16**).
4. Mount the slide at the PALM MicroBeam system and select all cells or tissue areas of interest by means of the PALM RoboSoftware (*see* **Note 17**).
5. Apply 10 µL of LPC solution into the cap of a special PALM tube, mount it at the cap holder of the microscope, and bring it into the appropriate position; then start directly with the automatic laser cutting and catapulting of the selected cells.
6. After 15 min of sampling stop the PALM MicroBeam system and pipet immediately 190 µL of lysis buffer into the tube and invert it directly to get the sample dissolved and therefore protected against RNA breakdown (*see* **Note 18**).
7. Collect the sample in the bottom of the tube by spinning down for 2 min at full speed using a tabletop centrifuge.
8. The samples can be stored at –80°C up to 2 mo until further processing.

3.3. RNA Isolation

Outlined is the procedure for RNA isolation (*see* **Note 19**) according to the RNeasy Micro Handbook (Qiagen).

1. Prepare RPE working solution before using the RNeasy Micro Kit the first time. Add 4 vol of 100% ethanol to the provided RPE concentrate.
2. Prepare DNase I stock solution by dissolving DNase I (1500 U of RNase-free DNase) in 550 µL of the RNase-free water provided and mix gently without vortexing. For long-term storage, divide the stock solution into single-use aliquots and store at –20°C for up to 9 mo. Thawed aliquots can be stored at 2 to 8°C for up to 6 wks. Do not refreeze the aliquots after thawing.
3. Prepare carrier RNA stock solution before using the kit for the fist time. Dissolve 310 µg of the carrier RNA in 1 mL RNase-free water provided. Store the stock solution at –20°C, and use it to make fresh dilutions for each set of RNA preparations.
4. Prepare carrier RNA working solution for 10 RNA preparations by adding 5 µL of the carrier RNA stock solution to 34 µL of lysis buffer and mix by pipetting. Take 6 µL of this diluted solution and add it to 54 µL of lysis buffer (final concentration of the working solution is 4 ng/µL).
5. Thaw the microdissected samples and adjust the volume to 350 µL with lysis buffer or by pooling several samples respectively. Incubate the samples for 5 min at 56°C with following vortexing for 1 min to assist lysis.
6. Add 5 µL of Carrier RNA working solution (4 ng/µL) per 350 µL of lysate to enhance RNA yield out of little starting material.

7. Pipet the lysate directly onto a QIAshredder Spin Column placed in a 2-mL collection tube and centrifuge for 2 min at maximum speed. Centrifugation through this column removes cell debris and simultaneously homogenizes the lysate (*see* **Note 20**).
8. Carefully transfer the supernatant of the flow-through fraction to a new microcentrifuge tube (not supplied) without disturbing the very small and transparent cell-debris pellet in the collection tube (*see* **Note 21**).
9. Add 350 µL (1 vol) of 70% ethanol to 350 µL of the homogenized lysate, to promote selective binding of the RNA to the RNeasy MiniElute membrane. Mix well by pipetting, and do not centrifuge (*see* **Note 22**).
10. Apply the sample, including any precipitate that may have formed, to an RNeasy MiniElute Spin Column in a 2-mL collection tube (supplied). Close the tube gently, and centrifuge for 15 s at ≥8000*g*. Discard the flow-through. Reuse the collection tube in **step 11**. If the volume of a sample exceeds 350 µL—for example, due to pooling of samples—reuse the QIAshredder spin column and the RNeasy MiniElute column and repeat the centrifugation steps until the entire RNA of one sample is bound to the silica-gel membrane.
11. Add 350 µL buffer RW1 to the RNeasy MiniElute spin column. Close the tube gently, and centrifuge for 15 s at ≥8000*g* to wash the column. Discard the flow-through. Reuse the collection tube in **step 12**.
12. Add 10 µL DNase I stock solution to 70 µL RDD buffer per sample. Mix by gently inverting the tube. DNase I is especially sensitive to physical denaturation. Mixing should therefore only be carried out by gently inverting the tube. Do not vortex.
13. Pipet 80 µL of the DNase I incubation mix directly onto the RNeasy MiniElute silica-gel membrane, and place on the benchtop at room temperature for 15 min. Make sure to pipet the DNase I incubation mix directly onto the RNeasy MiniElute silica-gel membrane. DNase digestion will be incomplete if part of the mix adheres to the walls or the O-ring of the RNeasy MiniElute Spin Column (*see* **Note 23**).
14. Pipet 350 µL buffer RW1 into the RNeasy MiniElute spin column, and centrifuge for 15 s at ≥8000*g*. Discard flow-through and the collection tube.
15. Transfer the RNeasy MiniElute spin column into a new 2-mL collection tube (supplied).
16. Pipet 500 µL buffer RPE onto the RNeasy MiniElute spin column. Close the tube gently and centrifuge for 15 s at ≥8000*g* to wash the column. Discard the flow-through. Reuse the collection tube in **step 17**.
17. Add 500 µL of 80% ethanol to the RNeasy MiniElute Spin Column. Close the tube gently and centrifuge for 2 min at ≥8000*g* to dry the silica-gel membrane. Following the centrifugation, remove the RNeasy MiniElute Spin column from the collection tube carefully so the column does not come into contact with the flow-through, as this will result in carryover of ethanol.
18. Transfer the RNeasy MiniElute spin column into a new 2-mL collection tube (supplied).
19. Open the cap of the spin column and centrifuge in a microcentrifuge at full speed for 5 min. Discard the flow-through and collection tube (*see* **Note 24**).
20. To elute, transfer the spin column to a new 1.5-mL collection tube (supplied). Pipet 14 µL RNase-free water (supplied) directly onto the center of the silica-gel membrane and incubate for 2 min at room temperature (*see* **Note 25**).
21. Centrifuge for 1 min at maximum speed to elute the RNA. The dead volume of the RNeasy MiniElute Spin Column is 2 µL; elution with 14 µL of RNase-free water results in an eluate with a volume of 12 µL.
22. Store RNA at –80°C. Avoid repeated freeze-thaw cycles, as this can compromise RNA integrity tremendously.

3.4. Examination of RNA Yield and Integrity

Lab-on-a-Chip technology using the Agilent 2100 bioanalyzer combined with the RNA 6000 Pico LabChip kit is the only technique available so far to analyze as little as 200 pg of total RNA not only for RNA concentration measurement but also to inspect RNA integrity (*see* **Note 26**).

3.4.1. Setting Up the Assay Equipment, Solutions and Bioanalyzer

1. Before beginning the chip preparation protocol, ensure that the Chip Priming Station (supplied with the bioanalyzer) and the Agilent 2100 bioanalyzer are set up and ready to use.

2. Make sure that you start the software before loading the chip because the chip must be measured within 5 min after loading, as reagents might evaporate, leading to poor results.
3. Allow all reagents supplied with the RNA 6000 Pico LabChip kit to equilibrate to room temperature. Protect the dye concentrate from light by wrapping the tube with aluminum foil.
4. Initial electrode cleaning (once a day):
   a. Slowly fill one of the wells of an electrode cleaner (supplied with the kit) with 350 μL RNase-free water.
   b. Open the lid and place it in the bioanalyzer.
   c. Close the lid and leave it closed for 5 min.
   d. Open the lid and remove the electrode cleaner. Label it and keep it for further use.
   e. Wait another 30 s for the water on the electrodes to evaporate before closing the lid.
5. Prepare the diluted RNA 6000 ladder:
   a. Place 5 μL of RNA 6000 ladder in a 1.5-mL tube and heat at 70°C for 2 min. Spin down the tube and keep the ladder on ice.
   b. Add 745 μL of RNase-free water, vortex, and spin down the tube.
   c. Store diluted RNA ladder in 10 μL aliquots at –80°C.
6. Prepare the gel:
   a. Apply 550 μL of RNA 6000 Pico gel matrix on a spin filter (supplied).
   b. Place the spin filter in a microcentrifuge and spin at 1500*g* ± 20% for 10 min.
   c. Aliquot 65 μL filtered gel into 0.5 ml RNase free tubes and store them at 4°C. Use the aliquots up within one month of preparation.
7. Prepare the gel-dye mix:
   a. Add 1 μL of RNA 6000 Pico Dye Concentrate to a 65-μL aliquot of filtered gel (*see* **Note 27**).
   b. Vortex the mixture thoroughly and visually inspect the proper mixing of gel and dye.
   c. Spin the tube at 13,000*g* for 10 min at room temperature.
   d. Protect the gel-dye mix from light by wrapping the tube with aluminum foil.

3.4.2. Loading of RNA 6000 Pico LabChip

The loading of the chip should be performed efficiently to avoid evaporation of the solutions (*see* **Note 28**).

1. Take a new RNA Pico chip out of its sealed bag and place it on the Chip Priming Station. Make sure that the Chip Priming Station base plate is in the correct position (C) before loading the gel-dye mix. Make also sure that the adjustable syringe clip is set to the upper position.
2. Pipet 9 μL of the gel-dye mix at the bottom of the well marked by a "G" within a blackcircle and dispense the gel-dye mix.
3. Set the timer to 30 s. Make sure that the plunger is at 1 mL, and then close the Chip Priming Station. The lock of the latch will click when the Priming Station is closed correctly.
4. Press the plunger until it is held by the syringe clip and wait for exactly 30 s.
5. Release the plunger with the clip release mechanism, wait for 5 s, and then slowly pull back the plunger to the 1 mL position.
6. Open the Chip Priming Station and turn over the chip to check the capillaries for air bubbles (*see* **Note 29**).
7. Pipet 9 μL of the gel-dye mix in each of the wells marked with an uncircled "G" and discard remaining gel-dye mix.
8. Pipet 9 μL of the RNA 6000 Pico Conditioning Solution into the well marked with "CS".
9. Pipet 5 μL of the RNA 6000 Pico Marker into the ladder well marked with the ladder symbol and each of the 11 sample wells. Do not leave any wells empty or the chip will not run properly. Add 6 μL of the RNA 6000 Pico Marker to every unused sample well.
10. Pipet 1 μL of the diluted RNA 6000 ladder into the ladder well marked with the ladder symbol. Do not heat the diluted ladder and use thawed diluted ladder within 1 d.
11. Pipet 1 μL of each sample into each of the 11 sample wells. To minimize secondary structure, you may heat denature the samples at 70°C for 2 min before loading onto the chip.

12. Place the chip in the adapter of the vortex mixer (supplied with the Agilent 2100 bioanalyzer). Vortex for 1 min at the IKA vortexer setpoint.
13. Place the chip in the Agilent 2100 bioanalyzer (*see* **Notes 30** and **31**).
14. Start the chip run within 5 min by pressing the start button on the software screen. One assay run takes 30 min to be completed. As result you get an electropherogram including a gel-like picture of every sample well and from the ladder well (*see* **Notes 32** and **33**).

3.4.3. Cleaning Up After Every Chip Run

1. When the assay is completed, immediately remove the used chip from the Agilent 2100 bioanalyzer and dispose of it according to the guidelines established by your laboratory safety officer.
2. To clean the electrodes slowly fill one of the wells of the electrode cleaner with 350 μL of fresh RNase-free water and place it in the bioanalyzer, then close the lid for about 30 s.
3. Remove the electrode cleaner and wait another 30 s for the water on the electrodes to evaporate before closing the lid.

4. Notes

1. DEPC is highly toxic and should be handled in a fume hood. Wear appropriate protective clothing.
2. Fixation and the subsequent dehydration of the tissue in preparation for the embedding process leads at first to the fast inhibition of RNase activity by the coagulation of proteins and, further, to the complete inactivation of metabolism by additional dehydration. To allow the fixative to permeate the tissue as fast as possible, even one dimension of the tissue should not exceed 5 mm. As an alternative to tissue fixation using alcohol-based fixatives, plant samples may also be snap-frozen in liquid nitrogen as is commonly done with animal and human samples. During freezing RNA integrity is sustained but tissue damage may occur by the frequent formation of ice crystals, which can be a problem especially with plant samples (*1*).
3. Vacuum infiltration is especially important with tissues containing large amounts of air-filled intercellular spaces, e.g., mature leaf tissue. These air-filled spaces impede liquid infiltration into the tissue. This problem can be overcome by applying a vacuum that facilitates the exchange of air against liquid. For other tissues this treatment may not be necessary or even bad, because extensive vacuum infiltration can also lead to tissue damage, especially with young and flimsy tissues. To prevent tissue damage it is therefore important to apply only a smooth vacuum using a water driven pump or a pump with adjustable vacuum pressure. Pay attention that no visible air bubbles are built during application and release of the vacuum, and release the vacuum slowly.
4. The given incubation times for tissue fixation and dehydration are minimum times. Prolonged incubation is possible and also recommended dependent on the tissue of interest.
5. Eosin staining will facilitate the recovery of the tissue within the cured paraffin wax blocks.
6. Xylene should be handled under a fume hood.
7. According to experience, paraffin wax as embedding medium preserves very good tissue morphology and, in addition, the embedded specimens become durable for a longer period of time (at least for several months or even longer if stored at 4°C together with a desiccant). At the same time paraffin wax itself can cause chemical degradation of RNA; furthermore it is also likely that physical degradation takes place by the need of several heat incubation steps during the embedding procedure. For these reasons RNA integrity is naturally compromised by this method. To minimize these negative influences of paraffin wax embedding and to simultaneously optimize the preservation of tissue morphology, the incubation steps at 60°C outlined below should be kept as short as possible.
8. A delay in boiling can occur very easily while evacuating hot paraffin wax.
9. If the tissue must be embedded with a special alignment to enable sectioning in the desired orientation, arrangement of the specimen should be carried out at 60°C using prewarmed molds and an appropriate oven or heating plate. In most cases it is much easier to embed the tissue with random alignment and just trim the hardened paraffin block later by means of a blade or scalpel.

10. PEN slides are special glass slides covered with a PEN (polyethylene naphthalate) membrane, which facilitates laser cutting. The membrane is cut together with the sample and acts as a stabilizing backbone during catapulting. It is important not to injure the membrane while handling since any leakage of liquid between the slide and the membrane impedes laser-cutting completely. Very recently, RNase-free PEN slides have also become available (PALM MembraneSlides NF, PALM) that can minimize the risk of leakage by omitting the procedure to remove RNases.
11. The thickness of sections depends strongly on the used tissue as well as on the size of the cells of interest. Sections should not be thinner than two cell layers of the cells of interest to make it more feasible to get undamaged cells.
12. The storage of tissue sections within paraffin wax is possible for several weeks at 4°C under desiccating conditions without loss of RNA yield and integrity.
13. Due to the fact that paraffin wax disturbs laser cutting as well as subsequent RNA isolation, it firstly has to be removed from the sections. Because sections without paraffin wax tend to get rehydrated by sheer air humidity, it is strongly recommended to remove paraffin wax just directly before starting with laser microdissection to prevent any loss of RNA yield and integrity.
14. Recently, laser capture microdissection (LCM) ***(2)*** has been established to collect tissue-specific samples from embedded plant tissue sections ***(3–5)***. In LCM, a laser beam is used to melt a thermoplastic film onto a selected area of tissue. The film is attached to the lid of a microfuge tube cap and harvested cells sticking to it can be plucked from their neighbors when the cap is removed. In contrast, the PALM MicroBeam system is based on a pulsed UV-A laser (337 nm) with the advantage of minimized heat formation during the cutting process and the contact-free collection of microdissected cells by pressure catapulting that minimizes the risk of contaminations (for details of the technology see www.palm-mikrolaser.com). Both techniques allow the collection of comparably large amounts of tissue-specific samples from material fixed in a defined physiological state.
15. β-ME is toxic and should be handled in a fume hood. Wear appropriate protective clothing.
16. RLT buffer contains guanidine thiocyanate, which can form highly reactive gaseous compounds when combined with bleach. Do not add bleach or acidic solutions directly to the RNA isolation waste. Lysis buffer is stable at room temperature up to 1 mo.
17. The most time-consuming step in LMPC is the selection of cells of interest. To save time it is therefore recommended to use all the possibilities and advantages provided by the software, such as marking cells or cell populations destined for different samples in different colors all at once and cutting and collecting them afterwards successively.
18. The most critical point during the laser microdissection process is the collection of dissected cells into LPC solution. Although RNasin is added to the LPC solution as an RNase inhibitor, the aqueous milieu can endanger RNA yield and integrity. Another problem is the relatively fast evaporation of the LPC solution associated with a loss of collection efficiency. To minimize these risks, the duration of sample collection should be restricted to 15 minutes per sampling period. Very recently, adhesive caps have become available (PALM AdhesiveCaps, PALM). These caps substitute any catapulting solution and therefore allow for prolonging the collection time far beyond 15 min, thus increasing the number of single cells that can be collected per tube.
19. RNA of high quality is the most important prerequisite for subsequent analysis, such as like RNA profiling using quantitative reverse transcription-polymerase chain reaction (RT-PCR), Northern analysis, or microarray hybridization. The isolation of high-quality RNA from microdissected tissue is especially challenging because of the naturally low sample amounts and the need for extended handling of the tissue during the embedding and also the microdissection processes. Therefore it is absolutely important to proceed as fast as possible through the RNA isolation protocol. Moreover, column-based RNA isolation methods should be used, since phenol-based extraction procedures cause relatively high absolute yield losses when starting sample amounts are limited. The RNeasy Micro technology described here is based on a silica-gel-based membrane showing selective binding properties for RNA. The membrane is located in a microspin column that provides an easy handling.

20. Lysis of tissue and the subsequent homogenization of the lysate are among the most critical steps concerning RNA yield and integrity. Plant samples are sometimes especially problematic because of the rigid cell walls and owing to cellular compounds like polysaccharides and phenolics, which may disturb proper lysis and homogenization. The RLT buffer provided by the RNeasy Micro Kit contains guanidine thiocyanate, which simultaneously inhibits ribonuclease activity and disrupts the cells effectively. In addition, use of the QIAshredder columns is absolutely essential to effectively homogenize plant tissue samples.
21. Use only this supernatant in subsequent steps, as a carryover of the pellet definitely compromises RNA yield.
22. If some lysate is lost during homogenization, reduce the volume of 70% ethanol accordingly. A precipitate may form after the addition of ethanol, but this will not affect the RNeasy procedure.
23. DNA contaminations of RNA samples may affect down-stream applications such as quantitative RT-PCR or microarray hybridizations. The selectivity of the silica-gel membrane for RNA ensures that only traces of DNA may be copurified. To remove these DNA traces, DNase I treatment directly on the membrane is strongly recommended.
24. To avoid damage to the open caps during centrifugation, place the columns into the centrifuge with at least one empty position between columns. Align the caps opposite to the rotation of the rotor (i.e., if the rotor rotates in a clockwise direction, orient the caps in a counterclockwise manner). It is important to dry the silica-gel membrane, as residual ethanol may interfere with downstream reactions. Centrifuging with open caps ensures that no ethanol is carried over during elution.
25. Smaller volumes of RNase-free water can be used to obtain a higher total RNA concentration, but this will influence the overall yield. The yield will be approx 20% less when using 10 μL RNase-free water for elution. Elution with less than 10 μL may not be enough to sufficiently hydrate the silica-gel membrane and is therefore not recommended.
26. Integrity of RNA samples is essential in the context of gene expression analysis via microarray technology, RT-PCR, Northern blotting, construction of cDNA libraries, and other applications ***(6)***. If results from different samples should be compared, it is extremely important to obtain samples with similar RNA quality. Assessing RNA quality in samples derived from microdissection poses a particular challenge because of the extremely low amounts of RNA they contain. As in *Arabidopsis* an RNA yield of 2 to 4 pg total RNA per cell can be estimated ***(5)***, standard techniques for RNA analysis such as denaturing gel electrophoresis or UV absorption are normally not applicable for microdissected samples. To overcome this, one option would be the collection of tens of thousands of cells to yield RNA amounts in the ng–μg dimension, but this is time-consuming and in many cases not practicable at all. A reproducible and easy alternative for analyzing as little as 200 pg of total RNA (equivalent to 100–200 cells) is the Lab-on-a-Chip technology using the Agilent 2100 bioanalyzer combined with the RNA 6000 Pico LabChip kit of Agilent Technologies (USA). The basis is a capillary electrophoresis technology manufactured on a small chip, which drastically reduces sample consumption (for detailed information about Lab-on-a-Chip technology see www.agilent.com/chem/labonachip).
27. No data are available addressing the mutagenictiy or toxicity of the dye/DMSO reagent. Because the dye binds to nucleic acids, it should be treated as a potential mutagen and used with appropriate care. The DMSO stock solution should be handled with particular caution, as DMSO is known to facilitate the entry of organic molecules into tissues. It is therefore strongly recommended to wear double gloves when handling the dye concentrate.
28. The RNA 6000 Pico LabChips includes a total of 16 small wells, 11 of them being sample wells, whereas one well is designated to the ladder and 4 wells for the gel dye mix and the Conditioning Solution (supplied), respectively. At the bottom side of the chip the vials are connected by microchannels where the separation by electrophoresis and detection by fluorescence takes place (for details see www.agilent.com). When pipetting the gel-dye mix make sure not to draw up particles that may sit at the bottom of the gel-dye mix vial. Insert the tip of the pipet to the bottom of the chip well when dispensing. This prevents a large air bubble forming under the gel-dye mix. Placing the pipet at the edge of the well may lead to poor results.

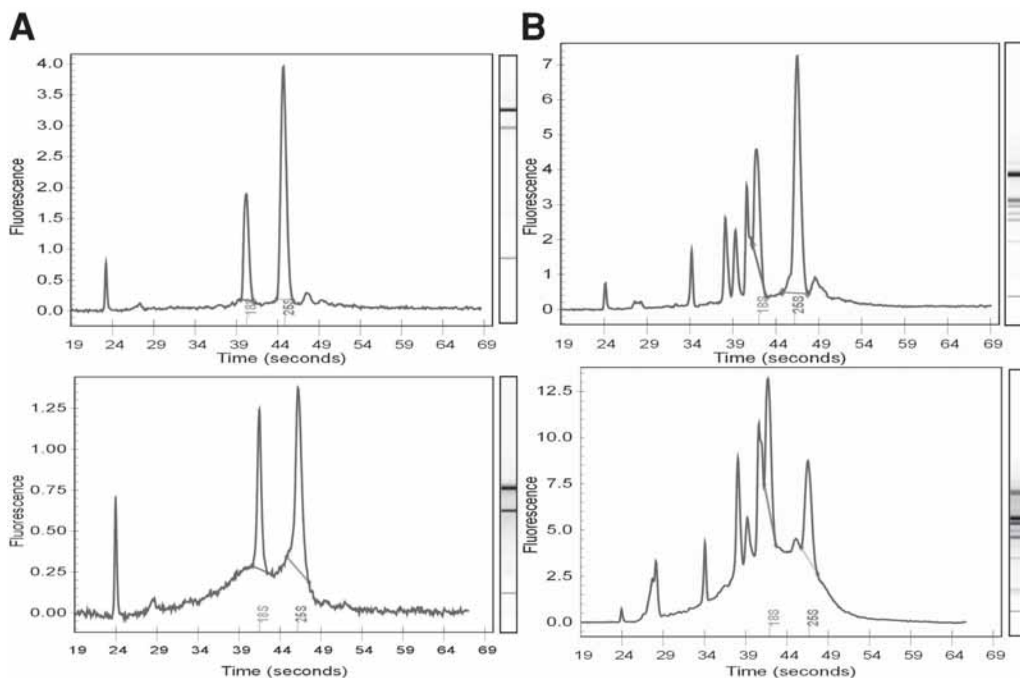


Fig. 1. RNA results derived from bioanalyzer measurements. Fluorescence spectra combined with gel-like pictures of intact RNA (upper charts) and slightly degraded RNA (lower charts) of *Arabidopsis* root **(A)** and leaf **(B)** tissues.

29. If the rare case occurs that air bubbles are visible, discard the chip and repeat the loading of the gel dye mix with a new one.
30. The chip fits only one way into the designated receptacle. Do not use any force to place the chip in its right position.
31. Do not force the lid closed or electrodes may be damaged. The bioanalyzer software screen shows that you have inserted a chip and closed the lid by displaying the chip icon at the top left of the screen.
32. The RNA results derived from bioanalyzer measurements characterize RNA for amount and integrity and are displayed as fluorescence spectra as well as gel-like pictures (**Fig. 1**). In comparison to RNA analysis via UV absorption, the quantitative results of bioanalyzer data may differ because the fluorescence bound to polynucleotides is measured, whereas UV-absorption measures every substance (short fragments and even single nucleotides) absorbing light at 260 nm. Characteristics of good RNA quality are the ratio of 25S and 18S rRNA species (in plant systems), an evenly size-distributed mRNA population, and no contaminations by genomic DNA. The 25S:18S rRNA ratio is usually used for assessing RNA quality, although the coherence between rRNA profile and mRNA integrity is still unclear ***(7)***. In general, a low baseline between the 18S and 5S rRNA should be visible in addition to a 25S:18S rRNA ratio of >1 in high-quality RNA.
33. As a basic principle, it is assumed that significance can be achieved only when RNA populations showing the same quality (or level of degradation) are compared. The quantification of RNA integrity via "Degradation Factor" using the "Pico Arabidopsis Degradometer" software can help to evaluate RNA samples of different origin. The software is a version of the Degradometer software ***(8)*** intended to analyze mammalian RNA and was kindly customized and provided by Karl Kornacker for the use with plant RNA (for detailed information and support for the Pico Arabidopsis Degradometer software please contact Karl Kornacker kkornack@columbus.rr.com, or Herbert Auer, auer-2@medctr.osu.edu at Ohio State University, Columbus, OH).

References

1. Kehr, J. (2003) Single cell technology. *Curr. Opin. Plant Biol.* **6,** 617–621.
2. Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., et al. (1996). Laser capture microdissection. *Science* **274,** 998–1001.
3. Asano, T., Masumura, T., Kusano, H., Kikuchi, S., Kurita, A., Shimada, H., and Kadowaki, K.-I. (2002) Construction of a specialized cDNA library from plant cells isolated by laser capture microdissection: toward comprehensive analysis of the genes expressed in the rice phloem. *Plant J.* **32,** 401–408.
4. Kerk, N. M., Ceserani, T., Tausta, S. L., Sussex, I. M., and Nelson, T. M. (2003) Laser capture microdissection of cells from plant tissues. *Plant Physiol.* **132,** 27–35.
5. Nakazono, M., Qiu, F., Borsuk, L. A., and Schnable, P. S. (2003) Laser capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes differentially expressed in epidermal cells or vascular tissues of maize. *Plant Cell* **15,** 583–596.
6. Xiang, C. C., Chen, M., Ma, L., Phan, Q. N., Inman, J. M., Kozhich, O. A., and Brownstein, M. J. (2003) A new strategy to amplify degraded RNA from small tissue samples for microarray studies. *Nucleic Acids Res.* **31,** e53.
7. Palmer, M. and Prediger, E. (2004) Assessing RNA quality. *Ambion TechNotes Newsletter* Vol. **11,** 13–14.
8. Auer, H., Lyianarachchi, S., Newsom, D., Klisovic, M. I., and Marcucciand Kornacker, K. (2003) Chipping away at the chip bias: RNA degradation in microarray analysis. *Nat. Genet.* **35,** 292–293.

31

Protein Isolation and Second-Dimension Electrophoretic Separation

Bernhard Schlesier and Hans-Peter Mock

Summary

Protocols for the proteome analysis of *Arabidopsis* tissues based on 2 D gel electrophoresis are given for leaves and roots as well as for seeds. The protocols contain a detailed description of the sample preparation step and of the solubilization of proteins. We then describe our protocols for the isoelectric focusing, the transfer of the strips from the first to the second dimension, SDS-PAGE as the second dimension of electrophoretic separation, staining of the gels using colloidal Coomassie blue, and finally drying of the gels between cellophane sheets. Notes contain additional background information on critical steps. Our protocols allow the analysis of leaves, roots, or seeds from different *Arabidopsis* ecotypes, from mutants or from plants subjected to various environmental conditions.

Key Words: 2D gel electrophoresis; *Arabidopsis*; colloidal Coomassie blue; proteome.

1. Introduction

With the full sequencing of its genome, large-scale transcript profiling has become very popular to study many molecular aspects of the model plant *Arabidopsis*. At the same time, the availability of vast nucleotide sequence information is also a prerequisite for the identification of unknown proteins of interest. Owing to the development of sensitive techniques for protein identification based on mass spectrometry and by using improved techniques for separation of complex protein mixtures, approaches for profiling the protein complement of organelles, cells, tissues, or whole organisms are more and more used for many species, including model and crop plants ***(1)***. Currently, the separation of complex protein extracts is still mostly performed by 2D gel electrophoresis, despite considerable limitations in separation capability. Chromatographic techniques have been tested as an alternative approach, but also provide considerable technical challenges. In comparison with nucleic acids, proteins are less uniform in their biochemical properties, e.g., ranging from hydrophilic to extremely hydrophobic, rendering extraction and separation often quite difficult by one technique. The separation capacity for proteins is hampered not only by the limited resolution of 2D gels, but also by the restricted dynamic range of staining methods for detection of proteins. Despite these considerable difficulties at present, it is a necessity to include approaches on the protein complement in many targeted investigations or global functional genomic studies, as many examples exist in which the levels of transcripts and proteins do not correspond very well (e.g., ***2,3***). Moreover, posttranslational modifications multiply the complexity of the protein complement in a cell and are important determinants in many cellular responses and regulatory mechanisms ***(4)***.

In this chapter, we will focus on the separation of protein extracts from *Arabidopsis* tissues based on 2D gel electrophoresis as part of proteome approaches. We will not further mention separation by 1D gel electrophoresis, as this technique is a standard procedure widely established in many laboratories. Instead, we will give detailed protocols for protein extraction and outline staining techniques for 2D gels.

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

2. Materials

1. Refrigerated microliter centrifuge (30,000*g*, e.g., MICRO 22R, Hettich).
2. Vacuum centrifuge.
3. Ultrasonic bath.
4. Thermomixer for microcentrifuge tubes.
5. Centrifugal filter device (e.g., Amicon Ultrafree-MC, 0.45 µm, Durapore membrane; maximum 12,000*g*).
6. Equipment for isoelectric focusing on immobilized pH gradients.
7. Vertical slab gel electrophoresis.
8. Gel drying frames and cellophane (e.g., Invitrogen).
9. Precipitation solution (to be made fresh): 10% (w/v) trichloroacetic acid (TCA), 0.07% (w/v) 2-mercaptoethanol in acetone.
10. Washing solution (to be made fresh): 0.07% (w/v) 2-mercaptoethanol in acetone.
11. Rehydration stock solution: 8 *M* urea, 2% CHAPS, 0.005% bromophenol blue in water. Store in aliquots at –20°C.
12. Rehydration solution according to Drs. Job's laboratory ***(7)***: 7 *M* urea, 2 *M* thiourea, 4% (w/v) CHAPS, 1% (v/v) Pharmalyte (Amersham Bioscience, pH range corresponding with the IPG strip), 18 m*M* Tris-HCl, 14 m*M* Tris base, protease inhibitor cocktail "complete mini" (Roche Diagnostics), 53 U/mL DNase I, 4.9 Kunitz U/mL RNase A, 0.2% (v/v) Triton X-100.
13. 2D Quant Kit (Amersham Biosciences).
14. IPG-buffer (Amersham Biosciences) or other highly purified carrier ampholytes.
15. IEF strips with immobilized pH gradients.
16. Mineral oil (e.g., IPG cover fluid, Amersham Biosciences).
17. Equilibration buffer (stock): 50 m*M* Tris-HCl, pH 8.8, 6 *M* urea, 30% (v/v) glycerol, 2% (w/v) sodium dedecyl sulfate (SDS), 0.005% bromophenol blue in water. Store in aliquots at –20°C).
18. DTT stock: 1 *M* DTT in water. Store in aliquots at –20°C.
19. Iodacetamide.
20. Separation gel buffer: 0.375 M Tris-HCl, pH 8.8, 0.8% (w/v) SDS in water.
21. Stacking gel buffer: 0.5 *M* Tris-HCl, pH 6.8, 0.4% (w/v) SDS in water.
22. Running buffer: 23 m*M* Tris 192 m*M* glycine, pH 8.3, 0.1% (w/v) SDS in water.
23. Acrylamide solution: 30% (w/v) acrylamide, 0.8% (w/v) N,N'-methylenebisacrylamide in water (store at 5°C).
24. Persulfate solution: 10% (w/v) ammonium persulfate in water. Prepare weekly; store at 5°C.
25. Colloidal Coomassie G250 (e.g., Gel Code Blue Stain Reagent, Pierce).
26. Drying solution: 20% (v/v) ethanol, 20% (w/v) glycerol in water.

3. Methods

The proteome of an organism or a cell is highly dynamic, reflecting the influence of environmental stimuli and developmental programs. To mininize the biological variation in a given set of experiments, it is therefore necessary to thoroughly control the culture conditions of the plants. We have introduced a hydroponic culture system allowing us to grow *Arabidopsis* under controlled environmental conditions, which also enables us to harvest root material easily ***(5)***. A number of similar systems are available (*see* references in **5**); when using plants grown in soil, it is also necessary to tightly control the environmental conditions, such as light and temperature. We will describe a protocol successfully used for *Arabidopsis* leaf tissue ***(6)***. As an example, 2D gel protein patterns from leaves of Arabidopsis ecotypes Col-0 and Ws-2 are shown in **Fig. 1**. This protocol has also been used for root tissue ***(5,6)*** as shown in **Fig. 2**. For seed tissue, a specific protocol has been published ***(7)*** which we have adopted for our laboratory (**Fig. 3**). The reader is also referred to recent protocol books on proteome methodology (e.g., ***8–10***).

3.1. Protein Preparation

Convenient protein isolation and sample preparation are indispensable preconditions for good 2D results. To prevent proteolysis, the freshly harvested sample is frozen in liquid nitrogen and processed immediately or stored at –80°C. Before protein extraction material is dry-

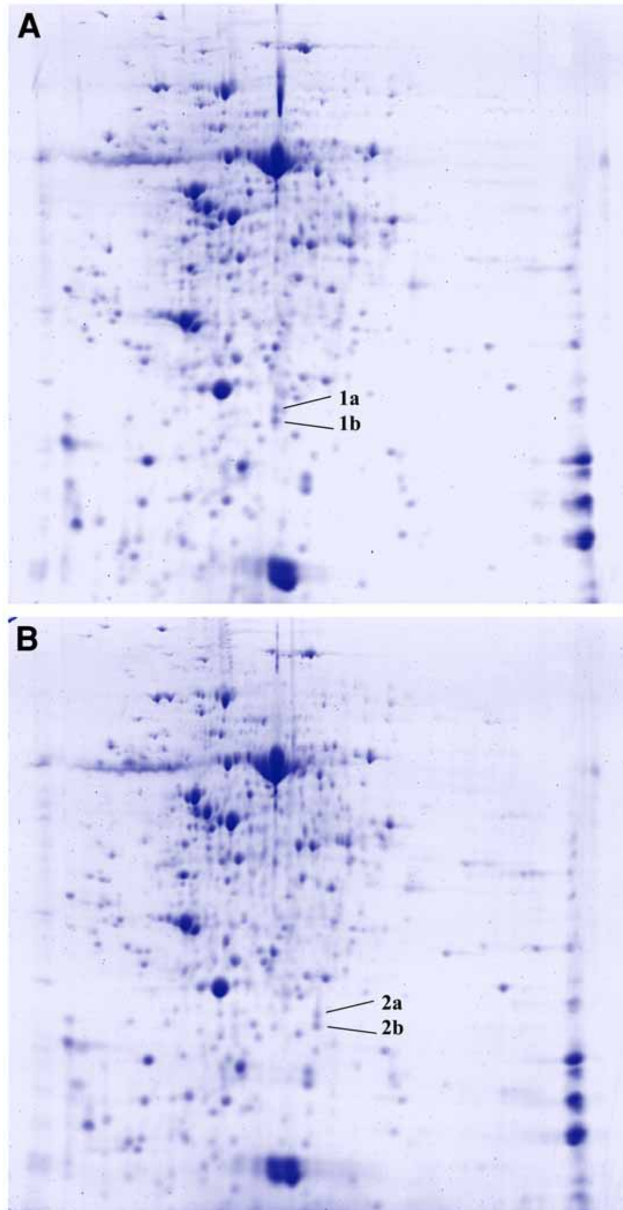


Fig. 1. 2D electrophoresis of leaf proteins from *A. thaliana* ecotypes Col-0 and Ws-2. After separation of 100 µg proteins from Col-0 (a) and Ws-2 (b) on 13-cm IPG strips, (pH 3.0–10.0) gels were stained by colloidal Coomassie G250. Two highly homologous germin-like proteins that differ only in one amino acid residue are identified by peptide mass fingerprints and post source decay spectra. The ecotype Col-0 expresses the sequence gi|15242028 (indicated by 1a,b), the ecotype WS2-1 the sequence gi| 1755154 (indicated by 2a,b).

crushed in liquid nitrogen. Proteins in the ground sample are precipitated by TCA-acetone ***(11)***. This step inactivates proteases very effectively.

Isoelectric focusing, the first step in 2D separation of proteins, is very sensitive to sample contamination by salts and other small ionic molecules, ionic detergents, nucleic acids, polysaccharides, lipids, phenolic compounds, and insoluble material. Separation of proteins from such

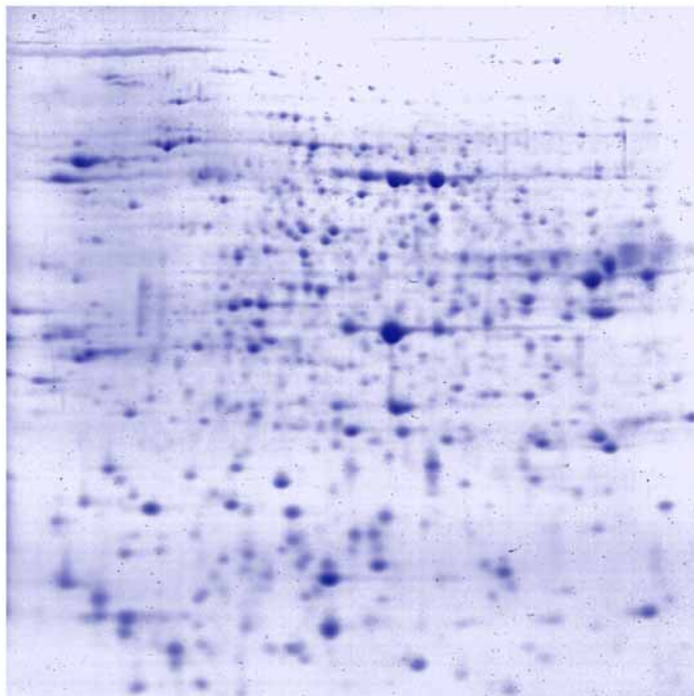


Fig. 2. 2D electrophoresis of root proteins from *A. thaliana* ecotype Col-0. After separation of 100 µg of protein on 13-cm IPG strips (pH 3.0–10.0) and SDS-PAGE, gels were stained by colloidal Coomassie G250.

contaminants is an absolute necessity. However, one must be aware that each purification step can alter the protein pattern of a sample. In order to achieve a high resolution isoelectric separation, all proteins should be completely disaggregated and stably solubilized. Any artifactual modification of polypeptide chains must be prevented. Therefore the lysis buffer always includes high concentrations of urea (and thiourea), at least one nonionic or zwitterionic detergent, and a reductant. Carrier ampholytes in the sample solution enhance sample solubility.

3.1.1. Precipitation of Proteins

TCA in acetone precipitates and denaturates instantaneously all proteins (*see* **Note 1**). Subsequent resolubilization of the protein precipitate in urea-containing buffer does not seem to yield to any reactivation of proteases.

3.1.1.1. Leaf and Root

1. Freeze tissue in liquid nitrogen.
2. Grind deep-frozen plant material under liquid nitrogen into a fine powder using a mortar with a pestle. Transfer 2.0 g of this sample into a 50-mL Falcon tube.
3. Immediately add 20 mL of precipitation solution and vortex for 10 s.
4. Transfer 1.7 mL aliquots of the well-mixed suspension into 2-mL microcentrifuge tubes. Use a 1-mL pipet tip and cut it to have a larger opening.
5. For quick and uniform cooling down, insert the tubes with the suspension in liquid nitrogen for 30 s.
6. Transfer the samples to –20°C for at least 45 min but not longer than 2 h. Mix the suspension by hand after 5, 10 and 15 min for better exchange of precipitation solution in the sediment.
7. Centrifuge the suspension (20,000*g* or higher, 4°C, 15 min). Carefully remove and discard the supernatant immediately after the centrifugation.

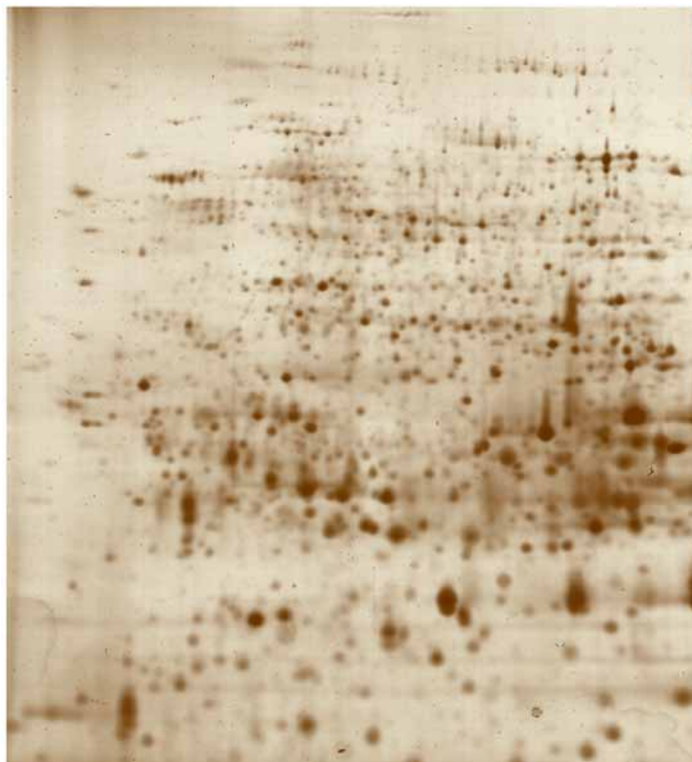


Fig 3. 2D electrophoresis of proteins from *A. thaliana* seeds. After separation of 200 µg of protein on 13-cm IPG strips (pH 4.0–7.0), gels were silver-stained ***(13)***.

8. Wash the sediment with 1.5 mL of washing solution. Resuspend the precipitate by vortexing, using a sonication bath or a glass rod. The homogenous distribution of the pellet is crucial for removal of TCA. Freeze in liquid nitrogen for 30 s and keep for 30 min at –20°C.
9. Repeat **steps 7** and **8**.
10. Carefully discard the supernatant and remove the residual acetone from the pellet using a vacuum centrifuge.
11. Store the samples at –20°C or –80°C.

3.1.1.2. Mature Seeds

In a first step most of the seed coat material is removed, as well as lipids and other low molecular-weight substances, which especially interfere with isoelectric focusing (*see* **Note 1**).

1. Transfer 100 mg of seeds into a microcentrifuge tube and freeze it in liquid nitrogen.
2. Cool down a mortar and a pestle using liquid nitrogen. Immediately after evaporation of the liquid nitrogen, carry over the seeds without liquid nitrogen and quickly grind deep-frozen seeds into a fine powder (if liquid nitrogen is added you will lose many seeds).
3. Transfer the powder with 1.5 mL acetone in a microcentrifuge tube and place it for 15 min in an ultrasonic bath.
4. After a short time most of the dark-brown seed coat material is sedimented. Carry over the yellowish supernatant in a new microcentrifuge tube and collect the fine-grained light-colored material by short centrifugation. Discard the supernatant.
5. Wash the crude sediment from previous step by vortexing with 1.5 mL acetone. Separate the crude and fine material as before and add the light material from the supernatant to the corresponding fraction from **step 4**.
6. Repeat **step 5**.

7. Remove acetone by vacuum centrifugation.
8. Store the samples at –20°C or –80°C.

3.1.2. Solubilization of Proteins for 2D Electrophoresis

In this step, proteins are dissolved in a buffer for isoelectric focusing and thereby separated from cell wall components and other substances, which are coprecipitated by either TCA-acetone or acetone (in the case of seeds) together with the proteins (*see* **Note 2**).

3.1.2.1. Leaf and Root

1. Prepare the rehydration solution just prior to use from rehydration stock solution by adding 0.5 to 2.0% of IPG buffer (use the correct pH range according to the IPG strip) and 20 to 100 m*M* DTT.
2. Add rehydration solution to TCA-acetone precipitated sample (50 µL/mg). Disperse sample by vortexing, collect material from wall by short centrifugation, and continue resuspension for 5 min in an ultrasonic bath at room temperature (RT).
3. Resolve the proteins by incubation at 37°C with continuous shaking for 1 h.
4. Centrifuge the suspension (30,000*g* RT, 15 min) and retain both the supernatant and the pellet. If the supernatant is not clear, transfer it to a new reaction vessel and repeat centrifugation or centrifuge the supernatant through a 45-µm centrifugal filter device (*see* **Note 3**).
5. Use the supernatant or filtrate immediately for isoelectric focusing or store it at –20°C or –80°C.

3.1.2.2. Mature Seeds

1. Just prior to use, prepare the rehydration solution ***(7)***.
2. Add this rehydration solution to acetone-treated seed flour (25 µL/mg). Disperse sample by vortexing, collect material from the wall by short centrifugation, and continue resuspension for 5 min in an ultrasonic bath at RT.
3. Resolve the proteins by incubation at RT with continuous shaking for 30 min. After 10 min, add 14 m*M* DTT.
4. Centrifuge the suspension (30,000*g* RT, 15 min) and retain both the supernatant and the pellet. If the supernatant is not clear, transfer it to a new reaction vessel and repeat centrifugation or centrifuge the supernatant through a 45-µm centrifugal filter device (*see* **Note 3**)
5. Use the supernatant or filtrate immediately for isoelectric focusing or store it at –20°C or –80°C.

3.1.3. First-Dimension Protein Separation: Isoelectric Focusing

Isoelectric focusing (IEF) separates the amphoteric proteins according to their isoelectric point. Driven by an electric field, each individual protein will move in a pH gradient to that position where its net charge is zero.

This step of the 2D procedure involves the quantification of protein in the sample (*see* **Note 4**), the rehydration of an appropriate immobilized pH gradient, rehydration of the IPG strip, application of the sample, and the separation of proteins according to their isoelectric points using optimized current-voltage protocols. To optimize the separation protocol for new applications we frequently use IPG strips with a length of 7 cm for the first dimension and minigels for the second dimension, which saves time and is less expensive. Upscaling to larger gel formats in the following is rather easy.

Here we refer to immobilized pH gradients using the IPGphor system (Amersham). Other equipment (e.g., from Bio Rad) will give similar results. Detailed instructions are supplied from the manufacturers.

1. Quantify the protein content directly in the IEF sample using the 2D Quant Kit. On 7- and 13-cm IPG strips we routinely separate about 50 or 200 µg protein, respectively, and visualize by staining with colloidal Coomassie G250.
2. Apply sample by rehydration (*see* **Note 5**): Mix an aliquot of sample with rehydration solution to get the correct volume for one strip (125 or 250 µL for 7- and 13-cm strips, respectively). Deliver the correct sample volume into the strip holder in the region between the lateral wells.

Table 1
Separation Conditions for IPG Strips

Length	Voltage (V)	Time (h:min)	
7 cm	0	14:00	rehydration
	500	1:00	gradient
	3000	0:30	gradient
	3000	4:40	
13 cm	0	14:00	rehydration
	500	2:00	gradient
	4000	1:00	gradient
	4000	5:30	

3. Remove the protective foil from the IPG strip. Pay attention to the direction of the pH gradient and insert the strip with the gel side down into the strip holder beginning at the anodic end. Assure complete wetting of the gel by lifting, lowering, and sliding the strip. Do not trap air bubbles.
4. To prevent evaporation, pipet IPG cover fluid (mineral oil) dropwise into both ends of the strip holder until the IPG strip is completely covered.
5. Place the cover on the strip holder, transfer it on the IPGphor unit, close the safety lid, and start the program (*see* **Note 6**). The programs for 7- and 13-cm IPG strips at the pH ranges 3.0 to 10.0 and 4.0 to 7.0 (**Table 1**) can be used as a starting point to optimize the separation conditions for samples prepared according to the above-described method.
6. After the end of IEF transfer the IPG strips in screw-cap tubes (support film toward the wall) and continue with second-dimension immediately or store the strips at –70°C.

3.1.4. Second-Dimension Protein Separation: SDS-PAGE

Proteins and the anionic detergent SDS form complexes having a roughly constant negative charge per mass unit. For this reason polypeptides are separated in SDS-polyacrylamide gel electrophoresis (PAGE) mainly according to their molecular weights.

The method of SDS-PAGE analysis described here based on the procedure described by Laemmli (*12*). Contrary to most other laboratories performing 2D we prefer a system with stacking gel (*see* **Note 7**).

3.1.4.1. Casting Slab Gels

Different types of vertical and horizontal electrophoresis equipments can be used. Details for the handling can be found in the manuals. Here we provide our protocol for vertical devices.

1. Assemble the gel sandwich made from two glass plates and 1.0-mm spacers and transfer it to the casting stand. Place the casting stand on a horizontally adjusted surface.
2. Prepare the separation gel solution (**Table 2**) and pour it in the gel cuvet up to 1.5 cm below the top of the glass plate.
3. Overlay the gel solution with water-saturated n-butanol (at least 5 mm in height) to create a flat surface and to protect the gel from oxygen. After about 15 to 30 min a sharp refractive line near the top of the gel indicates that polymerization has occurred.
4. Remove the overlay, rinse two times with water, and carefully remove the residual water.
5. Prepare the stacking gel solution (**Table 2**) and pour a 1-cm layer on the separation gel.
6. Overlay the stacking gel in the same way as the separation gel.
7. After polymerization remove the overlay and rinse the surface two times with running buffer. Leave the buffer over the gel. This assists the bubble-free insertion of IPG strips. Now the gel should be used without much time delay, as the diffusion of buffers for separation, stacking, and running reduces the concentration effect of the discontinuous buffer system. Therefore at this time the equilibration of the IPG strips should be completed.

Table 2
Homogeneous SDS Gels

Separation gel solution Acrylamide (%)		10.00	11.25	12.50
Separation gel buffer	(mL)	1.5	1.5	1.5
Acrylamide (30/0.8)	(mL)	4.0	4.5	5.0
18 MΩ water	(mL)	6.5	6.0	5.5
Ammonium persulfate	(μL)	120	120	120
TEMED	(μL)	6	6	6
Stacking gel solution Acrylamide (%)		6.0		
Stacking gel buffer	(mL)	1.0		
Acrylamide (30/0.8)	(mL)	0.8		
18 MΩ water	(mL)	2.2		
Ammonium persulfate	(μL)	25		
TEMED	(μL)	3		

Table 3
Separation Conditions for Second-Dimension Gels

Gel size (width × length)	stacking gel	separation gel		
Mini gel 8 cm × 7 cm	0:30	75 V	1:00	150 V
Medium gel 14 cm × 14 cm	0:30	75 V	4:30	150 V

3.1.4.2. Equilibration of IPG Strips and Running the Second Dimension

Prior to SDS electrophoresis the proteins/polypeptides in the IPG strip must be saturated with SDS (*see* **Note 8**).

1. Prepare equilibration buffer 1 by adding 650 μL 1 *M* DTT to 10 mL equilibration stock buffer.
2. Use 2.5 mL (5 mL) for each 7 cm (13 cm) strip. Equilibrate for 15 min on a rocking shaker.
3. Only for silver staining prepare equilibration buffer 2 by dissolving 250 mg iodacetamide in 10 mL equilibration stock buffer.
4. Use the same conditions as under **step 2**.
5. Transfer the strip on top of the SDS gel and start the separation using constant voltage (**Table 3**) at room temperature (*see* **Note 9**).

3.1.4.3. Protein detection

All methods for visualization of proteins on SDS gels can be used for 2D gels. The most popular method is staining with Coomassie blue. We routinely use a colloidal preparation of G250 (GelCode Blue Stain Reagent, Pierce; *see* **Note 10**).

1. After electrophoresis wash gels two times in deionized water with gentle shaking for 15 min.
2. Wash with 5% phosphoric acid with gentle shaking for 15 min.
3. Stain 1 h or longer with colloidal Coomassie G250. Use 25 mL per mini gel or 100 mL per medium-sized gel.
4. Wash in water for 5 min.
5. Replace water and store gels at 5°C.

3.1.4.4. Gel Drying

Gels dried between cellophane sheets are ideal for permanent storage.

1. Incubate the gel in gel drying solution on a shaker for 10 min. Use 30 or 100 mL for a mini gel or medium gel, respectively.

2. Soak the cellophane sheets for 5 min in gel drying solution.
3. Place one cellophane sheet on the base plate of the drying frame and add some gel drying solution.
4. Put the gel on this cellophane sheet and add some gel drying solution on the gel.
5. Carefully lay the second sheet of cellophane over the gel so that no bubbles are trapped anywhere between the cellophane and the gel.
6. Align the frame and push the clamps at top, left, and right sides.
7. Stand the gel dryer upright for 2 min for draining off the excess solution. Afterwards, push the bottom clamp. Avoid placing the gel dryer in a fume hood or near to a heat source, as accelerated drying can result in gel cracking.
8. After 12 to 36 h the gel is dried. Remove it from the frame and press it for several days between the pages of a catalog.

4. Notes

Always use gloves to protect yourself from hazardous chemicals. This also helps to diminish contamination of the experiment by keratins, which can be a serious problem during spot identification by mass spectrometry.

1. Other possibilities to inhibit proteases are protein extraction with basic buffers (Tris base, carbonate, or basic carrier ampholytes) or addition of protease inhibitors to the extraction buffer. We prefer the TCA-acetone precipitation because in the same step many low-molecular weight-substances that interfere especially with isoelectric focusing are removed. Extraction of proteins from seeds without previous treatment with acetone results in an extensive lipid layer above the protein solution after centrifugation.
2. For first experiments use 8 *M* urea, 2% CHAPS, 0.5% IPG buffer, and 20 m*M* DTT, which has been found to be effective for solubilizing a wide range of samples. For proteins hard to be solubilized include higher amounts of urea (up to 9.8 *M*), add 2 *M* thiourea (for solubility thiourea requires also the presence of at least 5 *M* urea), use more reductant, more ampholytes, and higher concentrations of one or more detergents. For crude fractionation of proteins the pellet after the first solubilization can be serially extracted with other buffers.
3. Because of the high density of most buffers used in protein solubilization for IEF, the undissolved material forms only a soft pellet. Carefully remove the supernatant without disturbing the pellet.
4. Separation of a defined amount of protein is a prerequisite for quantitative work. Most protein determination methods fail if buffers contain high concentrations of urea, detergents, and reducing agents like rehydration buffers for IEF. A procedure that uses only the quantitatively precipitated proteins as the sample while leaving contaminants in the discarded solution circumvents these limitations (e.g., 2-D Quant Kit, Amersham Bioscience).
5. Sample application during rehydration is the most popular method for protein loading. This technique allows the loading of high protein amounts. Rehydration loading is performed either without an electric field or under low voltage (30–120 V). Damp electrode pads can be applied between electrodes and IPG strips to absorb excess water and improve the results. Samples can also be applied after strip rehydration either by special sample caps, by pipetting the sample into the lateral wells of the strip holder, or by paper-bridge loading. This technique allows sample application near one of the electrodes. In some cases this results in improved separations.
6. Because of low conductivity in both pH gradient and correctly prepared samples, isoelectric focusing runs at low currents (at maximum 50 μA per IPG strip) and high voltages. During IEF all charged compounds migrate to their isoelectric point. This results in decreasing current over time and allows the application of increasing voltages. Generally, the program is started at low voltage. Voltage is than gradually increased to the final value, which is held up to the end of the run. A program for complete focusing of a specific sample must consider sample composition and application, IPG strip length, and pH gradient. Recording of the actual current and voltage values over time is very helpful during method optimization and for identification of problems. We use voltage gradients in the initially steps. The final voltage in each step is limited to a value about 5 to 10% lower as the maximal value observed in control experiments. By this means all programmed voltages are arrived at the predetermined time.

7. Commonly for a first orientation a homogeneous gel with 12.5% acrylamide is used. Single concentration gels are easier to prepare than gradient gels and offer good resolution for a particular molecular weight window. We routinely prefer 11.25% acrylamide in separation gel. This is a compromise between the separation ranges of 10% (14,000 to 200,000) and 12.5% (14,000 to 100,000). In our hands the complete SDS system according to Laemmli results in better separations as the system without stacking gel. The IPG strip is applied to the stacking gel without additional agarose. We have never observed moving or floating of any strip in the running buffer. Separation gels can be stored for a number of days if their surface is rinsed and overlaid with separation gel buffer at the same concentration as in the gel. The gels must be protected against drying.
8. During equilibration the proteins are saturated with SDS. This reaction needs a basic pH value that is compatible with the electrophoretic system. Urea and glycerol reduce electroendosmosis induced by the presence of fixed charges on the pH gradient and improve the protein transfer to the second dimension. DTT preserves the reduced state of cysteine residues in the proteins. To prevent artifacts in silver staining, the DTT is removed by a second equilibration step with iodacetamide. Bromophenol blue moves near to the ionic front and allows monitoring of separation.

   The equilibration should be performed under strongly controlled conditions. SDS must enter the gel strip to form complexes with proteins. On the other hand prolonged incubation with the SDS containing buffer easily elutes proteins from IPG strips and gives rise to loss of proteins.
9. Constant voltage prevents overheating of the gel because the current decreases in the course of separation in a discontinuous buffer system. Run the gels at room temperature to prevent SDS precipitation in the cold.
10. Stain intensity reaches a maximum within approximately 1 hour. Gels may be stained overnight without increasing background. The final washing in water enhances stain sensitivity. Too many changes of water during storage can decrease staining because the equilibrium is disturbed between protein bound stain and the small amount of stain in solution. The additional washing step with phosphoric acid allows to reuse the staining solution once.

Acknowledgments

The skilful technical assistance of Annegret Wolf and Petra Linow is gratefully acknowledged. We also thank the DFG for financial support (Mo 479/4-1 and Mo 479/4-2). We are grateful to Claudette and Dominique Job for introducing B.S. to their protocol for seed analysis.

References

1. Cánovas, F. M., Dumas-Gaudot, E., Recorbet, G., Jorrin, J., Mock, H.-P., and Rossignol, M. (2004) Plant proteome analysis. *Proteomics* **4,** 285–298.
2. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19,** 1720–1730.
3. Ideker, T., Thorsson, V., Ranish, J. A., et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292,** 929–934.
4. Mann, M. and Jensen, O. N. (2003) Proteomic analysis of post-translational modifications. *Science* **21,** 255–261.
5. Schlesier, B., Bréton, F., and Mock, H.-P. (2003) A hydroponic culture system for growth of *Arabidopsis thaliana* plantlets under sterile conditions. *Plant Mol. Biol. Rep*. **21,** 1–8.
6. Schlesier, B. and Mock, H.-P. (2004) Proteome analysis differentiates between two highly homologues germin-like proteins in *Arabidopsis thaliana* ecotypes Col-0 and Ws-2. *Phytochemistry* **65,** 1565–1574.
7. Gallardo, K., Job, C., Groot, S. P. C., Puype, M., Demol, H., Vandekerckhove, J., and Job, D. (2002) Proteomics of *Arabidopsis* seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiol.* **129,** 823–837.
8. Link, A. J., (ed.) (1999) *2-D protome analysis protocols*. Humana Press, Totowa, NJ.
9. Rabilloud, T., (ed.) (2000) *Protome Research: Two-Dimensional Gel Electrophoresis and Identification Methods*. Springer Verlag, Heidelberg, Berlin, New York.
10. Smith, R. D., (ed.) (2003) *Proteome Characterization and Proteomics.* Academic Press, San Diego, CA.

11. Damerval, C., de Vienne, D., Zivy, M., and Thiellement, H. (1986) Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. *Electrophoresis* **7,** 52–54.
12. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* ***227,*** 680–685.
13. Blum, H., Beier, H., and Gross, H. J. (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8,** 93–99.

32

Isolation of Nuclear Proteins

Tomasz T. Calikowski and Iris Meier

Summary

Proteome analysis is becoming a powerful tool of discovery-driven research, with investigations ranging from whole organisms to specific subcellular compartments. Especially for the latter, efficient and robust methods for protein purification are the prerequisite for obtaining meaningful proteomic data. The plant nucleus is the repository of critical components of the genetic and biochemical machinery and therefore of great interest as source material for proteomic studies. Although *Arabidopsis thaliana* is the prime model system for plant molecular genetics, proven protocols for biochemical fractionation are often best established for other plant species. In this chapter, we present methods for the isolation of nuclei, nuclear proteins, and nuclear protein fractions that have been adapted for *Arabidopsis*. Two different protocols for the isolation of nuclei and nuclear extracts from *Arabidopsis* plants and suspension cells are described. Discovering that developmental mutant phenotypes are based on chromatin-associated and chromatin-altering proteins has spiked a growing interest in *Arabidopsis* chromatin proteins. We describe a method for obtaining an *Arabidopsis* protein chromatin fraction and the entire histone complement.

Key Words: Nuclear proteins; isolation protocol; histone; chromatin.

1. Introduction

As we enter the postgenomic era of proteome analysis, it becomes increasingly important to utilize efficient and robust protocols for the isolation and purification of subfractions of cellular proteins. The nucleus is a crucial organelle of eukaryotic cells and the repository of critical components of the genetic and biochemical machinery. It is therefore of great interest as a source material for proteomic studies. *Arabidopsis thaliana* has an excellent record as a model plant organism *(1,2)*. Because of its fully sequenced genome, it is currently the plant of choice for mass-spectrometry-based protein identification. However, many commonly used protocols for the isolation of nuclei and nuclear proteins have been established for other plant systems, including soybean embryos, tobacco leaves, and tomato fruit *(3–6)*, and these may not be best for *Arabidopsis*. The following procedures have been adapted for *Arabidopsis*, taking into account the most common problems encountered with efficient nucleus isolation. Those result from an increased production of proteolytic enzymes, a high polycarbohydrate content, and the small size of the nucleus compared with tobacco, maize, or wheat (**Subheadings 3.1.** and **3.2.**).

Isolated nuclei serve as a source material for the resolution of nuclear polypeptides (**Subheading 3.3.**). This in turn allows the study of their expression patterns, modifications, and dynamic physiological responses using two-dimensional gel electrophoresis and mass-spectrometric evaluation. Complex patterns of the total nuclear proteins can then be studied, as well as more simplified subfractions of nuclear proteins (e.g., the nucleolar component), as separated by the variety of biochemical, electrophoretic, or chromatographic methods. Examples of this approach include nuclear matrix studies *(7,8)* and the analysis of changes of the nuclear

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

proteome under stress conditions ***(9)***. Because the study of plant chromatin is gaining importance, **Subheading 3.4.** describes a protocol for the isolation of this nuclear DNA-bound protein fraction from *Arabidopsis* plants. The protocol is based on the method described for wheat seedlings ***(3)*** and circumvents the need for separate nucleus isolation prior to the release of the nuclear proteins bound to DNA ***(4)***. The chromatin fraction can be processed to yield core and linker histones ***(10)***, which may in turn provide a convenient marker for the quality and purity of the nuclear proteins (e.g., by immunoblotting) or for the investigation of "histone code-type" modifications and their epigenetic effects ***(11,12)***. **Subheading 3.5.** provides a brief discussion of such a procedure from *Arabidopsis* plants grown axenically in liquid medium ***(13,14)***.

2. Materials

2.1. Nuclei Isolation

All solutions need to be freshly prepared, either directly or as stock solutions (5X concentrated), without protease inhibitors and polyamines (which should be added fresh before use, after the dilution of the original buffer). The same procedure can be used to obtain intact nuclei of good quality from *Arabidopsis thaliana* suspension cultured cells (*see* **Subheading 2.2.** on conditions for growth and harvesting of cells and the preparation of protoplasts). Thiodiglycol can be difficult to obtain in some countries, and can be substituted by 20 m*M* β-mercaptoethanol.

1. Nuclei isolation buffer 1 (NIB 1): 20 m*M* KCl, 20 m*M* HEPES, pH 7.4, 0.6% (v/v) Triton X-100, 13.8% (v/v) hexylene glycol, 1% (v/v) thiodiglycol, 50 μ*M* spermine, 125 μ*M* spermidine, 1 m*M* phenylmethylsulfonyl fluoride (PMSF), 2 μg/mL aprotinin.
2. Nuclei isolation buffer 2 (NIB 2): 20 m*M* KCl, 20 m*M* HEPES, pH 7.4, 13.8% (v/v) hexylene glycol,1% (v/v) thiodiglycol, 50 μ*M* spermine, 125 μ*M* spermidine, 1 m*M* PMSF, 2 μg/mL aprotinin.
3. Nuclei isolation buffer 2/50% glycerol: NIB 2 plus 50% (v/v) glycerol.
4. Percoll gradient solutions: For 80% (v/v) Percoll (Amersham Biosciences), mix 8 parts of 100% Percoll with 2 parts of NIB 1; for 60% (v/v), 6 parts of Percoll with 4 parts of NIB 1; for 30% (v/v), 3 parts of Percoll with 7 parts of NIB 1. Keep all Percoll solutions sterile by filtering them through a 0.22-μm bottle-top filter unit attached to a vacuum line (Nalgene cat. no. Z370541-12EA) placed under the laminar flow. In addition, keep all Percoll-containing solutions refrigerated when not in use, to prevent contamination.
5. Urea/NaCl solution: 5.5 *M* urea, 2.2 *M* NaCl.
6. Micrococcal nuclease (MNase, Fermentas International, Burlington, Ontario, Canada, cat. no. EN0181).
7. MNase digestion buffer: 1 *M* sorbitol, 0.1 *M* $CaCl_2$, 10 m*M* PIPES, 0.1 m*M* PMSF.
8. MNase stop buffer: 20 m*M* Tris-HCl, pH 7.6, 15 m*M* EDTA, 15 m*M* EGTA, 150 m*M* NaCl, 0.3% (v/v) SDS.

Other equipment, materials, and solutions required: 500-mL Erlenmeyer flasks, liquid Gamborg's B5 medium, 70% (v/v) ethanol, 50% bleach dilution (10% final concentration of sodium hypochlorite), steel coffee blender, Miracloth (Calbiochem) filtration material, absorbent paper tissues (e.g., Kleenex), spectrophotometer, liquid nitrogen, 50-mL polypropylene round-bottom Oakridge tubes (Nalgene, cat. no. C239-4), rotary shaker.

2.2. Protoplast Preparation from Arabidopsis *Suspension-Cultured Cells*

Grow *Arabidopsis* suspension-cultured cells in 50 mL of Gamborg's B5 medium (Serva, Germany, cat. no. 47306.04) ***(15)*** supplemented with 3% (w/v) sucrose, 1.1 mg/mL 2,4-D, and 0.5 g/L MES at 22°C under continuous fluorescent white light (60 μmol/m^2/s). The subculturing is performed every 7 d with a 10-fold dilution of the medium, with cell harvesting after 4 to 5 d. In order to obtain nuclei, protoplast isolation is performed first, as described below ***(16)***. For other methods for protoplast preparation, *see* also Chapter 20 in this book.

1. Enzyme solution: Cellulase "Onozuka" R-10 –1% (w/v), (Yakult Honsha, Tokyo, Japan), Macerozyme R-10–0.1% (w/v), (Yakult Honsha), 0.4 *M* mannitol, 20 m*M* 2-(N-morpholino)ethanesulfonic acid (MES), pH 5.5.

2. To prepare the enzyme solution, stir for at least 1 h at 4°C and filter-sterilize using the 0.22-μm bottle-top filter unit attached to a vacuum line (Nalgene, cat. no. Z370541-12EA) placed under the laminar flow. The solution works best when fresh, but can be aliquoted in 50-mL tubes and stored at –20°C. Repeated freezing or thawing should be avoided.

Other solutions/materials required: 5-d-old culture of suspension cells (100 mL), sterile 50- mL centrifuge tubes, centrifuge with swinging bucket rotor (room temperature [RT] and 4°C), sterile 250-mL flask and platform shaker (150–200 rpm), hemacytometer and microscope, 0.4 *M* mannitol/20 m*M* MES, pH 5.5 (RT), 0.4 *M* mannitol (ice-cold).

2.3. Alternative, Short Protocol for Isolation of Plant Nuclei

1. Homogenization buffer (HB): 25 m*M* PIPES, pH 7.0, 10 m*M* NaCl, 5 m*M* EDTA, pH 8.0, 250 m*M* sucrose, 0.15 m*M* spermine, 0.5 m*M* spermidine, 1% (v/v) thiodiglycol, 0.1% Nonidet P-40, 0.2 m*M* PMSF.
   Spermidine, spermine, Nonidet P-40, and PMSF must be added to the solution immediately before use. Thiodiglycol can be difficult to obtain in some countries and might be substituted for by 20 m*M* β-mercaptoethanol (added immediately before use).
2. Nuclei resuspension buffer (NRB): 50 m*M* HEPES, pH 7.6, 110 m*M* KCl, 5 m*M* $MgCl_2$, 50% (v/v) glycerol, 1 m*M* DTT, 0.5 μg/mL leupeptin, 50 μg/mL antipain.
   Instead of individual protease inhibitors, a tablet-type cocktail can be used (e.g., BCL, cat. no. 1836 170). Prepare the protease inhibitor and DTT stocks in sterile water and add them immediately before use.

Other materials required: large mortar and pestle, plastic funnel (10 cm od), large-bore sterile pipet (e.g., Gilson 5-mL pipet), Nitex (Sephar America, Depew, NY) membranes (mesh sizes 300, 100, and 56 μm), liquid nitrogen.

2.4. Nuclear Extract Preparation

1. Nuclear lysis buffer (NLB): 50 m*M* HEPES, pH 7.6, 2.5 *M* KCl, 5 m*M* $MgCl_2$, 20% glycerol, 1 m*M* DTT, protease inhibitor cocktail (or 0.5 μg/mL leupeptin, 50 μg/mL antipain).
2. Dialysis Buffer (DB): 20 m*M* HEPES, pH 7.9, 100 m*M* KCl, 0.1 m*M* EDTA, 10% (v/v) glycerol.

Other materials required: dialysis tubing, Centricon plus-20 concentrator (Millipore, cat. no. UFC2 LGC 02), Quick Start Bradford Protein Assay (Bio-Rad, cat. no. 500-0201), bovine serum albumin (BSA, Sigma-Aldrich, cat. no. A2153), liquid nitrogen.

2.5. Isolation of Chromatin from Arabidopsis Seedlings

1. TBT buffer: 0.5% (w/v) Triton X-100, 0.05 *M* Tris-HCl (pH 8.1), 15 m*M* β-mercaptoethanol, 0.1 m*M* PMSF, 12 m*M* $NaHSO_3$.
2. 3 *M* Ammonium sulfate.
3. TBM buffer: 0.01 *M* Tris HCl (pH 8.1), 15 m*M* β-mercaptoethanol, 0.1 m*M* PMSF, 12 m*M* $NaHSO_3$, 2 m*M* $MgCl_2$.
4. Final purification solution: 1.7 *M* sucrose, 0.01 *M* Tris-HCl (pH 8.1), 0.5% (w/v) Triton X-100, 15 m*M* β-mercaptoethanol, 0.1 m*M* PMSF, 12 m*M* $NaHSO_3$.

Other equipment and materials required: Sorvall Omnimixer or Waring blender, a Potter homogenizer with a Teflon pestle, Miracloth (Calbiochem) filtering material, liquid nitrogen.

2.6. Extraction of Total Histones from Chromatin

1. Guanidine hydrochloride solutions (GP): 40 or 6% (w/v) guanidine hydrochloride, 50 m*M* KH_2PO_4, 50 m*M* K_2HPO_4 (pH 6.8, adjust with KOH).
2. BioRex 70 cation exchange resin (Bio-Rad Laboratories, Richmond, VA), made by mixing 10 g of BioRex 70 and 60 mL of 6% GP solution, pH 6.8.
3. Dialysis buffer: 0.01 *M* H_2SO_4, 0.1 m*M* PMSF.
4. Trichloroacetic acid (100% w/v).

5. 100 m*M* potassium phosphate buffer, pH 6.8 (made by mixing 49.7 mL 1 *M* K_2HPO_4 and 50.3 mL 1 *M* KH_2PO_4).

Other equipment and materials required: Sonicator (e.g., UP 200 200H, Dr. Hielscher GmBH, Germany), refractometer, dialysis tubing, vacuum centrifuge, liquid nitrogen.

3. Methods

3.1. Nuclei Isolation

1. For the seed sterilization, use 150 µL of packed *Arabidopsis* seeds, add 70% ethanol to seeds in an Eppendorf tube and immediately remove ethanol.
2. Add 50% bleach (10% sodium hypochlorite in final solution) and mix by inverting the tube a few times for 5 min, spin the seeds down, wash with sterile water, spin down, and resuspend in water.
3. Start the *Arabidopsis thaliana* plant culture by taking a desired number (depending on the scale of the experiment and the available space in the growth chamber) of 500-mL Erlenmeyer flasks. To obtain approx 40 to 50 g of wet mass per Erlenmeyer flask, approx 40 mg of seeds are needed. Take sterilized seeds and place in approx 250 mL of Gamborg's medium. Maintain under appropriate light conditions (22°C, 16 h light/8 h dark cycle) with vigorous shaking. Maintain shaking at the approximate speed of 200 rpm.
4. Harvest plants after approx 3 wk of culture. The seeds should have germinated to produce a thick, dark-green mass of plant tissue, yielding approx 40 to 50 g of wet mass per flask.
5. Remove excess liquid medium from plants, using commercially available absorbent paper tissues (e.g., Kleenex). Working quickly in the cold room, fragment the mass of plant tissue manually, tearing small bits of plant material away. Drop fragments promptly into a prepared container with liquid nitrogen until all plant material has been processed. *See* also **Note 1**.
6. Precool a steel motorized blender (a coffee bean grinder that can withstand a temperature of liquid nitrogen can be used; *see* **Note 2**) and add liquid nitrogen. Add crushed plant material and homogenize at full speed until all tissue resembles fine, white-green powder (*see* **Note 3**).
7. Resuspend the homogenized plant material (or the *Arabidopsis* protoplasts) in 200 mL of ice-cold NIB 1. For *Arabidopsis* protoplasts obtained from 100 mL of a 5-d culture (usually less than 10 mL slurry), reduce the volume of added ice-cold NIB 1 to 40 mL. Stir with a glass rod until the solution becomes homogenous and less granular and when no ice crystals (formed upon the addition of very cold sample) are present. Stirring with a glass rod is not necessary when isolating nuclei from the protoplasts; in that case, invert the tube a few times by hand.
8. Filter the homogenate through Miracloth and reapply the filtrate for a new filtering step as before (*see* **Note 4**).
9. Centrifuge the filtrate for 10 min at 4°C, at 1000*g* in a swing-out rotor (e.g., Eppendorf A-4-62). Resuspend the dark-green pellet in 10 mL of 80% (v/v) Percoll gradient solution and place this sample at the bottom of the 50-mL centrifuge tube. On top of this sample, layer equal volumes (10 mL each) of Percoll gradient solutions, as follows (from bottom up): 80%, 60%, 30%, 0% (only NIB 1).
10. Carefully centrifuge the gradient tubes in a swing-out rotor (Sorvall SW-40 or equivalent) for 2 h at 4°C at 1000*g*. The nuclei separate from other cellular components during the centrifugation and migrate upward through the Percoll layers. After centrifugation, the nuclei are a whitish band between 30% and 60% Percoll layers, whereas the dark-green pellet contains chloroplasts. A pale-green band between the pellet and the nuclei band contains broken chloroplasts and some broken nuclei (**Fig. 1**; *see* also **Fig. 3**).
11. Collect the whitish pale green nuclear layer in a separate Falcon tube by a careful aspiration of the upper layers, using a disposable sterile plastic Pasteur pipet, and add up to 50 mL of NIB 2 to wash the nuclei. Centrifuge the nuclei for 10 min at 850*g* at 4°C and resuspend in a small volume (e.g., 1 mL) of NIB 2/50% glycerol. Freeze the sample in liquid nitrogen and store at –80°C.
12. Check for DNA absorbance at 260 nm by resuspending 5 µL of the nuclear sample in 500 µL of urea/NaCl solution. Use urea/NaCl solution as a blank. This step is intended to standardize serial nuclear preparations, allowing for their direct comparisons, as judged by equal DNA content. However, this step can be omitted and the quality of the nuclear preparations can be established as described below (**step 13**).

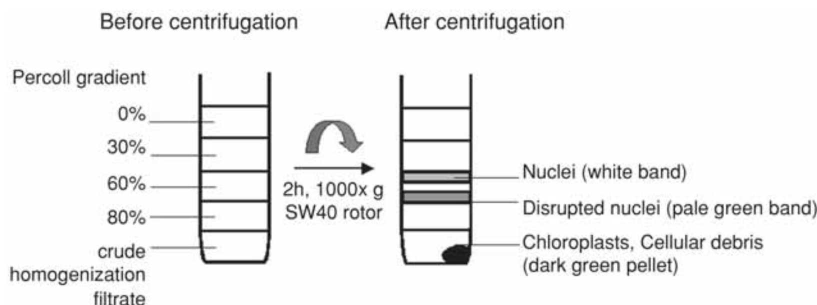


Fig. 1. Percoll gradient centrifugation applied to obtain *Arabidopsis* nuclei. Before the swing-out rotor centrifugation (on the right), the gradient consists of four layers of Percoll placed on top of the crude homogenization filtrate (left image). After centrifugation is complete, for 2 h at 1000*g*, the cellular components migrate upward, and the intact nuclei separate as a whitish band between 30 and 60% Percoll layer above the broken nuclei band and the cellular debris (right image).

13. Check the quality of nuclei preparations by 4',6-diamidino-2-phenylindole (DAPI) staining under an epifluorescent microscope (for the estimation of the ratio of intact to broken nuclei and for chromatin fiber release), by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (for the analysis of protein complement; **Fig. 3**) and by micrococcal nuclease (MNase) digestion (for the preservation of the intact chromatin organization). For the MNase digestion assay, resuspend the nuclear pellet in 15 mL MNase digestion buffer and centrifuge, to pellet again the nuclei, for 40 min at 18,000*g* in an ice-cold Sorval SS34 rotor. Resupend the washed nuclei in 1.2 mL of the MNase digestion buffer with two pestle strokes of the Potter homogenizer and aliquot the suspension into samples of equal volume (the number of samples depends on the number of desired time points; *see* **Subheading 3.2.**). Preheat the chromatin samples at 37°C for 5 min and add 10 U of MNase, incubating the samples at 37°C for a strictly defined and controlled time. Use time points of 1, 3, 5, 10 (or, additionally, 30) min as well as a "zero minute" control time point, stopping the reaction by the addition of the equal amount of the stop buffer and by chilling at the ice bath temperature. Analyze the digestion reaction products on 1.2% agarose gel with ethidium bromide added. Remember to include an undigested chromatin sample, as a negative control of the digestion process. *See* **Note 6** on the appearance of the nuclei.

3.2. Protoplast Preparation from Arabidopsis Suspension Cultured Cells

1. Centrifuge 100 mL of 5-d-old cultures of *Arabidopsis* suspension cells in two 50-mL tubes at 300*g* in a swinging bucket rotor for 10 min at room temperature.
2. Wash the cells by resuspending each tube in 50 mL of 0.4 *M* mannitol/20 m*M* MES, pH 5.5, and centrifuge as before. Cool centrifuge to 4°C after this step.
3. Resuspend the cells in 50 mL of protoplast enzyme solution in each tube and transfer to 250-mL flask. Shake at room temperature for 60 min with gentle agitation at 150–200 rpm. (Gently swirl the cells occasionally to keep them suspended.)
4. Check for protoplast formation under the microscope, using 10 µL of the suspension in a hemacytometer. Keep shaking if necessary.
5. Centrifuge the protoplasts in 50 mL tubes at 300*g* for 10 min at 4°C.
   Wash the protoplasts twice by resuspending in 50 mL of ice-cold 0.4 *M* mannitol in each tube and centrifuge at 300*g* for 10 min at 4°C. To proceed to the nuclei isolation protocols, continue with **step 7** of **Subheading 3.1.**

3.3. Alternative, Short Protocol for the Isolation of Plant Nuclei

This protocol can also be successfully adapted to isolation of nuclei from other plant species (e.g., pea, tobacco). It provides a more simple procedure than that in **Subheading 3.1.**, and is quicker to follow. Work quickly in a cold room, using sterile solutions and materials.

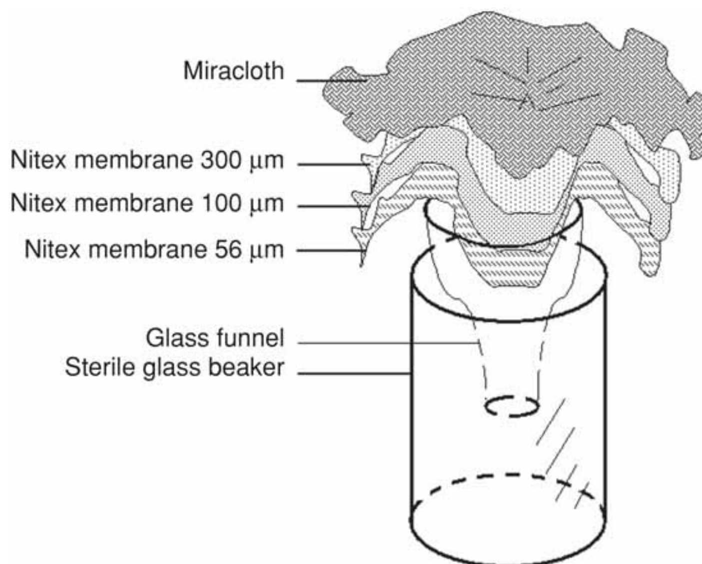


Fig. 2. The layering of the filtration materials on top of the glass funnel placed in a sterile glass beaker.

1. Collect small soil-grown *Arabidopsis* plants, cutting the aerial parts away with sharp scissors (50 g), freeze in liquid nitrogen, and store at –80°C unless processed immediately (*see also* **Note 6**).
2. Transfer the samples to a large mortar, prechilled with liquid nitrogen. Grind in liquid nitrogen for approx 5 min until a fine powder. Transfer the frozen leaf powder into a glass beaker filled with 300 mL of homogenization buffer and mix either with a glass rod or by magnetic stirring. Continue mixing until all lumps of frozen material break up.
3. When the homogenate has almost thawed, filter it through Miracloth and three layers of Nitex nylon membrane, placed in such a way that a 300 µm membrane is the outermost layer, a 100-µm is in the middle, and a 56-µm one faces the inside of the tube (**Fig. 2**). Use a sterile glass or plastic funnel placed on top of the large glass beaker for this purpose. Once all the homogenate has been filtered, remove the Miracloth layer and squeeze out the remaining homogenate into the Nitex membranes inside the funnel.
4. Centrifuge the homogenate at 4640*g* in a Sorvall SLA-1000 rotor for 20 min at 4°C.
5. Remove the green supernatant and resuspend the dark-green pellet containing a crude nuclear fraction in 5 mL of homogenization buffer by gently pipetting up and down with a large-bore sterile pipet. To wash the nuclei for the first time, transfer the homogenates into a 50-mL polypropylene Oakridge tube, and fill the volume to 25 mL with homogenization buffer.
6. Centrifuge the nuclear preparations at 1940*g* in a Sorvall SA-300 rotor for 10 min at 4°C. Repeat the previous washing step for the pellet, and centrifuge nuclei at 1480*g* for 10 min at 4°C, discarding the supernatant again. Repeat the last washing step twice, until the nuclear pellets appear in a shade of yellow, with as little green coloration as possible.
7. Resuspend the nuclear pellet of each batch of starting material in a small volume (1 mL) of nuclear resuspension buffer by gentle pipetting. Freeze the sample in liquid nitrogen and store at –80°C.
8. Check the quality of the preparation by epifluorescence microscopy, after DAPI staining, and by SDS-PAGE for protein analysis, as described in **Subheading 3.1.**

3.4. Nuclear Extract Preparation

1. Thaw the nuclear suspension (prepared according to **Subheading 3.1.** or **3.3.**) on ice, and determine the exact volume before transferring with a pipet to a 50-mL polypropylene Oakridge tube.
2. Add an amount of the nuclear lysis buffer to give a final concentration of KCl of 0.47 *M* (190 µL of buffer per mL of suspension).

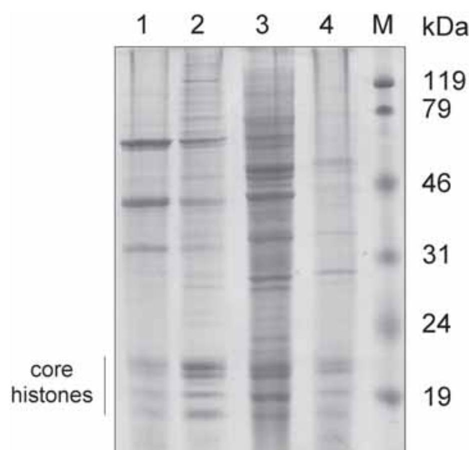


Fig. 3. SDS-polyacrylamide gel electrophoresis (PAGE) analysis of the *Arabidopsis* nuclear protein fractions obtained after the procedure in **Subheading 3.1.** (lanes 1–3) and after the procedure in **Subheading 3.4.** (lane 4): lane 1—lower band after Percoll gradient centrifugation, containing mostly broken nuclei from whole *Arabidopsis* seedlings; lane 2—upper band after Percoll centrifugation, containing intact nuclei from *Arabidopsis* seedlings; lane 3—upper band after Percoll gradient centrifugation containing intact nuclei from *Arabidopsis* suspension cells; lane 4—chromatin isolated from *Arabidopsis* seedlings. The 12% gel was stained with colloidal Coomassie; indicated on the left are the core histones; molecular weight markers (kDa) are shown on the right.

3. Place the suspension on a roller and gently mix at 4°C for 30 min.
4. Add 3 mL of dialysis buffer and mix gently by hand. Centrifuge the lysed nuclei at 48,400*g* in a Sorvall SA-300 rotor for 20 min at 4°C to sediment the chromatin.
5. Remove the supernatant without disturbing the viscous, whitish pellet, which contains genomic DNA, and transfer the supernatant into dialysis tubing. Dialyze the supernatant against several changes (e.g., 5 × 800 mL) of dialysis buffer over 3 to 4 h each at 4°C.
6. Precipitate the proteins by freezing the dialysate in liquid nitrogen, and by thawing on ice. Centrifuge at 14,600*g* for 15 min at 4°C. This step is intended to reduce the likelihood of clogging the concentrator during the next stage of the procedure. However, this step can be omitted to increase the protein yield.
7. Transfer the supernatant to a Centricon Plus-20 concentrator and place on top of the 50-mL tube. Centrifuge in a swing-out rotor (e.g., Beckman SW30 rotor) at 300*g* until the sample is about one-tenth of the original volume. About 4 h are required to complete this step.
8. Mix the extract while still in the concentrator by pipetting, and aliquot it into batches of 25 µL. Freeze the aliquots in liquid nitrogen and store at –80°C. Determine the protein concentration by using a Bio Rad protein assay kit, with bovine serum albumin (BSA) as the standard. Dilute BSA in dialysis buffer.

3.5. Isolation of Chromatin from Arabidopsis *Seedlings*

This procedure is based on methods developed by Simon and Becker ***(3)***, modified by Spiker ***(4)*** for wheat germ or wheat seedlings, and further modified for *Arabidopsis* seedlings. All procedures should be carried out at 4°C.

1. Grow young *Arabidopsis* seedlings in soil (7–8 d after germination). Harvest them in a convenient tube until needed (shock-frozen in liquid nitrogen and stored at –80°C).
2. Homogenize two 10-g samples for 1 min at full speed in a Sorvall Omnimixer in 40 mL TBT each. A glass-enclosed Waring blender can also be used (make sure the solution and tissue cover the steel blades on the bottom).

3. Pool both homogenates, add 20 mL TBT, and stir for 10 min with a magnetic bar.
4. Centrifuge for 5 min at 4000*g*. Discard the white-yellow pellets and collect the supernatant.
5. Filter the supernatant through one layer of Miracloth and add dropwise 3 *M* ammonium sulfate to a final concentration of 0.05 *M*. To ease the filtering, it is best to use a glass or plastic funnel placed in a cold beaker. Use a magnetic plate and a stirring bar to mix the supernatant during ammonium sulfate protein precipitation.
6. Centrifuge the supernatant for 10 min at 10,000*g*.
7. Resuspend the pellet in 40 mL TBT plus 0.05 *M* ammonium sulfate, using a Potter homogenizer with a Teflon pestle. It is best to resuspend the pellet in a smaller volume of TBT with ammonium sulfate in a Potter homogenizer, and to fill with the remainder of the buffer after a transfer to a glass beaker of a suitable volume.
8. Centrifuge the suspension for 10 min at 10,000*g*, as before.
9. Wash the white, distinct pellet by three cycles of suspension in 20 mL TBM with 0.05 *M* ammonium sulfate and centrifugation at 10,000*g* for 10 min, as performed previously.
10. Purify the resuspended chromatin by centrifugation through "Final purification solution" for 30 minutes at 16,000*g*. During this step, layer final purification solution on the bottom of the tube and slowly add the resuspended chromatin. A convenient volume for this procedure, when using the 50-mL polypropylene Oakridge tubes, is 20 mL of the bottom final purification solution and 10 mL of the top sample layer.
11. Wash the gelatinous pellet by two cycles of suspension in TBM, followed by centrifugation for 10 min at 15,000*g*. Finally, resuspend the pellet in a small volume (~1 mL) of TBM and freeze promptly in liquid nitrogen. Store at –80°C, until needed. This material can also be used as a starting base for the preparation of histone octamers, as described by Moehs et al. ***(6)***.

3.6. Extraction of Total Histones from Chromatin

For a detailed discussion of the difficulties of isolating intact chromatin histones from *Arabidopsis* plants grown from seeds in an axenic liquid medium, as opposed to soil-grown material ***(13)***, the reader is referred to an excellent publication on the topic ***(14)***. The liquid medium culture provides a high-yield source of material, with well-defined growth conditions and a convenient transformant selection ***(1)***, making liquid-grown *Arabidopsis* a useful avenue for the histone isolation procedure. However, the commonly experienced degradation of the core and linker histones found in the case of liquid-grown *Arabidopsis* plants can pose a problem. The degradation was attributed to higher polycarbohydrate content, an increased production of proteolytic enzymes, and their ability to bind to DNA ***(14)***. In order to selectively extract histones from *Arabidopsis* grown in liquid media, a method has been developed using extraction with guanidine hydrochloride followed by selective binding on BioRex70 resin in the batch system. To obtain intact histone proteins with this technique, the nuclear pellet, as obtained under **Subheading 3.1.** or **3.2.**, is used.

1. Resuspend the nuclear pellet in 15 mL of 40% GP buffer and sonicate 4 × 20 s in an ice bath, e.g., in a UP 200 200H sonicator with a 14 mm probe (amplitude set at 125 µm and power at 105 W/cm^{-2}). Centrifuge the solution for 10 min at 19,000*g* in a Sorvall SS34 rotor and discard the pellet.
2. Add concentrated HCl to the supernatant to a final concentration of 0.25 *M*, mix the solution on ice overnight, and centrifuge as before. Add 100 m*M* potassium phosphate buffer, pH 6.8, to the supernatant, to make the refractive index the same as that of 6% GP buffer, and add about 70 mL of BioRex 70 resin slurry to the supernatant. Gently agitate the mixture overnight at room temperature.
3. Allow the resin to settle and wash it with 6% GP buffer. Add an equal volume of 40% GP buffer to the slurry and mix for 1 h at room temperature. Centrifuge in a SS34 rotor at 500*g* for 5 min at 4°C. Carefully remove the supernatant, placing it in the dialysis tubing. Proceed with dialysis for 6 h, with three changes of dialysis buffer, and make the solution 25% in respect to trichloroacetic acid (TCA) by adding 100% (w/v) TCA.
4. Stir the solution overnight in the cold room and centrifuge for 45 min at 43,000*g* in a SS34 rotor. Wash the precipitate in 20 mL cold acetone and centrifuge for 30 min at 38,000*g* in an SS34 rotor.

5. Finally, the pellet is dried under cold air, dissolved in a small volume of cold water and concentrated in a vacuum centrifuge. *See* **Note 7** on the electrophoretic behavior of histone H1.

4. Notes

1. This method of freezing has proved far superior to batch freezing in large quantities, when aqueous medium trapped between the plants freezes as well, which leads to unnecessary dilution of the sample and to lower quality of the isolated nuclei. With all material frozen in liquid nitrogen, one has either the option of storing it at –80°C until needed, or processing it at once. It is apparent from our observations that the material stored at –80°C can retain good potential for yielding intact nuclei of desired quality. However, the material is much more prone to shearing and to breakage (as assessed by DAPI staining for the chromatin fiber release), compared to the freshly processed nuclei. Ideally, if storing of the plant material is warranted, one should contemplate storage in liquid nitrogen; for instance, in small cryopreservation containers of stainless steel, kept at –80°C. However, this storage method might not be feasible in all laboratories and requires constant monitoring for nitrogen evaporation.
2. In our experience, the Moulinex type 505 coffee bean grinder was the best for this procedure, but any grinder sturdy enough for extremely low temperatures of liquid nitrogen should work for some time. When working with liquid nitrogen, follow safety procedures; in particular, protect your eyes with safety goggles.
3. It is not advisable, from our own experience, to further disrupt the tissue as suggested in some older protocols (e.g., by the Potter-type homogenizers). We have found that this leads to breakage of the nuclear envelope and spillage of chromatin.
4. The purity of nuclei can be increased by filtering on a Nitex (Sephar America, Depew, NY) membrane series (100, 50, 30 μm). This can be achieved during the early stage of the procedures (after **step 7**) or at a late stage (after resuspension in NIB 2 without glycerol—**step 10**). However, as a disadvantage, such a procedure might lead to a lower yield of nuclei.
5. Optimally, the nuclei observed under epifluorescence should resemble round spheres with a diameter of about 4 to 5 μm, with few or no broken or torn structures found nearby. However, it is also common to observe small particles (starch granules) with no fluorescent properties adjacent to the nuclei. Such granules pose little or no trouble in most biochemical analyses, but can be easily removed by filtration through a nylon membrane (*see* **Note 5** above) or by a Percoll gradient centrifugation.
6. If needed, whole soil-grown plants can be used as well. Grow the seeds on a single layer of cheesecloth placed on top of soil and pull out the plants together with their roots. The roots should then be washed with sterile water and dried on a piece of absorbent paper.
7. As discussion of the different selective procedures for histone extraction exceeds the scope of this chapter, the reader is referred to an informative volume in the *Methods in Enzymology* series ***(17)***. It is worth noting, for instance, that plant histone H1 exhibits a weak staining with Coomassie stain, leading often to an underestimate of the amount of this protein isolated from plant material ***(18)***.

Acknowledgments

We would like to thank Maciej Kotlinski (Laboratory of Plant Molecular Biology of the University of Warsaw, Warsaw, Poland) for the excellent collaboration during preparation of nuclear extracts of *Arabidopsis* seedlings, and Andrzej Jerzmanowski (Institute of Biochemistry and Biophysics, Polish Academy of Sciences and Laboratory of Plant Molecular Biology of the University of Warsaw, Warsaw, Poland) for help with obtaining reagents and general support with funding from the Polish Committee for Scientific Research. We thank J. Marcela Hernandez for critical reading of the manuscript and Diane Furtney for expert manuscript editing (both at the Plant Biotechnology Center, The Ohio State University). Financial support by the National Science Foundation (MCB-0079577 and MCB-0209399) and the US Department of Agriculture (Plant Growth and Development no. 2001-01901) to I.M. is greatly acknowledged.

References

1. Weigel, D. and Glazebrook, J. (2002) *Arabidopsis. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

2. Bowman, J., ed. (1994) *Arabidopsis. An Atlas of Morphology and Development*. Springer, New York, NY.
3. Simon, J. H. and Becker, W. M. (1976) A polyethylene glycol/dextran procedure for the isolation of chromatin proteins (histones and nonhistones) from wheat germ. *Biochim. Biophys. Acta* **454,** 154–171.
4. Spiker, S. (1984) High-mobility group chromosomal proteins of wheat. *J. Biol. Chem.* **259,** 12,007–12,013.
5. Cox, K. H. and Goldberg, R. B. (1988) Analysis of plant gene expression, in Plant Molecular Biology. A Practical Approach, Shaw, C. H., ed. IRL Press, Oxford, Washington, DC, pp. 9– 15.
6. Moehs, C. P., McElwain, E. F., and Spiker, S. (1988) Chromosomal proteins of *Arabidopsis thaliana. Plant Mol. Biol.* **11,** 507–515.
7. Spiker, S. and Thompson, W. F. (1996) Nuclear matrix attachment regions and transgene expression in plants. *Plant Phys.* **110,** 15–21.
8. Calikowski, T. T., Meulia, T., and Meier, I. (2003) A proteomic study of the *Arabidopsis* nuclear matrix. *J. Cell. Biochem.* **90,** 361–378.
9. Bae, M. S., Cho, E. J., Choi, E.-Y. and Park, O. K. (2003) Analysis of the *Arabidopsis* nuclear proteome and its response to cold stress. *Plant J.* **36,** 652–663.
10. Moehs, C. P., Baxevanis, A. D., Moudrianakis, E. N., and Spiker, S. (1992) Enhanced stability of histone octamers from plant nucleosomes: role of H2A and H2B histones. *Biochemistry* **31,** 10,844–10,851.
11. Bastow, R., Mylne, J. S., Lister, C., Lippman, Z., Martienssen, R. A., and Dean, C. (2004) Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature* **427,** 164– 167.
12. Fransz, P. F. and de Jong, J. H. (2002) Chromatin dynamics in plants. *Curr. Opin. Plant Biol.* **5,** 560–567.
13. Karesch , H., Bilang, R., Mittlesten Scheid, O., and Potrykus, O. (1991) Direct gene transfer to protoplasts of *Arabidopsis thaliana. Plant Cell Rep.* **9,** 571–574.
14. Przewoka, M. R. and Jerzmanowski, A. (1997) Improved method of total histone isolation from *Arabidopsis thaliana. Biol. Plantarum* **39,** 299–302.
15. Gamborg, O. L., Miller, R. A., and Ojma, K. (1968) Nutrient requirements of suspension cultures of soybean root cells. *Exp. Cell Res*. **50,** 151–158.
16. Blackhall, N. W., Davey, M. R., and Power, J. B. (1994) Isolation, culture and regeneration of protoplasts, in *Plant Cell Culture. A Practical Approach*, Dixon, R. A. and Gonzales, R. A., eds. IRL Press, Oxford, UK, pp. 27–39.
17. von Holt, C., Brandt, W. F., Greyling H. J., et al. (1989) Isolation and characterization of histones, in *Methods in Enzymology*, Vol. 170. (Wassarman, P. M. and Kornberg, R. D., eds.), Academic, San Diego: pp. 431–524.
18. Mazzolini, L., Vaeck, M., and Montagu, M. (1989) Conserved epitopes on plant H1 histones recognized by monoclonal antibodies. *Eur. J. Bioch.* **178,** 719–819.

33

Purification and Fractionation of Membranes for Proteomic Analyses

Anne Marmagne, Daniel Salvi, Norbert Rolland, Geneviève Ephritikhine, Jacques Joyard, and Hélène Barbier-Brygoo

Summary

Proteomics is a very powerful approach to link the information contained in sequenced genomes, such as *Arabidopsis*, to the functional knowledge provided by studies of plant cell compartments. However, membrane proteomics remains a challenge. One way to bring into view the complex mixture of proteins present in a membrane is to develop proteomic analyses based on (1) the use of highly purified membrane fractions and (2) fractionation of membrane proteins to retrieve as many proteins as possible (from the most to the less hydrophobic ones). To illustrate such strategies, we choose two types of membranes, the plasma membrane and the chloroplast envelope membranes. Both types of membranes can be prepared in a reasonable degree of purity from different types of tissues: the plasma membrane from cultured cells and the chloroplast envelope membrane from whole plants. This article is restricted to the description of methods for the preparation of highly purified and characterized plant membrane fractions and the subsequent fractionation of these membrane proteins according to simple physicochemical criteria (i.e., chloroform/methanol extraction, alkaline or saline treatments) for further analyses using modern proteomic methodologies.

Key Words: Plasma membrane; chloroplast envelope membranes; two-phase partitioning; membrane purity; membrane proteins; membrane lipids; pigments.

1. Introduction

Membrane proteins play a crucial role in many cellular and physiological processes. They are essential mediators of material and information transfer between cells and their environment, and between compartments within cells. The functional diversity of proteins in a cell actually is strongly related to the diversity of their physicochemical properties. This is even more obvious in membranes because of their hydrophobic nature. Ion channels or receptors, for instance, are "integral" or "intrinsic" membrane proteins, often containing several transmembrane α-helices linked together by loops located outside the membrane in an aqueous environment. Such proteins are amphipathic, in that they contain both hydrophobic and hydrophilic regions, their overall hydrophobicity relying on the proportion between loops and α-helices. In some cases, amino acids in the loops are modified by oligosaccharides thus increasing their hydrophilicity. The secondary structure of a few membrane proteins consist of β-sheets, thus forming β-barrels through which hydrophilic molecules can cross the membrane. Porins are the most conspicuous example of this type of membrane proteins, which are much less hydrophobic than proteins containing α-helices. Not all membrane proteins have transmembrane domains. Some proteins are embedded within only one bilayer of the membrane (monotopic proteins). Other types of proteins are anchored to the membrane owing to a hydrophobic moiety (fatty acid or isoprenoid chain, for instance) that is embedded in the lipid phase of the membrane. Both nontransmembrane proteins and integral proteins may be more or less tightly bound to

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

other membrane proteins, the so-called class of "peripheral membrane proteins," through ionic or hydrophobic interactions.

A membrane therefore is an extremely complex mixture of proteins. Some are very hydrophobic while others are hydrophilic; there are basic and acid proteins, low and high molecular mass proteins, and they may be present at high or low abundance. Membrane proteins are extremely difficult to separate from one another and to analyze for further functional studies, essentially because of the presence of lipids. Therefore, innovative tools and methods should be developed for the study of membrane proteins. One way to bring such proteins into view is to develop proteomic analyses based on subcellular compartmentation and/or physicochemical criteria. The purpose of this article is to describe rather simple procedures that have been developed to set up membrane proteomic studies in plants and especially in *Arabidopsis* (*1–5*). To illustrate such strategies, we choose two types of membranes, the plasma membrane and the chloroplast envelope membranes, each one providing a unique lipid environment to membrane proteins. Furthermore, both types of membranes can be prepared in a reasonable stage of purity from different types of tissues: the plasma membrane from cultured cells and the chloroplast envelope membrane from whole plants. Interestingly, Ferro et al. (*2*) identified in *Arabidopsis* chloroplast envelope membranes a series of new (and often very minor) transporters that were very hydrophobic and had alkaline isoelectric points. None of them could be expected to be identified by classical proteomic approaches. This article is restricted to the description of methods for the preparation of highly purified and characterized plant membrane fractions and the subsequent fractionation of these membrane proteins according to simple physicochemical criteria for further analyses using modern proteomic methodologies (*6*).

2. Materials

2.1. Arabidopsis *Cell Culture*

1. Sterile beakers (500 mL, 1 L, 5 L).
2. Sterile pipet (10 mL).
3. Sterile cylinders (100 mL).
4. Sterile Erlenmeyer flasks with plastic tops.
5. Sterile hood for cell culture.
6. Autoclave.
7. Rotary shaker (INFOR, type AG, or equivalent).
8. Cell culture media:
    a. Phosphate buffer: 0.03 m*M* NaH_2PO_4, 0.06 m*M* Na_2HPO_4, 0.37 m*M* KH_2PO_4.
    b. Macronutrients: 19.43 m*M* KNO_3, 0.45 m*M* $MgSO_4$, 0.90 m*M* $CaCl_2$.
    c. Micronutrients: 30.1 μ*M* H_3BO_3, 1.5 μ*M* KI, 39.6 μ*M* $MnSO_4$, 11 μ*M* $ZnSO_4$, 0.3 μ*M* Na_2MoO_4, 0.03 μ*M* $CuSO_4$, 0.03 μ*M* $CoCl_2$, 5 μ*M* $FeSO_4$, 5 μ*M* Na_2 EDTA.
    d. Vitamins: 4.1 μ*M* nicotinic acid, 2.4 μ*M* pyridoxine, HCl (B6), 1.2 μ*M* thiamine, HCl (B1).
    e. Carbon and nitrogen sources: 548.8 μ*M* myo-inositol, 26.6 μ*M* glycine, 0.1 g/L casein acid hydrolysate from bovine milk, Hy-Case Amino, 43.86 m*M* (1.5%) sucrose.
    f. Hormone: 1 μ*M* 1-naphthalene acetic acid (NAA).

2.2. Plasma Membrane Preparation

1. Beakers (500 mL, 1 L, 5 L).
2. Ice and ice buckets.
3. Pipet (1 mL, 10 mL).
4. Pyrex filter funnel with fritted disk (90 cm diameter, 100 to 160 μm fritted glass porosity).
5. 1 vacuum flask (2 L).
6. 1 spatula.
7. 1 cylinder (50 mL).
8. Pasteur pipet.
9. Disposable plastic transfer pipet (3 or 5 mL).
10. Automatic pipets (20, 200, and 1000 μL).

11. Small blender with a unique speed (200 mL; such as type D56 from Moulinex or equivalent).
12. Potters and Teflon pestles (2 and 8 mL).
13. 50-mL centrifuge tubes (e.g., polyethylene tubes, Falcon).
14. PEG 3000 (Polyethyleneglycol; ICN).
15. Dextran T500 (Amersham).
16. Cell washing: 200 m*M* KCl, 50 m*M* EDTA, add KOH up to pH 5.5 (*see* **Note 1**).
17. Cell grinding: 50 m*M* Tris-HCl, 500 m*M* sucrose, 10% (v/v) glycerol, 20 m*M* EDTA, 20 m*M* EGTA, 50 m*M* sodium fluoride, 5 m*M* β-glycerophosphate, 1 m*M* phenanthroline, 0.6% (v/v) polyvinylpyrrolidone, 10 m*M* ascorbic acid, 1 *M* MES down to pH 8.0 and at the last moment: 5 m*M* dithiothreitol, 0.5 μg/mL leupeptine hemisulfate salt, dessicate 2 m*M* PMSF.
18. Microsome resuspending medium: 330 m*M* sucrose, 2 m*M* dithiothreitol, 5 m*M* KH_2PO_4, pH 7.8.
19. 27 g PEG-dextran phases for two-phase partitioning plasma membrane purification: 6.4% (w/w) polyethylene glycol (PEG) 3350, 6.4% (w/w) dextran T-500 (*see* **Note 2**), 5 m*M* KH_2PO_4, 3 m*M* KCl, 330 m*M* sucrose, add H_2O for a total weight of 27 g (*see* **Note 3**).
20. K-PO_4 salt washing: 0.39 mol K_2HPO_4 (43.55 g/L anhydrous) + 0.32 mol KH_2PO_4 (68 g/L anhydrous), pH 7.0.
21. Plasma membrane washing: 10 m*M* Tris-HCl, 10 m*M* ascorbic acid, 9 m*M* KCl, 300 m*M* sucrose, 10 μ*M* $MgCl_2$, adjust to pH 8.3, and add just before use: 5 μ*M* EDTA, 5 μ*M* EGTA, 0.5 μg/mL leupeptine hemisulfate salt, dessicate 5 m*M* dithiothreitol.
22. Plasma membrane resuspending medium: 50 m*M* MOPS/NaOH, 0.1 m*M* dithiothreitol, pH 7.8.
23. Refrigerated centrifuge (Jouan E96, Star B, 4 × 50 mL, or equivalent).
24. Refrigerated superspeed centrifuge (Sorvall RC5B), with the following rotors (and corresponding tubes): fixed angle rotor GSA (6 × 250 mL polycarbonate bottles), or equivalent.
25. Refrigerated preparative ultracentrifuge (Beckman L8-70M), with a SW 45 Ti rotor (6 × 70 mL polycarbonate tubes and aluminium caps), or equivalent.

2.3. Chloroplast Envelope Preparation

1. Muslin or cheesecloth, 80 cm large.
2. Nylon blutex, 50 μm aperture (Tripette et Renaud, Sailly Saillisel, France).
3. Beakers (500 mL, 1 L, 5 L).
4. Ice and ice buckets.
5. Pipet (1 mL, 10 mL).
6. Leaf grinding medium: 0.45 *M* sorbitol, 20 m*M* Tricine-KOH, pH 8.4, 10 m*M* EDTA, 10 m*M* $NaHCO_3$, 0.1 bovine serum albumin (BSA, defatted) (*see* **Note 4**).
7. Chloroplast isolation and washing medium: 0.30 *M* sorbitol, 20 m*M* Tricine-KOH, pH 7.6, 5 m*M* $MgCl_2$, 2.5 m*M* EDTA.
8. Solution for Percoll gradients: Mix 1 vol Percoll (Pharmacia, Uppsala) with 1 vol of medium containing 0.60 *M* sorbitol, 40 m*M* Tricine-KOH, pH 7.6, 10 m*M* $MgCl_2$, and 5 m*M* EDTA, to obtain a 50% Percoll/0.3 *M* sorbitol solution.
9. Hypotonic medium for chloroplast lysis: 10 m*M* MOPS-NaOH, pH 7.8, 4 m*M* MgCl2, 1 m*M* PMSF, 1 m*M* benzamidine and 0.5 m*M* ε-amino caproic acid
10. Sucrose gradients for chloroplast fractionation: 10 m*M* MOPS-NaOH, pH 7.8, 4 m*M* $MgCl_2$, with 0.3, 0.6, or 0.93 *M* sucrose.
11. Chloroplast envelope washing medium: 10 m*M* MOPS-NaOH, pH 7.8, 1 m*M* PMSF, 1 m*M* benzamidine, and 0.5 m*M* ε-amino caproic acid.
12. Motor-driven blender, 3 speed, 1 gallon (3.785 L) (Waring blender).
13. Superspeed refrigerated centrifuge (Sorvall RC5), with the following rotors (and corresponding tubes): fixed-angle rotors GS-3 (6 × 500 mL plastic bottles) and SS34 (8 × 50 mL polypropylene tubes); swinging bucket rotor HB-6 (6 × 50 mL polycarbonate tubes) or equivalent.
14. Preparative refrigerated ultracentrifuge (Beckman L7), with a SW 41 Ti rotor (6 × 13.2 mL Ultraclear tubes) or equivalent.

2.4. Enzymatic Markers

1. ATPase activity, nitrate sensitive (EC 3.6.1.35), marker for tonoplast (*see* **Note 5**):

a. Reaction medium: 50 m*M* (MES), 160 m*M* sucrose, 0.1 m*M* Na_2MoO_4, 5 m*M* $MgSO_4$, 1 m*M* NaN_3, 0.02% Brij 58, 1 m*M* dithiothreitol, and 50 m*M* KCl or inhibitor 50 m*M* KNO_3 adjusted to pH 8.0 with Tris-HCl (*see* **Note 6**).
b. 23.7 m*M* Tris - ATP.
c. Fiske and Subbarow reagent: solution A, 4.3 m*M* $(NH_4)_6Mo_7O_{24}$ and 7.4 m*M* SDS solubilized in 5 *N* SO_4H_2; solution B, 2.15 % (w/v) Fiske and Subbarow reagent (*see* **Note 7**).

2. ATPase activity, vanadate sensitive (EC 3.6.1.35), marker for plasma membrane (*see* **Note 5**):
   a. Reaction medium: 50 m*M* MES, 160 m*M* sucrose, 0.1 m*M* Na_2MoO_4, 5 m*M* $MgSO_4$, 1 m*M* NaN_3, 25 m*M* K_2SO_4, 0.02% Brij 58, 1 m*M* dithiothreitol ± inhibitor 0.6 m*M* Na_3VO_4, adjusted to pH 6.5 with Tris-HCl.
   b. 23.7 m*M* Tris-ATP.
   c. Fiske and Subbarow reagent: solution A, 4.3 m*M* $(NH_4)_6Mo_7O_{24}$ and 7.4 m*M* sodium dodecyl sulfate (SDS) solubilized in 5 *N*, SO_4H_2; solution B, 2.15 % (w/v) Fiske and Subbarow reagent (*see* **Note 7**).
3. Cytochrome c oxidase (EC 1.9.3.1), marker for the inner mitochondrial membrane (*see* **Note 5**). Assay medium: 50 m*M* Na_2HPO_4, pH 7.5, 0.3% Triton X-100.
4. Fumarase (EC 4.2.1.2), marker for mitochondrial matrix (*see* **Note 8**). Assay medium: 50 m*M* Tricine/NaOH, pH 7.5, 50 m*M* malate.
5. Hydroxypyruvate reductase (EC 1.1.1.81), marker for peroxisomes (*see* **Note 8**). Assay medium: 50 m*M* Mes-NaOH, pH 6.4, 200 m*M* NADH, 1 m*M* hydroxypyruvate.

2.5. Immunological Markers

1. Anti-H^+-ATPase (P-type) antibody ***(7)*** raised against the plasma membrane H^+-ATPase of *Nicotiana plumbaginifolia* (used at 1/250).
2. Anti-TIP antibody ***(8)*** raised against a tobacco (*Nicotiana tabacum*) tonoplast protein (used at 1/2000).
3. Anti-Nad 9 antibody ***(9)*** raised against an extrinsic protein of the wheat (*Triticum aestivum*) mitochondrial inner membrane (used at 1/2000).
4. Anti-TOM 40 antibody (formerly MOM42; ***10***) recognizes an outer membrane protein from *Vicia faba* mitochondria (used at 1/2000).
5. Anti-T subunit of the glycine-decarboxylase complex ***(11)*** recognizes a matrix protein from pea (*Pisum sativum*) mitochondria (used at 1:10,000).
6. Anti-E 37 antibody ***(12)*** raised against a protein from the inner envelope membrane of spinach (*Spinacia oleracea*) chloroplast (used at 1/20,000).
7. Anti-ceQORH antibody ***(13)*** raised against a protein from the inner envelope membrane of Arabidopsis chloroplast (used at 1/10000).
8. Anti-LHCP antibody ***(14)*** raised against a thylakoid membrane protein from *Chlamydomonas reinhardtii* chloroplast (used at 1/5000).

2.6. Lipids and Pigment Markers

1. 80% (v/v) acetone in water.
2. Methanol/chloroform (2:1, v/v) mixture.
3. Chloroform/methanol/water (65:25:4, v/v/v) mixture.
4. Chloroform/acetone/methanol/acetic acid/water (50:20:10:10:5, v/v) mixture.
5. 0.2% anilinonaphtalene sulfonate in NaOH.
6. Silica gel precoated TLC plate (Merck) (*see* **Note 9**).
7. Rectangular chromatography chamber.
8. Ultraviolet (UV) light.
9. Oven.

2.7. Membrane Treatment

2.7.1. Detergent Extraction

1. Solubilization solution: 50 m*M* MOPS/NaOH, pH 7.8, 1 m*M* DTT, containing either 1% (v/v) Triton X-100 or 0.1 *M* CHAPS (*see* **Note 10**).

2. 50 m*M* MOPS-NaOH, pH 7.8, 1 m*M* DTT.

2.7.2. Chloroform/Methanol Extraction

1. Chloroform/methanol mixtures in the following proportions: 0:9, 1:8, 2:7, 3:6, 4:5, 5:4, 6:3, 7:2, 8:1, 9:0 (v/v).
2. Cold (–20°C) acetone for a 80% final concentration, in water.

2.7.3. Alkaline or Salt Washing of Membranes

1. Na_2CO_3: 0.1 *M* final concentration (1 *M* stock solution).
2. NaOH: 0.1 *M* or 0.5 *M* final concentration (2 *M* stock solution).
3. NaCl: 1 *M* final concentration (2 *M* stock solution).
4. Sonicator.

2.7.4. Equipment

1. Centrifuge (Eppendorf centrifuge 5415D or equivalent) placed in a cold room with 1.5-mL plastic tubes.
2. Branson sonifier model 250 (or equivalent), with 3-mm microtip and ice bucket.
3. Nitrogen (or argon) gas supply (cylinder) with gas pressure regulator connected to a Pasteur pipet via a plastic tube.

2.8. SDS-PAGE and Protein Transfer on Nitrocellulose

1. Acrylamide stocks: store in amber bottles at 4°C.
   a. 30% (w/v) acrylamide–0.8% bisacrylamide: 300 g acrylamide, 8 g bisacrylamide, H_2O to 1 L.
   b. 60% (w/v) acrylamide–0.8% bisacrylamide: 600 g acrylamide, 8 g bisacrylamide, H_2O to 1 L.
2. SDS stock solution: 10% (w/v) SDS: 10 g SDS, H_2O to 1 L; store at room temperature.
3. 10% ammonium persulfate: 10 g ammonium persulfate, H_2O to 100 mL; store at 4°C, prepare fresh every month.
4. Gel buffers:
   a. 4X Laemmli stacking gel buffer (0.5 *M* Tris-HCl, pH 6.8): 363 g Tris-HCl H_2O to 900 mL, adjust to pH 8.8 at 25°C with concentrated HCl, make up volume to 1 L and store at room temperature.
   b. 8X Laemmli resolving gel buffer (3 *M* Tris-HCl, pH 8.8): 60.6 g Tris-HCl, H_2O to 900 mL, adjust to pH 6.8 at 25°C with concentrated HCl, make up volume to 1 L and store at room temperature.
5. Stacking gel (5% acrylamide): 5 ml 30% acrylamide–0.8% bisacrylamide stock solution, 7.5 mL 4X Laemmli stacking gel buffer, 17.1 mL H_2O, 40 µL TEMED, 4 mL 10% ammonium persulfate; total volume: 30 mL.
6. Single acrylamide concentration gels (10, 12 or 15% acrylamide):
   a. For 10% acrylamide gel: 33.3 mL 30% acrylamide–0.8% bisacrylamide stock solution, 12.5 mL 8X Laemmli resolving gel buffer, 54 mL H_2O, 20 µL TEMED, 0.2 mL 10% ammonium persulfate; total volume: 100 mL.
   b. For 12% acrylamide gel: 40 mL 30% acrylamide–0.8% bisacrylamide stock solution, 12.5 ml 8X Laemmli resolving gel buffer, 47.3 mL H_2O, 20 µL TEMED, 0.2 mL 10% ammonium persulfate, total volume: 100 mL.
   c. For 15% acrylamide gel: 50 mL 30% acrylamide–0.8% bisacrylamide stock solution, 12.5 mL 8X Laemmli resolving gel buffer, 37.3 mL H_2O, 20 µL TEMED, 0.2 mL 10% ammonium persulfate; total volume: 100 mL.
7. Protein solubilization: 4X stock solution: 200 m*M* Tris-HCl, pH 6.8, 40% (v/v) glycerol, 4% (v/v) SDS, 0.4% (v/v) bromophenol blue, 100 m*M* dithiothreitol.
8. Gel reservoir buffer: 38 m*M* glycine, 50 m*M* Tris, 0.1% SDS (about 400 mL in each reservoir).
9. Gel staining solution: 10% (v/v) acetic acid, 25% isopropanol, 2.5 g/L Coomassie brilliant blue R250, in water.

10. Gel destaining solution: 7% (v/v) acetic acid, 40% ethanol, in water
11. Mr markers: prestained SDS-polyacrylamide gel electrophoresis (PAGE) markers low range from Bio-Rad or equivalent.
12. Protein transfer medium (for Western blots): Gel reservoir buffer (see above) diluted with ethanol to obtain 20% (v/v) final ethanol concentration. Final concentration: 30.4 m*M* glycine, 40 m*M* Tris-HCl 0.08% SDS (about 800 mL).

2.9. Other

1. UV-visible spectrophotometer (Kontron, Uvikon 810, or equivalent), with 1-cm (disposable, glass or UV silica) cuvet, for enzymatic assays and/or pigment analyses.
2. Nitrocellulose membranes (BA85, Schleicher & Schuell or equivalent), for Western blot.
3. Gel electrophoresis apparatus (Bio Rad Protean 3 or equivalent), with the different sets of accessories (a) for protein separation by electrophoresis (combs, plates, and casting accessories) and (b) for protein transfer on nitrocellulose membranes (central core assembly, holder cassette, nitrocellulose filter paper, fiber pads, cooling unit).

3. Methods

3.1. Growing Cell Suspensions

Arabidopsis thaliana cells (*see* **Note 11**) are cultured in growth rooms using a rotary shaker.

1. Suspend 30 mL of cells from a previous cell culture in fresh culture medium (300 mL final volume) in 500-mL Erlenmeyer flasks under sterile conditions (*see* **Note 12**).
2. Place the Erlenmeyer flasks on their supports and start the rotary shaker. Use the following controlled conditions, i.e., continuous light (40 μmol/m^2/s), 23°C, and 125 rpm.
3. After 5 d of culture (during the exponential phase) collect cells. Filter the cell suspension through a fritted glass filter to harvest the cells.

Six flasks are expected to provide around 150 to 200 g of cells (fresh weight) for plasma membrane preparation (*see* **Note 13**). For more information on cell suspensions, *see also* Chapter 5 in this book.

3.2. Plasma Membrane Preparation (Fig. 1)

3.2.1. Microsome (see **Note 14**) Preparation

All operations are carried out at 0 to 5°C.

1. Collect 5-d-old cells from six Erlenmeyer flasks by filtration on a fritted glass filter.
2. Wash with 2 L washing medium directly through the filtration unit; dry the cells gently under vacuum until they are packed like a "cake." One should get approx 150 to 200 g of cells.
3. Add 375 to 500 mL grinding medium (*see* **Note 15**) to the cells and plasmolysis them for 10 min at 4°C. Then collect the cells on the fritted glass filter. Keep the filtrate; it will be used as a grinding medium for the next steps.
4. Resuspend the cell material in grinding medium (*see* **Note 16**) and homogenize two times 30 s in the 200-mL blender. Distribute the homogenate to 2 × 250 mL polycarbonate bottles (*see* **Note 17**). Centrifuge for 10 min at 1500*g* (Sorvall RC5B, fixed-angle rotor GSA). Distribute the two supernatants equally in 2 × 250 mL polycarbonate bottles. Mix the two pellets and homogenize again in grinding medium (*see* **Note 16**) four times 30 s (*see* **Note 18**). Transfer in the 2 × 250 mL polycarbonate bottles and centrifuge again for 10 min at 1500*g*. Add the two new supernatants to the previous ones in the 2 × 250 ml polycarbonate bottles (*see* **Note 19**). Centrifuge for 10 min at 10,500*g* (Sorvall RC5B, fixed-angle rotor GSA).
5. Distribute the supernatants in 6 × 70 mL polycarbonate tubes. Equilibrate the tubes; centrifuge in a refrigerated ultracentrifuge for 36 min at 35,000*g* (Beckman L8-70M, SW 45 Ti rotor) (*see* **Note 20**). Eliminate the supernatant; gather together the pellets and homogenize the suspension using a 10-mL mortar by adding microsome resuspending buffer (*see* **Note 21**). The suspension constitutes the membrane microsomal fraction.

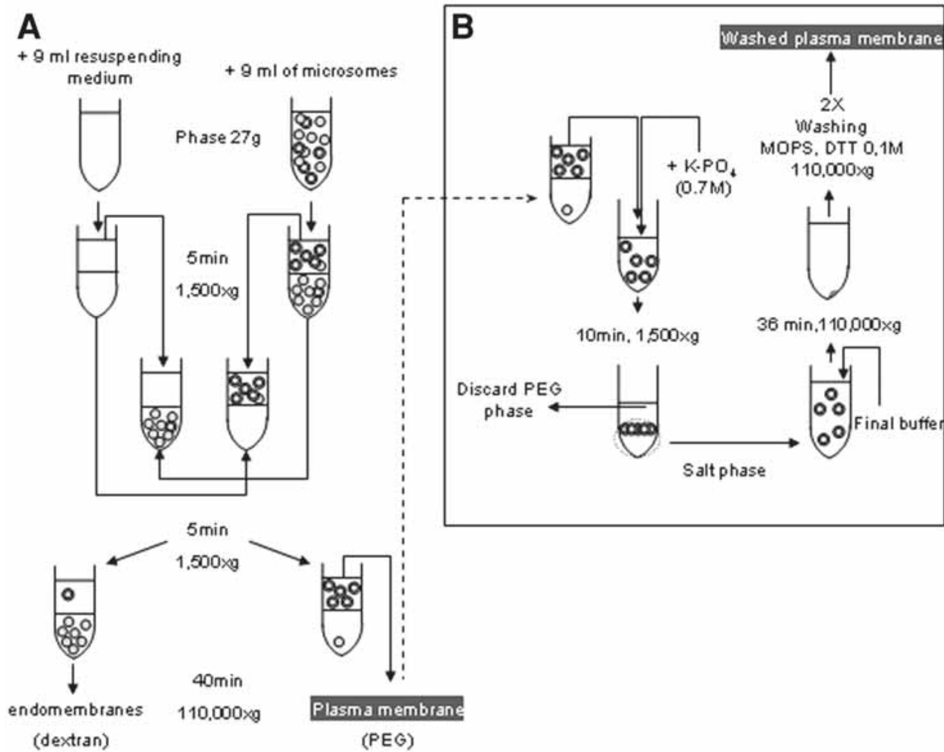


Fig. 1. Purification of *Arabidopsis* plasma membrane by two-phase partitioning. **(A)** Plasma membrane fractions are isolated from Arabidopsis cell suspensions using a PEG-dextran two-phase partitioning method ***(20)***. For *Arabidopsis* cell microsomes, the phases contain 6.4% PEG and Dextran. **(B)** Washing of the plasma membrane fractions prepared by two-phase partitioning. Removal of PEG from the membrane fraction is essential for further solubilization by organic solvents ***(4)***. A mixture containing 1.22 mol K_2HPO_4/mol KH_2PO_4 is added to the upper phase (PEG), obtained from step **A**. A new PEG–K-PO_4 two-phase partition occurs, resulting in the removal of PEG from the plasma membrane ***(4)***.

Microsomes prepared from ~200 g of cells should lead to ~200 mg of microsomal proteins, on average.

3.2.2. Plasma Membrane Purification (see **Note 22***)*

All operations are carried out at 0 to 5°C.

1. Prior to the experiment, take out eight deep-freeze “27 g phases” tubes (*see* **Note 23**).
2. For the preparation of fresh phases, add 9 mL of resuspending microsome medium to 27 g of a PEG-dextran mixture (*see* **Note 24**) and centrifuge at 1500g for 5 min (Jouan E96 or equivalent). Put aside the fresh upper and lower phases in separate tubes (*see* **Note 23**).
3. For the preparation of membrane two-phase partitioning, add 9 mL of the microsomal fraction to 27 g of a PEG-dextran mixture (*see* **Note 24**) and centrifuge at 1500*g*, 5 min (Jouan E96). The upper phase (PEG) is enriched in plasma membrane vesicles and the lower phase (dextran) contains vesicles from all the cell membrane systems (tonoplast, endoplasmic reticulum, Golgi, membranes from broken chloroplasts, and mitochondria) (*see* **Note 25**).
4. Transfer with a disposable plastic pipet the “white” PEG containing the plasma membrane in the fresh lower phase (*see* **Note 24**) 20 and centrifuge at 1500*g*, 5 min (Jouan E96) (*see* **Notes 26**, **27**).

5. To eliminate the PEG adsorbed at the vesicle's surface, PM is washed by a salt solution. Collect the different upper phases in new separate tubes; they represent around 17.5 mL each. Add to each phase crystals of both K_2HPO_4 (43.55 g/L anhydrous) and KH_2PO_4 (68 g/L anhydrous). Mix and centrifuge at 1500*g*, for 5 min (Jouan E96). Repeat these operations a second time to solubilize well the salts. The salt-washed plasma membrane is located at the interface.
6. Remove the upper phase containing the PEG. Mix the lower phase and the PM and pool all these homogenates in one 70 mL polycarbonate centrifuge tube. Complete with plasma membrane washing buffer and centrifuge in a refrigerated ultracentrifuge for 36 min, at 110,000*g* (Beckman L8-70M, SW 45 Ti rotor) (*see* **Note 28**). Eliminate the supernatant; collect the pellet and homogenize in a 2-mL potter by adding plasma membrane resuspending medium (*see* **Note 29**), and store at –80°C.

From about 200 mg of microsomal proteins, around 2 to 4 mg plasma membrane proteins are obtained.

3.3. Chloroplast Envelope Preparation (Fig. 2)

3.3.1. Purification of Arabidopsis *Chloroplasts*

All operations are carried out at 0 to 5°C.

1. Sow seeds at a high density (around 30 mg of seeds for a whole box). Four to six large (30 cm × 45 cm) plastic cases containing 3 to 4 wk-old *Arabidopsis* plantlets are expected to provide 400 to 500 g of rosette leaves (fresh weight) (*see* **Note 30**). For more information on conditions for *Arabidopsis* growth, *see also* Chapter 1 in this book.
2. Prior to the experiment, prepare six tubes containing 30 mL of a 50% Percoll/0.3 *M* sorbitol solution. Preform Percoll gradients for chloroplast purification by centrifugation at 38,700*g* for 55 min (Sorvall SS-34 rotor) (*see* **Note 31**). Store the tubes containing preformed Percoll gradients in a cold room until use.
3. Harvest (*see* **Note 32**) 400 to 500 g of rosette leaves. Wash them with deionized water. Blot the washed leaves on paper tissue, and transfer them in a cold room for the next step.
4. Homogenize the leaf material (400 to 500 g leaves, 2 L of grinding medium) two times for 2 s in a Waring blender at low speed (*see* **Note 33**). Filter the homogenate rapidly through 4 to 5 layers of muslin and one layer of nylon blutex.
5. Distribute the filtered suspension equally into six bottles for centrifugation (500 mL each) and centrifuge them at 2070*g* for 2 min (Sorvall GS 3 rotor) (*see* **Note 34**).
6. Aspirate the supernatant with a water trump and carefully resuspend each pellet, containing a crude chloroplast fraction, by addition of a minimal volume (36 mL, final volume) of washing medium (using a spatula).
7. Load the chloroplast suspension (6 mL per tube) on the top of the preformed Percoll gradients. Centrifuge the gradients at 13,300*g* for 10 min (Sorvall swinging HB-6 rotor) (*see* **Note 35**). At the conclusion of this step, aspirate the upper part of the gradient (*see* **Note 36**), and then recover intact chloroplasts (a broad dark-green band in the lower part of the gradient; *see* **Note 37**) with a pipet.
8. Dilute the intact chloroplast suspension 3 to 4 times with 200 to 300 mL washing medium. Centrifuge the suspension at 2070*g* for 2 min (Sorvall SS-34 rotor).
9. Recover each pellet, containing washed purified intact chloroplasts, for chloroplast envelope preparation.

At this stage, the yield of intact chloroplasts is 50 to 60 mg protein (*see* **Note 38**).

3.3.2. Purification of Envelope Membranes from Arabidopsis *Chloroplasts*

All operations are carried out at 0 to 5°C.

1. Prior to the experiment, prepare six 13.2-mL, tubes (Ultraclear, Beckman) for sucrose gradients containing three layers of 0.93 *M* (3 mL), 0.6 *M* (2.5 mL), and 0.3 *M* (2 mL) sucrose. Each layer is carefully overlaid with a pipet (*see* **Note 39**) on top of the other, starting with the denser one (0.93 *M*, at the bottom) and finishing with the lighter one. Store the tubes in a cold room until use.

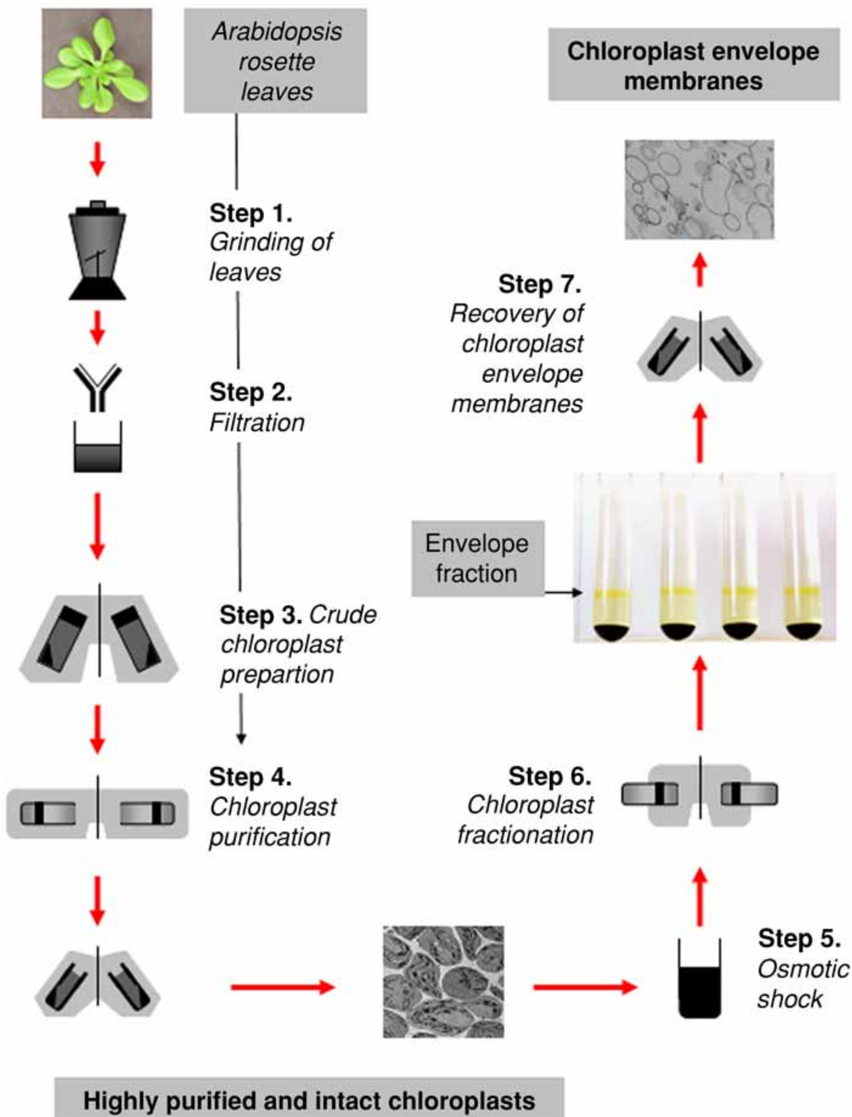


Fig. 2. Purification of envelope membranes from *Arabidopsis* chloroplasts. The first step (left) is the purification, using Percoll gradients, of intact *Arabidopsis* chloroplasts from *Arabidopsis* rosette leaves ***(1)***. Then, the envelope membranes are prepared (right) after a gentle swelling of the chloroplasts in a low osmolarity medium to break the intact organelles. The chloroplast constituents (envelope, stroma, thylakoids) are then separated by centrifugation through a sucrose gradient ***(1)***.

2. Lyse purified and washed intact chloroplasts (obtained as described in **Subheading 3.2.1.**) by adding to the pellets hypotonic medium (adjust for a final volume of 21 mL) containing protease inhibitors (10 m*M* MOPS-NaOH, pH 7.8, 4 m*M* $MgCl_2$, 1 m*M* PMSF, 1 m*M* benzamidine, and 0.5 m*M* ε-amino caproic acid).
3. Load lysed chloroplasts (3.5 mL per tube) on top of the sucrose gradients. Centrifuge the tubes at 70,000*g* for 1 h (Beckman SW41-Ti rotor) (*see* **Note 40**) to purify envelope membranes from the lysate. After centrifugation, they are present as a yellow band at the 0.93/0.6 *M* interface (*see* **Note 41**).

4. Remove carefully the upper part of the gradient by aspiration with a water trump, then recover the yellow band containing the envelope with a pipet, dilute the suspension 3 to 4 times in 10 m*M* MOPS-NaOH pH 7.8 buffer (containing protease inhibitors), and concentrate the membrane as a pellet by centrifugation at 110,000*g* for 1 h (Beckman SW 41 Ti rotor).
5. Add a minimum volume of washing medium (containing protease inhibitors) to the envelope pellet. Take an aliquot for protein determination (*see* **Note 42**). Store envelope membrane preparations in liquid nitrogen.

From such preparations, an average of ~30 mg of stroma proteins, ~20 mg thylakoid proteins and ~300 µg envelope proteins can be obtained (*see* **Note 43**).

3.4. Assessment of Organelle or Membrane Purity (see Note 44)

On a routine basis, three types of markers are used to characterize the different fractions (e.g., organelles, membranes) prepared: enzymatic markers, immunological markers and lipid/pigment markers. Pigment (chlorophyll and carotenoids) are the most conspicuous markers from chloroplast membranes.

3.4.1. Enzymatic Markers

1. ATPase activity, nitrate-sensitive, marker for tonoplast (***15***; *see* **Note 5**), measuring phosphate release resulting of the hydrolysis of ATP at 660 nm in reaction medium ± 50 m*M* KNO_3 and after addition of 23.7 m*M* Tris-ATP and 175 µL of Fiske and Subbarow reagent.
2. ATPase activity, vanadate-sensitive, marker for plasma membrane (***15***; *see* **Note 5**), measuring phosphate release resulting of the hydrolysis of ATP at 660 nm in reaction medium ± 0.6 m*M* Na_3VO_4 and after addition of 23.7 m*M* Tris-ATP and 175 µL of Fiske and Subbarow reagent.
3. Cytochrome c oxidase, marker for the inner mitochondrial membrane (***15***, *see* **Note 5**). Measured following oxidized cytochrome c at 550 nm in 50 m*M* Na_2HPO_4, pH 7.5, after addition of 0.45 m*M* of reduced cytochrome c.
4. Fumarase (EC 4.2.1.2), marker for mitochondrial matrix (*see* **Note 8**), measured according to the method of Hill and Bradshaw ***(16)***, following fumarate synthesis at 240 nm in 50 m*M* Tricine/NaOH, pH 7.5, and after addition of 50 m*M* malate.
5. Hydroxypyruvate reductase (EC 1.1.1.81), marker for peroxisomes (*see* **Note 8**), measured according to the method of Tolbert et al. ***(17)*** following NADH oxidation at 340 nm in 50 m*M* MES-NaOH, pH 6.4, and after addition of 200 m*M* NADH and 1 m*M* hydroxypyruvate.

3.4.2. Immunological Markers

Western blots are performed after separation of membrane proteins by SDS-PAGE (see below for a description of the method). After gel migration, the proteins are transferred to a nitrocellulose membrane using a Gel transfer apparatus (Bio Rad Protean 3 Mini Trans-Blot module or equivalent).

1. Prepare the cassette as follows: add successively 1 fiber pad, 3 nitrocellulose filter papers, the gel, a nitrocellulose membrane (BA85, Schleicher & Schuell or equivalent), 3 nitrocellulose filter papers, 1 fibber pad, and then insert the sandwich in the holder cassette (the membrane should be placed beside the + electrode).
2. Insert the cassette in the central core assembly unit (together with the cooling unit).
3. Perform the transfer for 2 h at 80 V in protein transfer medium.
4. Recover the nitrocellulose membrane.
5. Follow the instructions for saturation and incubation of the membrane with primary and secondary antibodies (*see* **Note 45**) provided by the manufacturers.

3.4.3. Lipids and Pigments

3.4.3.1. Determination of Chlorophyll Content (*see* Note 46) of a Fraction

Add 10 µL of the extract to be analyzed to 1 mL 80% (v/v) acetone in a 1-mL Eppendorf tube. Vortex and incubate for 15 min on ice and in the dark. Centrifuge for 15 min at 16,000*g*.

Pour into a 1-mL spectrophotometer glass cuvet. Measure the absorbance at 652 nm against a tube containing 80% (v/v) acetone for the zero. A ratio of OD652/36 = 1 corresponds to 1 mg chlorophyll/mL.

3.4.3.2. Lipid (and Pigment) Extraction and Analyses

Lipid and Pigment Extraction (adapted from ref. ***18***).

1. In order to form one liquid phase and subsequently extract the lipid, mix 200 µL of membrane suspension with 750 µL of a methanol/chloroform (2:1, v/v) mixture. Homogenize with a vortex, then add 250 µL water and 250 µL chloroform. Homogenize with a vortex.
2. Centrifuge the mixture for 10 min at 14,000*g* in order to get a two-phase system. Discard the upper phase with a pipet.
3. Remove the lower phase (*see* **Note 47**) by aspiration with a Pasteur pipet. Dry it under a stream of argon (or nitrogen). The residue is dissolved in a minimal volume of chloroform or 80% acetone.

TLC Lipid Analyses:

1. Prior to the experiment, heat a silica gel precoated TLC plate in an oven at 110°C for at least 1 h.
2. Load, with the tip of a Pasteur pipet, the chloroform lipid extract (prepared as in **step 1**) onto the silica gel precoated TLC plate (Merck) (*see* **Note 9**).
3. After spotting, place the plate in a rectangular chromatography chamber containing 100 to 150 mL of the chloroform/methanol/water (65:25:4, v/v/v) mixture for the first dimension of the chromatography. Allow the solvent system to run to the front of the plate. Remove the plate and place it in a glass chamber under a slow stream of argon (or nitrogen) for 1 to 2 h.
4. Turn the plate in order to have the lipids at the bottom; place it in a rectangular chromatography chamber containing 100 to 150 mL of the following mixture: chloroform/acetone/methanol/acetic acid/water (50:20:10:10:5, v/v) for the second dimension of the chromatography (perpendicular to the first dimension). Remove the plate when the solvent reaches the top of the plate, and place it in a glass chamber under a slow stream of argon (or nitrogen) until dry or staining.
5. Locate lipid spots by spraying with anilinonaphthalene sulfonate (0.2% in NaOH), and viewing under UV light. Detect glycolipids (*see* **Note 48**) by using α-naphthol, and phospholipids with a molybdate reagent.

When necessary, quantification of the amount of marker lipids in each sample is done by gas chromatography (for details, *see* refs. ***21,22***).

Pigment Analyses:

1. Dissolve the lipid extract (prepared as in **Subheading 3.3.3.2.**) in 80% acetone (1 mL, final volume). Pour the solution into a 1-mL spectrophotometer cuvet.
2. Record the absorption spectrum between 350 and 750 nm. Carotenoids are responsible for a series of peaks in the 400 to 500 nm region of the spectrum, whereas chlorophylls show in addition a sharp peak with a maximum in the 650 to 700 nm region (*see* **Note 49**).

3.5. Differential Extraction of Membrane Proteins (**see Note 50*****)***

3.5.1. Protein Solubilization With Detergents

1. Dilute the membrane proteins (0.2 mg) in 0.2 mL of solubilization solution (*see* **Note 51**).
2. After 30 min incubation on ice, centrifuge the mixture for 15 min (4°C) at 15,000*g* in 1.5 mL microfuge tubes to separate two fractions: the supernatant containing proteins solubilized by the treatment and the pellet containing the insoluble proteins.
3. Solubilize the insoluble protein pellets in 50 µL of 50 m*M* MOPS-NaOH, pH 7.8, 1 m*M* DTT.
4. Analyze the proteins by SDS-PAGE (*see* below).

*3.5.2. Membrane Protein Solubilization with Chloroform/Methanol Mixtures (*see **Note 52***)*

1. Dilute slowly one volume of the membrane preparation (0.5 to 1 mg protein in 0.1 mL of original buffer) (*see* **Note 53**) in 9 volumes of cold chloroform/methanol (5/4, v/v) mixture in 1.5 mL Eppendorf tubes (*see* **Note 54**).

2. Store the resulting mixtures for 15 min on ice before centrifugation 4°C for 15 min at 15,000*g* (Eppendorf).
3. Recover the organic phase (the white pellet containing fewer hydrophobic proteins is discarded). The pellet contains the chloroform/methanol-insoluble proteins (or organic solvent insoluble fraction). The supernatant contains the chloroform/methanol-soluble proteins (or organic solvent soluble fraction).
4. Evaporate (*see* **Note 55**) the organic phase under nitrogen (to 200 μL for large amounts of proteins or 100 μL when original protein concentration is limited). Directly precipitate the proteins by adding 4 vol (800 μL or 400 μL) of cold (–20°C) acetone (80% final acetone concentration) directly to the remaining volume of chloroform/methanol.
5. Store the resulting mixtures for 15 min on ice before centrifugation at 4°C for 15 min at 15,000*g* (Eppendorf).
6. Eliminate the organic supernatant; dry the protein pellet (*see* **Note 56**) on the bench and not under nitrogen. Be sure that there is no more acetone (*see* **Note 57**). Resuspend (*see* **Note 58**) the protein pellets in 20 μL of concentrated (4X) SDS/PAGE buffer and store the protein mixtures in liquid nitrogen.
7. Analyze the proteins by SDS-PAGE (various volumes on separate lanes).

3.5.3. Alkaline or Salt Washing of Membrane Fractions

1. Dilute slowly 1 vol of the membrane preparation (0.5 to 1 mg protein in 0.1 mL) to 0.5 mL with Na_2CO_3, NaOH, or NaCl stock solutions to obtain 0.1, 0.5, or 1 *M* final concentrations, respectively (*see* **Note 59**).
2. Sonicate the resulting mixtures 2 to 5 times 10 s, the power set at 40% duty cycle, output control 5, in ice.
3. Store the mixtures for 15 min on ice before centrifugation at 4°C for 20 min at 15,000*g* in 1.5- mL microfuge tubes.
4. Insoluble proteins are recovered as pellets (*see* **Note 60**). Resuspend them in 20 μL of (4X) SDS/PAGE buffer. Store the protein extracts in liquid nitrogen.
5. Analyze the proteins by SDS-PAGE (*see* below).

3.6. Separation of Membrane Proteins by 1D SDS-PAGE (see Note 61)

1. Prior to the experiment, prepare slab gels for protein electrophoresis (*see* **Note 62**):
   a. Prepare the gel apparatus according to the manufacturer's specifications (*see* **Note 63**).
   b. Prepare the different gel solutions (stacking gel, 10, 12, or 15% separation gel). The volumes to be used are determined by gel dimensions, and therefore by the specifications of the apparatus.
2. Heat the protein samples at 95°C for 5 min to solubilize the proteins. Add bromophenol blue dye in the samples. Place 20 μL protein samples into gel slots by means of a pipet. Mr markers are placed in another slot.
3. Set the conditions for the electrophoresis at 150 V. Run gels for 1 h at room temperature (until the bromophenol blue dye reaches the lower part of the gel) (*see* **Note 64**).
4. After electrophoresis, remove the gels; place them in plastic boxes in the presence of staining solutions. Shake the box gently for 30 min. Pour off the staining solution and replace it by destaining solution. Shake the box gently for 15 min. Repeat the washing step once or twice.
5. Perform in gel protein digestion for proteomic analyses (*see* **Note 65**), *see* Chapter 34 in this book.

4. Notes

1. All buffers stored in a deep-freezer should be thawed the day before use and stored in the cold room the night before.
2. Dextran stock solution: 20% (w/v) dextran T-500. Aqueous solutions of dextran T-500 are not easy to manipulate and it is almost impossible to prepare the exact 20% concentration. The accurate concentration of the stock solution is measured by a polarographic method. Prepare a stock solution of dextran T-500 in water around 25% (w/v). Prepare two dilutions (4 and 5 times) of the stock solution in water (w/w). Measure the optical radiation at 589 nm of the two

samples at room temperature (23°C). From the optical variation value of each sample, calculate the percentage in Dextran using the parameters specific for each Dextran batch as indicated by the manufacturer.

3. "27 g PEG-dextran phases": All the quantities and concentrations are calculated for final 36-g phases. All the stock solutions necessary for that mixture are kept at –20°C; defreeze just before use. Calculate the weights of PEG 3350 (40%) and dextran T-500 (*x*% determined by polarography) in order to obtain a 6.4% mixture in the final 36-g phases. Directly on the balance, using disposable plastic transfer pipet, fill the 50 mL tubes with the appropriate weights of PEG 3350 and dextran T-500; add potassium salts, sucrose, and water up to 27 g. Mix the tubes vigorously and place the tubes in ice as soon as they are ready. A series of "27 g phase" tubes can be prepared and stored in a deep freeze for further experiments.
4. BSA (Fraction V) from Sigma (St. Louis, MO) can be obtained fatty acid-free (Ref. A-4503). The preparation of defatted BSA from normal BSA can be achieved by successive washing of 100 g BSA by 1 L of ethanol and 1 L of acetone; the defatted BSA is then recovered by filtration on a Buchner funnel.
5. Marmagne et al. (5) have estimated the activity of three marker enzymes in plasma membrane preparation compared with crude microsomal fraction. The ratio between the activity of the nitrate-sensitive ATPase in the plasma membrane fraction and the activity of the microsomal fraction was 0.9 ± 0.1. The ratio between the activity of the vanadate-sensitive ATPase in the plasma membrane fraction and the activity of the microsomal fraction was 3.7 ± 0.3. The ratio of the vanadate-sensitive ATPase activity in the plasma membrane to the total ATPase activity in the same fraction was 91.6. This suggests that the plasma membrane contamination by tonoplasts and other membranes displaying ATPase activities (endoplasmic reticulum, mitochondria, chloroplasts) was at most 8%, a likely overestimation considering the relative lack of specificity of the inhibitors used in the enzymatic assays. The ratio between the activity of cytochrome c oxidase in the plasma membrane fraction and the activity of the microsomal fraction was $2.9 \times 10^{-3} \pm 2\ 10^{-∞3}$, corresponding to an extremely low level of mitochondrial contamination in the plasma membrane fraction.
6. Reaction media ± inhibitor can be stored at –20°C.
7. Mix solutions A and B just before use in the ratio 2:1 (v/v).
8. The activity of the mitochondrial marker fumarase was not detected in purified chloroplast from *Arabidopsis* leaves (no detected cross-contamination). The specific activity of the mitochondrial marker fumarase in *Arabidopsis* mitochondria purified from cell cultures ***(4)*** is around 400 nmol/min/mg proteins. This represents an enrichment of 10 when compared to the specific activity of the same marker in crude protoplast extract (about 40 nmol min-1 mg-1 proteins). The specific activity of the peroxisomal marker hydroxypyruvate reductase in chloroplasts purified from *Arabidopsis* leaves was 6.8 nmol/min/mg proteins for an original specific activity of 256 nmol/min/mg proteins in crude leaf extract (2–3% cross-contamination). By comparison, the specific activity of hydroxypyruvate reductase in mitochondria purified from *Arabidopsis* cell cultures was 67 nmol min-1 mg-1 proteins for an original specific activity of 103 nmol/min/mg proteins in crude protoplast extract deriving from *Arabidopsis* cell cultures ***(4)***.
9. The lipids dissolved in chloroform are applied at the lower right-hand corner of the plate, about 3 cm from each edge. This is done under a stream of nitrogen to accelerate solvent evaporation. One hundred to 500 µg lipids allows the detection of lipid contaminants.
10. A wide variety of detergents can be used: Triton X-100, CHAPS, Triton X-114, etc. ***(19)***.
11. *Arabidopsis* Columbia (Col) or Wassilewskija (WS) ecotypes have been generally used for ESTs or genomic sequencing projects and also to generate T-DNA insertion mutants using *Agrobacterium tumefaciens*. For these reasons, they might also be considered as reference ecotypes for *Arabidopsis* proteomic analyses. We used both *ecotypes*, the whole procedure for plasma membrane preparation was set up using Col ecotype whereas the WS ecotype was used for chloroplast preparation. The procedures we describe in this article are likely to be used with any ecotype.
12. A hood and sterile materials and medium are necessary (*see* **Subheading 2.**).
13. Cultured *Arabidopsis* cells can be used for preparation of many different types of membranes and organelles. For instance, Brugière et al. ***(4)*** purified mitochondria from isolated *Arabidopsis* cell suspensions.

14. Microsomes are an heterogenous population of vesicles derived from various cell membranes and organelles.
15. The volume of grinding medium added to the cell cake corresponds to about 2.5 times the amount of cells.
16. Add the grinding medium to the cell cake until getting a mixture that is neither too thick nor too liquid.
17. Rinse the blender with the grinding buffer. The total volume (homogenate plus washing) should not exceed 80 to 100 mL/bottle.
18. A second grinding of the pellet increases the yield of cell breakage and therefore the amount of material released in the medium. However, further grinding should be avoided, as the purity of any membrane preparation is always higher when low-yield procedures are favored.
19. There is no filtration in this process: Separation of broken cells from the cell extract is achieved by centrifugation.
20. It is essential to equilibrate each tube one by one on a balance (with the corresponding cap) prior to centrifugation. Equilibration is performed by adding fresh grinding medium until reaching a common weight for each.
21. The volume of microsome resuspending buffer to be added to microsomal pellets should be a multiple of 9 (for instance, 27 mL or 36 mL): it will be added to a 27 g PEG–dextran mixture.
22. This two-phase partitioning method quickly produces highly purified plasma membranes from *Arabidopsis* cells, with a reasonable yield. Contamination of plasma membranes with other cellular membranes, such as mitochondria, plastid and nuclear membranes, tonoplasts, endoplasmic reticulum, lysosomes, and Golgi, is reduced considerably when using the two-phase partition method, as compared with other methods ***(20)***.
23. If you have less than 150 g of cells, use six “27 g phase” tubes. Half of them will be used for preparing fresh phases, the other half for membrane two-phase partitioning. Both operations are done at the same time.
24. The tube is gently mixed with an up-and-down movement until the mixture gets homogenous color and aspect.
25. Note that the two phases are easy to distinguish as they display different aspects and colors, milky-white for upper phase and thick and pale green for lower phase. If this is not the case, the “27 g phases” were likely not cold enough.
26. The purification step can be repeated once or twice for improving the purification of the plasma membrane. The number of fresh phases will be prepared according the number of purification steps. At each purification round, the yield of plasma membrane decreases whereas the purity of the preparation increases. The quality of purified plasma membrane fractions was assessed by enzymatic and immunological assays (see below), and by mass spectrometry analyses ***(5)***.
27. Plasma membrane contaminating endomembranes in the “green” lower phase (Dextran phase) can be removed by the same procedure and the resulting upper phase is treated like the others.
28. The lower phase (dextran with endomembranes) can also be washed with the same buffer during the same centrifugation. The endomembrane fraction is usually used to estimate the quality of the plasma membrane preparation (purity and enrichment) by comparison with plasma membrane and microsomal fractions.
29. To concentrate the plasma membrane suspension, use a buffer volume as small as possible (usually around 1 mL).
30. Leaves are a much better material than isolated cells for preparing *Arabidopsis* chloroplasts.
31. Vertical rotors can easily be used to preform Percoll gradients and subsequently purify chloroplasts ***(21)***.
32. The number of starch granules present in chloroplasts is critical for the preparation of intact chloroplasts; chloroplasts containing large starch grain will generally be broken during centrifugation ***(21)***. Therefore, prior to the experiment, the plants should be kept in a dark and cold room (4°C) to reduce the amount of starch. A good way to proceed is to place the plants under such conditions the day before the extraction (we usually do this at the beginning of the afternoon prior to the day of the experiment).
33. It is critical the grinding process be as short as possible. Longer blending improves the yield of recovered chlorophyll, but increases the proportion of broken chloroplasts.

34. It is essential to equilibrate the different tubes two by two on a balance prior to centrifugation.
35. It is recommended to disconnect the brake or to use the automatic rate controller (if available) to prevent mixing of the gradients at the critical stage of deceleration.
36. It is important to carefully remove the top of the tube by aspiration with a water trump, then to recover the intact chloroplasts with a pipet.
37. Broken chloroplasts are present in the upper part as a broad band. A small pellet containing cell pieces and large debris is found at the bottom of the tube.
38. The excellent purity of the *Arabidopsis* chloroplasts prepared by Percoll purification step was confirmed through proteomic analysis: only 5% (6 of 112) of the *Arabidopsis* proteins identified by Ferro et al. ***(2)*** may correspond to nonplastid proteins. Among them, one protein appeared to correspond to a previously characterized major plasma membrane component; four proteins may indicate contamination by major tonoplast proteins and one from glyoxysomes. Considering the high sensitivity of current mass spectrometers it is not surprising to detect minute amounts of these few extra-plastidial contaminants, which are major proteins in their respective subcellular compartment. It is also important to notice that none of the proteins identified in *Arabidopsis* envelope membranes by Ferro et al. ***(2)*** appears to derive from mitochondria, a classical contaminant of plastid preparations.
39. The use of a peristaltic pump to prepare the gradients is recommended, as some expertise is needed to load the different layers by hand.
40. It is recommended to disconnect the brake to prevent mixing of the gradients at the critical stage of deceleration.
41. Thylakoids are concentrated as a dark-green pellet at the bottom of the tube, whereas the soluble fraction containing the stroma remains on top.
42. Protein contents of membrane fractions are estimated using the Bio-Rad protein assay reagent ***(22)***.
43. The use of Percoll-purified chloroplasts is very efficient to limit contamination of envelope membranes by extraplastidial membranes, as demonstrated by the absence of phosphatidylethanolamine and of different marker enzymes or proteins ***(21)***. Therefore, at this stage, the major possible contaminants of envelope preparations are soluble stroma proteins and small pieces of thylakoid membranes. Such cross-contamination has been extensively analyzed by Ferro et al. ***(2)***. Being the most likely source of membrane contamination of the purified envelope fraction, thylakoid cross-contamination needs to be precisely assessed. The yellow color of purified envelope vesicles first indicates that this membrane system contains almost no chlorophyll and therefore very few contaminating thylakoids. Indeed, by Western blot analyses using antibodies raised against light-harvesting, chlorophyll protein (LHCP), Ferro et al. ***(2)*** demonstrated that several independent *Arabidopsis* envelope preparations appeared to contain between 1 and 3% thylakoid proteins.
44. A thorough study of membrane purity is essential for a precise determination of the subcellular localization of the proteins of interest. An example of a protein previously expected to be located in the plasma membrane, but actually residing to the inner envelope membrane, is given by Ferro et al. ***(1)***.
45. Several dilutions of the primary antibodies should be tested to identify the best signal-to-noise ratio.
46. The chlorophyll content was 170 mg per mg protein in chloroplasts purified from *Arabidopsis* leaves and 84 mg per mg protein in crude leaf extract (enrichment of 2). By comparison, chlorophyll concentration in crude protoplast extract is about 4.5 mg chlorophyll per mg protein ***(4)***.
47. The chloroformic (lower) phase contains lipids and pigments.
48. Galactolipids are markers for plastid membranes, whereas phospholipids such as phosphatidylethanolamine are markers for extraplastidial membranes ***(21)***. However, in some very specific conditions, such as phosphate deprivation, phosphate present in phospholipids is mobilized and nonphosphorus membrane lipids, such as digalactosyldiacylglycerol (DGDG), increase in nonplastidial membranes (*see,* e.g., ***23***).
49. When correctly prepared, chloroplast envelope membranes do not contain chlorophylls, but only carotenoids. Plasma membranes, when highly purified, are expected to contain no trace of chlorophyll or carotenoids.

50. Because of the high functional value of a precise subcellular localization, we therefore focus in this article on the proteins that are the most tightly associated with the membranes. Therefore, in all cases, we analyze fractions containing the most hydrophobic proteins, i.e., the chloroform/methanol soluble proteins or the proteins remaining in the membrane after its treatment by NaOH. The discarded fractions contain a large variety of rather hydrophilic proteins, some of high interest. However, because many of them are also present in the cytosol, or in the chloroplast stroma, or any soluble extract from plant tissues, their subcellular localization cannot be precisely determined. They are of strong interest in several cases—for instance, for analyses of the protein content of the thylakoid lumen ***(24–26)***.
51. A wide variety of detergents can be used: Triton X-100, CHAPS, Triton X-114, sulfobetains, and the like. The reader is referred to articles by Santoni et al. ***(19,27)*** for detailed analyses of membrane treatment by detergents.
52. First, be sure that the membrane preparation does not contain too many hydrophilic proteins deriving from contamination of the membrane fraction with soluble compartments (this protocol is not to be used on a crude cell extract, for example, but can be used on a crude membrane extract). Hydrophilic proteins will precipitate during the process. A too-large amount of these hydrophilic proteins would coprecipitate your hydrophobic proteins.
53. Most of the time, we use 10 m*M* MOPS, pH 7.0, as a buffer.
54. The volume ratio between chloroform and methanol for an optimal extraction can be determined by comparing the polypeptide profile of the organic phase soluble proteins prepared as follows: membranes (5 mg proteins in 1 mL storage buffer) are divided into 10 fractions of 0.1 mL (in 1.5-mL Eppendorf tubes). The membrane fraction is then slowly diluted by addition of 0.9 ml of cold chloroform/methanol solutions (0:9, 1:8, 2:7, 3:6, 4:5, 5:4, 6:3, 7:2, 8:1, and 9:0, v/v). In general, the total volume of the mixture is 1 mL. If necessary, this can be increased to a much higher value when more membrane material is available.
55. Do not completely dry the sample.
56. Due to acetone precipitation (and removal of pigments), the pellet turns white; be careful in order not to lose it.
57. Trace amounts of solvent strongly limit protein resuspension.
58. Be patient; wash tube walls and avoid bubbles.
59. Treatment of membranes with these various compounds does not result in the extraction of the same proteins ***(2,4,5)***. Na_2CO_3 or NaCl extract proteins that are rather weakly associated with the membrane, whereas NaOH removes proteins that are more tightly associated. It is therefore recommended to try several of these compounds to achieve more comprehensive analyses.
60. The supernatant contains the proteins removed from the membrane by alkaline or salt treatment, i.e., the less hydrophobic membrane proteins.
61. Classical proteomic methodologies based on the use of 2D gel electrophoresis proved to be rather inefficient on membrane proteins. In general, almost no highly hydrophobic proteins, as defined by average hydrophobicity values, are found on 2D gel separations of membrane proteins. Adessi et al. ***(28)*** observed that loading 2D gels with high amounts of membrane proteins, resulted in the severe loss of hydrophobic proteins and therefore in the artifactual enrichment of the less hydrophobic components (and hydrophilic contaminants of the purified membrane fraction). In this case, hydrophobic proteins probably precipitated at their isoelectric point in the first dimension, thus preventing any further migration and separation in the second dimension ***(28)***. In contrast, 2D gel electrophoresis is very efficient to analyze peripheral membrane proteomes (*see* ***24–26,29***). Strategies for membrane proteomics based on 2D-electrophoresis combined with a wide diversity of detergents have been extensively analyzed by Santoni et al. ***(19,27)***.
62. We use routinely the procedure described by Chua ***(30)*** to separate membrane proteins by SDS-PAGE. This article describes in detail all stock solutions, media for stacking, and separation gels.
63. We used a Bio-Rad apparatus, with 7-cm-long gels.
64. For some analyses, protein migration can be stopped just between the stacking and the separating gels so that proteins are concentrated in a very thin band for further nanoLC-MS/MS analyses.

65. Gel pieces can be stored in the cold room until proteomic analyses. Different methods for mass spectrometry (MALDI-ToF, nanoLC-MS/MS, etc.) are described in Chapter 34 of this book. Readers are also referred to the description by Ferro et al. *(3)* for detailed conditions for in gel digestion and further MS analyses.

References

1. Ferro, M., Salvi, D., Rivière-Rolland, H., et al. (2002) Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. *Proc. Natl. Acad. Sci. USA* **99,** 11,487–11,492.
2. Ferro, M., Salvi, D., Brugière, S., et al. (2003) Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. *Mol. Cell Proteomics* **2,** 325–345.
3. Ferro, M., Seigneurin-Berny, D. Rolland, N., Chapel, A., Salvi, D., Garin, J., and Joyard, J. (2002) Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. *Electrophoresis* **21,** 3517–3526.
4. Brugière, S., Kowalski, S., Ferro, M., et al. (2004) The hydrophobic proteome of mitochondrial membranes from *Arabidopsis* cell suspensions. *Phytochemistry*, in press
5. Marmagne, A., Rouet, M. A., Ferro, M., et al. (2004) Identification of new intrinsic proteins in Arabidopsis plasma membrane proteome. *Mol. Cell. Proteomics*, in press (published online).
6. Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422,** 198–207.
7. Morsomme, P., Dambly, S., Maudoux, O., and Boutry, M. (1998) Single point mutations distributed in 10 soluble and membrane regions of the *Nicotiana plumbaginifolia* plasma membrane PMA2 H^+-ATPase activate the enzyme and modify the structure of the C-terminal region. *J. Biol. Chem.* **273,** 34,837–34,842.
8. Gerbeau, P., Guclu, J., Ripoche, P., and Maurel, C. (1999) Aquaporin Nt-TIPa can account for the high permeability of tobacco cell vacuolar membrane to small neutral solutes. *Plant J.* **18,** 577–587.
9. Combettes, B. and Grienenberger, J. M. (1999) Analysis of wheat mitochondrial complex I purified by a one-step immunoaffinity chromatography. *Biochimie* **81,** 645–653.
10. Perryman, R. A., Mooney, B., and Harmey, M. A. (1995) Identification of a 42-kDa plant mitochondrial outer membrane protein, MOM42, involved in the import of precursor proteins into plant mitochondria. *Arch. Biochem. Biophys.* **316,** 659–664.
11. Vauclare, P., Macherel, D., Douce, R., and Bourguignon, J. (1998) The gene encoding T protein of the glycine decarboxylase complex involved in the mitochondrial step of the photorespiratory pathway in plants exhibits features of light induced genes. *Plant Mol. Biol.* **37,** 309–318.
12. Joyard, J., Grossman, A. R., Bartlett, S. G., Douce, R., and Chua, N.-H. (1982) Characterization of envelope membrane polypeptides from spinach chloroplasts. *J. Biol. Chem.* **257,** 1095–1101.
13. Miras, S., Salvi, D., Ferro, M., Grunwald, D., Garin, J., Joyard, J., and Rolland, N. (2002) Non-canonical transit peptide for import into the chloroplast. *J Biol. Chem.* **277,** 47,770–47,778.
14. Vallon, O., Wollman, F. A., and Olive, J. (1986) Lateral distribution of the main protein complexes of the photosynthetic apparatus in *Chlamydomonas reinhardtii* and in spinach: an immunocytochemical study using intact thylakoid membranes and a PSII enriched membrane. *Photobiochem. Photobiophys.* **12,** 203–220.
15. Briskin, D. P., Leonard, R. T., and Hodges, T. K. (1987) Isolation of the plasma membrane: membrane markers and general principles. *Methods Enzymol.* **142,** 542–558.
16. Hill, R. L. and Bradshaw, R. A. (1969) Fumarase. *Methods Enzymol.* **13,** 91–99.
17. Tolbert, N. E., Yamazaki, R. K., and Oeser, A. (1970) Localization and properties of hydroxypyruvate and glyoxylate reducteses. *J. Biol. Chem.* **245,** 5129–5136.
18. Bligh, E. G. and Dyer, W. J. (1959) A rapid method of total lipid extraction and purification. *Can. J. Med. Sci.* **37,** 911–917.
19. Santoni, V., Kieffer, S., Desclaux, D., Masson, F., and Rabilloud, T. (2000) Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis* **21,** 3329–3344.
20. Larsson, C., Widell, S., and Kjellbom, P. (1987) Preparation of high-purity plasma membranes in, *Methods in Enzymology* (Packer, L. and Douce, R., eds.), vol 148, Academic Press, London, pp. 558–568.
21. Douce, R. and Joyard, J. (1982) Purification of the chloroplast envelope, in *Methods in Chloroplast Molecular Biology* (Edelman, M., Hallick, R., and Chua, N.-H., eds.), Elsevier/North-Holland, Amsterdam, pp. 139–256.

22. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72,** 248–254.
23. Jouhet, J., Maréchal, E., Bligny, R., Joyard, J., and Block, M. A. (2003) Transient increase of phosphatidylcholine in plant cells in response to phosphate deprivation. *FEBS Lett*. **544,** 63–68.
24. Peltier, J. B, Emanuelsson, O., Kalume, D. E., et al. (2002) Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. *Plant Cell* **14,** 211–236.
25. Peltier, J. B., Friso, G., Kalume, D. E., Roepstorff, P., Nilsson, F., Adamska, I., and van Wijk, K. J. (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell* **12,** 319–341.
26. Schubert, M., Petersson, U. A., Haas, B. J., Funk, C., Schroder, W. P., and Kieselbach T. (2002) Proteome map of the chloroplast lumen of *Arabidopsis thaliana. J. Biol. Chem.* **277,** 8354–8365.
27. Santoni, V., Molloy, M., and Rabilloud, T. (2000) Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21,** 1054–1070.
28. Adessi, C., Miège, C., Albrieux, C., and Rabilloud, T. (1997) Two-dimensional electrophoresis of membrane proteins: a current challenge for immobilized pH gradients. *Electrophoresis* **18,** 127–135.
29. Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B., Rudella, A., Sun, Q., and Wijk, K. J. (2004) In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell* **16,** 478–499.
30. Chua, N.-H. (1980) Electrophoretic analysis of chloroplast proteins. *Methods Enzymol*. **69,** 434–436.

34

Protein Microsequencing

Alberto Paradela and Juan Pablo Albar

Summary

Large-scale and high-throughput approaches play an essential role in the study of the highly complex biological systems. In plant biology, obtaining the complete genome sequence of the model plant *Arabidopsis thaliana* has definitively encouraged the incorporation of a completely new set of tools for the identification, characterization, and quantification of its proteome. Proteomics is a new branch of the protein sciences that makes use of these tools and the set of experimental methods specifically designed for them. In this chapter, we describe a complete method for the isolation, fractionation, and identification of a set of proteins obtained from *Arabidopsis thaliana* cells, stressing methodological aspects that could be of potential interest.

1. Introduction

The systematic analysis and documentation of all protein species of an organism or a specific tissue type is termed *proteomics* ***(1–2)***. Improvements in mass spectrometry, as well as the development of new bioinformatic tools and the sequencing of the entire *Arabidopsis thaliana* genome, among other factors, have led to a dramatic increase in the characterization of its proteome. This new era has allowed the study of the proteome in terms not only of presence/absence and intracellular distribution of proteins, but also the interactions, posttranslational modifications, and their quantitative changes in response to environmental alterations ***(3)***.

In studying the proteome of a given organism, the choice of a specific method is of key importance to obtain a set of reliable results. The characteristics of the sample to be analyzed (for example, soluble or membrane proteins) and the goal of the analysis are the main parameters to be considered. These parameters will also determine the method of protein separation/fractionation needed prior to the analysis. The growing number of approaches with which to study the *A. thaliana* proteome responds to an increasing number of specific problems to be resolved ***(4–10)***.

Since its introduction in 1975 ***(11)***, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) has been the most popular method for separation and characterization of complex protein mixtures, owing to its unique resolving capacity. This technique sorts proteins according to two different, independent properties. The first-dimensional step, isoelectric focusing (IEF), separates proteins according to isoelectric points; the second-dimensional step, sodium dodecyl sulfate-PAGE (SDS-PAGE), separates proteins according to molecular weight (*see* Chapter 31 in this book). The number of proteins that can be detected as individual spots in a single analysis ranges from a few hundred to 5000 to 10,000. This method, however, is not suitable for certain sets of proteins, such as membrane proteins, very low molecular mass proteins, and very acidic and basic proteins ***(12)***.

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

For protein sequencing, spots are isolated and digested, usually using trypsin. The peptide mixture generated is then analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF-MS) to obtain a "fingerprint." Protein identification is performed using specific algorithms based on correlating measured peptide masses with experimentally calculated peptide masses derived from proteins in existing sequence databases ***(13–16)***. The use of ion-trap mass spectrometers is highly recommended when fingerprinting is not informative enough to identify the protein of interest, or in cases in which detailed characterization (i.e., posttranslational modifications) of specific peptides is required. Ion-trap spectrometers also permit *de novo* sequencing, through manual and software-aided interpretation of collision-induced dissociation (CID) spectra.

A common alternative proteomic approach for the characterization and identification of proteins in complex mixtures is the multidimensional protein identification technique (MudPIT) ***(17)***. In this case, samples are trypsin-digested and the peptide mixture obtained is fractionated using strong cationic exchange SCX and reversed-phase columns, coupled to tandem mass spectrometry (MALDI-ToF/ToF or ESI-MS/MS). Protein identifications are obtained through the analysis of independent peptide fragmentation spectra.

This chapter offers a detailed description of a method for the extraction and 2D-PAGE separation of a complex protein mixture from *A. thaliana* leaves, their isolation, in-gel tryptic digestion, and, finally, identification of individual spots by two different mass spectrometry (MS) approaches. The method should be considered a general (and useful) approach to the *A. thaliana* proteome. Readers are invited to find specific methods more adapted to other purposes elsewhere (***3***, ***4–10***, ***18–19***; *see also* Chapters 31 and 33 in this book).

2. Materials

2.1. Sample Preparation

1. Extraction buffer: 62.5 m*M* Tris-HCl, pH 7.5, 1 m*M* DTT, 1 m*M* PMSF, protease inhibitor cocktail (Sigma-Aldrich).

2.2. Sample Precipitation

1. Precipitation solution: 50% (v/v) trichloroacetic acid (TCA) in water.
2. Washing solution: 80% v/v acetone in water.

2.3. IEF Electrophoresis

1. Sample solution: 8.4 *M* urea, 2.4 *M* thiourea, 5% (v/v) CHAPS (3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate), 2 m*M* TCEP (Tris-carboxy-ethyl-phosphine)-HCl
2. Rehydration buffer: 7 *M* urea, 2 *M* thiourea, 4% (v/v) CHAPS, 50 m*M* DTT (add prior to use), 0.002% (w/v) bromophenol blue.
   Store in aliquots at –20 °C.

2.4. 2D Electrophoresis

1. Equilibration buffer: 50 m*M* Tris-HCl, pH 8.8, 6 *M* urea, 30% (v/v) glycerol, 2% (w/v) SDS, 0.002% (w/v) bromophenol blue. Store in aliquots at –20°C.

2.5. Silver Staining

1. Fixing solution: Ethanol/glacial acetic acid/water 40:10:50.
2. Sensitizing solution: mix 75 mL ethanol, 10 mL sodium thiosulfate (5% w/v) and 17 g sodium acetate. Make up to 250 mL with distilled water.
3. Silver solution: 25 mL silver nitrate solution (2.5% w/v). Make up to 250 mL with distilled water.
4. Developing solution: Add 6.25 g sodium carbonate and 100 µL formaldehyde (37% w/v) (add just prior to use), to make up 250 mL with distilled water.
5. Stop solution: 3.65 g EDTA-$Na_2 \cdot 2H_2O$ in 250 mL of water.

6. Preserving solution: Add 75 mL ethanol and 11.5 mL glycerol (87% w/w), to make up 250 mL with distilled water.

2.6. In-Gel Protein Digestion

1. Soaking solution: 50 m*M* NH_4HCO_3.
2. Reducing solution: 10 m*M* dithiotreitol (DTT), 25 m*M* NH_4HCO_3
3. Alkylating solution: 100 m*M* iodoacetamide (add prior to use), 50 m*M* NH_4HCO_3.
4. Digestion solution: 25 m*M* NH_4HCO_3.
5. Sequencing-grade modified porcine trypsin (Promega, Madison, WI)
6. Investigator ProGest automatic digestor (Genomic Solutions, Cambridgeshire, UK).

2.7. MALDI-ToF-MS

1. TA solution: 33% acetonitrile, 0.1% trifluoroacetic acid (TFA) in water.
2. Matrix solution: 1 µg/µL saturated solution of DHB (2,5-dihydroxybenzoic acid) in TA solution.
3. Unijet II Refrigerated Aspirator (Uniequip).
4. Bruker Reflex IV mass spectrometer (Bruker-Franzen Analytic GmbH, Bremen, Germany).
5. MS-Biotools 2.2 software (Bruker Daltonics GmbH, Bremen, Germany).
6. Licensed version of the Mascot search engine (MatrixScience, London, UK).

2.8. LC-ESI MS/MS

1. Buffer A: 0.5% acetic acid in water.
2. Buffer B: 0.5% acetic acid in acetonitrile:water 90:10.
3. Famos autosampler (Lcpackings, Amsterdam, Netherlands)
4. 300 µm ID × 5 mm PepMap C18, 5 µ*M*, 100 Å, µPrecolumn cartridge (Lcpackings).
5. Home-made 100 µm ID × 100 mm nanocolumn, packed with Kromasil C18, 5 µ*M*, 100 Å stationary phase (Eka Chemicals, Bohus, Sweden).
6. Automated Esquire 3000^{plus} ion-trap mass spectrometer (Bruker Daltonics).
7. DataAnalysis 3.1 software (Bruker Daltonics).

3. Methods

3.1. Sample Preparation

This process should ideally result in the complete solubilization, disaggregation, denaturation, and reduction of the proteins in the sample. There is nonetheless no single method of sample preparation that can be universally applied to all sample types. For example, when the protein(s) of interest is found only at very low levels and/or in specific locations inside the cell, it is desirable to enrich the sample using methods such as differential centrifugation prior to protein extraction and solubilization. However, as a rule of thumb, it is desirable to keep the method as simple as possible. Additional separation steps might improve the final quality of the sample, but also result in undesired protein losses. The next method described is of general application, suitable for extraction of soluble proteins.

1. Freeze *Arabidopsis thaliana* leaves (approx 0.2 g) in liquid nitrogen to disrupt cell membranes and organelles, and resuspend them in 750 µL of extraction buffer in a 1.5-mL Eppendorf tube (*see* **Note 1**).
2. Homogenize the suspension (3 pulses, 15 s each, 4000 rpm) using an electric homogenizer (Glas-Col).
3. Add 750 µL of extraction buffer to the homogenized solution. Centrifuge in a microfuge (3 min, 3600 rpm, 4°C) to pellet debris. Freeze the supernatant overnight at –80°C and centrifuge again (10 sec, 14000 rpm, 4°C).
4. Recover the supernatant, divide it in two, and precipitate proteins using the TCA method (*see* **Note 2**). Briefly, add precipitation solution to the tube to obtain a final TCA concentration of 15% v/v. Incubate this solution overnight at 4°C and pellet proteins by microfuge centrifugation (10 min, 14000 rpm, 4°C). After two washes with washing solution, allow pellets to dry until moist and keep at –80 °C until use.

3.2. IEF Electrophoresis with Immobilized pH Gradient Strips

The sample solution must contain certain components to ensure complete protein solubilization and denaturation. It should contain always urea/thiourea and one or more nonionic or zwitterionic detergents. Proteins should remain in sample solution at room temperature (RT) for 1 to 2 h for full denaturation and solubilization.

1. Briefly, resuspend the pellet in 100 µL of sample solution containing 1.5% (v/v) IPG buffer 3–10 L (Amersham-Pharmacia Biotech) and incubate with agitation (2 h, RT; *see* **Note 3**). After this step, take 5 µL to measure sample concentration using a reductor/detergent-compatible assay (Bio Rad RC-DC protein assay) (*see* **Note 4**).
2. For 1D Electrophoresis (IEF), mix 50 µL of sample solution containing 125 µg of total protein extract with 300 µL of rehydration buffer containing 0.5% (v/v) IPG buffer 4–7 L (Amersham Pharmacia Biotech). Apply samples by rehydration of immobilized pH gradient (IPG) strips (pH 4–7, linear, 18 cm length; Pharmacia Biotech, Uppsala, Sweden) (*see* **Note 5**). Carry out focusing on an IPGPhor (Pharmacia Biotech) using a voltage stepwise increase of 30 V for 7 h, 60 V 7 h, 120 V 1h, 250 V 1 h, 500 V 1 h, 1000 V 30 min, gradient increase from 1000 to 8000 V for 30 min, and constant 8,000 V until total Vh (~30,000) is reached.

3.3. 2D Electrophoresis

1. After IEF separation, equilibrate strips 2 × 15 min with equilibration buffer. The first equilibration buffer contains 2% 2,4-dithio-threitol (DTT) (add prior to use), whereas 4% iodoacetamide (without DTT) is added to the buffer in the second equilibration step (*see* **Note 6**).
2. The equilibrated IPG strip is positioned on the surface of the second-dimension gel. Extreme care needs to be taken in pushing the IPG strip down so that the entire lower edge of the strip is in contact with the top surface of the gel; the absence of bubbles trapped between the IPG strip and the gel has to be tested. Second-dimension SDS-PAGE is performed using 1-mm-thick, 16 × 15 cm, 12.5% homogeneous polyacrylamide gels (recommended for the 14–100 kDa range), and electrophoresis is carried out overnight at constant current (5 mA/gel) and temperature (5°C).

3.4. Mass Spectrometry-Compatible Silver Gel Staining

Silver staining methods are about 25 to 100 times more sensitive than the more common Coomassie blue gel staining techniques. Consequently, they are the method of choice when very small amounts of proteins are to be detected (*see* **Note 7**). When MS is the method used for sample analysis (as in this case), a compatible silver staining method must be used (*see* **Note 8**) but its poorer sensitivity (up to 5 ng) compared to traditional silver staining methods (500 pg–1 ng) must be taken into account.

The silver staining method described below follows, with some minor modifications, the protocol described in "Silver Staining Kit Protein Instructions" (Amersham-Pharmacia Biotech); reagents used are from the Silver Staining Kit. All steps should be performed with gentle shaking of the staining tray.

1. Soak gel in fixing solution for 30 min (*see* **Note 9**).
2. Remove fixing solution and add sensitizing solution. Shake gel gently for at least 30 min.
3. Remove sensitizing solution and wash gel three times (5 min each) with distilled water.
4. Remove water, and add silver solution to the gel, and shake for 20 min.
5. Remove silver solution and wash twice (1 min each time) with distilled water.
6. Remove water and add developing solution. Shake for 2 to 5 min.
7. Remove developing solution and add stop solution. Shake for 10 min.
8. Remove stop solution and wash the gel three times with distilled water for 5 min each time.
9. For short-term preservation, keep gels in distilled water. For longer storage, use preserving solution.

3.5. Protein Spot Picking

Protein spots of interest (N1–5, Fig. 1) were excised manually (*see* **Note 10**) using a home-made gel spot-picker (*see* **Note 11**), then placed in a 96-well plate with holes (Bruker Daltonic).

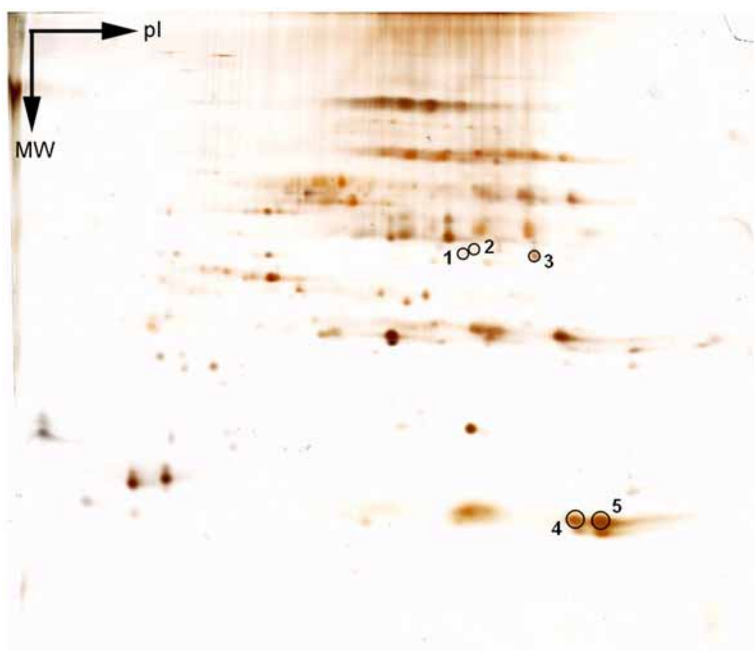


Fig. 1. 2D-PAGE map of *Arabidopsis thaliana* leaf proteins. The first dimension was a pH 4.0 to 7.0 gradient (*see* **Subheading 2.**). Numbered circles mark protein spots that were picked, trypsin-digested, and analyzed by MALDI-ToF or ion-trap MS/MS.

Plate holes are necessary to allow the automatic digestor to change solutions used for tryptic digestion. Deionized water (20 µL) was added to move gel pieces in the bottom of the wells.

3.6. In-Gel Protein Digestion

In-gel digestion was performed automatically (*see* **Note 12**) in an automatic digestor. The digestion protocol was essentially as described by Shevchenko et al. ***(20)***, with minor modifications.

1. Soak gel spots for 20 min in soaking solution, followed by 10 min incubation with acetonitrile to dehydrate the gel piece.
2. Raise digestor temperature to 56°C and reduce proteins in reducing solution. After 10 min, turn off heaters and leave samples in reducing solution for an additional 20 min.
3. Alkylate samples (15 min, RT) in alkylating solution and wash them 10 min with soaking solution. After the washing step, dehydrate gel pieces for 15 min with acetonitrile.
4. Add trypsin to the samples at a final concentration of 16 ng/µL in digestion solution. Incubate 6 h, 37°C.
5. Wash gel pieces with soaking solution and extract peptides using, sequentially, digestion solution (15 min), 0.1% trifluoroacetic acid (TFA) in water (15 min), and acetonitrile (15 min), to a final volume of 100 µL.

3.7. Sample Preparation for MALDI-ToF-MS Analysis

1. Dry samples in a refrigerated aspirator and resuspend in 10 µL of TA solution.
2. Spot 5% of each sample (0.5 µL) manually in a sample carrier and allow to dry at RT.
3. Add 0.5 µL of matrix solution and allow to dry at RT.
4. Samples are automatically analyzed in a MALDI-ToF mass spectrometer. Calibrate each spectrum internally using signals corresponding to well-defined tryptic autolysis peptides. Set mass accuracy to ± 30 ppm.

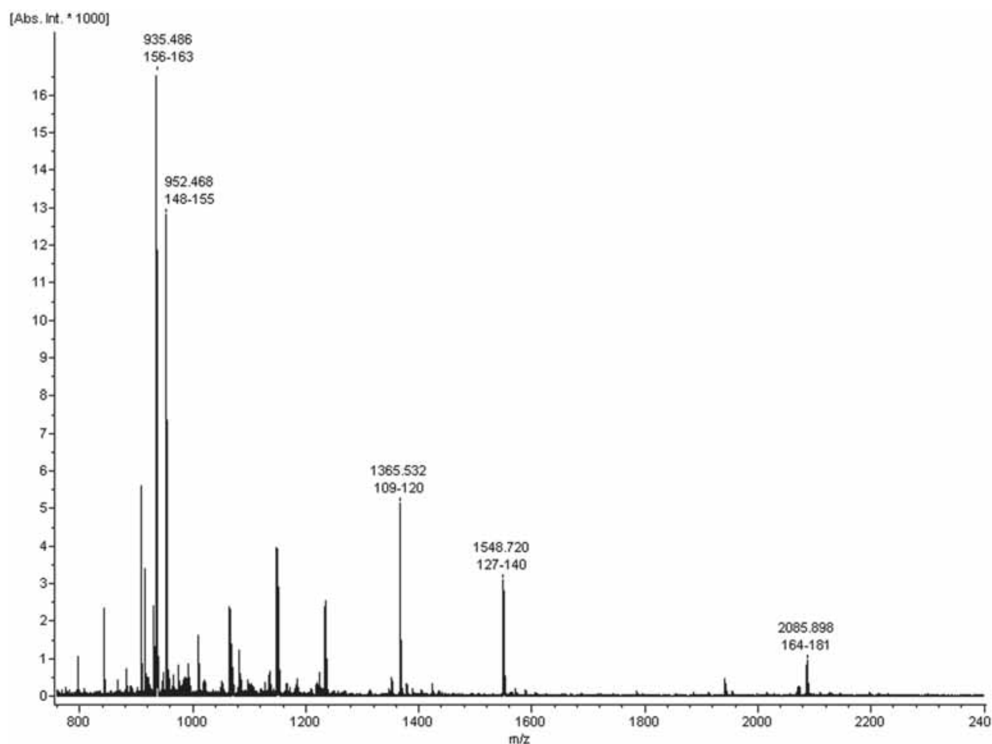


Fig. 2. Fingerprint spectrum corresponding to spot 5 (*see* **Fig. 1**). Database search identified the protein as ribulose-bisphosphate carboxylase small-chain A1 precursor. The upper number in each pair indicate monoisotopic mass ($M+H^+$); lower numbers indicate the respective protein fragment. Sequences of each fragment are as follows: 109–120: EHGNSPGYYDGR; 127–140: LPLFGCTDSAQVLK; 148–155:EYPNAFIR; 156–163: IIGFDNTR; 164–181: QVQCISF VAYKPPSFTG.

5. Send each digestion spectrum (fingerprint) automatically through MS-Biotools 2.2 software to a licensed version (*see* **Note 13**) of the Mascot search engine. Perform searching through MSDB, a comprehensive, nonidentical protein sequence database designed specifically for MS applications and maintained by the Proteomics Department, Hammersmith Campus, Imperial College London (http://csc-fserve.hh.med.imperial.ac.uk/msdb.html). Restrict organism to *Arabidopsis thaliana* and establish 200 kDa as the upper limit for molecular mass. Allow one missed tryptic cleavage (*see* **Note 14**) and define carbamidomethylation of cysteines as a fixed modification (*see* **Note 15**). Fix accuracy at ± 70 ppm (*see* **Note 16**). The minimum number of tryptic peptides allowed for the adjustment is 4.

Only protein identifications with score values higher than that signaling identity or extensive homology are accepted. Nonetheless, each identification has to be analyzed manually to check for possible misidentification. **Figure 2** shows the fingerprint spectrum obtained from spot N5.

3.8. LC ESI-MS/MS Analysis

Ion-trap mass spectrometers are specifically designed to generate CID collision-induced dissociation MS/MS spectra from ions (peptides) isolated from a complex mixture (i.e., a tryptic mixture). Their use in protein identification is recommended when MALDI-ToF analysis of a tryptic digestion yields no results, or when post-translational modifications must be confirmed (*see* **Note 17**). When coupled to liquid chromatography, this is termed LC-MS/MS (*see* **Note 18**).

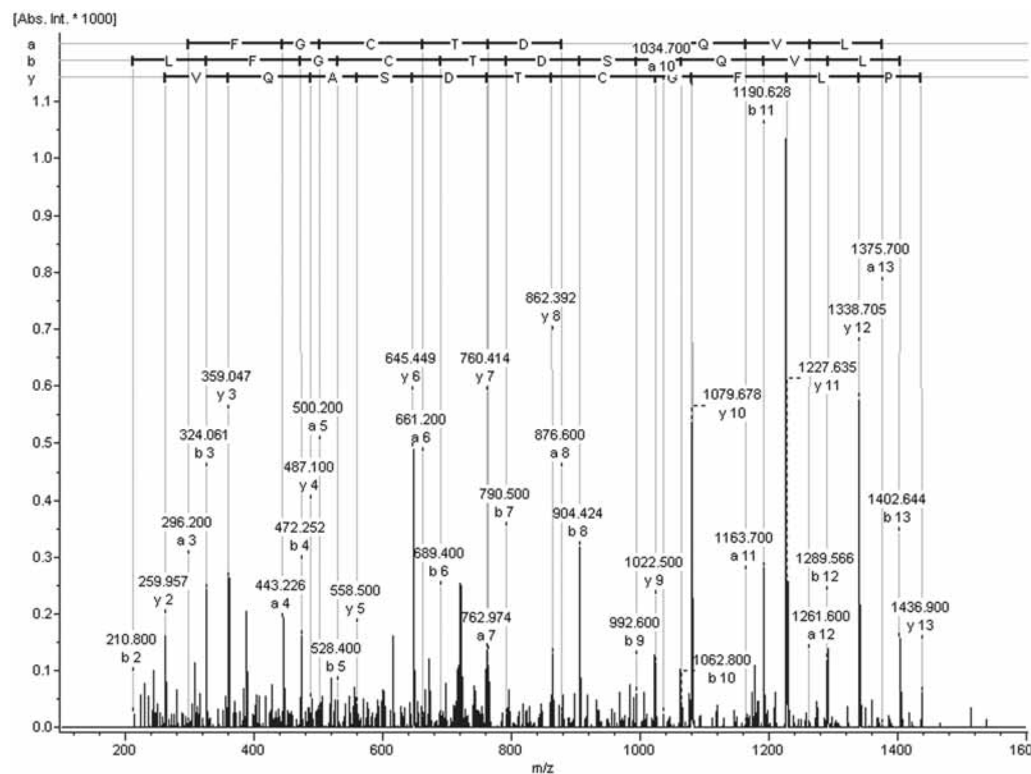


Fig. 3. CID MS/MS spectrum of the double-charged parent ion (m/z 775.4) of a tryptic peptide from the ribulose-bisphosphate carboxylase small-chain A1 precursor protein. The peptide sequence (LPLFGCTDSAQVLK) corresponds to the protein fragment 127–140. Signals corresponding to the main fragmentation series (*a*, *b* and *y*) are highlighted.

1. Dry trypsin-digested spots and resuspend them in 22 μL of buffer A.
2. Inject 20 μL of each sample automatically using a Famos autosampler in a μPrecolumn cartridge.
3. Pass 450 μL of buffer A through the cartridge at a flow rate of 30 μL/min.
4. After this washing step, apply a linear gradient ranging from 2 to 35% of buffer B for 55 min to separate the peptides in a home-made (*see* **Note 19**) nanocolumn, packed with Kromasil C18, 5 μ*M*, 100 Å stationary phase. Flow is 500 nL/min (0.5 μL/min).

Peptide fragmentation by CID is carried out in an automated ion-trap mass spectrometer. The most abundant peptide in each full MS spectrum is automatically isolated and fragmented. Dynamic exclusion is used to avoid unnecessary recurrence of the most abundant peptides in the sample (*see* **Note 20**). Crude data are processed using DataAnalysis 3.1; processed data are used to search databases through BioTools 2.2 and the Mascot search engine. The organism is restricted to *A. thaliana*, one missed tryptic cleavage is allowed, and carbamidomethylation of cysteines (*see* **Note 15**) is defined as a fixed modification. Mass tolerances are set to 1 and 0.8 DaHon for the parental peptide and fragments, respectively. Each positive identification is checked individually to further confirm the result. **Figure 3** shows fragmentation spectra of a tryptic peptide from spot 5 (**Fig. 1**).

4. Notes

1. Cell or tissue disruption releases proteases that can severely affect the sample. To avoid this, disruption should be performed at as low a temperature as possible in a denaturing solution. Addition of protease inhibitors is strongly recommended.

2. Methods that allow selective precipitation of proteins should be considered if the presence of contaminants (such as ionic detergents, salts, lipids, nucleic acids, etc.) are suspected to affect the final result. Interfering substances usually affect the 1D step (IEF) and are commonly detected as horizontal streaking caused by inappropriate protein focusing. Precipitation should be considered after one initial set of experiments in which no precipitation step has been done. No precipitation technique is 100% efficient, as some selectively precipitate some proteins to the detriment of others.
3. Never heat (over 37°C) a sample after adding urea. High temperatures cause urea hydrolysis to isocyanate, which modifies proteins by carbamylation.
4. Most colorimetric assays used to quantify proteins are sensitive to the presence of detergents and reducing reagents in the sample, which can cause considerable distortion in quantification. It is therefore desirable to use reductor/detergent-compatible protein assays.
5. Recipes to prepare home-made IPG gels are found elsewhere, but we strongly recommend those described in "Two-dimensional electrophoresis with immobilized pH gradient for proteome analysis" by Angelika Görg et al. (http://www.weihenstephan.de/blm/deg).
6. Iodoacetamide introduced in the second equilibration step alkylates thiol groups on proteins, preventing their reoxidation during electrophoresis. Reoxidation during electrophoresis can result in streaking and other artifacts.
7. A new generation of dyes has been developed and marketed in recent years. In some cases, these dyes offer better sensitivity (up to 100 to 200 pg) and a wider dynamic range, making them very useful for quantitative studies. However, they are also considerably more expensive and require more hardware and expertise than the traditional silver-staining methods.
8. Many of the common sensitizing reagents (e.g., glutaraldehyde) cannot be used as they modify proteins, making them very difficult to analyze by MS.
9. Freshly made solutions (not older than 24 h) give best results. Add components marked (*) immediately before use. For 12.5 × 26 cm gels, 250 mL of solutions are recommended; use 125 mL for smaller gels. All steps should be performed at RT with gentle shaking of the staining tray.
10. A piece cut from a blank region of the gel serves as a control, which should be processed as the rest of the sample pieces.
11. The home-made spot picker consists of a metallic pen-shaped device, containing a 3-mm ID tube for picking the spots and an inner rod that pushes the piece of gel to the sampling tube. A similar commercial device can be purchased from The Gel Company, cat no. P2D3.0.
12. All reagents should be prepared fresh prior to use. Automatic digestion is strongly recommended to minimize contaminations (keratin, etc.) common to manual digestion. Manual digestion protocols can be found elsewhere ***(18)***.
13. Mascot has become established as the cross-platform standard for protein identification using mass spectrometry data. Although Mascot is freely accessible on the Web, some issues, such as confidentiality and high-throughput work, strongly recommend its licensing (www.matrixscience.com\products.html).
14. Experience shows that tryptic digestion of complex mixtures usually includes peptides with missed cleavage sites. In these cases a setting of 1 is recommended. Higher values increase the number of calculated masses to be matched against the experimental data and thus computing time.
15. Iodoacetamide alkylation of free (reduced) cysteines during the in-gel tryptic digestion procedure is done to avoid the undesired formation of secondary protein structures that would reduce tryptic digestion efficiency.
16. This value stands for the error window on experimental mass values and depends on the mass spectrometer and the calibration method used.
17. Protein absence in the databases and poor-quality spectra are the main reasons why MALDI-ToF fingerprint spectra do not yield satisfactory identifications.
18. As tryptic digestions of protein spots are not very complex samples, online liquid chromatography (LC-MS/MS) is not mandatory, but is highly advisable to automate the process, especially if it is coupled to an autosampler. It also minimizes sample handling.
19. Several packing methods can be found in the scientific literature. Our method follows that published by Robins and Guido ***(21)***.

20. Dynamic exclusion is a tool to avoid more MS/MS data then we need and to search minor compounds in the presence of major compounds. The software automatically excludes during a user-defined period of time (e.g., 1 min) all those masses (compounds) that have been previously analyzed.

References

1. Anderson, N. G., Matheson, A., and Anderson, N. L. (2001) Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* **1,** 3–12.
2. Blackstock, W. P. and Weir, M. P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol*. **17,** 121–127.
3. Kersten, B., Bürkle, L., Eckehard, J. K., et al. (2002) Large-scale plant proteomics. *Plant Mol. Biol.* **48,** 133–141.
4. Yamazaki, D., Motohashi, K., Kasama, T., Hara, Y., and Hisabori, T. (2004) Target proteins of the cytosolic thioredoxins in *Arabidopsis thaliana. Plant Cell Physiol.* **45,** 18–27.
5. Borderies, G., Jamet, E., Lafitte, C., et al. (2003). Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: a critical analysis. *Electrophoresis* **24,** 3421–3432.
6. Zabrouskov, V., Giacomelli, L., Van Wijk, K. J., and McLafferty, F. W. (2003) A new approach for plant proteomics: characterization of chloroplast proteins of *Arabidopsis thaliana* by top-down mass spectrometry. *Mol. Cell. Proteomics* **2,** 1253–1260.
7. Froehlich, J. E., Wilkerson, C. G., Ray, W. K., McAndrew, R. S., Osteryoung, K. W., Gage, D. A., and Phinney, B. S. (2003) Proteomic study of the *Arabidopsis thaliana* chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. *J. Proteome Res*. **2,** 413–425.
8. Ferro, M., Salvi, D., Brugiere, S.,et al. (2003) Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. *Mol. Cell. Proteomics* **2,** 325–345.
9. Luche, S., Santoni, V., and Rabilloud, T. (2003) Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics* **3,** 249–253.
10. Gallardo, K., Job, C., Groot, S. P., Puype, M., Demol, H., Vandekerckhove, J., and Job, D. (2002) Proteomics of *Arabidpsis* seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiol.* **129,** 823–837.
11. O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250,** 4007–4021.
12. Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., and Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21,** 1037–1053.
13. Henzel, W.J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* **90,** 5011–5015.
14. Mann, M., Hojrup, P., and Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* **22,** 338–345.
15. Pappin, D. J. C., Hojrup, P., and Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3,** 327–332.
16. Yates, J. R. III, Speicher, S., Griffin, P. R., and Hunkapiller, T. (1993) Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* **214,** 397–408.
17. Link, A. J., Eng, J., Schieltz, D. M., et al. (1999). *Nature Biotechnol.* **17,** 676–682.
18. Simpson, R. J. (2003) *Proteins and Proteomics. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Pennington, S. R. and Dunn, M. J. (2001) *Proteomics: From Protein Sequence to Function*. BIOS Scientific Publishers, Oxford, UK.
20. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68,** 850–858.
21. Robins, R. H. and Guido, J. E. (1997) Design, construction and application of a simple packed capillary high performance liquid chromatography/electrospray mass spectrometry system. *Rapid Commun. Mass Spectrom*. **11,** 1661–1666.

35

Peptide-Based Phosphoproteomics With Immobilized Metal Ion Chromatography

Thomas S. Nühse and Scott C. Peck

Summary

The identification of protein phosphorylation sites has always been a challenging task, traditionally involving large amounts of radioactive phosphorus and high-performance liquid chromatography separation and Edman sequencing of phosphopeptides. The rapid development of mass spectrometric methods has advanced protein research significantly, and the identification of in vivo post-translational modifications of even rare proteins is now possible. Even with the new generation of machines, however, phosphopeptides do not lend themselves easily to mass spectrometric analysis. In complex mixtures of peptides, phosphopeptides are often difficult to detect because of suppression effects during ionization. This problem can be largely solved by affinity purification/enrichment of the phosphopeptides, and immobilized metal ion affinity chromatography (IMAC) on chelated Fe^{3+} or other metal ions has emerged as the simplest and most useful method. IMAC has been useful to identify several in vivo phosphorylation sites of individual proteins, but is easier to apply to complex mixtures. We describe a complete protocol as it has been used for *Arabidopsis* plasma membranes, and note where it can be adapted for soluble protein mixtures.

Key Words: phosphoproteomics; phosphopeptide; IMAC; posttranslational modification; mass spectrometry; phosphorylation.

1. Introduction

Protein phosphorylation is perhaps the most widespread and by far the best understood posttranslational modification. To understand what effect phosphorylation of a particular site has on a protein of interest and to dissect the signaling pathway leading to the modification, it is essential to identify the phosphorylation site(s). This has always been a challenging task, traditionally involving large amounts of radioactive phosphorus and high-performance liquid chromatography (HPLC) separation and Edman sequencing of phosphopeptides. The rapid development of mass spectrometric methods has advanced protein research significantly, and the identification of in vivo posttranslational modifications of even rare proteins is now possible. Even with the new generation of machines, however, phosphopeptides do not lend themselves easily to mass spectrometric analysis. In complex mixtures of peptides, phosphopeptides are often difficult to detect because of suppression effects during ionization. That is, mass spectrometers only detect ionized peptides; however, in complex mixtures with unphosphorylated peptides, phosphopeptides do not ionize well. Therefore, they tend to "disappear" from the analysis. One solution to this problem is the use of immobilized metal ion affinity chromatography (IMAC) on chelated Fe^{3+} or other metal ions to affinity-purify/enrich the phosphopeptides. Although IMAC has been useful to identify several in vivo phosphorylation sites of individual proteins (*1–3*), there has existed concern that nonspecific binding of highly acidic peptides

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

would contaminate these preparations rendering analyses of complex samples impossible. Using the method described in this chapter, however, we found that contaminating peptides showed no bias towards more acidic pI nor did they contain more clustered glutamic and aspartic acid residues ***(3)***. Rather, contaminating peptides tended to come from highly abundant proteins. Thus, not only is it possible to use IMAC for complex mixtures, but we also often find it is easier to apply the method to complex mixtures than to individual proteins (*see* **Note 1**). We describe a complete protocol as it has been used for *Arabidopsis* plasma membranes ***(3,4)***, and note where it can be adapted for soluble protein mixtures. Unless more complicated quantitative techniques are used ***(5)***, the procedure described here will only give a static picture of phosphorylation sites. Nevertheless, a vast amount of new insights into signalling principles can be gained from such studies ***(4)***.

2. Materials

1. Homogenization buffer: 250 m*M* sucrose, 50 m*M* HEPES/KOH, pH 7.5, 50 m*M* sodium pyrophosphate, 25 m*M* sodium fluoride, 10 m*M* EDTA, 1 m*M* sodium molybdate, 5% glycerol, 0.5% polyvinyl pyrrolidone (PMSF added to 1 m*M* before use from 100 m*M* stock in ethanol; DTT added to 3 m*M* after first centrifugation).
2. 1 m*M* K252a in DMSO (store at –20°C).
3. 100 µ*M* Calyculin A in DMSO (store at –20°C).
4. Phenol solution, buffered (Sigma P4557).
5. Phenol back-extraction buffer: 100 m*M* Tris-HCl, pH 8.5, 100 m*M* KCl, 50 m*M* EDTA and 0.5% v/v (β-mercaptoethanol).
6. 100 m*M* ammonium acetate in methanol.
7. 100 m*M* sodium carbonate.
8. Ammonium hydrogen carbonate.
9. Porcine modified trypsin (Promega).
10. SAX buffer: 30% acetonitrile, 25 m*M* ammonium hydrogen carbonate.
11. Acetic acid.
12. POROS® chromatography materials (Self Pack OligoR3, MC 20, HQ 20) (Applied Biosy-stems).
13. 100 m*M* $FeCl_3$ in 100 m*M* acetic acid.
14. GELoader Tips (Eppendorf).
15. 50 m*M* sodium phosphate, pH 9.0.
16. Ammonia (25% aqueous solution).
17. Sodium chloride.
18. Formic acid.

3. Methods

The IMAC method for phosphopeptide isolation is, in principle, universally applicable for complex protein mixtures and single proteins. We will give protocols for total soluble protein and membrane proteins. Any special cases can be adapted from the basic protocols. After digestion with trypsin, the peptides can be either used directly for IMAC or, if the mixture is very complex, fractionated by anion exchange chromatography before IMAC (**Fig. 1**).

3.1. Protein Extraction

1. Homogenize cell culture or leaf material on ice in homogenization buffer (2 mL/g fresh weight; *see* **Note 2**). Separate the homogenate from debris by centrifugation (10,000*g*; 10 min). For separation of soluble and microsomal proteins, centrifuge the supernatant of the first spin (100,000*g*; 60 min).
2. Strip peripheral proteins from the microsomal pellet by resuspending in 100 m*M* sodium carbonate and incubate on ice for 10 min. Collect by centrifugation (100,000*g*; 30 min). Repeat washes with 500 m*M* and then 50 m*M* ammonium hydrogen carbonate.
3. Dialyse soluble proteins against 50 m*M* ammonium hydrogen carbonate and determine the protein concentration.
4. Alternatively, precipitate the soluble protein with phenol ***(6)***: add 1 vol of phenol solution to 1 vol of protein extract, vortex, and keep on ice for 5 min. Centrifuge (10,000*g*; 10 min at room temperature), carefully remove the aqueous upper phase and reextract the phenol- and inter-

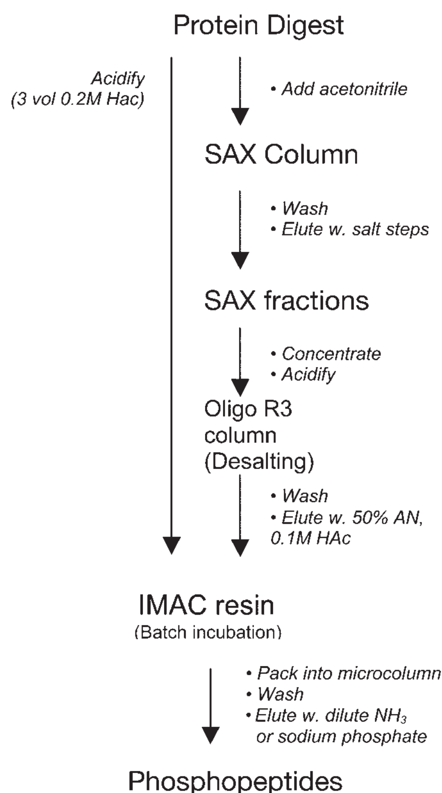


Fig. 1. Phosphopeptide purification flow chart.

phase (with the precipitated protein) twice with phenol extraction buffer. After removal of the last upper phase, add at least three volumes of ammonium acetate/methanol, mix and keep at –20°C for 30 min to overnight. Pellet the precipitated protein by centrifugation (20,000*g*; 15 min at 4°C) and wash the pellet twice with ammonium acetate/methanol and once with 80% acetone before redissolving in 50 m*M* ammonium hydrogen carbonate and determining the protein concentration.

3.2. Trypsin Digest

Dissolve the lyophilized trypsin either in 20 µL 50 m*M* acetic acid (delivered with the trypsin; if only small amounts are used) or directly in 50 m*M* ammonium hydrogen carbonate. Add to the protein at a ratio of 1:100, or up to 1:20 for smaller amounts of protein.

3.3. Anion Exchange Chromatography

Protein digests of low complexity can be used directly for IMAC by adjusting the solution to 200 m*M* acetic acid (then proceed to **Subheading 3.4.**). For highly complex mixtures, strong anion-exchange chromatography (SAX) is a useful fractionation step prior to IMAC. The scale described below works for up to 1 mg of trypsin-digested protein.

1. Add acetonitrile (final concentration 30%) to the protein digest, vortex, and spin 5 min at maximal speed in a microcentrifuge.
2. Prepare a SAX column as follows: pack approx 30 µL (= 60 µL of a 50% slurry) of POROS HQ20 into a constricted GELoader tip (*see* **Subheading 3.5.**) and wash with 3 bed volumes of SAX buffer.
3. Slowly load the digest (up to 1 mg of trypsin-digested protein) onto the SAX column. Wash with 3 bed volumes of SAX buffer.

4. Elute stepwise with 2 bed volumes of SAX buffer plus increments of 40 to 240 m*M* NaCl (4 to 10 steps, depending on the complexity of the sample and the scale of the experiment).
5. To prepare the eluate fractions for IMAC, concentrate them in the speed-vac to evaporate the acetonitrile (5 min, avoid drying them completely), acidify with formic acid (final 5%) and load onto a microcolumn packed with a 5 to 10 µL bed volume of POROS OligoR3. Wash twice with 10 µL 100 m*M* acetic acid and elute with 20 µL 50% acetonitrile, 100 m*M* acetic acid.

3.4. Immobilized Metal Ion Affinity Chromatography

1. Prepare POROS MC material (*see* **Note 3**) with a series of batch washes in a centrifuge tube as follows: wash once with 50 m*M* EDTA, once with water, three times with 100 m*M* acetic acid, once with 100 m*M* $FeCl_3$ in 100 m*M* acetic acid, once in 100 m*M* acetic acid, once in 1 *M* NaCl, 100 m*M* acetic acid and at least five times in 100 m*M* acetic acid. Due to the hydrophobicity of the beads, complete settling is easier to achieve with 30 to 40% acetonitrile in every step. Store as a 50% slurry in 100 m*M* acetic acid (stable at 4°C for at least 2 mo).
2. Pipet 5 to 10 µL of Fe^{3+}-POROS MC slurry (for approx 100 µg protein digest or one SAX fraction) into a centrifuge tube. Acidify the digest with 3 volumes of 200 m*M* acetic acid and add to the IMAC resin or, if SAX was used, add the desalted SAX fraction. Incubate with shaking (~1200 rpm) for 10 min at room temperature.
3. Pack slurry plus sample into a constricted GELoader tip (*see* **Subheading 3.5.**) and wash twice with 30% acetonitrile, 100 m*M* acetic acid and once with 100 m*M* acetic acid, each 10 to 20 µL.
4. Elute slowly with 10 to 20 µL dilute ammonia, pH 11, or 50 m*M* sodium phosphate, pH 9.0. Yields are optimal if the IMAC column is "equilibrated" with half of the eluant and left for 10 min without pressure before pumping the rest of the eluant through the column (*see* **Note 3**). Acidify to 5% formic acid before liquid chromatography-mass spectrometry (LC-MS).

3.5. Microcolumns

Small amounts of peptides can be quickly and efficiently handled in microscale chromatography columns packed into capillary pipet tips. Although there are commercial products for concentrating peptide samples (ZipTip®), self-made columns are cheaper, more versatile (any material can be packed and the bed volume is adaptable), and closer to "real" chromatography because of their small diameter-to-length ratio. The principle has been described in ref. *7*.

1. Prepare a 1-mL syringe with adapter: Trim the ends of a 200-µL pipet tip on both sides with a razor blade, so that the wide end can be fitted tightly onto the syringe tip and the narrow end into the top of a GELoader tip. The adapter should not protrude more than ~2 mm into the GELoader tip.
2. Constrict the capillary end of the GELoader tip about 1 mm from the very end. The easiest way is to place the capillary onto a hard, smooth work surface and roll the edge of a pen over it with some force (**Fig. 2**).
   Pipet 10 µL water into the wide part of the GELoader tip. When holding the tip upright, the liquid should not leak. Ideally, only application of full pressure with the syringe will, after a few seconds, open the constriction. It will take some practice to apply just the right amount of pressure with the pen to properly constrict the GELoader tip. If the test liquid leaks out immediately, the chromatography material will later not be safely retained in the tip.
3. Prepare a 50% (v/v) slurry of the desired chromatography material in acetonitrile (OligoR3), or appropriate buffers (HQ in SAX buffer, MC in 100 m*M* acetic acid). Pipet into the GELoader tip, flick (i.e., snap wrist) to force resin into the capillary end, and apply pressure with the 1-mL syringe plus adapter. The beads will pack into the capillary tip and be retained by the constriction. For small-scale desalting of peptides (e.g., in matrix-assisted laser desorption/ionization [MALDI]), an OligoR3 column a few mm long is sufficient ***(7)***. IMAC columns for 100 µg-scale digests with 5 to 10 µL beads will pack the full narrow end of the tip plus a bit more.
4. Wash solutions can be placed into the wide part of the tip and squeezed through the packed column without concern about the column running dry. For washing steps in columns packed only into the narrow end of the tip, running dry for a brief moment does not seem to affect the chromatographic performance. "Midi columns" that have been packed into the wide part of the GELoader

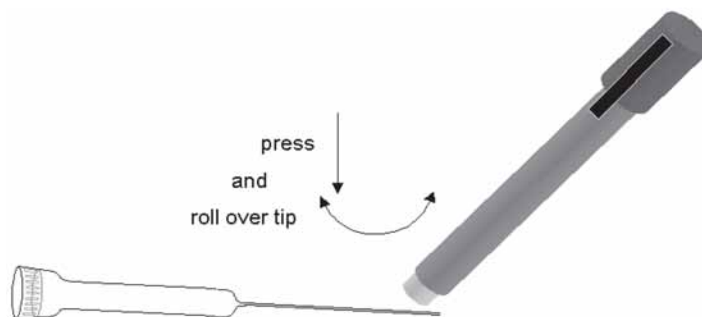


Fig. 2. Packing microscale chromatography columns into capillary pipet tips.

tip should not be allowed to run dry. To prepare freshly packed POROS OligoR3 columns for desalting, the column is washed once each with acetonitrile and 5% formic acid in water.

3.6. Identification of Phosphorylation Site(s) by MS/MS

The specific settings for MS/MS analysis will vary depending on the type of machine being used. We have analyzed samples prepared as described in this protocol on both Q-TOFs and ion traps from different companies and found that all machines performed well. Because of the purity of these phosphopeptide samples, we found no advantage in using negative ion mode or neutral loss as prescans to detect phosphopeptides, although these scans can be useful during initial experiments to confirm the presence of phosphopeptides. Thus, it is generally sufficient to acquire MS/MS spectra and then allow for phosphorylated residues (S, T, Y) in the search. An important consideration, however, is that all protein/phosphorylation site identifications are based on single peptides. As a rule, we manually inspect the spectra to confirm the assignment.

4. Notes

1. IMAC is not the ultimate answer for phosphoproteomics. The method described works well to enrich phosphopeptides from complex mixtures, but it is more difficult to apply to single proteins. Depending on the stoichiometry of phosphorylation and protein amount, some optimisation of the protein-resin ratio may be necessary. Not all phosphopeptides will be isolated this way ***(8)***, and it is worthwhile to optimize a number of parameters (e.g., chromatographic material, metals, pH of binding, and elution buffers) for isolation of specific phosphopeptides. Methyl esterification of peptides has been reported to improve the specificity for phosphorylated over nonphosphorylated peptides ***(9)***. In our experiments, this modification did not work and, in fact, turned out to be unnecessary. It is possible, however, that it may be beneficial for specific applications.
2. The homogenization buffer contains sufficient amounts of competitive phosphatase inhibitors to preserve the phosphorylation status of the sample. For small volumes of extracts, 10 n*M* calyculin A can be added for complete inhibition of phosphatases 1 and 2a. If it is crucial that not only dephosphorylation, but also *de novo* phosphorylation is completely inhibited after homogenization, at 100 n*M* K252a can be added to the buffer.
3. A large variety of IMAC materials have been published for enrichment of phosphopeptides, including metal-chelating Sepharose or agarose. Although those may be suitable in some cases, the microcolumn method works only with HPLC-grade material with small, uniform particles. The options are POROS MC with iminodiacetic acid (IDA) as functional group and nitrilotriacetic acid (NTA-) silica. The latter is used in spin columns for the purification of His-tagged proteins (Qiagen). The material can be retrieved by opening the columns with a razorblade. Alternatively, larger amounts (grams) of the NTA silica are available from the company upon request. The specificity of both materials is similar, and the hydrophilicity of the NTA silica may decrease contamination by nonphosphorylated peptides. However, POROS

MC material is easier to handle because of its lower back-pressure in microcolumns and because it is less likely to leak particles.

Only a small number of metal ions are suitable for phosphopeptide affinity, and Fe^{3+} is the most widely used and best in our hands. Whether different types of peptides are retained by other metals (e.g., gallium *(10)* has not been tested.

4. We found it essential that the whole column is equilibrated in elution buffer for approx 10 min to give the phosphopeptides enough time for complete desorption. To determine how much eluant is needed for this step, a dummy column without sample can be treated in the same way while continuously monitoring the pH of the eluting liquid by spotting it onto pH indicator paper. The transition from acidic wash solution to the basic eluant should be quite sharp. When eluting a real sample, of course, the complete eluate is collected.

Acknowledgments

We would like to thank Drs. Ole N. Jensen and Allan Stensballe (Odense, Denmark) for a successful collaboration and continued discussions. This work was supported by BBSRC grant 83/C17990 (TSN, SCP), the Gatsby Charitable Foundation (TSN, SCP) and an EMBO Short-Term fellowship (TSN).

References

1. Fuglsang, A. T., Visconti, S., Drumm, K., et al. (1999) Binding of 14-3-3 protein to the plasma membrane H(+)-ATPase AHA2 involves the three C-terminal residues Tyr(946)-Thr-Val and requires phosphorylation of Thr(947). *J. Biol. Chem.* **274,** 36,774–36,780.
2. Stensballe, A., Andersen, S., and Jensen, O. N. (2001) Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* **1,** 207–222.
3. Nühse, T. S., Stensballe, A., Jensen, O. N., and Peck, S. C. (2003) Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol. Cell Proteomics* **2,** 1234–1243.
4. Nühse, T. S., Stensballe, A., Jensen, O. N., and Peck, S. C. (2004) Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. *Plant Cell* **16,** 2394–2405.
5. Blagoev, B., Ong, S. E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* **22,** 1139–1145.
6. Peck, S. C., Nühse, T. S., Hess, D., Iglesias, A., Meins, F., and Boller, T. (2001) Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. *Plant Cell* **13,** 1467–1475.
7. Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R., and Roepstorff, P. (1999) Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* **34,** 105–116.
8. Haydon, C. E., Eyers, P. A., Aveline-Wolf, L. D., Resing, K. A., Maller, J. L., and Ahn, N. G. (2003) Identification of novel phosphorylation sites on *Xenopus laevis* Aurora A and analysis of phosphopeptide enrichment by immobilized metal-affinity chromatography. *Mol. Cell Proteomics* **2,** 1055–1067.
9. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., et al. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae. Nat. Biotechnol.* **20,** 301–305.
10. Posewitz, M. C. and Tempst, P. (1999) Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* **71,** 2883–2892.

36

Metabolite Profiling in *Arabidopsis*

Oliver Fiehn

Summary

Metabolite profiling is the multiparallel relative quantification of a mixture of compounds or compound classes using chromatography and universal detection technologies (GC-MS, LC-MS). In this respect it is an extension of classical single-target methods from which it can be distinguished by its broader view on profiling major biochemical events. This broader scope of analysis outweighs the disadvantages by making compromises in method development and the reduced accuracy for specific metabolites. This chapter exemplifies the strategies in metabolite profiling of polar compounds by gas chromatography-mass spectrometry (GC-MS). It gives experimental details on the basic steps: harvest, homogenization, extraction, fractionation, concentration, derivatization, data acquisition, raw data processing and result data tranformation.

Key Words: Mass spectrometry; GC-MS; metabolomics; data mining.

1. Introduction

Metabolite profiling is an analytical method for relative quantitation of a number of metabolites from biological samples ***(1)***. Commonly, these samples have been garnered from a specific tissue or a part of a tissue of interest, but, depending on the biological question, also either from a larger mixture of different organs (such as whole shoots) or, conversely, on a micro scale from single cells or purified organelles. Metabolite profiling is distinguished from other analytical procedures by its scope: (1) Target analysis is constrained to one or a very few target compounds (such as phytohormones). Such targets are usually quantified in an absolute manner using calibration curves and/or stable isotope labeled internal standards. (2) metabolite profiling restricts itself to a certain range of compounds or even to screening a predefined number of members of a compound class. Within these constraints, a single analytical platform may be sufficient. Examples might be the analysis of xantophyll cycle intermediates by high-performance liquid chromatography/diode array ultraviolet detection (HPLC-UV), or sugars and hydroxy- and amino acids by fractionation and gas chromatography/mass spectrometry (GC-MS), or membrane lipid profiling by HPLC-MS/MS. Quantification in metabolite profiling is usually carried out relative to comparator samples, such as positive and negative controls. (3) Metabolomics seeks a truly unbiased quantitative and qualitative analysis of all biochemical intermediates in a sample. It must not be restricted by any physicochemical property of the metabolites, such as molecular weight, polarity, volatility, electrical charge, chemical structure, and others. Since there is currently no single technology available that would allow such comprehensive analysis, metabolomics is characterized by the use of multiple techniques and unbiased software. Metabolomics also uses relative quantification. In addition, it must include a strong focus on de novo identification of unknown metabolites whose presence is demonstrated. (4) Metabolite fingerprinting is different from the other three approaches in that it does not aim to physically separate individual metabolites. Instead, spectra from full

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

sample extracts are acquired by a single instrument (such as ^{1}H-nuclear magnetic resonance [NMR]). Spectra are then compared by multivariate statistics in order to find spectral regions that discriminate samples by their biological origin. In some instances, these regions may again point toward specific metabolites; in general, however, one-dimensional methods are restricted in resolving complex mixtures.

Metabolite profiling therefore stands between classical target analysis and cutting-edge metabolomics. As it often aims at chemically very different compounds, there are various methods published for acquiring metabolite profiles from a certain tissue. Each procedure will be a compromise among several parameters such as compound stability, solubility, influence of the cellular matrix, time needed to carry out the protocol, available devices for tissue harvesting, homogenization, extraction, fractionation, submission to analytical instruments, raw data analysis and statistics. For example, a protocol found to be well suited for the analysis of lipophilic leaf membrane lipids and cuticle waxes will be very much different from one that aims at hydrophilic sugars and amino acids. Any protocol on metabolite profiling will therefore have validation criteria that are different from target analysis: (1) the reproducibility of the method is far more important than the absolute recovery; (2) the robustness and achievability of a method is more important than its absolute precision; (3) the comprehensiveness of a method is more important than the inclusion of a certain metabolite that is missed; (4) the overall dynamic range for the majority of compounds is more important than the detection limit for a specific substance; and (5) the ability to include important known key metabolites is more important than the detection of unidentified peaks that might be biochemical side-products of enzymes with low substrate specificity. In this respect, the most important validation criterion for a metabolite profiling protocol is the exact definition of its scope ***(2)***, and here, the identity and analytical reproducibility for each selected compound in a given biological matrix.

In this chapter, the acquisition of metabolite profiles for hydrophilic and semipolar compounds is presented. It is based on the use of inexpensive and robust technologies (such as GC/quadrupole MS) that are found in many biological laboratories, and not on more sophisticated, pilot-type instrumentation. The basic steps in the process (**Fig. 1**) can be summarized as:

1. Design randomized plots for plant growth, with an experimental design according to the biological question.
2. Harvest and weigh the plant tissue quickly and immediately freeze it in liquid nitrogen.
3. Homogenize the plant tissue in the frozen state.
4. Extract the tissue in a comprehensive and mild way concomitant with enzyme inactivation and add internal standards.
5. Fractionate the extract into a polar and a lipophilic fraction.
6. Dry down the polar fraction.
7. Derivatize the polar fraction by first adding methoxyamine pyridine, and then a trimethylsilylating agent.
8. Analyze the derivatized sample by GC-MS.
9. Process the raw GC-MS data.
10. Normalize and transform the result data, and perform statistical evaluations.

The basic theoretical considerations behind this process are quite simple: The measured metabolite levels should reflect the in vivo state. Therefore, any artifactual formation by chemical processes, or any postharvest biochemical activity, must be prevented. Biochemical inactivation can be ensured by coagulation of enzymes, using either heat-shock or cold-shock methods, with the help of organic solvents such as chloroform or acetonitrile that force protein precipitation. Conversely, chemical artifact formation depends on the stability of each specific compound and is therefore hard to predict. Generally, any harsh treatment of the metabolome mixture should be avoided. Instead, conditions for extraction, storage, chemical derivatization, and analysis should be as mild as possible, and, on the other hand, as comprehensive and universal as possible.

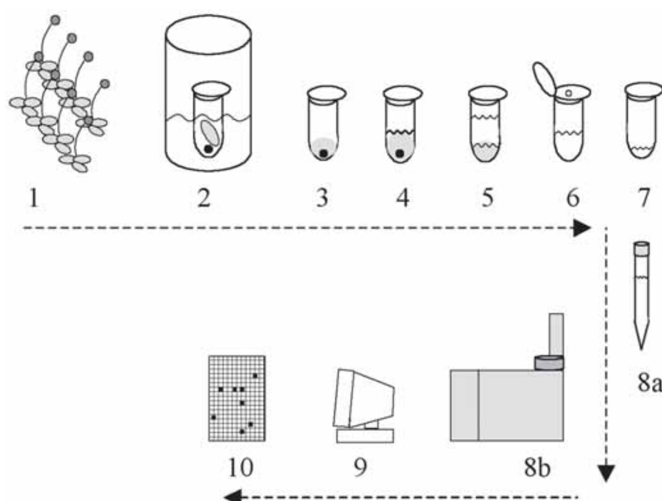


Fig.1. Scheme of the process of metabolite profiling. (1) Statistical design of plant growth, (2) harvest into liquid nitrogen, (3) homogenization, (4) extraction, (5) fractionation, (6) concentration to dryness, (7) derivatization, (8a) transfer to GC-MS vial, (8b) GC-MS data acquisition, (9) raw data processing, (10) data matrix transformation and statistics.

2. Materials

2.1. Harvesting Plant Tissues

1. Digital camera.
2. Scissors and tweezers.
3. Balance with a precision of ±0.1 mg.
4. Round-bottomed, uncolored, and prelabeled 2-mL microcentrifuge tubes, equipped with 5-mm id metal balls.
5. Liquid nitrogen.

2.2. Homogenization and Extraction

1. Ball mill (available, e.g., from Retsch, Germany).
2. Degassing device (such as vacuum/ultrasonic bath, or pure argon or nitrogen gas bombs).
3. Liquid cooling system.
4. Extraction mixture: freshly prepared, chilled (–15°C) and degassed mixture of chloroform, methanol, and water (1:2.5:1 [v/v/v]). Each solvent should be >99% ultrapure HPLC-MS gradient-grade purity. Stored the mixture at room temperature in the dark.
5. Liquid nitrogen.

2.3. Derivatization

1. Speed vacuum concentrator (e.g., Heraeus, Germany).
2. Methoxyamine HCl (20 mg/mL in pyridine) freshly prepared.
3. Silylating agent: N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) from freshly opened 1-mL bottles.
4. Thermoshakers at 28°C and 37°C.

2.4. Mass Spectrometric Analysis

1. Quadrupole mass spectrometer equipped with autosampler and electron impact ionization.
2. Quality control samples (e.g., reagent blanks, method blanks, reference compound mixture, reference sample).

3. Standard glass liners with glass wool.
4. 10-µL injection needles (e.g., Hamilton).
5. 30-m Long, 0.25-mm id, and 0.25-µm (5%-phenyl)-methylpolysiloxane column with an additional 10 m integrated guard column.

3. Methods

*3.1. Harvesting (*see Note 1*)*

1. Take digital pictures of the plants of interest, indicating the plants that are going to be harvested.
2. Cut the organs of interest with scissors or a cork borer.
3. Collect the plant organ with tweezers and weigh it. The range, which is validated with respect to the amount of 1-mL extraction solvent, is 10 to 50 mg FW.
4. Store the organ in a labeled 2-mL microcentrifuge tube that contains a metal ball.
5. Close the tube and place it into liquid nitrogen.

*3.2. Homogenization (*see Note 2*)*

1. Prechill two ball mill tube holders in liquid nitrogen.
2. Quickly take out six samples from liquid nitrogen and place them into tube holders (three samples per holder).
3. Fix tube holders in ball mill and mill at 25/s for 30 s.
4. Quickly take out the six samples and return them to liquid nitrogen.
5. Repeat **steps 1–4** until all samples are homogenized.

*3.3. Extraction and Fractionation (*see Note 3*)*

1. Take out homogenized samples one by one and add 1 mL of cold extraction mixture (–15°C, degassed) to each (*3*).
2. Add internal standards, e.g., U-13C-Sorbitol (200 ng per vial) for normalization.
3. Shake the samples in batches of 10 for 5 min in a 4°C room. When taking out the samples, place them in an ice bath.
4. Centrifuge samples at 20,800 rcf for 2 min.
5. Collect the liquid supernatant of each sample and store in a clean micro centrifuge tube. The metal balls can be reused after cleaning. The cell debris pellet can be discarded.
6. Repeat **steps 1–5** until all samples are extracted.
7. For storage, samples must be deoxygenated with a gentle stream of nitrogen or argon gas for 1 min prior to tube closure. Tubes can then be stored in the dark at –80°C for about 4 wk.
8. Add 400 µL of pure water to each sample, and vortex for 10 s.
9. Centrifuge samples at 20,800 rcf for 2 min.
10. Collect the upper phase of each sample (a mixture of water and methanol, the 'polar phase') and store it in a new microcentrifuge tube. The lower phase (a mixture of chloroform and methanol, 'lipophilic phase') can be used for lipid metabolite profiling or discarded.
11. Close the polar phase sample tubes with spare tube caps that have been punched with 50-µm *od* needles.
12. Dry the polar phase samples in a speed vacuum concentrator to complete dryness and remove the punched tube caps.
13. For storage, deoxygenate samples with a gentle stream of nitrogen or argon gas for 1 min before closing the tubes. Tubes can then be stored in the dark at -80°C for at least four weeks.

*3.4. Derivatization (*see Note 4*)*

1. Take out dried samples from store and allow them to warm up to room temperature for at least 15 min before opening.
2. Add 20 µL of methoxyamine solution (20 mg/mL in pyridine) to each sample, and immediately close each sample.
3. Shake samples for 90 min at 28°C.
4. Centrifuge samples at 14,000 rcf for 30 s.

5. Add 180 µL silylating agent MSTFA to each sample, and immediately close each sample after methoxyamine addition.
6. Shake samples for 30 min at 37°C.
7. Transfer sample reaction solutions to glass vials suitable for the GC-MS autosampler. Immediately close each sample with crimps that contain a Teflon rubber seal. Wait 2 h before injecting the first sample into the GC-MS.

*3.5. Data Acquisition by GC-MS (***see Note 5***)*

1. The mass spectrometer must be tuned according to the manufacturer's manuals for optimal parameters for ion lenses, detector voltage and other settings. Usually, this can be performed in autotune operation.
2. Change or clean the liner every 50 samples.
3. Check that the manufacturer's recommended maintenance routines have all been carried out.
4. Inject 1 µL of each sample in splitless or split mode, depending on the metabolite concentrations and eventual signal-to-noise ratios in the GC-MS profiles. Injection temperature is set to 230°C. Injection programs must include syringe washing steps before and after the injection, a sample pumping step for removal of small air bubbles, and an air buffer for complete sample removal during injection.
5. Separate metabolites using a GC temperature ramping program. Reasonable values are: GC start conditions at 80°C, 2 min isothermal, ramp with 5°C/min up to 330°C, 5 min isothermal, cool down to initial conditions. The ion source should be turned off during the solvent delay.
6. Detect metabolites by setting the ion source filament energy to 70 eV. Scan a mass range of at least 83–500 Daltons, or 40 to 500 Daltons, if low mass-to-charge (m/z) fragment ions are to be recorded. At least two scans per second should be recorded in full scan mode.
7. Transfer raw GC-MS profile chromatograms to a server station.

*3.6. Data Analysis (***see Note 6***)*

1. For raw data processing, use appropriate software. First choice is the GC-MS manufacturer's software. For data deconvolution, the freely available software AMDIS is recommended (http://chemdata.nist.gov/mass-spc/amdis/) ***(4)***.
2. Define target peaks that are to be included in the metabolite profiles.
3. Define optimal peak finding thresholds and quantification ion traces for each target compound. Peak identifications must be carried out by matching retention indices and mass spectral similarity against a user-defined metabolite library (**Fig. 2**).
4. Quantify metabolite peaks by area of target ion traces. Export result peak tables of all chromatograms to a database or a PC office table calculation software (e.g., MS Excel 6.0).
5. Organize peak area results in a matrix of metabolites vs chromatograms.
6. Count the number of detected metabolites per chromatogram. If one or a few chromatograms show an unexplainable large deviation in the number of detected peaks, check the chromatograms visually and delete them from the result table.
7. For each target metabolite, count the number of chromatograms in which the metabolite could be positively identified. If one or a few metabolites show an unexplainable large deviation in the number of positive peak findings, check the chromatograms visually, especially for the thresholds that were used for peak finding. Delete metabolites from the result table that have many negative peak findings (missing values).
8. For each chromatogram, divide all peak areas by the area of the internal standard (e.g., U-^{13}C-sorbitol) and the sample weight. Log10 transform all data to reduce weight outliers and ensure a more Gaussian-type frequency distribution.
9. Calculate univariate statistics (e.g., *t*-test in Excel, analysis of variance [ANOVA] in MatLab).
10. Calculate multivariate statistics. Often, such calculations do not accommodate missing values for metabolites so suitable strategies must be employed for dealing with such occurrences. The results of multivariate statistics from two strategies must be compared: (a) calculations that were carried out on all metabolites that had no missing values; (b) calculations that were carried on all metabolites after replacement of missing values.

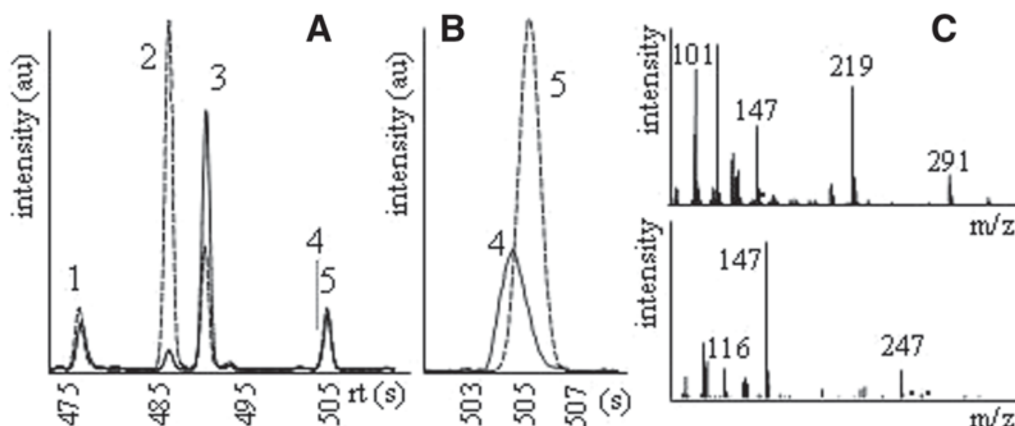


Fig. 2. Example of differential analysis of two *Arabidopsis thaliana* accessions under cold stress situation by GC-MS metabolite profiling for polar leaf extracts. Total ion chromatogram for overview purposes **(A)**: solid line: Cvi; dotted line: C24. Peak #1 = internal standard, #2- fumarate, #3-serine, #4 threonate lactone, #5 threonine. Ion trace chromatograms **(B)**: quantification of peak #4 at m/z 247 (solid line) is achievable despite coelution of abundant peak #5, which can be quantified at m/z 320 (dotted line). Mass spectrum at retention time 504.5 s **(C)**: betabolite identification before (upper panel) and after (lower panel) peak deconvolution. Automated peak finding and identification of low-intensity peak #4 as threonate lactone (lower panel) is impossible without mass spectral deconvolution, owing to coelution of the highly abundant peak #5 (threonine) at this retention time (upper panel).

4. Notes

1. Metabolite profiling starts with the experimental design of plant growth and harvest. It is a rather inexpensive technique, compared with proteomics or transcript expression analysis. Therefore, larger numbers of individual analyses can be carried out, allowing rigid statistical assessments of the quantitative results. This permits the adequate addressing of the issue of natural biological variability, which usually contributes more to the standard deviation of metabolite mean levels than technical errors. In this respect, the first issue to consider is the randomization of plant growth and the accuracy of controlling and monitoring physiological conditions. Plants should be grown in complete randomization or in appropriate block designs, e.g., Latin square. Although biological variation has been known for a great length of time, it is often underestimated due to the tradition in other branches of biology aiming at average values and qualitative yes/no results *(5)*. For example, metabolite levels vary dramatically over different zones of a leaf, and also between young, growing leaves and mature, fully expanded leaves. Depending on the biological question underlying a study, pooling strategies must be designed to counteract this variability. For example, all rosette leaves of a single plant might be pooled and compared with other plants, or, alternatively, small disks of many individual plants could be collated and analyzed in batches. Similarly, plants grown in climate chambers or greenhouses should be grown in plots that are adopted from agronomical field trials. The reasoning is that microclimate conditions in climate chambers are not uniformly distributed, especially with respect to air circulations that lead to location effects by differences in individual plant transpiration rates.
2. For comprehensive extraction, a complete breakdown of plant cell walls is needed to ensure that the extraction solvent reaches all metabolites, independent of subcellular organization. Plant leaves include some 30 different cell types, all of which may have different internal metabolite levels. Homogenization therefore enables randomization over organs, which is a prerequisite for repeatability of analysis. For the scheme presented here, milling of frozen Arabidopsis leaves with a ball mill is proposed. However, for other plants or other organs, ball

milling may not be sufficiently disruptive. Then, other devices may be used, e.g., an Ultra-Turrax, that disrupt tissues with small rotating razor blades. Generally, also a simple pestle and mortar may be used (under liquid nitrogen). However, the throughput of this procedure is believed to be too low for preparing enough samples for accurate statistics. Still, if a ball mill or an Ultra-Turrax is not available, pestle and mortar are definitely regarded as valid tools for homogenization.

If you are using a ball mill, certain pitfalls should be avoided: first, it is wise to have more than one dewar available. Use one for chilling the tube holders, one for the samples that have not been homogenized, and one for samples that are already done. If you don't want to lose your order of samples, you may also use paper boxes with separating inserts. The boxes would then be placed on dry ice and be filled with liquid nitrogen. The most important factor during homogenization is that the samples must not thaw. For this reason, each ball mill tube holder is filled not with five but with only two or three samples; it is simply quicker and more convenient. If tubes are milled at too high a frequency (>25/s) the balls might disrupt the microcentrifuge tube caps, resulting in lost samples.

3. Extraction is the most critical step in metabolite profiling. As stated above, it makes compromises among the demands for best possible metabolite recoveries, total metabolic comprehensiveness and time needed to perform the extraction. The protocol suggested here puts more emphasis on comprehensiveness and throughput, and less on recovery. For example, each sample might be extracted two or three times instead of using a single extraction step. As given by Nernst's law, a duplicate or triplicate extraction would give a better recovery and accuracy. However, more time would be needed per sample. Using the proposed single-extraction method, repeatability was found to be around 10% CV for most compounds. This is ensured by a well-defined ratio of solvent volume to fresh weight of tissue (50:1). Comprehensiveness of extraction is given by the simultaneous use of highly polar solvents (water) and highly hydrophilic solvents (chloroform), with methanol being the mediator to avoid phase separation. Note that chloroform will diffuse through microcentrifuge tube plastic faster than methanol or water will, if samples are stored for prolonged periods at –80°C. Chloroform/methanol at low temperatures will also help precipitating proteins, thereby ensuring the integrity of the metabolic composition. As important as stopping any enzymatic activity is avoiding oxidation. Solvents will contain huge amounts of oxygen if they are not degassed by vacuum/ultrasonicator or by bubbling inert gases through it (argon or nitrogen are most convenient). If deoxygenation is performed by gas exchange, great care must be taken to use ultrapure gases and clean bubble tips (e.g., rinsed Pasteur pipet). Somewhat less important is the avoidance of light: some metabolites, such as catecholamines, will decompose if exposed to light for too long. For *Arabidopsis* leaves, polar metabolite profiles will tell you whether you have carried out the protocol in a safe way, avoiding biochemical or physical alterations of metabolite compositions: (a) Low-abundance cysteine must be present. It will disappear if there is oxygen left in the solvents. The accuracy of redox state preservation may also be checked by comparing the ratio of ascorbate to dehydroascorbate using this protocol and target assays for these compounds. (b) Low-abundance glucose-6-phosphate and fructose-6-phosphate must be present. These compounds will disappear if enzymatic activity is not immediately stopped, and also through prolonged high-temperature exposure (if heat-shock enzyme inactivation is used). (c) Classical compounds such as fumarate, malate, citrate, serine, threonine, aspartate, glutamate, glucose, and sucrose tend to represent the most abundant peaks in *Arabidopsis* profiles. If any of these are missing (or are in low abundance in comparison with others), the analysis has gone badly wrong.

   When drying samples in a speed vacuum concentrator, caution should be taken to avoid sample losses, spilling, or cross-contamination due to boiling retardation. This is the reason why extracts are dried with puntured plastic tube caps.

4. The most critical point is to avoid any water or moisture during derivatization. The silylating step is highly especially vulnerable. Problems can be detected through occurrence and abundance of polysiloxanes in GC-MS chromatograms. Such degradation (hydrolysation) products are recognized by their typical spectra with abundant ions m/z 221 and m/z 281. Generally, it is not needed to perform the derivatization in completely dry atmosphere. However, water

condensation due to early opening of sample tubes after cold storage must be avoided, as well as storage of derivatized samples in refrigerators or freezers. If, through bad luck, samples cannot be injected after derivatization, they should be stored in the dark at room temperature. After injections, sample vials will have received an amount of water, and seals may have been compromised. Reanalysis of samples that have already been used is not recommended, therefore. Temperatures and times of derivatization steps can be kept flexible, because they again present a compromise between completeness of reaction, time and efforts needed to perform the reactions, and breakdown of certain compounds (e.g., chemical conversion of glutamine to oxoproline). Pyridine serves as a catalyst in the methoximation procedure, which protects carbonyl moieties. It does not seem to be replaceable by other aprotic polar solvents. The volume ratios of pyridine methoxyamine to MSTFA are again flexible: we here propose a ratio of 1:9, but other ratios such as 1:2 or 1:1 have also been reported in the literature. Generally, lower amounts of pyridine will give better peak shapes for early eluting metabolites if splitless injections are carried out.

5. Take care to randomize your injection sequence: you must not inject the samples in the order of your underlying biological question, because there might always be subtle machine drifts that would obscure statistical analysis. Problems that are seemingly attributed to the machine are usually neither the gas chromatograph nor the mass spectrometer (6): in >80% of cases the injection is to blame. Here, problems may occur due to dirt ("matrix") injected into the liner, the injector body, and the first centimeters of the column. As for the column, problems may be recognized by decreasing intensity of trisaccharides. Shortening the column by 10 cm will help, but take care to readjust the total length of the column in the GC-MS front end, as this value will be taken to adjust the gas flow. With respect to the liners, glass wool will prevent the majority of nonvolatile matrix constituents from reaching the injector body or the column. However, ultimately the injector body itself will also become contaminated, which will pyrolyse and build up spots with catalytic or adsorption properties, disabling high-quality GC-MS runs. The level of contamination in the injector body is related to the type of liner used, and also to the presence of so-called cold spots in the specific injector type. A prolonged heating at 330°C for 6 h (without column) will cure this problem in most instances. If you plan long sample sequences that may even involve column changes, you need to refer to retention indices instead of retention times. Retention indices are calculated from retention times of internal marker peaks: usually alkanes are added to the samples to serve as retention anchor points. In the protocol presented above, an unusual scan range of 85 to 500 Daltons is proposed for mass spectrometric detection. Reasons for this choice are found in the properties of silylated compounds, which often have characteristic ions between 100 and 370 Daltons. Additionally, at m/z 79, bleeding of pyridine may infer mass spectra of low-boiling compounds. For almost all peaks, m/z 73 is the most abundant ion–although this is helpful for lower limits of detection for pure compounds, m/z 73 does not have any selectivity power in metabolic profiles.
6. Metabolite profiles normally result in complex chromatograms that contain numerous overlapping peaks. For ensuring routine peak identification in high-throughput operations (i.e., without manual inspection of chromatograms), mass spectral deconvolution is mandatory, especially for low-abundant metabolites that might coelute with abundant major peaks (**Fig. 2**). Deconvolution software will also suggest model ions that best discriminate a peak from its coeluting neighbor compounds; hence, such model ions are already a good choice for calculating peak areas. Defining thresholds for peak finding is a difficult task. If the thresholds are set too high, many peaks will not be taken as metabolite targets, although these are actually present in the chromatograms: such instances are called false negatives and would result in missing values in the resulting experiment data matrix. Vice versa, if the thresholds for mass spectrum matching, retention index windows, abundance, and peak widths are set too low, peaks might be falsely taken as true target metabolites although these are actually be absent from the chromatograms (false positives). In any way, as much metainformation about a peak should be acquired as possible to ensure correctness of peak annotations in metabolite profiling databases. Arguably, the resulting data matrix will still contain missing values. Certain statistical tools such as principal component analysis will require complete data matrices without missing values; therefore, these must be filled. Whatever is filled into these missing value cells in an automatic mode will carry a

larger error than truly detected targets that passed all peak finding thresholds. Therefore, multivariate statistical analysis should be done twice: first with all metabolites that do not contain a single missing value, and second also including metabolites for which empty cells have been replaced. In literature, there are many ways proposed of how to best perform such replacements. For less sophisticated approaches, a simple rationale might be to replace missing values by the arithmetic mean of each line (i.e., wild-type control line, mutant 1, etc.). For cases in which a target metabolite is positively detected fewer than 20% of the cases or in no chromatogram of the corresponding line at all, it may be suspected that it is indeed not present in this line and the actual peak findings are false positives. In this case, entering half the detection limit might be a sensible method. In any case, the better way to replace missing values is to investigate the chromatograms one by one and replace the empty cells in the data matrix manually.

Last, the protocol proposed here suggests using the log10 transformation for down weighting outliers and transforming the data matrix to a more normal frequency distribution. In transcript microarray experiments, often the natural logarithm is taken, but there are no theoretical considerations that clearly vote for one or the other alternative. However, in any case it must not be forgotten that data need to be retransformed when *x*-fold average values are to be computed, e.g., in mutant/wild-type line comparisons.

Acknowledgments

Leaf extracts from an *Arabidopsis* cold stress experiment were kindly provided by Dana Wiese and Dr. Dirk Hincha, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany. The author thanks Dr. Gareth Catchpole for proofreading and manuscript corrections.

References

1. Fiehn, O. (2002) Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **48,** 155–171.
2. Krull, I. S. and Swartz, M. (1999) Analytical method development and validation for the academic researcher. *Anal. Lett.* **32,** 1067–1080.
3. Weckwerth, W., Wenzel, K., and Fiehn, O. (2004) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* **4,** 78–83.
4. Stein, S. E. (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass. Spectrom.* **10,** 770–781.
5. Molloy, M. P., Brzezinski, E. E., Hang, J., McDowell, M. T., and van Bogelen, R. A. (2003) Overcoming technical variation in quantitative proteomics. *Proteomics* **3,** 1912–1919.
6. Oehme, M. (1998) *Practical Introduction to GC-MS Analysis with Quadrupoles*, 1st ed. Hüthig, Heidelberg, Germany.

37

Hormone Profiling in *Arabidopsis*

Axel Müller, Petra Düchting, and Elmar W. Weiler

Summary

A highly sensitive and accurate multiplex gas chromatography-mass spectrometry (GC-MS/MS)-technique is reported for the quantitative analysis of acidic phytohormones in *Arabidopsis thaliana* and other plant species. The optimized setup allows the routine processing and analysis of between 20 mg and 5 g of tissue. The protocol was designed and the equipment used was chosen to facilitate implementation of the method into other laboratories and to provide access to state-of-the-art analytical tools for the acidic phytohormones and related signaling molecules.

Key Words: Abscisic acid; chemical-ionization tandem mass spectrometry (CI-GC-MS/MS); diazomethane; indole-3-acetic acid; jasmonic acid; 12-oxo-phytodienoic acid; salicylic acid; solid-phase extraction microcolumn; stable isotopes; phytohormones; plant hormone analysis.

1. Introduction

The regulatory roles of hormones in plant growth and development have been known for many decades. It is nevertheless still very challenging to identify and quantify phytohormones in plant tissues. Using molecular genetic approaches, large collections of mutants and transgenic plants have been generated over the last two decades to study the functions of hormones and hormone signaling in much detail. This work has made it increasingly clear that many plant genes respond to more than just one hormone. Likewise, many physiological responses against changing environmental conditions cause changes in concentrations of several regulators and hormone biosynthetic pathways are regulated by other hormones, often several ones. This has led to the conclusion that extensive crosstalk occurs in hormone metabolism and signaling. Consequently, it is no longer advisable to study just a single hormone. Rather, focus must be laid on the whole group synoptically. This requires multiplex techniques for hormone analysis and quantitation that can cope—if required—with small tissue samples, i.e., those that are intrinsically highly sensitive ***(1–4)***.

Methods for hormone profiling nowadays use mass spectrometrical (MS) techniques that, when combined with the use of isotopically labeled internal standards, allow not only the unambiguous identification but also the precise quantitation of multiple analytes from one sample. Modern instruments possess tandem MS (MS/MS or MS^n) capabilities. By selection of a characteristic parent ion in the first stage of the MS process, fragmentation of this ion into daughter ions and recording of the daughter ion spectra, these instruments yield better separation of signal from matrix noise and allow generation of clean spectra from samples of a lower degree of purity than would be required for single-stage mass spectrometers. Group-selective protocols for sample preparation, an important prerequisite for any efficient multiplex process, can be used only in this case.

Ideally, for multiplex analysis, one would like to analyze a primary extract directly in order to avoid any discrimination in recoveries among a group of analytes due to differential losses

From: *Methods in Molecular Biology, vol. 323:* Arabidopsis *Protocols, Second Edition*
Edited by: J. Salinas and J. J. Sanchez-Serrano © Humana Press Inc., Totowa, NJ

during workup. This is not possible, however, and, thus, the main logistic problem to be solved in multiplex analysis is to develop protocols that, in a minimal number of efficient and simple steps, separate the analytes from any matrix contamination that will interfere in the analysis while, at the same time, being totally or nearly indiscriminate for all analytes of interest. Multisample processing capacity is another point high on the agenda, as is the suitability for processing microsamples, i.e., amounts of tissue in the milligram range.

The technique described in this chapter provides access to the parallel and quantitative analysis of several acidic phytohormones and related regulators in small tissue samples. It has been used successfully for the following plant species: *Arabidopsis thaliana, Oryza sativa, Nicotiana tabacum, Zea mays, Solanum tuberosum, Hordeum vulgare*, and *Pyrus communis* (leaves, roots, and seeds). The procedure yields quantitative data for the following phytohormones and related compounds: jasmonic acid, 12-oxo-phytodienoic acid, abscisic acid, indole-3-acetic acid, and salicylic acid (*see* **Note 1**). Many other acidic metabolites can also potentially be analyzed by the described technique. Others, such as the gibberellins, will require modifications of the procedure, which are beyond the scope of this chapter. For the analysis of the basic cytokinins and hormone conjugates, the reader is referred to the specialized literature ***(5–11)***.

2. Materials

2.1. Solvents and Reagents

The following reagents should be used in pro analysis or high-performance liquid chromatography (HPLC) grade:

1. Acetic acid.
2. Choroform.
3. Diethyl ether, dry, free of peroxides.
4. Ethyl acetate.
5. 36-38% (w/v) HCl.
6. Isohexane.
7. 2-Propanol.
8. Potassium hydroxide.
9. *N*-Nitrosomethyl urea.

2.2. Equipment

1. Centrifuge and corresponding rotors (Heraeus Biofuge Pico).
2. Speed-vac concentrator (Eppendorf Concentrator 5301).
3. Membrane vacuum pump (Vacuubrand MD 4C).
4. Ball mill with Eppendorf cup adaptor (Retsch MM300, modified and purchased from Qiagen, Hilden, Germany).
5. Stainless steel balls, 3 mm diameter (Retsch, Haan,Germany).
6. Ultrasonic bath (Sonorex RK 510S, Bandelin, Berlin, Germany).

2.3. Isotopic Labeled Internal Standards

The following internal standards are recommended:

1. $[^{2}H]_{4}$-salicylic acid ***(1)***, available from CDN Isotopes.
2. $[^{13}C]_{2}$-jasmonic acid ***(1)***.
3. $[^{2}H]_{2}$-indole-3-acetic acid, available from Isotec (Sigma-Aldrich).
4. $[^{2}H]_{6}$-abscisic acid ***(1,12,13)***.
5. $[^{2}H]_{5}$-12-oxophytodienoic acid ***(1)***.

It is possible to use other standards (*see* **Note 2**), provided (1) the isotopic enrichment is above 96%, (2) the standard is stable and does not detectably exchange the heavy isotope during sample processing and analysis, and (3) the mass difference between the standard and the analyte is not less than 2 AMU.

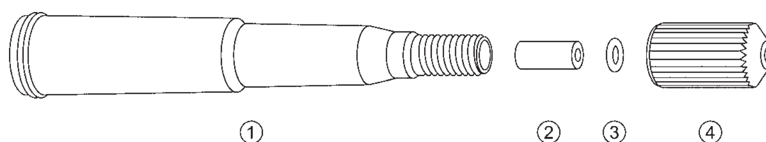


Fig. 1. Fabricating a capillary holder for mechanical pipets. The distal 3 cm of the shaft of a 1-mL Gilson Pipetman pipet are cut away and an M8 or similar external (male) thread is applied to the shortened shaft with a screw cutting machine (1). A 1-cm-long cylindrical plastic insert (4.9 mm od) bearing a central drilled hole of 2.5 mm id (2). This insert is glued tightly into the threaded end of the pipet shaft and serves as the guide for the glass capillaries. At the modified tip of the shaft, a shallow concave notch is then drilled to center the O-ring (3) when compressed. A cap nut with the appropriate internal (female) thread and a central borehole of 2.5 mm diameter (4) compresses the rubber O-ring (6 mm od, 2 mm id) when screwed up. The device ensures gas-tight sealing of the inserted capillary.

2.4. Assembly of Solid-Phase Extraction (SPE) Microcolumns

2.4.1. Devices and Materials

1. Cylindrical glass capillaries for the fabrication of SPE microcolumns (150 × 1.5 mm i.d., Hirschmann Laborgeräte, Eberstadt, Germany).
2. Cylindrical noncoated polyolefine plugs (2.5 × 1.5 mm i.d; Filtrona, Reinbek, Germany).
3. Polygoprep60-12NH2 silica-based aminopropyl matrix (Macherey-Nagel No. 711012.100, particle size 12 µm, Macherey-Nagel, Düren, Germany).
4. Capillary holder, custom-made by modification of the shaft of a Gilson P1000 Pipetman® mechanical pipet, allowing tight mounting of the glass capillaries (**Fig. 1**).
5. Filling capillaries are prepared by manually pulling Pasteur pipets (**Fig. 2**). The filling capillaries are designed to fit into the capillary holder (*see* **Subheading 2.4.1.**, item 4) for handling the SPE microcolumns.

2.4.2. Procedure

1. Compress one polyolefin plug on one side with a pair of tweezers and insert it 1.7 cm deep into the glass capillary (*see* **Subheading 2.4.1.**, item 1). Fire-polish the other end of the capillary and fix this end in the capillary holder.
2. Prepare a suspension using equal volumes of Polygoprep60-12NH2 and diethyl ether.
3. Dip the tip of the capillary into the swirled suspension and slowly draw the suspension into the tip until the capillary is filled with matrix up to the plug.
4. Disconnect the capillary from the holder, and place it, filled end upside, into a beaker. Allow the ether to evaporate (15 min).
5. Hit the capillary (with filled side up) on the table to compress the SPE matrix (2–3 mm of the top of the capillary should now be free of matrix).
6. Insert a second polyolefin plug (compressed on one side) into the tip of the capillary to secure the matrix.

The SPE microcolumn with a gel-bed of 15 mm × 1.5 mm id (**Fig. 3**) is washed with 0.5 mL diethyl ether (**Subheading 3.2.**) or isohexane (**Subheading 3.6.**) and is now ready to use. It should be used only once (*see* **Note 3**).

3. Methods

3.1. Preparation of Ethereal Diazomethane

1. Add 20 mL of 10 *N* aqueous KOH to 100 mL peroxide-free diethyl ether in an Erlenmeyer flask and chill for 30 min in an ice/salt mixture.
2. Add 2 g of *N*-nitrosomethyl urea and stir slowly (use PTFE magnetic stirrer bars) without cooling until *N*-nitrosomethyl urea has reacted. The ether phase acquires a yellow color (*see* **Note 4**).

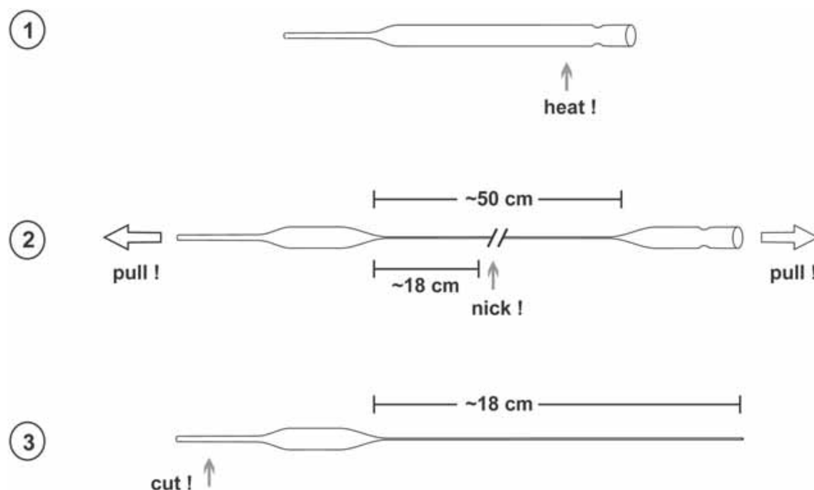


Fig. 2. Fabricating a filling capillary. (1) Heat glass near the wide end of a Pasteur pipet over a Bunsen flame until becoming soft. Take the heated pipet out of the flame. (2) Quickly and with a single stroke pull out the glass to a length of approximately 50 cm. Nick the drawn end of the capillary at about 18 cm from the base. The tip should be less than 0.05 mm wide. (3) Using a glass knife, shorten the original tip of the Pasteur pipet such that the remaining tip allows it to fit 1 cm deep into the capillary holder. This ensures gas-tight mounting into the holder. Flame-polish the cut end and let cool before insertion into the holder. The finished device has a microtip on one side, can be inserted with the other side into the capillary holder described in Fig. 1, and has a liquid reservoir that can be filled and emptied by suction using the mechanical pipet.

Fig. 3. SPE-microcolumn.

3. Decant the ethereal diazomethane onto 5 g of solid KOH and keep for a minimum of 30 min in an ice bath at a place without intensive light exposure (*see* **Note 5**).

3.2. Standard Workup Procedure

The single-stage purification protocol (**steps 1–9**) described here is designed for tissue samples ranging typically from 100 to 200 mg of fresh mass but not exceeding 400 mg (*see* **Note 6**). When larger amounts of tissue or seeds are to be analyzed, *see* **Subheading 3.4.** for adaptations of the standard procedure.

1. Freshly prepare the extractant: pure methanol containing the appropriate internal standards at the predetermined levels, which range mostly between 10 and 50 pmol/mL (*see* **Notes 7– 9**).
2. Collect the tissue (typically between 100 and 200 mg of fresh mass) in a 2 ml Eppendorf tube and add 1 mL of the extractant containing the standards (*see* **Note 10**), then add 4 stainless steel balls of 3 mm diameter and cap the tube.
3. Homogenize the tissue for 10 min (*see* **Note 11**) using a vibrating-ball micromill at a vibrating frequency of 25/s.

4. Remove the steel balls using a magnetic tool and centrifuge the sample for 5 min at 16,000*g*.
5. Decant the supernatant into a fresh Eppendorf tube and remove the solvent in a vacuum centrifuge at 60°C and 10 mbar (this requires ~60 min, depending on the total number of samples processed at the same time) (*see* **Notes 12, 13**).
6. Dissolve the dried crude residue in 30 µL of methanol and add 200 µL of diethyl ether and cap the tube. Brief sonication (5 min ultrasonic bath) will ensure fine dispersion of any particles. Remove any unsoluble material by centrifugation (2 min, 16,000*g*).
7. Using a filling capillary (**Subheading 2.4.1.**, item 5), transfer the cleared sample to a SPE-microcolumn equilibrated with diethyl ether (**Subheading 2.4.2.**). After passage of the sample through the column by gravity flow (*see* **Note 14**), the microcolumn is washed with 150 µL of $CHCl_3$: 2-propanol, 2:1 (by vol) (*see* **Note 15**), and the tip is cleaned on the outside using a strip of absorbing paper.
8. Elute the acidic hormones from the column into an fresh Eppendorf tube with 2 × 200 µL of diethyl ether containing 2% (by vol) of acetic acid.
9. Take the eluate to dryness in a vacuum centrifuge operating for 1 min at 100 mbar and then for a further 10 min at 10 mbar. Remove any residual acetic acid in a stream of nitrogen. Cap the tubes.

3.3. Variations of Standard Protocol to Cope with Impurities

When extracting lipid-rich or mucopolysaccharide-rich tissues, bulky residues may result after tissue extraction (**Subheading 3.2.**, **step 5**) which trap analytes and give low recovery rates. Also, such samples tend to clog the SPE-microcolumns. If these problems arise, use the following modifications after **step 4** of the standard procedure (**Subheading 3.2.**):

a. Decant the supernatant in equal parts into two clean Eppendorf tubes, determine the volume, and add, to each tube, equal volumes of 10 m*M* aqueous potassium hydroxide and chloroform. Close the tubes and briefly mix on a vortex shaker. Separate the phases by centrifugation (2 min, 16,000*g*). Discard the chloroform phases (lower phases). This removes lipids (*see* **Note 16**).
b. Add, to both tubes, another volume of chloroform and 10 µL of concentrated HCl (*see* **Note 17**) to lower the pH below 3.0 Close the tubes, mix, and centrifuge again. Collect and combine the chloroform phases (lower phases), which contain the acidic hormones, into one fresh Eppendorf tube.
c. Remove the solvent in a vacuum centrifuge operating for 1 min at 100 mbar and then for a further 10 min at 10 mbar. Remove any residual acid in a stream of nitrogen. Ensure that the sample is completely dried and no traces of water are left.

Continue with **step 6** of the standard procedure. If this refined protocol still does not yield interference-free MS data, follow the procedure in **Subheading 3.6.**

3.4. Variations of Standard Protocol to Cope With Larger Amounts of Tissue

The procedure in **Subheading 3.2.** is not intended to be used with samples of more than 400 mg of fresh tissue. The following modification of the standard procedure is used for amounts of tissue ranging from >400 mg to 5 g of fresh tissue.

Replace **steps 1–5** of the standard procedure (**Subheading 3.2.**) by the following:

1. Use as extractant 20 mL of methanol containing predetermined amounts of the required internal standards.
2. Collect soft tissue into the predelivered extractant in a 25-mL Erlenmeyer flask. Lignified tissue or seeds should be ground in liquid nitrogen before extraction (*see* **Note 11**).
3. Boil the sample for 2 min on a heating plate and store it for 1 h at room temperature.
4. Filter the supernatant through a standard paper filter and wash debris with additional 5 mL of methanol.
5. Adjust the methanolic extract to 20 mL volume, add equal volumes of water and chloroform as well as 50 µL concentrated HCl to bring the pH below 3.0 Collect the (lower) chloroform phase by using a separating funnel.
6. Dry the chloroform phase by adding 1 g of sodium sulfate, swirl, and filter the solution. Remove the solvent by vacuum evaporation using a rotary evaporator.

7. Dissolve the sample in 500 µL of methanol, transfer it to an Eppendorf cup, and remove the solvent in a vacuum centrifuge at 45°C and 10 mbar.

Continue with **step 6** of the standard procedure in **Subheading 3.2.** and optionally purify the methylated sample further using the protocol in **Subheading 3.6.**

3.5. Derivatizations Prior to GC-MS Analysis

All carboxyl-containing hormone compounds are analyzed by gas chromatograph/mass spectrometry (GC-MS) as methyl esters. These are prepared as follows:

1. Take the pre-cleaned residue of **Subheading 3.2.**, **step 9**, and dissolve it in 20 µL methanol (initial tissue samples up to 200 mg) or 50 µL methanol for larger samples and add 100 µL ethereal diazomethane solution (or 0.5 mL for larger samples). Diazomethane acts immediately.
2. Reduce the volume of the solvent to approx 50 µL in a gentle stream of nitrogen (this also destroys excess diazomethane), transfer the solution into a conical autosampler vial (Chromacol Uni-VI, Supelco 27312; *see* **Note 18**), and remove the rest of the solvent. Avoid spreading the sample across too much of the surface of the vial (*see* **Note 19**).
3. Dissolve the methylated sample in 7 µL chloroform and crimp the autosampler vials immediately to reduce evaporation. Avoid tilting of the ready-to-analyze samples.

3.6. Optional Purification of Methyl Esters

If GC-MS analysis of samples prepared according to **Subheading 3.2.**, **steps 1–9**, or **Subheading 3.3.** still shows a nonacceptable matrix background, the following extension of sample cleaning may be tried to eliminate the interference. For this protocol, use the methylated samples (**Subheading 3.5.**, **step 1**).

1. Remove the diethyl ether in a gentle stream of nitrogen and redissolve the methylated sample in 200 µL isohexane. Pass the solution by gravity flow through an SPE microcolumn equilibrated with isohexane (*see* **Subheading 2.4.2.**).
2. Wash the column with 50 µL isohexane.
3. Clean the tip of the microcolumn on the outside using a strip of absorbing paper.
4. Elute the bound methyl esters from the column with 200 µL ethyl acetate into an autosampler vial, and take the sample to dryness using a vacuum centrifuge at 30°C and 100 mbar.
5. Dissolve the residue in 7 µL chloroform and crimp the autosampler vial. Avoid tilting of the ready-to-analyze samples.

3.7. Gas Chromatography-Mass Spectrometry

Aliquots of 1 µL of each sample are injected into the GC-MS system for separation and mass fragment analysis.

3.7.1. GC Settings

Splitless injection (1 µL), splitter opening 1:100 after 1 min, injector temperature 260°C; separation on a ZB-50 fused silica capillary column (30 m, 0.25 mm id, 0.25-µm film thickness, Phenomenex, Aschaffenburg, Germany (*see* **Note 20**) using He carrier gas at 1 mL/min; temperature program: 1 min isothermally at 50°C, linear ramp at a rate of 40°C/min to 150°C, 6 min isothermally at 150°C, linear ramp at a rate of 20°C/min to 250°C, 4 min isothermally at 250°C; transfer line temperature 260°C.

3.7.2. MS Settings

The mass spectrometer should operate in CI-MRM mode with methanol as the reactant gas and positive ion detection. The method described is optimized for ion-trap mass spectrometers, set at a maximum reaction time of 128 ms, a maximum ionization time of 2 ms, a scan rate of 0.38 s/scan, a multiplier offset voltage of 300 V, an emission current of 30 µA and the resonant waveform type for MS/MS mode together with a parent-ion selection window of three atomic mass units. Parent-ion selections and excitation amplitudes are switched segmentally within each run to provide optimum fragmentation conditions for each analyte as it elutes.

Table 1
Characteristic Parent and Product Ions for Representative Hormones and Their Corresponding Internal Standards

Compound	Excitation voltage	Parent ion m/z	Product ion mass range m/z	Diagnostic product ion m/z
SA	0.6	153	50–70	121
$[^2H]_4$-SA		157		125
JA	0.5	225	110–250	207
$[^{13}C]_2$-JA		227		209
IAA	0.5	190	110–200	130
$[^2H]_2$-IAA		192		132
ABA	0.5	261	110–300	229
$[^2H]_6$-ABA		267		233+234
OPDA	0.5	307	270–315	275
$[^2H]_5$-OPDA		312		280

The indicated mass ranges selected for product ion recording contain the diagnostic product ion(s) used for quantitation as well as further characteristic ions which give information about the analyte as well as indicate the presence of potential impurities. Time segments for the elution of each analyte may vary caused by system dependent parameters and must be determined for the actual setup used by analysis of reference compounds (*see* **Note 20**).

The amounts of endogenous compounds are calculated from the signal ratios of the unlabeled over the corresponding stable isotope-containing mass fragments (**Table 1**).

4. Notes

1. The reported clean-up procedure is also useful for analyzing fatty acids such as linolenic acid, 1-aminocyclopropane-1-carboxylic acid, and gibberellins from plant extracts. Compounds containing hydroxyl and/or amino groups need additional derivatization and purification steps not described here.
2. $[^2H]_{2}$-, $[^2H]_5$, $[^2H]_7$, and $[^{13}C]_6$IAA, $[^2H]_6$-ABA, and $[^2H]_6$SA are commercially available from CDN-Isotopes (www.cdniso.com), ICON-Isotopes (www.iconisotopes.com), Cambridge Isotopes (www.isotope.com), Isotec (Sigma-Aldrich; www.isotec.com, www.sigma-aldrich.com), and Medical Isotopes (www.medicalisotopes.com). These and other labeled plant hormones are also sold by OlChemIm Ltd, Šlechtitelů 27, P. O. Box 22, 770 10 Olomouc, Czech Republic www.olchemim.cz. Labeled jasmonates are still not commercially available, and need to be synthesized by the researchers themselves.
3. Commercial microtips for solid-phase extraction are available, e.g., from Glygen Corp. Columbia, USA. Please note that the procedure given in this chapter has not been developed or optimized for such tips.
4. **Warning**: Strictly obey all safety precautions when preparing or handling ethereal diazomethane. All steps are to be carried out in a well-ventilated laboratory hood. Etheral diazomethane should be prepared from *N*-nitrosomethylurea recrystallized from methanol. Crude *N*-nitrosomethylurea is provided by Sigma-Aldrich. Diazomethane prepared from *N*-methyl-*N*-nitroso-*p*-toluenesulfonamide has proven less useful in our hands.
5. It is recommended to prepare diazomethane in quantities just sufficient for the immediate use. Excess diazomethane is destroyed by decanting the solution onto a few drops of acetic acid (the yellow color must disappear).
6. The procedure is suitable for manually processing up to 48 samples in parallel, all steps to be carried out at room temperature.
7. It is mandatory to determine the appropriate amount of standard in preliminary experiments for each tissue to be analyzed. In order to secure most accurate results, the amount of standard used should not be less than 1/5 and not more than five times the respective hormone present in the sample to be analyzed. Standards must be added, preferentially as a mix, to the extractant

before this, in turn, is added to the tissue in appropriate aliquots. This procedure ensures the most uniform delivery of standards within a group of samples to be processed at the same time, and it furthermore ensures contact of standard with the tissue from the very beginning of extraction.

8. If one or several of the analytes are present in very high levels, while the levels of others are exceedingly low, thus requiring large amounts of tissue to be extracted, internal standards for the high-level analytes may be used below their optimal range to save standard compound. Take care, however, to guarantee a proper standard signal with acceptable signal-to-noise ratio. Usually, 100 pmol of standard results in acceptable signals when using even high amounts of tissue up to 5 g fresh weight.
9. Accuracy of the hormone quantification depends on very well-defined standard stock solutions. Labeled standards should be weighed in sufficient amounts to allow accurate weight determination. Sufficiently sensitive high-precision balances must be used. If unavailable, standard solutions may be prepared and validated optically or by MS against accurately calibrated fresh solutions of unlabeled reference compounds. Solutions of standards should not be stored for prolonged periods of time. Rather, it is recommended to aliquot a calibrated reference solution, remove the solvent, and then store the dried aliquots under nitrogen in well capped or crimped vials at low temperatures. If unavoidable, then standard solutions to be kept over longer periods of time must be stored at –20°C in tightly closed vials with PTFE seals. They must be recalibrated in regular intervals against freshly and accurately made reference solutions of the corresponding unlabeled compounds.
10. Hormone compounds may be destroyed or released from conjugates by enzymatic activity or by chemical hydrolysis. Therefore, the extractant should penetrate the tissue quickly to ensure rapid enzyme inactivation. In case of very watery tissues, additional methanol needs to be added to keep the final methanol concentration above 85 to 90 %. It is possible to speed the extraction of analytes and enzyme inactivation by using heated methanol (60°C) for extraction. We have not observed any deleterious effects of this procedural variation on hormone recoveries.
11. Seeds and lignified tissue need longer vibration times. If no ball mill is available, in case of soft tissue such as roots and young leaves, extraction can be effected by allowing the sample to stand in extractant at room temperature for 1 to 2 h with intermittent shaking. Alternatively, the tissue can be ground in liquid nitrogen and then extracted in hot methanol (60°C). In this case, the powder must not thaw before the onset of methanol extraction.
12. Ensure that the sample is completely dried and no traces of water are left, but avoid unnecessarily long processing times. In our experience, there is no evaporative loss of analytes during vacuum centrifugation times of up to 2 h.
13. It is generally recommended to process samples and extracts immediately. If this is not possible, methanolic crude fractions as well as samples prior to methylation remain fairly stable when kept dried and sealed under nitrogen and stored at –20°C or lower for several weeks.
14. The speed of gravity flow through an SPE microcolumn should be approx 200 μL per 15 to 30 min. Faster flow rates will reduce the recovery of analytes. Washing and elution steps can be carried out faster by applying mild pressure using the mechanical pipet with capillary holder (*see* **Note 15**). Do not allow the gel bed to fall dry at any time.
15. Passage of the chloroform:isopropanol wash solution through the SPE microcolumn may be speeded up by applying mild pressure through the mechanical pipet with capillary holder. Prevent extrusion of the SPE matrix by placing the open end of the column against the inner wall of a vial. Alternatively, diethyl ether instead of chloroform:isopropanol may be used as a wash solution.
16. This step is dispensable for extracts low in lipids. In this case, the following acidic extraction step can be carried out directly.
17. Other acids, such as phosphoric acid or acetic acid, are not useful.
18. We have observed that the derivatization reaction is not working properly if executed directly in the autosampler vials.
19. Alternatively, a vacuum centrifuge can be used, which concentrates the sample at the bottom of the autosampler vial. However, avoid unnecessarily long centrifugation times and do not apply vacuum less than 20 mbar to minimize evaporative loss of the volatile methyl esters. Particularly, the SA-methyl ester may be lost under such conditions.

20. Capillary columns with polarity similar to that of ZB-50 (50% phenyl- 50% dimethyl-polysiloxane, low bleeding) are standard for acidic phytohormone profiling, however, other stationary phases (e. g., 5% or 35% phenyl) with lower polarity can also be used. Such columns may need a slightly different temperature program than that reported here to achieve separation of the analytes and impurities.

References

1. Müller, A., Düchting, P., and Weiler, E. W. (2002) A multiplex GC-MS/MS technique for the sensitive and quantitative single-run analysis of acidic phytohormones and related compounds, and its application to *Arabidopsis thaliana. Planta* **216,** 44–56.
2. Schmelz, E. A., Engelberth, J., Alborn, H. T., O'Donnell, P., Sammons, M., Toshima, H., and Tumlinson, J. H. 3rd. (2003) Simultaneous analysis of phytohormones, phytotoxins, and volatile organic compounds in plants. *Proc. Natl. Acad. Sci. USA* **100,** 10,552–10,557.
3. Birkemeyer, C., Kolasa, A., and Kopka, J. (2003) Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. *J. Chromatogr. A* **993,** 89–102.
4. Chiwocha, S. D., Abrams, S. R., Ambrose, S. J., Cutler, A. J., Loewen, M., Ross, A. R., and Kermode, A. R. (2003) A method for profiling classes of plant hormones and their metabolites using liquid chromatography-electrospray ionization tandem mass spectrometry: an analysis of hormone regulation of thermodormancy of lettuce (*Lactuca sativa L.*) seeds. *Plant J.* **35,** 405–417.
5. Prinsen, E., Van Laer, S., Oden, S., and Van Onckelen, H. (2000) Auxin analysis. *Methods Mol. Biol.* **141,** 49–65.
6. Astot, C., Dolezal, K., Moritz, T., and Sandberg, G. (1998) Precolumn derivatization and capillary liquid chromatographic/frit-fast atom bombardment mass spectrometric analysis of cytokinins in Arabidopsis thaliana. *J. Mass Spectrom.* **33,** 892–902.
7. Jameson, P. E. , Zhang, H., and Lewis, D. H. (2000) Cytokinins Extraction, separation, and analysis. *Methods Mol. Biol.* **141,** 101–121.
8. Prinsen, E., Redig, P., Strand, M., Galis, I., Van Dongen, W., and Van Onckelen, H. (1995) Quantifying phytohormones in transformed plants. *Methods Mol. Biol.* **44,** 245–262.
9. Tam, Y. Y., Epstein, E., and Normanly, J. (2000) Characterization of auxin conjugates in *Arabidopsis*. Low steady-state levels of indole-3-acetyl-aspartate, indole-3-acetyl-glutamate, and indole-3-acetyl-glucose. *Plant Physiol.* **123,** 589–596.
10. Ribnicky, D. M., Cooke, T. J., and Cohen, J. D. (1998) A microtechnique for the analysis of free and conjugated indole-3-acetic acid in milligram amounts of plant tissue using a benchtop gas chromatograph-mass spectrometer. *Planta* **204,** 1–7.
11. Kowalczyk, M. and Sandberg, G. (2001) Quantitative analysis of indole-3-acetic acid metabolites in *Arabidopsis*. *Plant Physiol.* **127,** 1845–1853.
12. These and other labelled ABA derivatives are provided by Suzanne R. Abrams PhD., Principal Research Officer, Plant Biotechnology Institute, National Research Council of Canada, 110 Gymnasium Place, Saskatoon SK S7N 0W9, Canada.
13. Netting, A. G. and Lidgard, R. O. (1999) A fragmentation of methyl abscisate and pentafluorobenzyl abscisate in methane electron capture negative ionisation tandem mass spectrometry. *J. Mass Spectrom*. **34,** 611–621.

Index

A

ABRC, database mining, 192, 193
Abscisic acid, *see* Hormone profiling
Aequorin,
  calcium affinity and reporting, 309, 310, 317, 318
  constructs, 309
  imaging in plants,
    calculations, 317, 321–323
    DNA assembly and transformation, 313, 315, 318, 319
    materials, 311, 313, 318
    transgenic line screening, 315, 316, 319, 321
  luminometry, 310, 318
  properties, 307, 309
  resolution of imaging, 310, 318
  sources, 307
Agrobacterium, see Transformation
Amplification of insertion mutagenized sites (AIMS),
  adapter ligation, 155
  DNA extraction and restriction digestion, 155
  gel electrophoresis of amplification products, 155, 158
  materials, 149, 152
  polymerase chain reaction, 155
AIMS, *see* Amplification of insertion mutagenized sites
Analysis of variance (ANOVA), differential gene expression detection, 364
ANOVA, *see* Analysis of variance
Aphid, *see* Pests
Aphidicolin, cell cycle synchronization, 49, 57
Arabidopsis lyrata,
  general characteristics, 27, 28
  growth,
    environment, 30, 36, 37
    germination, 31, 37
    materials, 29, 30
    optimization,
      fertilizer, 33
      light, 33
      temperature, 33, 36
      water and humidity, 33, 37
    planting, 31–33, 37
    reproductive growth, 35, 37
    seed stratification, 30
    vegetative growth, 34, 37
  pest management, 34
  seeds,
    harvesting and preservation, 36, 37
    production, 35–37

B

Bioinformatics, *see* Databases
Botrytis, see Gray mold
Brassica species,
  general characteristics, 27, 28
  growth,
    environment, 30, 36, 37
    germination, 31, 37
    materials, 29, 30
    optimization,
      fertilizer, 33
      light, 33
      temperature, 33, 36
      water and humidity, 33, 37
    planting, 31–33, 37
    reproductive growth, 35, 37
    seed stratification, 30
    vegetative growth, 34, 37
  pest management, 34
  seeds,
    harvesting and preservation, 36, 37
    production, 35–37

C

Calcium reporter, *see* Aequorin
Capsella species,

general characteristics, 27, 28
growth,
environment, 30, 36, 37
germination, 31, 37
materials, 29, 30
optimization,
fertilizer, 33
light, 33
temperature, 33, 36
water and humidity, 33, 37
planting, 31–33, 37
reproductive growth, 35, 37
seed stratification, 30
vegetative growth, 34, 37
pest management, 34
seeds,
harvesting and preservation, 36, 37
production, 35–37
Cell culture,
cell lines, 45, 46
cryopreservation,
liquid nitrogen, 55, 58, 59
materials, 48
regeneration, 55, 56, 59
viability testing, 56, 59
maintenance,
growth curves, 48, 57
MM1 cells, 48, 56
MM2d cells, 48, 56
materials, 46, 47
synchronization, *see* Cell cycle synchronization
transformation, *see* Transformation
Cell cycle synchronization,
aphidicolin-induced synchronization, 49, 57
assays,
bromodeoxyuridine incorporation, 51, 52, 57
flow cytometry of DNA content, 53
materials, 47
sucrose starvation, 49, 57
Chemically regulated gene expression, *see* XVE system
Chloroplast,
chlorophyll content determination, 412, 413, 417
envelope membrane purification, 410–412, 417
genetic engineering,
expression cassette construction, 250, 251
homologous flanking sequence cloning, 250
integration sites for transformation, 249
materials, 246–249, 257
maternal inheritance of transgenes, 256, 257
organogenesis,
principles, 253, 258, 259
tissue regeneration and selection, 254, 259
overview, 245, 246
particle bombardment,
consumables, 251, 252
DNA coating, 252, 257
gene gun preparation, 262, 253, 258
gold particle suspension preparation, 251
macrocarrier loading, 252, 258
plant tissue preparation, 252
polymerase chain reaction analysis, 255, 256
Southern blot analysis, 256
genome sequencing, 246, 247
purification, 410, 416, 417
trnI/trnA gene integration, 248
Chromatin, *see* Nuclear proteins
Chromosome painting,
materials, 17
overview, 175
Cryopreservation, *see* Cell culture
Cytogenetic analysis, *see* Chromosome painting; Fluorescence *in situ* hybridization; Immunostaining; Interphase nuclei

D

Databases,
ABRC, 192, 193
DNA microarray data,
databases, 197, 198
expression patterns in different experiments, 198, 199
expressed sequence tags,
differentially expressed genes, 201

Unigene clusters, 200, 201
literature database searching, 194, 195
map-based cloning databases and tools,
candidate genes in genetically defined intervals, 209
genetic markers for mapping, 207
polymorphisms between ecotypes for marker generation, 208
massively parallel signature sequencing data, 199, 200
MatDB, 191, 192
metabolomics,
AraCyc Pathways Database, 201, 202
metabolic gene expression changes, 202, 203
mining hardware and software, 191
mutants,
knockout mutations, 205, 206
mutants with similar phenotypes, 206, 207
NASC, 193
protein sequences,
similar sequences, 203, 204
similar structures and domains, 204
PubMed, 194
SIGnAL T-DNA Express Gene Mapping Tool, 193
TAIR,
Gene Ontology annotations, 196
gene finding, 195, 196
overview, 192
publication search, 194
TIGR *Arabidopsis* annotation database, 192
types and access, 188–190, 192
Diazomethane, *see* Hormone profiling
DNA microarray,
database mining,
databases, 197, 198
expression patterns in different experiments, 198, 199
experimental design, 359, 360
materials, 351, 352, 354, 355
optimization, 350, 351
platforms, 359
probes, 349, 350
sample preparation for GeneChip analysis,
cDNA preparation,
first-strand synthesis, 352
purification, 353, 355
second strand synthesis, 352
cRNA preparation,
biotin-labeled cRNA synthesis, 353, 355
fragmentation, 353, 354, 356
purification and quantification, 353, 355, 356
sample preparation for spotted microarrays, 354, 356, 357
statistical analysis,
cluster analysis, 360, 361
differential expression detection,
analysis of variance, 364
false discovery rate, 364
LIMMA, 364
t-test, 363, 364
ISS arrays and normalization, 361, 363
multivariate methods,
distance and similarity matrix, 364
hierarchical clustering, 364, 365
nonhierarchical clustering, 365
principal component analysis, 365
self-organizing maps, 365
overview, 360
two-color experiments, 361–363
two-color arrays,
background subtraction, 362
bad feature flagging, 362
normalization,
global intensity normalization, 362
intensity-dependent normalization, 362, 363
spot intensity quantification, 361, 362
DsRED,
discovery, 276
expression vectors,
cytosolic expression, 277, 286
endoplasmic reticulum targeting, 277
fusion proteins, 277, 280, 286, 287
transformation vectors, 280
fluorescence microscopy,
imaging, 286
leaf tissue, 285, 288, 289
materials, 276, 277
protoplasts, 282–285
stable expression, 281, 287

transient expression, 280
variants, 276, 279

E

EMS mutagenesis, *see* Ethyl methanesulfonate mutagenesis
ESTs, *see* Expressed sequence tags
Ethyl methanesulfonate (EMS) mutagenesis,
advantages in gene mapping, 116
materials, 102, 103
mechanisms, 101
overview, 101, 102
seeds,
mutagenesis, 102, 103
planting, 103
Expressed sequence tags (ESTs), database mining,
differentially expressed genes, 201
Unigene clusters, 200, 201
Extreme array mapping (XAM), principles, 143–145

F

False discovery rate (FDR), differential gene expression detection, 364
FDR, *see* False discovery rate
FISH, *see* Fluorescence *in situ* hybridization
Flow cytometry, DNA content of cells, 53
Fluorescence *in situ* hybridization (FISH),
chromosome spreading, 180, 184, 185
fiber-FISH,
fiber preparation, 184
hybridization, 184
materials, 179, 180
nuclei isolation, 184
fluorescence detection, 182, 185
hybridization, 181, 182, 184, 185
materials, 175, 177, 178, 184
overview, 175
probe labeling, 180, 181, 185
Fluorescent proteins, *see* DsRED; Green fluorescent protein
Forward genetic screening, *see* Insertional mutants
Fungus gnat, *see* Pests

G

GeneChip, *see* DNA microarray
Genetic analysis,
allelism tests, 68, 71
crossing, 67, 68, 70, 71
double mutant analysis for epistatic relationship analysis, 68, 69, 71
mapping, *see also* Genetic analysis; Multiplex polymerase chain reaction; Positional cloning; Single-feature polymorphisms,
fine genetic mapping, 70, 75, 76
linkage analysis, 66, 67, 69, 72, 73
linkage map construction, 70, 75
principles, 66, 67
recombination frequency and map distance estimation, 69, 70, 73–75
materials, 67
overview, 65, 66
quantitative trait loci analysis, *see* Quantitative trait loci
segregation analysis, 68, 71
Germination, *see* Growth,
GFP, *see* Green fluorescent protein
β-Glucuronidase (GUS),
advantages as reporter, 264, 266
applications, 266
assays,
histological staining,
embedding and sectioning, 270–272
whole-mount staining, 270–272
materials, 266, 267, 271
quantitative assays, 267, 269–271
chromogenic substrates, 264, 265
function, 263
structure, 263
Grafting,
materials, 40, 42
overview, 39
root-shoot grafts, 40, 42, 43
wedge graft, 40, 41
Y-graft, 41, 42
Gray mold,
detection and treatment, 24, 25
prevention, 24
Green fluorescent protein (GFP),
expression vectors,
cytosolic expression, 277, 286
endoplasmic reticulum targeting, 277
fusion proteins, 277, 280, 286, 287
transformation vectors, 280

fluorescence microscopy,
imaging, 286
leaf tissue, 285, 288, 289
materials, 276, 277
protoplasts, 282–285
fluorophore, 275, 276
stable expression, 281, 287
structure, 276
transient expression, 280
variants, 275, 276, 278
Growth,
fertilizer, 5, 11
harvesting, 6, 7
inflorescence bagging, 7, 8, 11
insecticides, 6, 11
materials, 3, 4, 10
optimization,
light, 5, 6
temperature, 6
watering and humidity, 6, 11
overview, 3
pest prevention, *see* Pests
planting and germination,
soils, 5, 10, 11
sterile conditions, 4, 5, 10
related species, *see Arabidopsis lyrata*; *Brassica* species; *Capsella* species
GUS, *see* β-Glucuronidase

H

Histones, *see* Nuclear proteins
Hormone profiling,
ethereal diazomethane preparation, 451, 452, 455
gas chromatography–mass spectrometry,
derivatization, 454, 456
settings, 454, 455, 457
materials, 450, 451, 455
methyl ester purification, 454
overview, 449, 455
sample preparation, 452–454, 456
solid-phase extraction microcolumn assembly, 451, 455

I

IMAC, *see* Immobilized metal affinity chromatography
Immobilized metal affinity chromatography (IMAC), phosphopeptide isolation,
anion-exchange chromatography, 433, 434
column chromatography, 434–436
materials, 432
microcolumns, 434, 435
overview, 431, 432, 435
protein extraction, 432–434
trypsin digest, 433
Immunostaining, cytogenetic analysis,
chromatin protein immunolocalization, 183
materials, 178, 179, 184
modified DNA immunolocalization, 183, 184
overview, 174
Indels, *see* Multiplex polymerase chain reaction
Indole-3-acetic acid, *see* Hormone profiling
Insertional mutants,
amplification of insertion mutagenized sites,
adapter ligation, 155
DNA extraction and restriction digestion, 155
gel electrophoresis of amplification products, 155, 158
materials, 149, 152
polymerase chain reaction, 155
confirmation and analysis of mutation–phenotype relationships,
in silico studies, 155, 156, 158, 159
materials, 152, 153, 157
reverse transcription-polymerase chain reaction, 156, 158, 159
RNA extraction, 156, 159
forward genetic screens,
DNA extraction, 154
materials, 149–151
mutant identification and segregation analysis, 154, 157, 158
Southern blot, 154, 158
generation, 147
polymerase chain reaction-based screening,
DNA pool construction, 165–168
materials, 164, 165
overview, 163, 164
screening, 167–171
probability of insertion within a given gene, 147, 157

T-DNA integration, 147, 148
transposon footprint identification by polymerase chain reaction,
amplification of empty donor sites, 157
DNA extraction, 157
gel electrophoresis of amplification products, 157, 159
materials, 153, 157
stabilized mutant allele identification, 157, 159
Interphase nuclei,
isolation, 178, 182, 183, 185
overview of studies, 173, 174

J–L

Jasmonic acid, *see* Hormone profiling
Laser microdissection,
applications, 367
dewaxing, 370
laser microdissection and pressure catapulting, 370, 374
materials, 367, 368
RNA isolation, 370, 371, 374, 375
RNA yield and integrity analysis, 371–373, 375, 376
RNase contamination prevention, 368, 373
specimen preparation,
paraffin wax embedding, 369, 373
sectioning and slide preparation, 369, 370, 374
tissue fixation and dehydration, 368, 369, 373
LIMMA, differential gene expression detection, 364
Linkage analysis, *see* Genetic analysis
LOD score, *see* Quantitative trait loci
Luciferase,
advantages as reporter, 293
applications, 293
half-life, 293
luminescence measurement,
CCD camera automated measurement,
analysis, 300, 304
setup, 298–300, 304
imaging, 295–297
materials, 297, 298, 303, 304
TopCount automated measurement,
analysis, 303
setup, 300–305
reporter constructs, 294
sources, 293, 294

M

Mapping, *see* Genetic analysis; Multiplex polymerase chain reaction; Positional cloning; Single-feature polymorphisms
Mass spectrometry (MS),
hormone profiling, *see* Hormone profiling
metabolomics, *see* Metabolomics
phosphorylation site determination with tandem mass spectrometry, 431, 435
protein microsequencing, *see* Protein microsequencing
proteomics, 422
Massively parallel signature sequencing (MPSS), database mining, 199, 200
MatDB, database mining, 191, 192
Membrane proteins,
isolation and characterization,
cell suspension culture, 408, 415
chloroplast preparation,
envelope membrane purification, 410–412, 417
purification, 410, 416, 417
differential extraction,
alkaline or salt washing, 414, 418
chloroform/methanol solubilization, 413, 414, 418
detergent solubilization, 413, 418
gel electrophoresis, 414, 418, 419
materials, 404–408, 414, 415
microsome preparation, 408, 409, 416
organelle/membrane purity assessment,
chlorophyll content determination, 412, 413, 417
enzyme markers, 412, 415
immunological markers, 412, 417
lipid and pigment analysis, 413, 417
overview, 403, 404
plasma membrane purification, 409, 410, 416
structure, 403
Metabolomics,
database mining,
AraCyc Pathways Database, 201, 202

metabolic gene expression changes, 202, 203
hormone profiling, *see* Hormone profiling
metabolite profiling,
data analysis, 443, 446, 447
gas chromatography–mass spectrometry, 443, 446
materials, 441, 442
overview, 439, 440
sample preparation,
derivatization, 442, 443, 445, 446
extraction and fractionation, 442, 445
harvesting, 442, 444
homogenization, 442, 444, 445
Microparticle bombardment,
chloroplast genetic engineering, *see* Chloroplast
transcription factor functional analysis,
DNA-coated gold particle preparation, 234
expression vectors,
reference gene, 232
reporter gene, 232, 235
transacting gene, 232
leaf bombardment, 234, 235
materials, 231, 232, 235
overview, 231
plant growth, 233, 234
relative transactivation analysis, 235
tissue preparation, 234, 235
Microsequencing, *see* Protein microsequencing
MPSS, *see* Massively parallel signature sequencing
MS, *see* Mass spectrometry
Multiplex polymerase chain reaction, low-resolution mapping of untagged mutations,
amplification reactions, 109, 112
DNA extraction, 107, 109, 112
genotyping, 109, 112
indels, 106
linkage testing and map distance estimation, 109
materials, 106–108
oligonucleotide mix preparation, 107–109, 112
overview, 105, 106
product electrophoresis, 109, 112
simple sequence length polymorphisms, 106

N

NASC, database mining, 193
Nuclear proteins,
isolation,
chromatin isolation from seedlings, 399, 400
histone extraction from chromatin, 400, 401
materials, 394–396
nuclear extract preparation, 398, 399
nuclei isolation, 395–398, 401
protoplast preparation, 397
proteomics, 393, 394
Nucleolus organizer, features, 173

O

Oligonucleotide arrays, *see* DNA microarray; Single-feature polymorphisms
Organogenesis, *see* Chloroplast

P

Particle bombardment, *see* Microparticle bombardment
PCR, *see* Polymerase chain reaction
Pests,
aphids,
detection and monitoring, 20
prevention, 20
species, 19, 20
treatment, 21
fungus gnats,
detection, 16
prevention, 6
treatment, 17, 18
management in related plants, 34
outbreaks,
control, 14, 15
early signs, 15
overview, 13
prevention,
dead plant material removal, 14
growth area, 14
growth conditions, 15, 16
growth supplies, 14
spider mites,

detection, 21
prevention, 21
treatment, 21
thrips,
detection and monitoring, 18
prevention, 18
treatment, 18, 19
Phosphoproteins,
immobilized metal affinity chromatography for phosphopeptide isolation,
anion-exchange chromatography, 433, 434
column chromatography, 434–436
materials, 432
microcolumns, 434, 435
overview, 431, 432, 435
protein extraction, 432–434
trypsin digest, 433
phosphorylation site determination, 431, 435
Planting, *see* Growth
Polymerase chain reaction (PCR), *see also* Multiplex polymerase chain reaction; TILLING,
chloroplast genetic engineering analysis, 255, 256
insertional mutant analysis, *see* Insertional mutants
Positional cloning,
databases and tools,
candidate genes in genetically defined intervals, 209
genetic markers for mapping, 207
polymorphisms between ecotypes for marker generation, 208
F2 seed generation for mapping population, 119, 121, 122
fine-scale mapping, 121, 124, 125
first-pass mapping,
DNA preparation, 119, 122
genetic linkage determination, 119, 122, 123
map refinement, 121, 123, 124
plant identification for genotyping, 119, 122
materials, 118, 119
mutation identification, 121, 125
overview of mapping techniques, 115, 116, 118
Powdery mildew,
detection, 23
features, 22, 23
prevention, 23
treatment, 23, 24
Protein microsequencing,
in-gel protein digestion, 425, 428
liquid chromatography–tandem mass spectrometry, 426–429
materials, 422, 423
sample preparation,
gel electrophoresis, 423, 427, 428
mass spectrometry, 425, 426, 428
spot picking, 424, 425, 428
two-dimensional gel electrophoresis and staining, 424, 428
Proteomics, *see* Membrane proteins; Nuclear proteins; Phosphoproteins; Protein microsequencing; Two-dimensional gel electrophoresis
Protoplast,
fluorescent protein imaging, 282–285
preparation from cell suspensions, 397
transfection assays,
applications, 237
auxin-responsive gene expression with transfected vs integrated reporter gene, 241, 242
cell culture protoplasts,
advantages and limitations, 239, 240
cell culture, 240, 241
isolation, 241, 243
transfection, 241, 243, 244
leaf protoplasts,
advantages and limitations, 239, 240
isolation, 240, 243
transfection, 240, 243, 244
materials, 238, 239, 243
principles, 237, 238
reporter gene assays, 241
PubMed, literature searches, 194

Q

QTL, *see* Quantitative trait loci
Quantitative trait loci (QTL),
detection, factors affecting,
number of observations per genotype, 84, 85

overview, 83, 84
population type and size, 84, 85
genome-wide linkage maps, 88, 92–94
Internet resources, 81
localization,
confidence intervals, 86
experimental factors affecting, 86, 87
LOD thresholds, 85, 86
mapping, 88, 89, 96, 97
mapping population construction, 87, 88, 91, 92
marker spacing, 85
materials, 82, 89
multiple loci models, 85
overview of analysis, 79–82
parental accessions for mapping, 87, 89–91
statistical concepts in analysis, 83

R

Recombinant inbred line (RIL), quantitative trait loci analysis, 80, 81, 88, 92–95
RIL, *see* Recombinant inbred line
RNA extraction,
applications, 345
large-scale preparation, 346–348
materials, 346, 347
overview, 345
small-scale isolation, 346, 347
RNA interference,
inducible activation tagging and functional cloning, 339–341
inducible conditional knockout, 340, 341
regulated in vivo cDNA generation, 339, 341
XVE system,
cloning, 338
phenotypic characterization and genetic analysis, 339
transformation and transgenic plant identification, 338
vectors, 338
RNA isolation, laser microdissection samples, 370, 371, 374, 375
Root-shoot graft, *see* Grafting

S

Salicylic acid, *see* Hormone profiling
Seeds,
Arabidopsis-related species,
harvesting and preservation, 36, 37
production, 35–37
collection, 7
drying, 8, 11
moisture content determination, 8, 9, 11
mutagenesis, *see* Ethyl methanesulfonate mutagenesis
storage, 9, 10
threshing and cleaning, 8, 11
viability testing, 10
Self-organizing maps (SOMs), DNA microarray analysis, 365
SFPs, *see* Single-feature polymorphisms
SIGnAL T-DNA Express Gene Mapping Tool, 193
Simple sequence length polymorphisms, *see* Multiplex polymerase chain reaction
Single-feature polymorphisms (SFPs), genotyping and mapping with high-density oligonucleotide arrays,
array mapping, 141, 143
deletion detection via clustering, 141
extreme array mapping, 143–145
file requirements for identification, 138, 139
hybridization, 138, 144
labeling, 138, 144
materials, 138, 144
principles, 137
quality control, 139
SAM *d*-statistic calculation and false discover rates, 139–141
SOMs, *see* Self-organizing maps
Southern blot,
chloroplast genetic engineering analysis, 256
forward genetic screening of insertional mutants, 154, 158
Spider mite, *see* Pests

T

TAIR,
DNA microarray data, 197, 198
Gene Ontology annotations, 196
gene finding, 195, 196
overview, 192
publication search, 194
Targeting induced local lesions in genomes, *see* TILLING
Thrip, *see* Pests

TIGR *Arabidopsis* annotation database, database mining, 192
TILLING,
CEL I digestion, 130, 131
gel electrophoresis,
gel preparation, 131, 134
imaging and analysis, 132, 133, 135
running conditions, 131, 132, 134, 135
sample loading, 131, 134
individual mutant analysis, 133, 135
materials, 128, 130, 133, 134
polymerase chain reaction and heteroduplex formation, 130, 134
principles, 127, 128
spin plate cleanup, 131, 134
Transcription factors, functional analysis with microparticle bombardment,
DNA-coated gold particle preparation, 234
expression vectors,
reference gene, 232
reporter gene, 232, 235
transacting gene, 232
leaf bombardment, 234, 235
materials, 231, 232, 235
overview, 231
plant growth, 233, 234
relative transactivation analysis, 235
tissue preparation, 234, 235
Transformation,
Agrobacterium-mediated transformation,
cocultivation, 53, 57
materials, 47, 48
transgenic calli recovery, 53, 54, 57, 58
transgenic liquid culture establishment, 54, 55
vacuum infiltration *in planta*,
Agrobacterium culture and preparation, 217, 220
growth conditions, 217, 220
infiltration and harvesting, 217, 219–221
materials, 216, 217, 220
principles, 215, 216
screening of transformants, 219, 221, 222
chloroplasts, *see* Chloroplast
insertional mutant forward genetic screening, *see* Insertional mutants
overview, 46, 215
transient expression assay by agroinfiltration of leaves,
Agrobacterium suspension preparation, 226
agroinfiltration, 227
dexamethasone-inducible promoter activation, 227, 228
histochemical staining, 227
materials, 226, 227
overview, 225, 226
phenotyping, 227
plant growth and leaf selection, 226, 227
sample collection, 227, 228
XVE system,
floral dipping, 334
root explants, 334, 335
Transient expression, *see* Protoplast transfection assays; Transformation
t-test, differential gene expression detection, 363, 364
Two-dimensional gel electrophoresis,
denaturing gel electrophoresis,
casting, 387
drying of gels, 388, 389
equilibration and running, 388, 390
stacking gel, 387, 390
staining, 388, 390
isoelectric focusing, 386, 387, 389
materials, 382
overview, 381, 421
protein microsequencing, *see* Protein microsequencing
protein preparation,
extraction, 382, 383
precipitation,
leaf and root proteins, 384, 385
mature seed proteins, 385, 386, 389
solubilization,
leaf and root proteins, 386, 389
mature seed proteins, 386, 389

V–Y

Vacuum infiltration, *see* Transformation
Wedge graft, *see* Grafting
XAM, *see* Extreme array mapping
XVE system,
adaptation for other plants, 331, 341

applications, 330
inducible DNA excision,
cloning and transformation, 337
phenotypic characterization and genetic analysis, 337
vectors, 336, 337
materials, 330, 331
overexpression studies,
cloning, 331, 334, 341
expression vectors, 331
transformation,
floral dipping, 334
root explants, 334, 335
transgene induction and phenotypic characterization, 335, 336, 341
principles, 329
RNA interference studies,
cloning, 338
phenotypic characterization and genetic analysis, 339
transformation and transgenic plant identification, 338
vectors, 338
transgene induction, 329, 330
Y-graft, *see* Grafting

METHODS IN MOLECULAR BIOLOGY™ • 323
SERIES EDITOR: *John M. Walker*

Arabidopsis Protocols

SECOND EDITION

Edited by

Julio Salinas

Dept. of Biotechnology, Instituto Nacional de Investigacion y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain

Jose J. Sanchez-Serrano

Dept. de Genetica Molecular de Plantas, Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco, Madrid, Spain

From Reviews of the First Edition...

"...recommend[ed]...provides a set of enduring methods and is comprehensive enough for instant use in the lab." **—TRENDS IN PLANT SCIENCE**

"...a 'must have' for every laboratory involved in Arabidopsis.*"* **—QUARTERLY REVIEW OF BIOLOGY**

"...a 'must' for those labs working with not just Arabidopsis *specifically but plants in general."*
—MICROBIOLOGY TODAY

"...a must for anybody moving into the Arabidopsis *field and setting up a new lab...extremely useful."*
—GENETICAL RESEARCH

Emulating the much-praised first edition of *Arabidopsis Protocols,* leading scientists have generated an up-to-date work that reflects recent advances in plant biology, the completion of the *Arabidopsis* genome sequence—essential for studying plant function—and the development of whole systems approaches that allow global analysis of gene expression, as well as protein and metabolite dynamics. The authors have included nearly all techniques developed in *Arabidopsis,* others recently adapted from more traditional work in crop species, and the latest using *Arabidopsis* as a model system. Highlights include the most recent methods—transcriptomics, proteomics, and metabolomics—and their novel applications (phosphoproteomics, DNA microarray-based genotyping, high throughput metabolite profiling, and single-cell RNA). Traditional protocols from the agricultural sciences and others developed in crop species (grafting and chloroplast transformation) have been adapted to exploit the advantages of the *Arabidopsis* model. The protocols themselves follow the successful *Methods in Molecular Biology™* series format, each offering step-by-step laboratory instructions, an introduction outlining the principles behind the technique, lists of the necessary equipment and reagents, and tips on troubleshooting and avoiding known pitfalls.

Indispensable and highly practical, *Arabidopsis Protocols, Second Edition* offers both novice and experienced plant biologists cutting-edge tools to explore new scenarios and gain an understanding of how this complex, multicellular organism works, how it copes with a sessile life style, and how these strategies compare with those developed in other organisms.

FEATURES

- **Readily reproducible methods that use the latest molecular techniques**
- **Utilization of the completed *Arabidopsis* genome to study plant function**
- **Global analysis of gene expression and protein and metabolite dynamics**
- **DNA microarray-based genotyping and single-cell RNA characterization**
- **High-throughput metabolite profiling**

Methods in Molecular Biology™ • 323
ARABIDOPSIS PROTOCOLS
SECOND EDITION
ISBN: 1-58829-395-5 E-ISBN: 1-59745-003-0
ISSN: 1064–3745 humanapress.com