

METHODS IN MOLECULAR BIOLOGY™

Volume 255

Bacterial Artificial Chromosomes

*Volume 1: Library Construction,
Physical Mapping,
and Sequencing*

Edited by

**Shaying Zhao
Marvin Stodolsky**

 HUMANA PRESS

BAC Library Construction

Kazutoyo Osoegawa and Pieter J. de Jong

1. Introduction

DNA cloning, especially large DNA cloning, is the first step in contemporary complex genome analysis. Cloning technology of high-molecular-weight DNA has been developed mainly using yeast and *Escherichia coli* as hosts. In the early stages of the Human Genome Project, yeast artificial chromosome (YAC) libraries have been generated and used for construction of a framework of the genome. The YAC cloning system has a great advantage of cloning of very large (>500 kb) DNA, thus facilitating construction of a physical map of the complex genome. The bacterial artificial chromosome (BAC) technologies matured later but proved to have so many advantages that the BAC libraries have been the primary input to contig assembly and the public sector human genome sequencing. BACs are easily purified as plasmid DNAs, have little if any chimerism, and are stable, with a very few interesting exceptions. Both BAC and bacteriophage P1-derived artificial chromosome (PAC) cloning systems have been developed, respectively, using the *E. coli* F-factor plasmid replication and bacteriophage P1 plasmid origin to maintain largeness (100–250 kb). Genomic DNA is subjected to partial digestion with a restriction endonuclease in order to break DNA into clonable size and size fractionated using pulsed-field gel electrophoresis (PFGE). The size-fractionated DNA is cloned into a BAC vector and transformed into *E. coli* by electrical shock. The transformants are arrayed into microtiter dishes and high-density replica filters are prepared to facilitate screening of the library. Human genome draft sequences were reported using two different (BAC clone-by-BAC clone and whole genome shotgun) approaches. For the clone-by-clone strategy, construction of a high-quality and highly redundant BAC library was a critical step to ensure

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

the almost complete representation of the genome. The library was distributed worldwide as an arrayed format that allows sharing of the data in the public domain. A contiguous BAC clone map has been assembled facilitating a selection of minimally overlapping clone sets to reduce sequence redundancy. In theory, construction of a BAC library does not appear to be a difficult task. In practice, construction of a high-quality library is an art. This chapter describes all the requirements for constructing a high-quality BAC library.

2. Materials

2.1. Preparation of Broadly Used Reagents

1. EDTA, pH 8.0, 0.5 *M* stock solution: 200 g of EDTA•4Na and 176.44 g of EDTA•2Na in 1600 mL of H₂O. Adjust the pH to 8.0 with NaOH palette and make up to 2 L with distilled deionized water. Autoclave at 121°C for 30 min.
2. Red blood cell (RBC) lysis solution (10X): Dissolve 9.54 g of NH₄Cl (1.78 *M* final) and 0.237 g of NH₄HCO₃ (0.03 *M* final) in sterile, distilled deionized water. Filtrate (sterilization filter unit cellulose nitrate membrane; cat. no. 28199-075; Nalgene) and store in the filter unit receiver at 4°C up to 1 mo.
3. Phosphate-buffered saline (PBS) (pH 7.4): 10X PBS is prepared as follows: Mix 80 g of NaCl (final conc.: 8%), 2 g of KCl (0.2%), 14.4 g of Na₂HPO₄ (1.44%), and 2.4 g of 0.24% KH₂PO₄ for a total volume of 1 L. Adjust the pH to 7.4 with HCl. Dilute 10 times with sterile, distilled deionized water prior to use.
4. *N*-Lauroyl sarcosine (cat. no. L-5125; Sigma, St. Louis, MO) (10% stock solution): Dissolve 10 g of *N*-lauroyl sarcosine in 100 mL of sterile, distilled deionized water. Filtrate (sterilization filter unit cellulose nitrate membrane, cat. no. 28199-075; Nalgene) and store at room temperature in the filter unit receiver.
5. Cell lysis solution: 10 mL of filtrated 10% *N*-lauroyl sarcosine (sodium salt; Sigma) (final concentration: 2%), 40 mL of 0.5 *M* EDTA (pH 8.0) (final concentration: 0.4 *M*), and 100 mg of proteinase K (cat. no. 1 092 766; Roche) (final concentration: 2 mg/mL). Prepare the solution just prior to use.
6. Phenylmethylsulfonyl fluoride (PMSF) (cat. no. P-7626; Sigma) (100 mM stock solution): Dissolve 174.2 mg in 10 mL of isopropanol and store at -20°C in small aliquots (200 µL).
7. Spermidine (cat. no. S-2501; Sigma) (0.1 *M* stock solution): Dissolve 0.255 g of spermidine trihydrochloride in 10 mL of sterile, distilled deionized water. Filtrate (Acrodisc, 0.2-µm syringe filters, 25 mm, 50/pack; cat. no. 4192; German Sciences) and store at -20°C in small aliquots (200 µL).
8. 10X *Eco*RI and *Eco*RI methylase buffer: 100 µL of 32 mM *S*-adenosyl-methionine (cat. no. B9003S; New England Biolabs), 80 µL of 1 *M* MgCl₂, 800 µL of 5 *M* NaCl, 2 mL of 1 *M* Tris-HCl (pH 7.5), 40 µL of 1 *M* dithiothreitol (DTT), and 980 µL of sterile-distilled deionized water. The total volume is 4 mL (*see* **Note 1**).
9. 10X *Mbo*I buffer without Mg⁺⁺ and DTT (1 *M* NaCl; 0.5 *M* Tris-HCl, pH 8.0): Mix 100 mL of 1 *M* Tris-HCl (pH 8.0), 40 mL of 5 *M* NaCl, and 60 mL of distilled deionized water in a 250-mL glass bottle. Autoclave at 121°C for 30 min.

10. Polyethylene glycol 8000 (PEG8000) solution (30% [w/v]) PEG8000; 10 mM Tris-HCl, pH 8.0; 0.5 mM EDTA): Dissolve 300 g of PEG8000 in 600 mL of water. Add 1 mL of 0.5 M EDTA and 5 mL of 1 M Tris-HCl (pH 8.0). Adjust the volume to 1 L and autoclave at 121°C for 30 min.
11. Gel-loading dye 1: 0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol. Weigh 0.125 g of bromophenol blue and 0.125 g of xylene cyanol FF in a 50-mL conical screw-cap polypropylene tube. Add 15 mL of glycerol and 35 mL of TE buffer (pH 8.0) and mix well. Store at 4°C.
12. Gel-loading dye 2, not containing xylene cyanol FF (0.25% bromophenol blue, 40% sucrose: Weigh 0.125 g of bromophenol blue and 20 g of sucrose in a 50-mL conical screw-cap polypropylene tube. Add TE buffer (pH 8.0) up to 50 mL and mix.
13. Chloramphenicol stock solution (20 mg/mL): Dissolve 1 g of chloramphenicol (C0378; Sigma) in 50 mL of 99.5% ethanol and filtrate (Acrodisc, 0.2- μ m syringe filters, 25 mm, 50/pack; cat. no. 4192, GermanSciences); into a 50-mL disposable centrifuge tube (Corning cat. no. 25325-50, or equivalent). Store at -20°C. The antibiotic is stable for 1 yr.
14. Kanamycin stock solution (25 mg/mL): Dissolve 1.25 g of kanamycin (cat. no. K-4000; Sigma) in 50 mL of sterile, deionized distilled water and filtrate. Aliquot 500 μ L into microcentrifuge tubes and store at either 4°C for short term or -20°C for long term. The antibiotic is stable for 1 yr at -20°C.
15. Ampicillin stock solution (100 mg/mL): Dissolve 1 g of ampicillin (cat. no. A-9518; Sigma) in 10 mL of sterile, deionized distilled water. Aliquot 500 μ L into 1.5-mL microcentrifuge tubes and store at -20°C; it is stable for 1 yr.
16. Ethidium bromide (EtBr) staining buffer: Stock solution (10 mg/mL) is diluted to 0.5 μ g/mL with 0.5X TBE buffer prior to staining gels.
17. Suspension buffer: 50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0). To prepare the solution, mix 50 mL of 1 M glucose, 25 mL of 1 M Tris-HCl (pH 8.0), and 20 mL of 0.5 M EDTA (pH 8.0). Autoclave at 121°C for 20 min. The solution can be stored at room temperature up to 1 yr.
18. Lysis solution: 0.2 N NaOH, 1% sodium dodecyl sulfate (SDS). To prepare the solution, add 3 mL of 10 N NaOH and 7.5 mL of 20% SDS in 139.5 mL of water.
19. Potassium acetate (pH 4.8) solution: Dissolve 147.21 g of potassium acetate in 400 mL of water, add 57.5 mL of glacial acetic acid, and adjust the volume to 500 mL. Filtrate the solution and store at room temperature.
20. CsCl solution: Dissolve 50 g in 50 mL of TE buffer (pH 8.0) and autoclave at 121°C for 20 min. Store at room temperature.

2.2. Preparation of Luria Bertani Plates Containing Antibiotics

1. Tryptone peptone (500 g) (pancreatic digest of casein; cat. no. 211705; Difco, Detroit, MI).
2. Bacto Yeast Extract (500 g) (cat. no. 212750; Difco).
3. NaCl (50 kg) (cat. no. S-9888, Sigma).
4. 5 N NaOH.

5. Bacto agar (2 kg) (cat. no. 214030, Difco).
6. Chloramphenicol stock solution (20 mg/mL).
7. Ampicillin stock solution (100 mg/mL).
8. Kanamycin stock solution (25 mg/mL).
9. Petri dish (cat. no. 351029, 100 × 15 mm style, 20/bag; Falcon).

2.3. Testing of Vector

1. *E. coli* DH10B cells containing pBACe3.6, pTARBAC1.3, and pTARBAC2.1 (1,2): in 15% glycerol stored at -80°C. (Contact pdejong@chori.org)
2. Luria Bertani (LB) plates containing antibiotics (see **Subheading 2.2.**).
3. Six-well green tubes for AutoGen740 machine or 15-mL snap-cap polypropylene tubes.
4. Orbital shaker, 37°C.
5. Automatic plasmid isolation machine (AutoGen740 if applicable).
6. *Bam*HI (50,000 U, 20,000 U/mL) (cat. no. R0136L; New England Biolabs). *Bam*HI reaction buffer: 150 mM NaCl, 10 mM Tris-HCl (pH 7.9), 10 mM MgCl₂, 1 mM DTT. Supplement with 100 µg/mL of bovine serum albumin (BSA).
7. *Eco*RI (50,000 U, 20,000 U/mL) (cat. no. R01011, New England Biolabs). *Eco*RI reaction buffer: 50 mM NaCl, 100 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 0.025% Triton X-100.
8. *Not*I (2500 U, 10,000 U/mL) (cat. no. R01891; New England Biolabs). *Not*I reaction buffer: 100 mM NaCl, 50 mM Tris-HCl (pH 7.9), 10 mM MgCl₂, 1 mM DTT. Supplement with 100 µg/mL of BSA.
9. *Apa*LI (2500 U, 10,000 U/mL) (cat. no. R0507S; New England Biolabs). *Apa*LI reaction buffer: 50 mM potassium acetate, 20 mM Tris-acetate (pH 7.9), 10 mM magnesium acetate, 1 mM DTT. Supplement with 100 µg/mL of BSA.
10. BSA (10 mg/mL) (cat. no. B9001S; New England Biolabs).
11. Flexible plate, 96-well (U-bottomed without lid; cat. no. 353911; Falcon).
12. Conventional agarose electrophoresis system, with 10-cm-long, 15-cm-wide gel tray.
13. Gel-loading dye 2 without xylene cyanol FF: 0.25% bromophenol blue, 40% sucrose.
14. EtBr staining buffer (0.5 µg/mL).
15. Alpha Innotech IS1000 digital imager.

2.4. Purification of Vector DNA

1. Cell suspension buffer: 50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0. Store at room temperature.
2. Lysis solution: 0.2 N NaOH, 1% SDS; prepare fresh solution prior to use.
3. Potassium acetate, pH 4.8. Store at room temperature.
4. CsCl (molecular biology grade) (cat. no. 15542-020; Invitrogen).
5. CsCl solution.
6. 50-mL Conical screw-cap polypropylene tube (cat. no. 430828; Corning).
7. Centrifuge tubes (polyallomer, Quick-Seal centrifuge tubes, 1 × 3½ in. or 25 × 89 mm; Beckman) and heating sealer.

8. VTi 50 rotor (minimum radius 60.8 mm, maximum radius 86.6 mm, maximum rotor speed 50,000 rpm; Beckman or equivalent).
9. Beckman L8-M Ultracentrifuge.
10. 3 mL single-use syringe with 18-gage needle (cat no. BD309580).

2.5. Removal of EtBr

1. Isoamyl alcohol (Fisher).
2. Refrigerated centrifuge with rotor and adapters for 50-mL tubes (Sorvall RT7 centrifuge with H-1000B swinging-bucket rotor or equivalent).
3. Dialysis tubing (Spectra/Pro Membrane MWCO: 8000, cat no. 132115).
4. Dialysis clip.
5. 2-L Glass beaker and magnetic stirring bar.
6. TE buffer, pH 8.0.
7. Gel-loading dye 2 without xylene cyanol FF: 0.25% bromophenol blue, 40% sucrose.
8. EtBr staining buffer (0.5 $\mu\text{g/mL}$).
9. Alpha Innotech IS1000 digital imager.

2.6. Digestion of Vector DNA With Restriction Enzymes

1. pBACe3.6, pTARBAC1.3, or pTARBAC2.1 vector DNA.
2. 10X NEBuffer 4, 10 mg/mL BSA, *Apa*LI (10 U/ μL) (New England Biolabs).
3. Enzyme dilution buffer for *Apa*LI diluent A: 50 mM KCl, 10 mM Tris-HCl, 0.1 mM EDTA, 1 mM DTT, 200 $\mu\text{g/mL}$ of BSA, 50% glycerol (pH 7.4 at 25°C) (cat. no. B8001S; New England Biolabs).
4. Enzyme dilution buffer for *Eco*RI diluent C: 250 mM NaCl, 10 mM Tris-HCl, 0.1 mM EDTA, 0.15% Triton X-100, 200 $\mu\text{g/mL}$ of BSA, 50% glycerol (pH 7.4 at 25°C) (cat. no. B8003S; New England Biolabs).
5. Calf intestinal alkaline phosphatase (CIP) (1 U/ μL) (Roche).
6. Phenol;chloroform;isoamyl alcohol (25:24:1) (P-2069; Sigma).
7. Chloroform (Fisher).
8. 3 M Sodium acetate, pH 5.2.
9. Glycogen (20 mg/mL) (Roche).
10. Isopropanol.

2.7. Purification of Digested Vector DNA by Electrophoresis

1. 0.5X TBE buffer: 45 mM Tris-borate, pH 8.3, 1 mM EDTA.
2. Gel-loading dye 2: 0.25% bromophenol blue, 40% (w/v) sucrose in TE.
3. Dialysis tubing: Spectra/Pro or equivalent.
4. Dialysis clip.
5. Submarine gel electrophoresis apparatus (Bio-Rad Sub-Cell GT DNA Electrophoresis Cell, 31 cm length and 16 cm width, or equivalent; Hercules, CA).
6. Centricon YM-100 device (Amicon).
7. 2-L Glass beaker and magnetic stirring bar.
8. TE buffer (pH 8.0).

9. EtBr staining buffer (0.5 $\mu\text{g/mL}$).
10. Alpha Innotech IS1000 digital imager.

2.8. Quality Control of Vector DNA

1. Petri dish (cat. no. 351029, 100 $\times$ 15 mm style, 20/bag; Falcon).
2. LB plates (100 $\times$ 15 mm) containing sucrose/chloramphenicol (*see Subheading 2.2.*).
3. Ampicillin stock solution (100 mg/mL).
4. Electromax DH10B T1 Phage-resistant cells (cat. no. 12033-015; Invitrogen).

2.9. Preparation of DNA Blocks From Leukocytes

1. Blood-drawing equipment.
2. Blood collection tubes containing EDTA with purple cap (Becton Dickinson).
3. Blood (~50 mL).
4. RBC lysis solution.
5. PBS.
6. Automated hematology counter or hemocytometer (VWR counting chamber) with microscope.
7. 50-mL Conical screw-cap polypropylene tube (cat. no. 430828; Corning).
8. Refrigerated centrifuge with rotor/adapters for 50-mL tubes (e.g., Sorvall RT 7 centrifuge with RTH-250 swinging-bucket rotor or equivalent).
9. Rotating mixer.

2.10. Preparation of DNA Blocks From Animal Tissue

1. Dissecting tools (scissors, forceps).
2. Dounce homogenizer.
3. 50-mL Conical screw-cap polypropylene tube (cat. no. 430828; Corning).
4. Disposable Petri dish (Falcon).
5. Equipment for euthanasia using CO₂ gas.

2.11. Embedding of Cells in Agarose

1. PBS.
2. InCert agarose (cat. no. 50123; Cambrex, www.cambrex.com).
3. Disposable DNA plug mold (10 $\times$ 5 $\times$ 1.5 mm) (cat. no. 1703706; BioRad).
4. Microwave.

2.12. Extraction of High-Molecular-Weight DNA in Agarose

1. Cell lysis solution.
2. 50-mL Conical screw-cap polypropylene tube (cat. no. 430828; Corning).
3. Water bath set at 50°C or rotating oven.
4. TE50: 10 mM Tris-HCl (pH 8.0), 50 mM EDTA.
5. PMSF (100 mM stock solution) (cat. no. P-7626; Sigma).
6. 0.5 M EDTA, pH 8.0.

2.13. Preelectrophoresis

1. DNA blocks stored in 0.5 M EDTA.
2. Petri dish (cat. no. 351029, 100 × 15 mm style, 20/bag; Falcon).
3. Sterile 0.5X TBE buffer.
4. 50-mL Conical screw-cap polypropylene tube (cat. no. 430828; Corning).
5. 20-Well 1.5-mm-thick comb, platform (14 × 13 cm), and a gel-casting stand (Bio-Rad).
6. Contour-clamped homogeneous electric field (CHEF) apparatus (Bio-Rad).
7. Ultrapure agarose (Invitrogen).
8. Microwave.
9. Low Range PFG marker (50 gel lanes) (cat. no. N0350S; New England Biolabs).
10. TE buffer, pH 8.0.
11. Alpha Innotech IS1000 digital imager.

2.14. Partial Digestion Using Combination of *EcoRI* and *EcoRI* Methylase

1. Preelectrophoresed DNA blocks stored in TE (pH 8.0).
2. Petri dish (cat. no. 351029, 100 × 15 mm style, 20/bag; Falcon).
3. *EcoRI* (50,000 U; 20,000 U/mL) (cat. no. R0101L; New England Biolabs).
4. *EcoRI* dilution buffer.
5. *EcoRI* methylase (40,000 U/mL) (cat. no. M0211L; New England Biolabs).
6. BSA 10 mg/mL BSA (cat. no. B9001S; New England Biolabs).
7. Spermidine, 0.1 M stock solution.
8. Proteinase K (cat. no. 1 092 766; Roche), 10 mg/mL stock solution in TE, stored at -20°C.
9. *N*-Lauroyl sarcosine (cat no. L-5125; Sigma), 10% stock solution.
10. EDTA, 0.5 M stock solution, pH 8.0.
11. TE50: 10 mM Tris-HCl (pH 8.0), 50 mM EDTA.
12. PMSF (cat no. P-7626; Sigma), 100 mM stock solution.
13. 10X *EcoRI* and *EcoRI* Methylase buffer.
14. 15-Well 1.5-mm-thick comb, platform (14 × 13 cm), and gel-casting stand (Bio-Rad).
15. CHEF apparatus (Bio-Rad).
16. Ultrapure agarose (Invitrogen).
17. Microwave.
18. Low Range PFG marker (cat. no. N0350S; New England Biolabs) (50 gel lanes).

2.15. Partial Digestion Using *MboI*

1. Preelectrophoresed DNA blocks stored in TE (pH 8.0).
2. 10X *MboI* buffer without Mg⁺⁺ and DTT.
3. DTT, 0.1 M stock solution.
4. Proteinase K (cat. no. 1 092 766; Roche), 10 mg/mL stock solution in TE, stored at -20°C.

5. *N*-Lauroyl sarcosine (cat. no. L-5125; Sigma), 10% stock solution.
6. EDTA, 0.5 *M* stock solution, pH 8.0.
7. TE50: 10 *mM* Tris-HCl (pH 8.0), 50 *mM* EDTA.
8. PMSF (cat. no. P-7626; Sigma), 100 *mM* stock solution.
9. Petri dish (cat. no. 351029, 100 – 15 mm style, 20/bag; Falcon).

2.16. Size Fractionation

1. Agarose blocks containing partially digested DNA with either *Eco*RI or *Mbo*I.
2. 20-Well 1.5-mm-thick comb, platform (14 × 13 cm), and gel-casting stand (Bio-Rad).
3. CHEF apparatus (Bio-Rad).
4. Ultrapure agarose (Invitrogen).
5. Microwave.
6. Low Range PFG marker (50 gel lanes) (cat. no. N0350S; New England Biolabs).
7. 15-mL Conical screw-cap polypropylene tubes (Corning).

2.17. Recovery of Insert DNA by Electroelution

1. Size-fractionated DNA stored in 0.5X TBE buffer.
2. Clean forceps.
3. Dialysis tubing: Spectra/Pro Membrane MWCO: 8000, or equivalent.
4. Dialysis clip.
5. 2-L Glass beaker and magnetic stirring bar.
6. TE buffer, pH 8.0.
7. Submarine gel electrophoresis apparatus (Bio-Rad Sub-Cell GT DNA Electrophoresis Cell, 31 cm length and 16 cm width, or equivalent).

2.18. Ligation and Transformation

1. 5X T4 DNA ligase buffer (Invitrogen): 2 *M* Tris-HCl (pH 7.6), 50 *mM* MgCl₂, 5 *mM* adenosine triphosphate, 5 *mM* DTT, 25% (w/v) PEG8000.
2. T4 DNA ligase (Invitrogen) (1 Weiss unit/μL).
3. Proteinase K.
4. PMSF, 100 *mM* stock solution.
5. Microdialysis filters (0.025-μm pore size) (Millipore, Bedford, MA): 25-mm diameter (cat. no. VSWP02500, 100/pack) for small-scale test ligation and 47-mm diameter (cat. no. VSWP04700, 100/pack 0.025-μm pore size, white, 47-mm diameter) for large-scale ligation.
6. Small (for test ligation) and large (for large-scale ligation) Petri dishes.
7. PEG8000 solution: 30% PEG8000 (w/v), 10 *mM* Tris-HCl (pH 8.0), and 0.5 *mM* EDTA.
8. Electromax DH10B T1 Phage-resistant cells (cat. no. 12033-015; Invitrogen).
9. Electroporation cuvette with a 0.15-cm gap (Invitrogen).
10. Electroporator (Cell Porator equipped with a voltage booster; Invitrogen).

11. 14-mL Snap-cap polypropylene tubes (cat. no. 2059, 25/pack; Falcon).
12. SOC medium (Invitrogen): 2% bacto-tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose.
13. LB plates containing 5% sucrose and antibiotics (see Colony Picking for preparation of this medium) in 100 × 15 mm Petri dish (Falcon).
14. Petri dish (cat. no. 351029, 100 × 15 mm style, 20/bag; Falcon).
15. Petri dish (cat. no. 351007, 60 × 15 mm style, 20/bag; Falcon).

2.19. Analyzing BAC Clones

1. Sterile toothpick.
2. LB medium containing antibiotics.
3. AutoGen740 or AutoGen960.
4. Flexible plate, 96-well, U-bottomed without lid (cat. no. 353911, 25/dispenser pack; Falcon).
5. 10X NEBuffer 3 (New England Biolabs): 1 M NaCl, 0.5 M Tris-HCl (pH 7.9), 0.1 M MgCl₂, 10 mM DTT.
6. BSA (New England Biolabs): 10 mg/mL in 20 mM phosphate buffer, 50 mM NaCl, 0.1 mM EDTA, 5% glycerol (pH 7.0 at 25°C).
7. *NotI* (10 U/μL) (cat. no. R0189L; New England Biolabs).
8. 45-Well, 21-cm-wide, 1.5-mm-thick comb (cat. no. 170-3645; Bio-Rad).
9. Wide/long combination casting stand, platform (21 × 14 cm), and gel-casting stand (cat. no. 170-3704; Bio-Rad).
10. Plastic seal (TR100, Thermo Seal Plate Film 2.0 PP, PK/100; Marsh Biomedical).
11. Low Range PFG marker (50 gel lanes) (cat. no. N0350S; New England Biolabs).
12. Gel-loading dye 1.
13. CHEF (Bio-Rad) or field inversion gel electrophoresis (FIGE) apparatus (cat. no. 170-3716; Bio-Rad).

2.20. Preparation of LB Plates Containing Sucrose and Antibiotics

1. Tryptone peptone (pancreatic digest of casein) (500 g) (cat. no. 211705; Difco).
2. Bacto yeast extract (500 g) (cat. no. 212750; Difco).
3. NaCl (50 kg) (cat. no. S-9888; Sigma).
4. Sucrose (2.5 kg) (cat. no. SX1075-3; EM Science).
5. 5 N NaOH.
6. Bacto agar (2 kg) (cat. no. 214030; Difco).
7. Chloramphenicol stock solution (20 mg/mL): Dissolve 1 g of chloramphenicol (cat. no. C0378; Sigma) in 50 mL of 99.5% ethanol and filtrate (Acrodisc, syringe filters, 25 mm, 50/pack; 0.2-μm cat. no. 4192; GermanSciences) into a 50-mL disposable centrifuge tube (Corning 25325-50 or equivalent). Store at -20°C. The antibiotic is stable for 1 yr.
8. Kanamycin stock solution (25 mg/mL): Dissolve 1.25 g of kanamycin (cat. no. K-4000; Sigma) in 50 mL of sterile, deionized distilled water and filtrate. Aliquot

500 μ L into microcentrifuge tubes and store at either 4°C for short term or -20°C for long term. The antibiotic is stable for 1 yr at -20°C.

9. Q-trays/covers (22.2 $\times$ 22.2 cm) (Genetix X6021, 20 plates/box).

2.21. Preparation of LB Medium Containing 7.5% Glycerol

1. Tryptone peptone (pancreatic digest of casein) (500 g) (cat. no. 211705; Difco).
2. Bacto Yeast Extract (500 g) (cat. no. 212750; Difco).
3. NaCl (50 kg) (cat. no. S-9888; Sigma).
4. Glycerol GR ACS (4 L) (cat. no. GX0185-5; EM Science).
5. 5 N NaOH.
6. Bottle-top filter (1 L) with 45-mm neck (cat. no. 430016, 12 filters/box, 0.45- μ m cellulose acetate, low-protein-binding membrane; Corning).
7. 10-L glass bottle.

2.22. Filling of LB Medium Containing 7.5% Glycerol into 384-Well Plates

1. 384-Well microtiter plates (cat. no. X6001, 160 plates/box; Genetix).
2. Q-Fill 2 (Genetix).
3. Chloramphenicol stock solution (20 mg/mL) or kanamycin stock solution (25 mg/mL).

2.23. Thawing 384-Well Plates

1. 384-Well plates from target library.
2. Kimwipes EX-L (cat. no. 34256; 38.1 $\times$ 42.6 cm).
3. Sterile, foil-wrapped blotter paper.
4. Dryers on stands: at least two; optimum, four (two per stand).
5. Large cart.

2.24. Library Replication

1. Thawed 384-well "replication master" plates from target library.
2. 384-Well labeled "copy" plates stacked on cart.
3. Four 384-pin hand tools.
4. Stainless steel dish.
5. 190-Proof ethanol.
6. Bunsen burner.
7. Fire extinguisher (keep a fire extinguisher within reach while replicating the library).
8. Large cart.
9. Laminar flow hood.

2.25. Preparation of LB Plates for High-Density Replica Filters

1. Tryptone peptone (pancreatic digest of casein) (500 g) (cat. no. 211705; Difco).
2. Bacto Yeast Extract (500 g) (cat. no. 212750; Difco).

3. NaCl (50 kg) (cat. no. S-9888; Sigma).
4. 5 *N* NaOH.
5. Ultrapure agarose (cat. no. 15510-027; Invitrogen).
6. Chloramphenicol stock solution (20 mg/mL): Dissolve 1 g of chloramphenicol (cat. no. C0378; Sigma) in 50 mL of 99.5% ethanol and filtrate (Acrodisc, 0.2- μ m syringe filters, 25 mm, 50/pack; cat. no. 4192; GermanSciences) into a 50-mL disposable centrifuge tube (Corning 25325-50 or equivalent). Store at -20°C . The antibiotic is stable for 1 yr.
7. Kanamycin stock solution (25 mg/mL): dissolve 1.25 g of kanamycin (cat. no. K-4000; Sigma) in 50 mL of sterile, deionized distilled water and filtrate. Aliquot 500 μ L into microcentrifuge tubes and store at either 4°C for short term or -20°C for long term. The antibiotic is stable for 1 yr at -20°C .
8. Square Bio Assay Dish (245 $\times$ 245 cm) (cat. no. 77776-742; Corning).

2.26. Setting of Nylon Filters on Agarose Plates

1. Kimwipes EX-L (cat. no. 34256; 38.1 $\times$ 42.6 cm).
2. Nylon filters (22 $\times$ 22 cm, 0.45- μ m pore size) (Schleicher & Schuell).
3. LB agarose plates containing antibiotics.

2.27. Gridding of Filters Using an Automatic Colony-Gridding Machine

1. Thawed 384-well plates from target library with bar codes attached. (Bar codes are to be attached to the narrow end of the Genetix 384-well plate with flat corners.) *See also Subheadings 2.23. and 2.24.*
2. Three BioBanks (each holder accommodates twenty-four 384-well plates): two for 48 thawed 384-well plates and one for a control clone 384-well plate.
3. Kimwipes EX-L (cat. no. 34256; 38.1 $\times$ 42.6 cm).
4. Sterile, foil-wrapped blotter paper.
5. LB agarose plates with 22 $\times$ 22 cm nylon filters.
6. Automatic colony-gridding machine (BioGrid, BioRobotics).

2.28. Processing of Filters

1. Chromatography papers, 3-mm 58 $\times$ 68 cm CHR paper (Whatman).
2. Alkaline solution: 0.5 *M* NaOH, 1.5 *M* NaCl. Dissolve 80 g of NaOH and 350.4 g of NaCl in deionized distilled water and adjust the volume to 4 L.
3. Neutralization solution: 0.5 *M* Tris-HCl, pH 7.5, 1.5 *M* NaCl. Prepare as follows:
 - a. Dissolve 484.4 g of Tris and 350.4 g of NaCl in 1.5 L of deionized water.
 - b. Adjust the pH to 7.4 with 5 *M* HCl ($\sim$ 800 mL).
 - c. Add up the remainder volume of deionized water to make it close to 4 L.
 - d. Adjust the pH to 7.4 again before autoclaving.
4. Pronase (Roche).
5. ProPK buffer: A 4-L solution contains 24.22 g of Tris, 74.40 g of DiEDTA, 23.36 g of NaCl, 40.00 g of *N*-lauroyl-sarcosine, and 3.7–3.8 g of NaOH, adjusted to pH 8.5.

6. NaOH, NaCl, and DiEDTA (Angus Buffers & Biochemicals).
7. Tris and *N*-lauroyl-sarcosine (Sigma).
8. HCl (Fisher).
9. Baking dishes (Pyrex).
10. Electronic timer with at least three channels (VWR).
11. Flat-headed short forceps (Millipore).
12. Water bath with electronic temperature control: Isotemp 220 (Fisher).
13. Square Bio Assay Dish (245 × 245 cm) (cat. no. 77776-742; Corning).
14. Ultraviolet (UV) crosslinker and GS Gene LinkerTM (Bio-Rad).

3. Methods

3.1. Preparation of BAC/PAC Vector for Cloning

This section contains procedures for cloning *EcoRI* partial-digest fragments using either the pBACe3.6 or pTARBAC2.1 vectors, and for cloning *MboI* partial-digest fragments using pBACe3.6, pCYPAC2, pPAC4, or pTARBAC1.3 (see **Note 2**). To reduce the fraction of nonrecombinant vector clones in the libraries, each of these vectors is digested with the cloning enzyme (*EcoRI* or *BamHI*) and an additional enzyme to cut the pUC-link fragment into unclonable pieces. With respect to this “background-reducing” enzyme, the BAC vectors (e.g., pTARBAC1.3) can be digested with *ApaLI* while the PAC vectors are treated with *ScaI*. After the initial digestion of vectors with the *ApaLI* or *ScaI*, the vector is further digested with either *BamHI* or *EcoRI*, as appropriate (see **Note 3**).

3.2. Preparation of LB Plates Containing Antibiotics

1. Add 1.5 g of tryptone peptone, 0.75 g of yeast extract, and 0.75 g of NaCl into 150 mL of deionized distilled water in a 250-mL flask.
2. Mix with a magnetic stirring bar until the powder is completely dissolved.
3. Adjust the pH to 7.2 with 5 *N* NaOH (~68 μ L).
4. Add 2.25 g of bacto agar and stir the solution for 5 min.
5. Cover the flask with aluminum foil.
6. Autoclave the medium still including the magnetic bar at 121°C for 20 min.
7. Once finished, carefully remove the bottle from the autoclave.
8. Stir the medium on a magnetic stirrer and cool the medium to 55°C.
9. Add both 150 μ L of 100 mg/mL ampicillin and 150 μ L of 20 mg/mL chloramphenicol for BAC vectors or 150 μ L 25 mg/mL for PAC vectors.
10. Stir the medium gently to avoid bubbles.
11. Pour approx 25 mL of medium/Petri dish (100 mm diameter).
12. Leave the plates at room temperature for about 45 min to solidify.
13. Wrap the plates in a plastic bag and store them upside down at 4°C. The plates can be kept up to 3 mo at 4°C.

3.3. Testing of Vector

1. Streak *E. coli* DH10B cells containing a vector on an LB plate containing appropriate antibiotics.
2. Incubate at 37°C overnight.
3. Inoculate 12 colonies into two six-well green tubes or twelve 15-mL snap-cap polypropylene tubes each containing 2 mL of LB medium with antibiotics, and incubate at 37°C with shaking at 200 rpm for 16 h.
4. Transfer 500 μ L of each culture into 1.5-mL microcentrifuge tubes containing 72 μ L of sterile 80% glycerol, mix well, and store the glycerol stocks at -80°C.
5. Purify the plasmid DNA from the remaining 1.5-mL culture using an AutoGen 740 or a standard alkaline lysis procedure.
6. Suspend the plasmid DNA in 100 μ L of TE (pH 8.0).
7. Transfer 4 μ L of DNA solution (~100 ng) into 96-well flexible plates, one each for *Apa*LI, *Bam*HI, *Eco*RI, and *Not*I digestion. Aliquot 16 μ L of *Apa*LI, *Bam*HI, *Eco*RI, and *Not*I restriction enzyme cocktail prepared as follows:
 - a. For 15 samples of *Apa*LI digestion, mix 206 μ L of sterile, deionized distilled water; 30 μ L of 10X NE Buffer 4; 3 μ L of 10 mg/mL BSA; and 1 μ L of *Apa*LI (10 U/ μ L) in a microcentrifuge tube on ice.
 - b. For 15 samples of *Bam*HI digestion, mix 206 μ L of sterile deionized distilled water; 30 μ L of 10X *Bam*HI reaction buffer; 3 μ L of 10 mg/mL BSA; and 1 μ L of *Bam*HI (20 U/ μ L) in a microcentrifuge tube on ice.
 - c. For 15 samples of *Eco*RI digestion, mix 209 μ L of sterile, deionized distilled water; 30 μ L of 10X *Eco*RI reaction buffer; and 1 μ L of *Eco*RI (20 U/ μ L) in a microcentrifuge tube on ice.
 - d. For 15 samples of *Not*I digestion, mix 206 μ L of sterile, deionized distilled water; 30 μ L of 10X NE Buffer 3; 3 μ L of 10 mg/mL BSA; and 1 μ L of *Not*I (10 U/ μ L) in a microcentrifuge tube on ice.
8. Incubate at 37°C for 1 h.
9. Weigh 0.7 g of agarose and add to 100 mL of 0.5X TBE buffer. Melt the agarose using a microwave, and cool at 50°C with stirring.
10. Prepare a 10-cm-long, 15-cm-wide gel tray. Wipe the tray with 95% ethanol and seal the edge of the tray with plastic tape.
11. Wipe a 33-well comb (14 cm long, 1.5 mm thick) with 95% ethanol.
12. Set the gel tray on a horizontal bench and place the comb on the tray. Adjust the height of the comb from the bottom of the gel tray using a 1.5-mm spacer.
13. Pour the 0.7% molten agarose in the tray and solidify at room temperature for at least 1 h.
14. Add 2 μ L of gel-loading dye 2 into the samples after 1 h incubation at step 9, and mix gently.
15. Load the samples into the wells of the 0.7% agarose gel (15 $\times$ 10 cm) in 0.5X TBE buffer and run at 6 V/cm for 1 h.
16. Stain the gel in EtBr solution for 30 min.

17. Take a picture and make sure that the vector is not rearranged.
 - a. pBACe3.6: Three bands (9.8, 1.3, 0.5 kb) for *Apa*LI digestion should be visible, and two bands (8.8, 2.8 kb) for *Eco*RI, *Bam*HI, and *Not*I should be visible in the case of complete digestion.
 - b. pTARBAC1.3: Three bands (11.7, 1.3, 0.5 kb) for *Apa*LI digestion should be visible, and two bands (10.7, 2.8 kb) for *Eco*RI, *Bam*HI, and *Not*I should be visible in the case of complete digestion.
 - c. pTARBAC2.1: The vector cannot be digested with *Bam*HI. Three bands (11.7, 1.3, 0.5 kb) for *Apa*LI digestion should be visible, and two bands (10.7, 2.8 kb) for *Eco*RI and *Not*I should be visible in the case of complete digestion.

3.4. Purification of Vector DNA

1. Inoculate 500 μ L of nonrearranged glycerol stock solution into two bottles of 2-L flasks containing 750 mL of LB medium with antibiotics.
2. Incubate at 37°C with shaking at 200 rpm for 20 h.
3. Transfer the culture into six 250-mL centrifuge tubes and close the caps tightly.
4. Centrifuge at 4000g (5200 rpm with an SLA-1500 Sorvall Centrifuge rotor) for 10 min at 4°C.
5. Add 10 mL of cell suspension buffer to each tube and resuspend the cells thoroughly. Combine the cells into two centrifuge tubes.
6. Add 60 mL of lysis solution, mix gently, and keep on ice for 5 min.
7. Add 45 mL of ice-cold potassium acetate (pH 4.8) solution, mix gently, and keep on ice for 10 min.
8. Centrifuge at 5500g (6000 rpm with an SLA-1500 Sorvall Centrifuge rotor) for 20 min at 4°C.
9. Transfer the supernatant into two, clean 250-mL centrifuge tubes; add 70 mL of isopropanol (0.6 times the volume); and keep at 4°C for at least 15 min. The sample can be kept at 4°C overnight.
10. Centrifuge at 12,000g (9000 rpm with an SLA-1500 Sorvall Centrifuge rotor) for 30 min at 4°C.
11. Discard the supernatant being careful not to disturb the pellet.
12. Add 100 mL of 70% ethanol, and rotate the tubes to rinse the pellet and the inside of the tubes.
13. Centrifuge at 12,000g for 3 min at 4°C.
14. Carefully remove the supernatant.
15. Dry the pellet in an air-circulating hood. It is difficult to dissolve the DNA if it is dried completely. Check the dryness every 5 min.
16. Add 2 mL of TE buffer (pH 8.0) to each tube and dissolve the pellet.
17. Combine the solution into a 50-mL conical tube and measure the volume.
18. Add 1 g of CsCl for each milliliter of solution and dissolve the salt completely. Incubating at 37°C in a shaking incubator facilitates dissolving of the CsCl into the solution.
19. Transfer the solution into two ultracentrifuge tubes using a 6-mL syringe with a G18 needle.

20. Add 400 μL of 10 mg/mL EtBr solution into each tube.
21. Add CsCl solution to fill the tubes near the top.
22. Balance the tubes within 0.03 g by adding mineral oil onto the solution.
23. Close the tubes using a heating sealer.
24. Set the tubes in a VTi 50 rotor and centrifuge at average 174,633g (maximum 205,235g; 46,000 rpm) in an L8-M model ultracentrifuge (Beckman) for at least 24 h at 23°C.
25. Remove the tube from the rotor being very careful not to disturb the gradient.
26. Observe DNA band under white light (*see Note 4*).
27. Place a 2-cm-long plastic tape on the side of the tube covering the DNA band. Place another plastic tape on top of the tube.
28. Using a needle, poke a hole on top of the tube through the plastic tape. Poke another hole on the side of the tube through the plastic tape, and using a G18 needle with a syringe, recover the DNA band from the tube.
29. Collect the recovered solution into a 15-mL tube with a screw cap.

3.5. Removal of EtBr

1. Add equal volume of isoamyl alcohol into the tube and mix by gentle inversion.
2. Centrifuge at 1864g at room temperature for 3 min.
3. Remove the top organic phase by pipet but do not disturb the lower aqueous phase. Discard the organic solution.
4. Repeat steps 1–3 until red color is completely removed.
5. Cut a piece of dialysis tubing 15 cm long and soak in a 200-mL glass beaker containing sterile, deionized distilled water.
6. Close one end of the dialysis tubing with a dialysis clip.
7. Transfer the solution into the dialysis tubing.
8. Close the other end with a dialysis clip leaving space for two times the volume expansion.
9. Dialyze in 2 L of TE (pH 8.0) for 48 h while exchanging TE buffer (pH 8.0) four times every 10 h.
10. Recover the solution from the tubing into a 15-mL tube for temporal storage.
11. Transfer 4 μL of DNA solution into 96-well flexible plates, one each for *Apa*LI, *Bam*HI, *Eco*RI, and *Not*I digestion. Digest with *Apa*LI, *Bam*HI, *Eco*RI, and *Not*I in 20 μL of reaction mixture by following **steps 7–14 in Subheading 3.3**.
12. Prepare 50, 40, 30, 20, 10, and 5 ng of standard DNA by mixing λ DNA, 2 μL of gel-loading dye 2, and 5 μL of 0.5X TBE buffer as indicated (*see Note 5*).
 - a. 50 ng: 5 μL of 10 ng/ μL standard DNA.
 - b. 40 ng: 4 μL of 10 ng/ μL standard DNA.
 - c. 30 ng: 6 μL of 5 ng/ μL standard DNA.
 - d. 20 ng: 4 μL of 5 ng/ μL standard DNA.
 - e. 10 ng: 5 μL of 2 ng/ μL standard DNA.
 - f. 5 ng: 2.5 μL of 2 ng/ μL standard DNA.

13. Load the sample as well as the standard DNA into 0.7% agarose gel (15 × 10 cm) in 0.5X TBE buffer and run at 6 V/cm for 1 h.
14. Stain the gel in EtBr solution for 30 min.
15. Take a picture using the digital imager. Estimate DNA concentration based on the intensity of the DNA band using λ DNA as a standard.
16. Aliquot the solution in microcentrifuge tubes to 1 mL each and store at -80°C.

3.6. Digestion of Vector DNA With Restriction Enzymes

1. Mix 12 μ g of vector DNA, 50 μ L of 10X NE Buffer 4, 5 μ L of 10 mg/mL BSA, and 10 μ L of *Apa*LI (1 U/ μ L). Adjust the volume to 500 μ L with sterile, deionized distilled water. Prepare four separate reactions in parallel.
2. Dilute *Apa*LI (10 U/ μ L) to 1 U/ μ L with enzyme dilution buffer prior to use. The amount of *Apa*LI can be reduced as long as complete digestion is achieved.
3. Incubate at 37°C for 15 min.
4. Add 3 U of CIP and incubate at 37°C for 1 h.
5. Keep on ice. Confirm complete digestion by following **steps 9–17 in Subheading 3.3**. Load 4 μ L of DNA into 0.7% agarose gel in 0.5X TBE buffer, and electrophorese at 6 V/cm for 1 h at **step 15 in Subheading 3.3**.
6. During electrophoresis, extract the solution with phenol/chloroform and centrifuge at 16,000g (maximum speed: 13,000 rpm) in a microcentrifuge at room temperature for 3 min.
7. Transfer the supernatant into a new 1.5-mL microcentrifuge tube and extract with 500 μ L of chloroform.
8. Centrifuge at 16,000g (maximum speed: 13,000 rpm) in a microcentrifuge at room temperature for 3 min, and transfer the supernatant into a new 1.5-mL microcentrifuge tube.
9. Add 50 μ L of 3 M sodium acetate (pH 5.2) and 1 μ L of 20 mg/mL glycogen and mix. Add 500 μ L of isopropanol and mix. Keep at -20°C for at least 2 h.
10. Centrifuge at 16,000g (maximum speed: 13,000 rpm) in a microcentrifuge at 4°C for 30 min. Carefully remove the supernatant and rinse the pellet with 70% ethanol twice. Do not dry thoroughly; it will be difficult to dissolve DNA for the next step.
11. Dry the pellet under a hood. Dissolve DNA in 440 μ L of water.
12. Set *Eco*RI digestion reactions as follows:
 - a. DNA: 12 μ g.
 - b. 10X *Eco*RI buffer: 50 μ L.
 - c. *Eco*RI (1 U/ μ L): 3, 5, 7, and 10 μ L.
13. Incubate at 37°C for 15 min.
14. Add 1 U of CIP (1 μ L) and incubate at 37°C for 1 h.
15. Inactivate the enzymes and recover the DNA by following **steps 6–10**.
16. Dissolve in 150 μ L of TE and keep on ice until the sample is loaded in a gel.

3.7. Purification of Digested Vector DNA by Electrophoresis

1. Melt 250 mL of 0.7% agarose in 0.5X TBE buffer with a microwave and cool at 50°C with stirring. Prepare four bottles of molten agarose.

2. Prepare a 25-cm-long, 15-cm-wide gel tray. Wipe the tray with 95% ethanol and seal the edge of the tray with plastic tape.
3. Prepare a 20-well comb (12.8 cm long, 1.5 mm thick). Seal 16 wells with autoclave tape to create a large preparative well in the middle, and wipe the comb with 95% ethanol.
4. Adjust the height of the comb from the bottom of the gel tray using a 1.5-mm spacer. Set the gel tray and the comb on a horizontal surface.
5. Pour 0.7% molten agarose in the tray and solidify at room temperature for at least 1 h. Prepare four gels.
6. Pour 1.8 L of 0.5X TBE buffer in a submarine gel electrophoresis tank (31 cm long, 16 cm wide).
7. Carefully remove the comb and plastic tape, and set the gel in the electrophoresis tank.
8. Add 20 μ L of gel-loading dye 2 in the sample and mix well. Carefully load the sample in the large well and a 1-kb DNA ladder marker in the most outer wells on each side of the gel.
9. Let sit for 10 min to allow the DNA to diffuse into the well homogeneously. Run at 100 V (3 V/cm) for 16 h at room temperature.
10. Place a ruler 2 mm inside from an edge of the preparative well and cut the gel. Place the ruler 2 mm inside from the other edge of the preparative well and cut the gel. Stain the outer portions of the gel with EtBr.
11. Store the remaining part (middle portion of the gel) at 4°C. Do not stain this part with EtBr.
12. Place the gels with a fluorescent ruler, the 0-cm position of which is adjusted at the well position of the gel, on an Alpha Innotech IS1000 digital imager. Take a gel image of the gels to identify the position of the vector fragment. Transfer the gels back into the EtBr solution.
13. Determine the position where the vector DNA is based on the picture. Slice off the vector portion from the unstained gel. Stain the remaining gel pieces that do contain vector DNA fragments together with the outer portion of the gels that are kept from **step 12** in 0.5 μ g/mL of EtBr solution for at least 30 min.
14. Cut a piece of dialysis tubing approx 13 cm long and rinse with sterile, distilled deionized water.
15. Close one end of the tubing with a dialysis clip, and place the sliced gel piece that contains vector fragment in the tubing.
16. Add 1 mL of 0.5X TBE buffer in the tube and remove bubbles thoroughly. Close the other end of the tubing with a dialysis clip.
17. Set the tubing in the middle of the tank, and orient the long axis of the tubing along the width so that the current is able to pass through the width of the gel without obstruction.
18. Electrophorese at 100 V (3 V/cm) for 3 h to elute the DNA from the gel slice.
19. Reverse the current for 30 s to release the DNA from the wall of the dialysis tubing.

20. Open one of the dialysis clips and remove the gel slice. Stain the gel slice together with the rest of the gel portions that are kept from **step 13** in 0.5 µg/mL of EtBr solution for at least 30 min.
21. Recover the solution from the tubing and transfer to a Centricon YM-100 device.
22. Centrifuge at 500g (2200 rpm in a Sorvall SM24 rotor) for 30 min at 4°C.
23. Add 2.0 mL of TE buffer and centrifuge at 500g (2200 rpm in a Sorvall SM24 rotor) for 30 min at 4°C.
24. Add 2.0 mL of TE buffer to the retentate and repeat centrifugation. Repeat the washing three times.
25. Assemble the gel pieces kept from **step 20** on the digital imager. Capture an image to ascertain whether vector DNA is sliced out from the gel correctly and DNA is eluted from the gel slice by electroelution.
26. Recover the retentate from the device and determine the DNA concentration as described in **steps 12–14** in **Subheading 3.5**.

3.8. Quality Control of Vector DNA

Quality control should be done prior to use for construction of a BAC library. It is extremely important to know the number of nonrecombinant clones per ligation. If the vector is digested with restriction enzymes at the correct restriction sites, clones that retain self-ligated vector will not grow on the medium containing sucrose. However, noninsert clones are often observed, which retain smaller vector size than regular ones, by analyzing clones with PFGE.

1. Mix 25 ng of pBACe3.6 vector (30 ng of pTARBAC or 50 ng of PAC vector) and 10 µL of 5X T4 DNA ligase buffer in a microcentrifuge tube.
2. Add sterile, deionized distilled water to bring the total volume to 49 µL and mix gently.
3. Add 1 Weiss unit of T4 DNA ligase (1 µL) and mix gently. Incubate at 4°C for 3 h for *EcoRI-EcoRI* ligation (pBACe3.6 or pTARBAC2.1) or 6 h for *MboI-BamHI* (pBACe3.6 or pTARBAC1.3) ligation.
4. Follow **steps 4–12** in **Subheading 3.10.1**.
5. Mix 8 µL of ligation mixture and 80 µL of electrocompetent cells in a microcentrifuge tube and keep on ice.
6. Place wet ice in an electroporation chamber, and set an electroporation cuvet in the chamber.
7. Prepare a 15-mL snap-cap polypropylene tube containing 2 mL of SOC medium.
8. Transfer 22 µL of ligation and electrocompetent cell mixture into the cuvet using a wide-bore pipet tip, placing the droplet between the electrode.
9. Deliver a pulse using the same conditions as in **step 17** in **Subheading 3.10.1**.
10. Collect the cells and transfer into the 15-mL snap-cap polypropylene tube containing 2 mL of SOC medium. Perform four transformations by repeating **steps 8–10**. Transfer the sample in the same tube each time.
11. Incubate at 37°C in an orbital shaker at 200 rpm for 1 h.
12. Clean a flow hood with 70% ethanol.

13. Take four LB plates containing sucrose/chloramphenicol from the plastic bag. To two of four plates add 400 μ L of sterile, deionized distilled water and 30 μ L of 100 mg/mL ampicillin; mix; and spread homogeneously. Dry these four plates in the hood for 40 min.
14. Soak a glass spreader in ethanol and flame. Keep in the hood.
15. Spread 500 μ L of cells on each of the plates.
16. Dry the plates and incubate at 37°C overnight; it takes 10–15 min to dry the plates.
17. Count the number of colonies on the sucrose/chloramphenicol plates and sucrose/chloramphenicol/ampicillin plates (*see Notes 6 and 7*).

3.9. Preparation of Insert DNA

Isolating chromosomal DNA is a critical step for constructing a genomic DNA library. To construct a large-insert (>150-kb) library, high-molecular-weight DNA has to be isolated from cells. It is difficult, if not impossible, to isolate a large (>100-kb) chromosomal DNA molecule in solution because of physical breakage during preparation. To isolate high-molecular-weight DNA without causing physical shearing, cells are embedded in agarose. The agarose blocks containing cells are treated in solution containing proteinase K, *N*-lauroyl sarcosine, and EDTA. The cells are lysed and most of the biologic components, such as protein and lipids, are removed in this solution. High-molecular-weight DNA is protected from nuclease digestion in a high concentration of EDTA and from physical shearing by embedding in agarose. Cultured cell lines are a good source for isolating DNA. Chromosome rearrangement might occur during cell culture. It is therefore desirable to obtain DNA from a live animal. The most convenient source for this purpose is to isolate DNA from circulating leukocytes. Although it is difficult to collect blood samples from small animals, such as mice, it is feasible to use tissue from a live animal.

3.9.1. Preparation of DNA Blocks From Leukocytes

The procedures described here are applicable to cultured cell lines by omitting the erythrocyte lysis step.

1. Obtain approx 50 mL of venous blood from a healthy animal using blood-drawing equipment in blood collection tubes containing EDTA. Mix well to avoid clot formation (*see Note 8*).
2. Divide the blood into two 50-mL conical screw-cap polypropylene tubes (approx 25 mL each), add 10 mL of ice-cold PBS, and mix gently.
3. Centrifuge at 1864*g* (equivalent to 3000 rpm using an RTH-250 rotor) for 5 min at 4°C. Remove the upper layer using a 10-mL disposable pipet but leave a small volume of upper layer so as not to remove any white cells from the fuzzy coat layer. A thin layer of white blood cells (the fuzzy coat layer) should be seen between the plasma (upper) layer and the RBC layer. The supernatant should be mixed with bleach and kept for at least a day prior to discarding into a sink.

4. Add 10 mL of ice-cold PBS and mix gently. Repeat wash 10 times. If the color of the upper layer becomes clear after washing five to six times, the washing step may be discontinued.
5. After the final wash, remove the upper layer as much as possible.
6. Mix the cell suspension well and divide into four 50-mL conical screw-cap polypropylene tubes.
7. Add 25 mL of 1X RBC lysis solution into each tube, and mix gently on a roller mixer at room temperature (*see Note 9*).
8. Carefully watch the color change from light red to dark red. The color change occurs suddenly and usually happens within 30 min.
9. Centrifuge the tubes for 10 min at 207g (equivalent to 1000 rpm using an RTH-250 rotor) at 4°C. A white pellet of leukocytes should be observed at the bottom of each tube.
10. Discard the supernatant by gentle inversion so as not to disturb the leukocyte pellet.
11. Rinse the inside of the tubes with 2 mL of ice-cold PBS, and remove the supernatant with a micropipet without disturbing the pellet.
12. Suspend the leukocytes in 10 mL of ice-cold PBS, and combine in one tube.
13. Centrifuge for 5 min at 207g at 4°C, and discard the supernatant by gentle inversion.
14. Repeat the washing step with ice-cold PBS until most of the red color is removed. A small amount of red color may stay with the pellet.
15. Suspend the cells in approx 2 mL of ice-cold PBS. Prepare a 20X dilution of cell suspension by mixing 2 μ L of cell suspension into 38 μ L of PBS to estimate the number of cells per milliliter.
16. Rinse the hemocytometer with 95% ethanol and wipe with a Kimwipe.
17. Place a cover slip on a hemocytometer. Apply 10 μ L of cell suspension between the cover slip and the hemocytometer allowing diffusion of the solution by capillary action.
18. Count the number of cells in the five middle-size (0.2-mm) squares using $\times 400$ magnification (ocular: $\times 10$; objective: $\times 40$) under a microscope (*see Note 10*).
19. Calculate the cell concentration as follows: Number of cells per five 0.2-mm squares $\times 5 \times 10^4 \times 20$ (dilution factor) = number of cells/mL.
20. Dilute the cell suspension to 1×10^8 cells/mL (~ 600 μ g of DNA/mL) and keep on ice.
21. Proceed to **Subheading 3.9.3**.

3.9.2. Preparation of DNA Blocks From Animal Tissue

1. Euthanize an animal by flushing CO₂ gas into a desiccator.
2. Dissect the animal from the abdomen using sharp scissors and remove the spleen, kidney, liver, and brain. Transfer each tissue onto a Petri dish that is on ice (*see Note 11*).
3. Rinse with ice-cold PBS and remove hairs with forceps. Remove connective tissues, which are like fibers, with scissors and forceps.

4. Cut each organ into small pieces and transfer them to a 15-mL Wheaton Dounce homogenizer with a “tight” pestle.
5. Add 2 to 3 mL of ice-cold PBS into the homogenizer and homogenize gently five times on ice.
6. Transfer the cell suspension into a 50-mL conical screw-cap polypropylene tube.
7. Repeat **steps 5 and 6** until the tissue is completely homogenized. Remove large debris with forceps.
8. Add ice-cold PBS to 50 mL and stand on ice for 3 min.
9. Transfer the supernatant into a new 50-mL conical screw-cap polypropylene tube by slowly slanting the tube paying attention not to transfer large debris.
10. Centrifuge at 207g (equivalent to 1000 rpm using an RTH-250 rotor) for 10 min at 4°C.
11. Discard the supernatant by inverting the tube gently.
12. Suspend the cells in residual solution by tapping gently on ice. Add 1 mL of ice-cold PBS and mix by gently pipetting up and down. Remove large debris that is not possible to disperse with the pipet.
13. Add 49 mL of ice-cold PBS and mix gently. Repeat **steps 10 and 11**.
14. Suspend the cells completely in residual solution by gentle tapping on ice.
15. For sample from the kidney, spleen, and liver, proceed to **steps 16–19**. For sample from the brain, go to **step 20**.
16. Add 1 mL of ice-cold PBS and mix gently. Prepare a 20X dilution of cell suspension by mixing 2 μ L of cell suspension into 38 μ L of PBS to estimate the number of cells per milliliter.
17. Estimate the cell concentration by following **steps 16–19 in Subheading 3.9.1**.
18. Use the estimation that 60% of the cells counted contain chromosomal DNA, assuming that 40% of the cells are erythrocytes that do not have chromosomal DNA.
19. Dilute the cell suspension to 1×10^8 cells/mL after subtracting the factor of erythrocytes and keep on ice.
20. Estimate a volume of brain sample. Add ice-cold PBS using the following ratio: 4 mL of brain: 1 mL of PBS. Keep on ice.
21. Proceed to **Subheading 3.9.3**.

3.9.3. Embedding of Cells in Agarose

1. Melt 0.1 g of InCert agarose in 10 mL of PBS in a microwave and keep at 50°C in a water bath.
2. Mix the cell suspension well and transfer 400 μ L into a clean microcentrifuge tube. Warm the tube by gripping with fingers for 3 min.
3. Add 400 μ L of 1% molten InCert agarose to the tube and mix by gently pipetting up and down taking care not to make bubbles. The final agarose concentration is 0.5% and the cell concentration is 5×10^7 /mL.
4. Load the cell-agarose mixture as quickly as possible into $10 \times 5.5 \times 1.5$ mm disposable block molds by pipet.
5. Place the molds on ice for 0.5–1 h to solidify the agarose (*see Note 12*).

3.9.4. Extraction of High-Molecular-Weight DNA in Agarose

1. Break a plastic piece that is used as a tool to push the DNA blocks out from the edge of a mold and peel off the white plastic tape from the bottom of the mold. Using the tool directly extrude the DNA blocks from the mold into 50-mL conical screw-cap polypropylene tubes containing 50 mL of cell lysis solution. It is desirable to treat less than 50 DNA blocks in 50 mL of lysis solution.
2. Incubate the tube containing the blocks at 50°C in a water bath with periodic mixing or in a rotating oven. Continue incubation for 24 h. Residual red color disappears within a couple of hours.
3. Discard the lysis solution, add fresh lysis solution and continue incubation at 50°C for another 24 h.
4. Remove the lysis solution and rinse the DNA blocks with sterile, distilled deionized water several times.
5. Add 50 mL of TE50 buffer and rotate on a roller mixer at 4°C for 24 h. Replace the TE50 buffer with fresh TE50 buffer at least twice during the rotating.
6. Rinse the DNA blocks with 50 mL of TE50 buffer containing 0.1 mM PMSF on the roller mixer at 4°C twice, for 2 h each, to inactivate proteinase K.
7. Rinse with TE50 buffer by rotating on the roller mixer at 4°C for 24 h. Replace the TE50 buffer with fresh TE50 buffer at least one time during the rotating.
8. Store the DNA blocks in 0.5 M EDTA (pH 8.0) at 4°C.

3.9.5. Preelectrophoresis

1. Pour agarose blocks off into a Petri dish and remove 0.5 M EDTA solution with a pipet. Agarose blocks may stick on the dish surface after removal of the EDTA solution.
2. Add 10 mL of sterile 0.5X TBE buffer into the dish and stir with a pipet tip gently to release DNA blocks from the dish surface.
3. Transfer the DNA blocks to a 50-mL conical screw-cap polypropylene tube and add sterile 0.5X TBE buffer up to 50 mL.
4. Dialyze the DNA blocks rotating the tube on a mixer for at least 2 h.
5. Cover 16 wells of a 20-well, 1.5-mm-thick comb with autoclave tape to create a large preparative slot that provides a sufficient space to array DNA blocks leaving 2 wells each on both sides (*see Note 13*).
6. Clean the comb, a platform (14 × 13 cm), and a gel-casting stand with 95% ethanol. Set the clean platform and comb in the gel-casting stand.
7. Using a microwave, thoroughly melt 1.5 g of agarose in 150 mL of 0.5X TBE buffer in a 500-mL glass bottle containing a magnetic stirring bar. Cool molten agarose to 55°C with stirring, and pour into the gel-casting stand. Allow the gel to solidify for 1 h at room temperature.
8. Pour 2 L of 0.5X TBE buffer in a CHEF apparatus tank, and equilibrate the unit at 14°C during dialysis and preparation of the gel.
9. Remove the comb gently from the solidified gel, and load the DNA blocks into the large preparative slot. Do not seal the well with molten agarose.

10. Load Low Range PFG marker in the outermost wells on each side of the gel.
11. Place the gel in the precooled unit and run at 4.0 V/cm for 10 h with a 5-s constant pulse time, 120° included angle.
12. Remove the DNA blocks from the well and transfer into a 50-mL conical screw-cap polypropylene tube containing 50 mL of TE buffer (pH 8.0). Dialyze the DNA blocks by rotating the tube on a mixer for at least 2 h.
13. Stain the gel in 0.5 µg/mL of EtBr solution for at least 30 min, and take a gel image on an Alpha Innotech IS1000 digital imager (*see Note 14*).

3.9.6. Partial Digestion Using Combination of *EcoRI* and *EcoRI* Methylase

1. Transfer preelectrophoresed DNA blocks into a Petri dish. Remove TE buffer with a pipet and cut the DNA blocks into four pieces.
2. Transfer the small DNA block pieces into four microcentrifuge tubes.
3. Add 25 µL of 10 mg/mL BSA; 50 µL of 10X *EcoRI* and *EcoRI* Methylase buffer; 13 µL of 0.1 M spermidine; and 390 µL of sterile, distilled deionized water in the tubes and mix well.
4. Add *EcoRI* and *EcoRI* Methylase in each tube as described next. *EcoRI* is diluted with enzyme dilution buffer prior to use.
 - a. Tube 1: 0 U of *EcoRI* and 0 U of *EcoRI* Methylase.
 - b. Tube 2: 1 U of *EcoRI* and 0 U of *EcoRI* Methylase.
 - c. Tube 3: 2 U of *EcoRI* and 50 U of *EcoRI* Methylase.
 - d. Tube 4: 2 U of *EcoRI* and 100 U of *EcoRI* Methylase.
5. Place the tubes on ice for 1 h to allow the enzymes to diffuse into the agarose blocks.
6. Incubate at 37°C for 2.5 h.
7. Add 150 µL of 0.5 M EDTA, 30 µL of 10 mg/mL proteinase K, and 75 µL of 10% *N*-lauroyl-sarcosine and mix well. Incubate at 37°C for 1 h. The partial digestion reaction is stopped and the enzymes are inactivated by proteinase K.
8. Remove the solution from each tube using a pipet. Add 1 mL of TE50 buffer and mix gently.
9. Remove the solution from each tube using a pipet.
10. Add 1 mL of TE50 buffer containing 100 µM PMSF and mix gently. Keep at room temperature for 20 min.
11. Wash with 1 mL of TE50 buffer containing 100 µM PMSF three times.
12. Rinse the DNA blocks with 1 mL of TE50 buffer twice. Partially digested DNA in agarose can be stored in TE50 buffer for at least 1 wk.
13. Clean a 15-well, 1.5-mm-thick comb; a platform (14 × 13 cm), and a gel-casting stand with 95% ethanol. Set the clean platform and the comb in the gel-casting stand.
14. Using a microwave, thoroughly melt 1.5 g of agarose in 150 mL of 0.5X TBE buffer in a 500-mL glass bottle containing a magnetic stirring bar. Cool the molten agarose to 55°C with stirring and pour into the gel-casting stand.

15. Allow the gel to solidify for 1 h at room temperature.
16. Pour 2 L of 0.5X TBE buffer in a CHEF apparatus tank, and equilibrate the unit at 14°C during the partial digestion procedure and preparation of the gel.
17. Move the comb gently from the solidified gel. Load the DNA blocks in the middle wells and Low Range PFG marker in the outermost lanes on each side of the samples.
18. Seal the remaining space in the wells with 1% molten agarose.
19. Place the gel in the precooled unit and run at 6 V/cm for 16 h with a 0.1 to 40-s pulse time, 120° included angle at 14°C.
20. Stain the gel in 0.5 µg/mL of EtBr solution, and take a gel image on an Alpha Innotech IS1000 digital imager.
21. Determine the optimal partial digestion condition. More partially digested DNA between 150–200 kb is a better partial digestion condition. The sample from tube 1 is a negative control; no smearing pattern should be observed.
22. Once the optimal partial digestion condition is determined, repeat **steps 1–12** except for **step 4** using two DNA blocks as a starting material. Add the optimal amount of enzymes per tube at **step 4**. Keep agarose blocks containing partially digested DNA in TE50 solution until starting size fractionation.

3.9.7. Partial Digestion Using *MboI*

1. Transfer preelectrophoresed DNA blocks into a petri dish. Remove TE buffer with a pipet and cut the DNA blocks into four pieces.
2. Transfer the small DNA block pieces into four microcentrifuge tubes.
3. Add 50 µL of 10X *MboI* buffer without Mg⁺⁺ and DTT, 5 µL of 0.1 M DTT and 420 µL water. Mix gently.
4. Keep on ice for 5 min and add *MboI* in each tube as follows. *MboI* is diluted with enzyme dilution buffer prior to use.
 - a. Tube 1: 0 U of *MboI*.
 - b. Tube 2: 1 U of *MboI*.
 - c. Tube 3: 2 U of *MboI*.
 - d. Tube 4: 4 U of *MboI*.
5. Place on ice for 1 h to allow the enzymes to diffuse into the agarose blocks.
6. Add 5 µL of 1 M MgCl₂ and keep on ice for 15 min.
7. Incubate at 37°C for 20 min and keep on ice.
8. Follow **steps 7–12** in **Subheading 3.9.6**.
9. Proceed to **steps 13–20** in **Subheading 3.9.6**.
10. Determine the optimal partial digestion condition that will allow obtaining the highest DNA concentration between 150 and 200 kb (*see Note 15*).
11. Once the optimal partial digestion condition is determined, repeat **steps 1–7** except for **steps 3** and **6** using two DNA blocks as a starting material. Add the optimal amount of *MboI* per tube at **step 3** or incubate at 37°C for the optimized time. Keep agarose blocks containing partially digested DNA in TE50 solution until starting size fractionation.

3.9.8. Size Fractionation

1. Cover six wells of a 20-well, 1.5-mm-thick comb with autoclave tape to create a large preparative slot that provides sufficient space to array DNA blocks leaving 8 wells each on both sides.
2. Clean the covered comb, a platform (14 × 13 cm), and a gel-casting stand with 95% ethanol. Set the clean platform in the gel-casting stand. Set the comb 1 cm away from the nearest edge of the gel in the gel-casting stand.
3. Using a microwave, thoroughly melt 1.5 g of agarose in 150 mL of 0.5X TBE buffer in a 500-mL glass bottle containing a magnetic stirring bar. Cool the molten agarose to 55°C with stirring and pour into the gel-casting stand. Allow the gel to solidify for 1 h at room temperature.
4. Pour 2 L of 0.5X TBE buffer in a CHEF apparatus tank, and equilibrate the unit at 14°C during dialysis and preparation of the gel.
5. Gently remove the comb gently from the solidified gel and array the DNA blocks in the large preparative slot. Eight small agarose blocks derived from two large agarose blocks should fit in the large preparative well.
6. Load the agarose blocks and Low Range PFG marker in the preparative well leaving an empty well on each side of the samples.
7. Cover all the wells (preparative, marker, and empty wells) with 1% molten agarose gel.
8. Place the gel in the precooled unit. Orient the gel so that DNA migrates from the wells toward the nearest gel edge (1 cm away from the wells). Run at 4.7 V/cm for 5 h with a 15-s constant pulse time, 120° included angle (*see Note 16*).
9. Discard the electrophoresis buffer, keep the gel in the tank, and wipe residual buffer off on the gel with a large Kimwipe.
10. Remove the agarose blocks as well as covered agarose from the preparative well using a γ -ray-sterilized inoculating loop. Suck out residual buffers in the well by pipet. Pour 1% molten agarose in the preparative well, and keep at room temperature for 5 min to solidify.
11. Pour 2 L of fresh 0.5X TBE buffer in a CHEF apparatus tank and equilibrate the unit at 14°C.
12. Rotate the gel 180° in the tank and run using the same conditions as in **step 8**. This returns all DNA fragments remaining in the narrow space of the gel to the original preparative well.
13. Repeat **steps 8–12** once more.
14. Apply new Low Range PFG marker to the outer wells of the original marker wells.
15. Fractionate the DNA molecule using the following electrophoresis conditions: 6 V/cm, 16 h, 0.1- (initial) to 40-s (final) linear pulse time, buffer temperature of 14°C, 120° included angle.
16. Place a ruler 2 mm inside from an edge of the preparative well and cut the gel. Place the ruler 2 mm inside from the other edge of the preparative well and cut the gel. The middle part of the gel contains size-fractionated DNA. The outer parts of

the gel contain original and second markers as well as size-fractionated DNA in the 2-mm space that make it feasible to assess the success of the partial digestion and size fractionation.

17. Wrap the middle portion of the gel with a plastic wrap and keep at 4°C. The size-fractionated DNA in the middle portion of the gel should not be stained with EtBr nor be exposed to UV light.
18. Stain the outer portions of the gel in 0.5 µg/mL of EtBr solution for at least 30 min. Place the gels with a fluorescent ruler, the 0-cm position of which is adjusted at the well position of the gel, on an Alpha Innotech IS1000 digital imager. Take a gel image of the gels. Keep the gels in the EtBr (*see Note 17*).
19. Prepare an agarose gel following **steps 13–16** in **Subheading 3.9.6**.
20. Determine the approximate position containing 150–300 kb of DNA fragments based on the picture.
21. Slice the stored gel (middle portion) by cutting horizontally at 0.3- to 0.5-cm intervals in the range of 150–300 kb. Stain the gel pieces that contain DNA fragments below 150 kb and above 300 kb together with the outer portions of the gel kept from **step 18** in 0.5 µg/mL EtBr solution for at least 30 min (*see Note 18*).
22. Assemble the gel pieces, which lack the middle part of the gel containing DNA fragments between 150–300 kb, on the digital imager. Capture an image with a fluorescent ruler to ascertain the size fractionation and cutoff point.
23. Cut a 1-mm-wide slice from each agarose slice. Load the 1-mm slices and Low Range PFG marker into the wells in the agarose gel prepared in **step 19**. Store the remaining gel slices in 15-mL conical screw-cap polypropylene tubes containing sterile 0.5X TBE buffer at 4°C.
24. Perform electrophoresis by following **steps 18–20** in **Subheading 3.9.6**.
25. Determine the size distribution of each gel slice.

3.9.9. Recovery of Insert DNA by Electroelution

1. Cut a piece of dialysis tubing 10 cm long and soak in a 200-mL glass beaker containing sterile, deionized distilled water.
2. Close one end of the tubing with a dialysis clip, and remove residual water from inside the tubing with a pipet.
3. Insert an agarose slice containing size-fractionated DNA using clean forceps, and add 300–400 µL of sterile 0.5X TBE buffer.
4. Remove air bubbles thoroughly and close the other end of the tubing with a dialysis clip. Orient the long axis of the gel parallel to the long axis of the tubing.
5. Add 1.6 L of 0.5X TBE buffer in a submarine gel electrophoresis tank, and immerse the tubing in a shallow layer of 0.5X TBE buffer. Pile up four to seven pieces of 1.5-mm-thick plastic combs on the dialysis clips to hold down the sample in the tank.
6. Pass electrical current through the short axis of the gel at 3 V/cm (equivalent to 100 V for a Bio-Rad Sub-Cell GT DNA Electrophoresis Cell) for 3 h at room temperature.
7. Reverse the polarity of electrophoresis for 30 s to release the DNA from the wall of the dialysis tubing.

8. Transfer the dialysis tubing still containing agarose slice into a 2-L glass beaker containing 1 L of TE buffer (pH 8.0), and dialyze for at least 2 h at 4°C (*see Note 19*).
9. Remove a dialysis clip and open the end of the dialysis tubing. Recover solution using a wide-bore pipet tip in a new 1.5-mL microcentrifuge tube. Keep at 4°C. Do not freeze the eluted solution.
10. Prepare a 0.7% agarose gel in 0.5X TBE buffer. Load 5 μ L of recovered sample and DNA concentration markers with various amounts (5–50 ng) of λ DNA in the wells and run at 6 V/cm for 1 h. Stain the gel in 0.5 μ g/mL of EtBr solution for at least 30 min. Take a picture using the digital imager. Estimate the DNA concentration using λ DNA as a standard.

3.10. Construction of a BAC Library

3.10.1. Ligation and Transformation

1. Mix 50 ng of insert DNA, 25 ng of pBACe3.6 vector (30 ng of pTARBAC or 50 ng of PAC vector), and 10 μ L of 5X T4 DNA ligase buffer in a microcentrifuge tube.
2. Add sterile, deionized distilled water to bring the total volume to 49 μ L and mix gently.
3. Add 1 Weiss unit of T4 DNA ligase (1 μ L) and mix gently. Incubate at 4°C for 3 h for *EcoRI-EcoRI* ligation or 6 h for *MboI-BamHI* ligation.
4. Add 1 μ L of 0.5 M EDTA (pH 8.0) and 2 μ L of 10 mg/mL proteinase K. Mix gently and incubate at 37°C for 1 h.
5. Add 1 μ L of 100 mM PMSF solution. Mix gently and keep at room temperature for 1 h. Mix 100 mM PMSF solution vigorously using a vortex prior to use until crystallized PMSF is dissolved completely.
6. Transfer the ligation mixture onto a 25-mm-diameter, 0.025- μ m-pore-size microdialysis filter floating on 10–15 mL of sterile, deionized distilled water in a Petri dish. Dialyze for 2 h at room temperature.
7. Using a wide-bore pipet tip, recover the solution carefully into a microcentrifuge tube. Transfer the microdialysis filter using forceps onto the cover of Petri dish that is placed upside down. Discard water from the dish and remove residual water using a pipet (*see Note 20*).
8. Add 10–15 mL of PEG8000 solution to the dish and transfer the filter recovered in **step 7** keeping the surface up on the solution.
9. Transfer the sample onto the filter and dialyze for at least 3 h at room temperature. The ligation mixture should be concentrated to approx 8 μ L from the 50 μ L ligation reaction in **step 3**.
10. Move the cover from the dish and place upside down. Transfer the filter paying attention not to lose the sample on the cover. Remove the residual PEG8000 solution around the filter with a pipet (*see Note 21*).
11. Recover the ligation mixture from the filter using a wide-bore pipet tip. Keep on ice.
12. Remove the required amount of electrocompetent cells from a –80°C freezer and thaw on ice. It takes approx 20 min for the frozen cells to thaw completely. Do not freeze extra cells; the titer will drop for the next transformation.

13. Mix 4 μL of ligation mixture and 40 μL of electrocompetent cells in a microcentrifuge tube and keep on ice (*see Note 22*).
14. Place wet ice in an electroporation chamber and set an electroporation cuvet in the chamber.
15. Prepare a 15-mL snap-cap polypropylene tube containing 1 mL of SOC medium. The volume (1 mL) of SOC medium is for two transformations. For large-scale transformation, increase the volume using a larger tube (50-mL conical screw-cap polypropylene tube).
16. Transfer 22 μL of the ligation and electrocompetent cell mixture into the cuvet using a wide-bore pipet tip, and place the droplet between the electrodes (*see Note 23*).
17. Deliver a pulse using the following conditions: voltage booster settings: resistance = 4000 Ω ; cell-porator settings: voltage = 1.95 kV (voltage gradient = 13 kV/cm), capacitance = 330 μF , impedance = low Ω , charge rate = fast.
18. Collect the cells and transfer into the 15-mL snap-cap polypropylene tube containing 1 mL of SOC medium. Repeat **steps 16** and **17**. Transfer the sample in the same tube.
19. Incubate at 37°C in an orbital shaker at 200 rpm for 1 h.
20. Clean a flow hood with 70% ethanol and dry 100 $\times$ 15 mm LB plates containing 5% sucrose and antibiotics (*see Subheading 3.11.*) for 40 min during the incubation.
21. Soak a glass spreader in ethanol and flame. Keep in the hood.
22. Spread 500 μL of cells on each of the plates. Spread 100 μL of cells for large-scale transformation.
23. Dry the plates and incubate at 37°C overnight.
24. Count the number of colonies on the plates for test ligation and transformation.
25. For large-scale ligation and transformation, repeat **steps 1–21** by increasing the number of samples. Add 80% glycerol to be 10% final glycerol concentration 1 h after incubation at **step 19**. Spread 100 μL of cells plus 400 μL of SOC medium on each of two plates at **step 22**. Freeze the remaining cells in the tube in ethanol–dry ice bath. Keep at –80°C until colony picking is scheduled.

3.10.2. Analysis of BAC Clones

1. Pick 42 colonies from each fraction with a sterile toothpick in six-well green tubes containing 1.5 mL of LB medium for AutoGen740 or in a 96-deep-well block containing 1 mL of LB medium for AutoGen960.
2. Incubate at 37°C with shaking at 200 rpm overnight.
3. Purify DNA using an automated plasmid isolation machine or a modified alkaline lysis method.
4. Dissolve DNA in 100 μL of TE buffer (pH 8.0) from 1.5 mL of culture in the tubes or 40 μL of TE buffer from the 96-deep-well block.
5. Transfer 10 μL of DNA solution into a flexible 96-well plastic plate.

6. For 100 samples, mix 775 μL of sterile, deionized distilled water; 200 μL of 10X NE buffer 3; 20 μL of 10 mg/mL BSA, and 5 μL of *NotI* (10 U/ μL) in a microcentrifuge tube on ice.
7. Aliquot 10 μL of the enzyme mixture in the DNA sample and mix gently. Cover tightly with a plastic seal.
8. Incubate at 37°C for 2 h.
9. Clean a 45-well, 21-cm-wide, 1.5-mm-thick comb; a platform (21 $\times$ 14 cm); and a gel-casting stand with 95% ethanol. Set the clean platform and comb in the gel-casting stand.
10. Using a microwave, thoroughly melt 2 g of agarose in 200 mL of 0.5X TBE buffer in a 500-mL glass bottle containing a magnetic stirring bar. Cool molten agarose to 55°C with stirring and pour into the gel-casting stand. Allow the gel to solidify for 1 h at room temperature.
11. Pour 2 L of 0.5X TBE buffer in a CHEF apparatus tank and equilibrate the unit at 14°C during dialysis and preparation of the gel.
12. Remove the comb gently from the solidified gel, and load Low Range PFG marker in the outermost wells on each side of the gel.
13. Add 2 μL of loading dye in the sample and mix gently.
14. Place the gel in the precooled unit and load the sample in the gel.
15. Run at 6 V/cm for 16 h with 0.1- to 40-s pulse time, 120° included angle at 14°C (see **Note 24**).
16. Stain the gel in 0.5 $\mu\text{g/mL}$ of EtBr solution for 30 min, and take a gel image on an Alpha Innotech IS1000 digital imager.
17. Determine the average insert size and insert size distribution.

3.11. Colony Picking

3.11.1. Preparation of LB Plates Containing Sucrose and Antibiotics

1. Add 15 g of tryptone peptone, 7.5 g of yeast extract, 7.5 g of NaCl, and 75 g of sucrose to a 2-L flask containing 1.5 L of deionized distilled water.
2. Mix with a magnetic stirring bar until the powder is completely dissolved.
3. Adjust the pH to 7.2 with 5 N NaOH (~680 μL).
4. Add 22.5 g of bacto agar and stir the solution for 5 min.
5. Cover the bottle with aluminum foil.
6. Autoclave the medium still containing the magnetic stirring bar at 121°C for 20–30 min.
7. Once the autoclave cycle is finished, carefully take the bottle out of the autoclave.
8. Stir the medium on a stirrer and cool the medium to 55°C (see **Note 25**).
9. Add 1.5 mL of 20 mg/mL chloramphenicol for BAC clones or 1.5 mL of 25 mg/mL kanamycin for PAC clones.
10. Stir the medium gently on the stirrer to avoid bubble formation.
11. Pour 300 mL of medium into a Q-tray using a 500-mL sterile cylinder (see **Note 26**).

12. Briefly flame the surfaces of the medium in the Q-tray with a Bunsen burner while lifting the cover slightly (*see Note 27*).
13. Leave the plates at room temperature for about 45 min to solidify.
14. Wrap the plates in a plastic bag and store them upside down at 4°C. The plates can be kept up to 1 mo at 4°C.

3.11.2. Preparation of LB Medium Containing 7.5% Glycerol

Thoroughly rinse all glassware in the procedures with deionized water and autoclave. Do not use any detergent.

1. Add 100 g of tryptone peptone, 100 g of NaCl, and 50 g of yeast extract to a 4-L beaker containing 2.5 L of deionized distilled water.
2. Mix with a magnetic stirring bar until the powder is completely dissolved.
3. Add 750 mL of glycerol and stir for 2 min (*see Note 28*).
4. Transfer the solution and the magnetic stirring bar into a 10-L glass bottle through a funnel.
5. Add deionized distilled water to 10 L and stir.
6. Adjust the pH to 7.2 with 5 N NaOH (3.5–4.0 mL).
7. Remove the magnetic stirring bar using a magnetic rod, and transfer the medium into two 5-L glass bottles.
8. Autoclave the medium at 121°C for 60 min (*see Note 29*).
9. Once the autoclave cycle is finished, carefully take the bottles out of the autoclave and keep at room temperature overnight.
10. Close the cap tightly when the bottles are cooled down to room temperature. The medium can be stored at room temperature for several weeks.
11. Attach the top filter to a 500-mL glass bottle in a flow hood and connect to a vacuum. An Erlenmeyer flask should be connected between the filter and the vacuum pump to trap the air-scattered medium.
12. Open the valve for the vacuum and pour the medium into the top filter.
13. Transfer the top filter onto another empty 500-mL bottle when the bottle is full.
14. Loosen the cap and autoclave the bottles at 121°C for 45 min.
15. Once the autoclave cycle is finished, carefully take the bottles out of the autoclave and keep at room temperature overnight.
16. Close the cap tightly. The medium can be stored at room temperature for several months.

3.11.3. Filling of LB Medium Containing 7.5% Glycerol into 384-Well Plates

1. Prepare a 384-well manifold, two sets of silicon tubing, and a cap with stainless steel pipes for Genetix QFill 2.
2. Rinse all the tools with deionized distilled water, dry, and wrap with aluminum foil.
3. Autoclave the tools and dry (*see Note 30*).

4. Add 500 μL of 20 mg/mL chloramphenicol for a BAC library or 500 μL of 25 mg/mL kanamycin for a PAC library to the 500 mL of filtrated and autoclaved LB medium containing 7.5% glycerol.
5. Follow the manufacturer's instructions for setting up the Genetix Q-Fill 2.
6. Place a 500-mL bottle containing the medium to the side of the Qfill2 apparatus.
7. Insert the stainless steel pipe into the medium and screw the dispensing bottle lid onto the bottle.
8. Connect a silicon tube to the air pipe on the Q-Fill2.
9. Set the volume setting to 0048 on the Q-Fill 2.
10. Purge some medium to force out air in the manifold and tubing.
11. Adjust the volume setting after filling the first plate. The medium should be 1 mm below the surface of the 384-well plate (*see Note 31*).
12. Once the volume is satisfied, continue to fill the plates until the level of medium is 2 mm above the tip of the stainless straw that is inserted into the medium.
13. When the level of medium in the 500-mL bottle becomes low, to maintain the same volume setting, use the same 500-mL bottle that is attached to the Q-Fill 2 by refilling it.
14. Stack seven plates together, wrap with plastic wrap, and leave at room temperature for at least 1 d in order to monitor contamination. During the storage at room temperature, the volume of the medium decreases about 10%. Store the filled 384-well plates at 4°C. It is not recommended to store longer than 2 wk because the medium evaporates.

3.11.4. Picking of Colonies

1. Thaw the frozen cell suspension on ice. It takes at least an hour for the suspension to thaw thoroughly.
2. Dry the LB agar plates containing sucrose and antibiotics that are poured into a 22 $\times$ 22 cm tray in an air-circulating hood for 30 min.
3. Dilute the cell suspension to approx 500–700 colonies/mL with LB medium that does not contain antibiotics prior to spreading cells.
4. Place 3 mL of diluted cell suspension and sterile glass beads onto the plates (*see Note 32*).
5. Shake the plates to spread the cells with the glass beads.
6. Dry the plates for 30–40 min.
7. Incubate at 37°C for 20 h.
8. Pick colonies into 384-well plates containing LB medium with glycerol and antibiotics using an automatic colony-picking machine (Q-bot or Q-Pix; Genetix).
9. Stack six inoculated plates, placing two empty plates on the top.
10. Enclose with Saran Wrap and place in a 37°C incubator.
11. Fill the 384-well plates with water using the Q-Fill 2.
12. Enclose with Saran Wrap and place in an incubator.
13. Incubate both the inoculated and water plates at 37°C for 20 h. Be sure to place a metal or pyrex dish filled with water in the incubator to maintain moisture.

14. Score empty wells by looking at each plate from the bottom.
15. Place 18 plates plus a water dummy in a freezer. The water dummy must be incubated at 37°C overnight prior to placing at -80°C.

3.12. Library Replication

3.12.1. Replication of a Library From an Original or Replication Master Copy

The result, if a library replication is done correctly, is that a previously determined number of identical copies are created. The replication could be of a single plate, the entire range of plates in the library, or any number of plates desired from the library. Two important terms used in this protocol are *R0* and *Master*. *R0* refers to the *original* copy of the library. This copy was generated by the use of colony-picking robots in the laboratory. In any case, the format is the 384-well plate, which is carried over to all copies. *Master* refers to a designated *copy* of the complete library to be used exclusively for the purposes of replication. It is not used for any other purpose and in this regard is insulated from mishaps such as contamination that would otherwise be subsequently passed to the copies generated from it. The procedures for replicating from a Replication Master and an *R0* differ slightly. The difference is, however, of great importance because it pertains to the protection of the *R0* from corruption. The nature of an *R0* copy is that it is unique and the first of its kind. There is no possibility of repair if an error in procedure leads to the damage, or contamination of the clones contained in the *R0* copy. The difference between the procedures is an additional sterilization step when replicating from *R0* copy. The tools are sterilized between inoculations of new copy plates. When replicating from a Replication Master copy, the tools are sterilized after all copies of a single template have been inoculated. The tool moves back and forth between the template and new copy plates without sterilizing. The objective of the additional sterilization step for *R0* replication is to ensure that only a sterilized tool enters the wells of an *R0* template.

3.12.2. Thawing of 384-Well Plates

1. Remove the plates from the freezer and set on a cart.
2. Remove excess ice powder from around the boxes of the plates using a freezer brush or Kimwipes.
3. Turn on dryers at opposite ends of the counter to be used for thawing.
4. Arrange the frozen plates on the counter in four rows of 12 plates each, for a total of 48 plates (*see Note 33*).
5. Lift the front edge of each plate lid and shift back approx 2 mm so that the plate lid remains propped open but does not expose the wells containing medium.
6. Repeat **steps 4** and **5** until the countertop is full.

7. Carefully examine all plates and lids (*see Note 34*).
8. Examine all plates for excess moisture on the deck of the plates (*see Note 35*).
9. Use the edge of *sterile* blot paper to wick up moisture. Rotate to use the four edges. Do not use the same edge twice. Use one edge for each row.
10. Use the flat surface of *sterile* blot paper and blot the deck of a plate as a whole. Do not slide side to side; use a simple up-and-down motion, coming down directly on top of the plate once and withdrawing quickly, and then discard the paper.
11. Once the plate deck and inside of the lid are free of excess moisture, close the lid and allow it to continue thawing (*see Note 36*).

3.12.3. Replication

1. Clean the surfaces in a laminar flow hood with 70% ethanol.
2. Arrange *template* plates in consecutively numbered stacks of six with the highest number on the bottom.
3. Arrange *template* plates on an adjacent counter so that they can be easily obtained from a seated position in front of the hood. New copy plates should have been stacked on a cart the day before. Arrange them also to be easily reached.
4. Arrange a steel dish filled to approx 8 mm with 190-proof ethanol, tools, and a Bunsen burner under the hood. The dish should be near the center and the burner off to one side.
5. Keep the burner at least one half the width of the hood away from the dish containing alcohol and light the burner.
6. Sterilize tools by setting them into the dish with the pins facing down, and remove to ignite in the flame one at a time (*see Note 37*).
7. Place a stack of six template plates under the hood near the center front.
8. Place the first stack of new copy plates next to the template plates (*see Note 38*).
9. Ensure that the template plate at the top of the stack matches (in name and plate number) the stack of copy plates. The only difference allowable is the “R” or “copy” number, which will necessarily be different because a subsequent copy is now being made.
10. Remove the lid of the template and set it aside.
11. Dip the tool into the wells of the template (*see Note 39*).
12. Remove the lid of the copy plate on top of the stack and set it aside. Dip the tool into the wells of the copy plate. Place the lid back onto the copy plate (*see Note 40*).
13. Place the used tool in an ethanol bath. Remove the tool previously placed in the bath (no tool if first inoculation) and *flame* the tool (*see Note 41*).
14. Repeat **steps 11–13** until the stack of new copy plates, matching the single template, has been inoculated.
15. Set the inoculated stack of new copy plates and single template aside, and obtain a fresh stack of new copy plates. This new copy stack should match the next template as in **step 1**. Repeat **steps 9–14** until the stack of templates is finished.
16. Obtain the next stack of templates and repeat **steps 1–15**. Do this until all new copy plates have been inoculated (*see Note 42*).

3.12.4. Cleaning

1. Place the tools in two stainless steel dishes with the pins facing down.
2. Place the dishes in a sink, run hot water over the tools, and soak.
3. Clean the entire surface of the tools using a brush and running water (*see Note 43*).
4. Rinse with deionized water and set the tools aside to dry.
5. Once the tools are dry, inspect for bent pins and straighten them using a 384-well plate as a guide.

3.13. Preparation of High-Density Replica Filters

High-density replica filters can be prepared for hybridization screening purposes. Each filter contains 36,864 colonies, which represents 18,432 independent clones that have been spotted in duplicate in a 4×4 clone array. The filter sets will vary in number in accordance with the number of plates in the library they represent. It is practical to construct a BAC library consisting of a number of plates that are a multiple of 48 for preparation of high-density replica filters. The procedure for preparing high-density filters is described using a Gridding Robot (BioRobotics). The procedure would differ if a different robot or software configuration were used.

3.13.1. Preparation of LB Plates for High-Density Replica Filters

1. Add 15 g of tryptone peptone, 7.5 g of yeast extract, and 7.5 g of NaCl to a 2 L flask containing 1.5 L of deionized distilled water.
2. Mix with a magnetic stirring bar until the powder is completely dissolved.
3. Adjust the pH to 7.2 with 5 N NaOH (~680 μ L).
4. Add 22.5 g of agarose and stir the solution for 5 min.
5. Cover the bottle with aluminum foil.
6. Autoclave the medium still containing the magnetic stirring bar at 121°C for 20–30 min.
7. Once the autoclave cycle is finished, carefully remove bottle from the autoclave.
8. Stir the medium on a stirrer and cool the medium to 55°C.
9. Add 1.5 mL of 20 mg/mL chloramphenicol for BAC clones or 1.5 mL of 25 mg/mL kanamycin for PAC clones.
10. Stir the medium gently on the stirrer to avoid bubble formation.
11. Pour 300 mL of medium into a Square Bio Assay Dish using a 500-mL sterile cylinder (*see Note 44*).
12. Briefly flame the surfaces of the medium in a Square Bio Assay Dish with a Bunsen burner while lifting the cover slightly.
13. Leave the plates at room temperature for about 45 min to solidify.
14. Wrap the plates in a plastic bag and store them upside down at 4°C. The plates can be kept up to 1 mo at 4°C.

Table 1
Correspondence of Plate Numbers to Filter
Numbers (see Subheading 3.12.2.)

Filter no.	Plate (numerical range)
1	1–48
2	49–96
3	97–144
4	145–192
5	193–240
6	241–288
7	289–336
8	337–384
9	385–432
10	433–480
11	481–528
12	529–576
13	577–624
14	625–672
15	673–720
16	721–768
17	769–816
18	817–864
19	865–912
20	913–960
21	961–1008
22	1009–1056
23	1057–1104
24	1105–1152
25	1153–1200
26	1201–1248
27	1249–1296
28	1297–1344
29	1345–1392
30	1393–1440

3.13.2. Setting of Nylon Filters on Agarose Plates

1. Remove a quantity of filter trays from a 4°C refrigerator sufficient to complete the desired number of high-density filters.

2. Remove the filter trays from plastic bags maintaining the inverted (upside down) orientation.
3. Tilt an inverted tray up from one side until a good grip can be achieved on the tray bottom under the hood (*see Note 45*).
4. Stack the uncovered tray bottoms, agar up, one on top of the other, rotated 45° each, to about 12 in a stack. Leave to dry until the bulk of the water present on the surface of the agar disappears.
5. Stand the lids on edges off to the side under the hood to dry. Remove excess water with large Kimwipes.
6. Obtain labeled 22 × 22 cm nylon filter sets (usually 12 per envelope lettered A–L). Across the top of the nylon filters will be the following beginning at left, then center, then right, respectively: date, accession no., source library designation, and filter number with letter. *See Table 1* to identify which numbered plates correspond to which numbered filter.
7. Return the filters to drying filter trays set out now free of standing water.
8. Position one filter tray in front of your body, still under the hood, in the space not taken up by stacks of drying trays.
9. Remove the filters from the envelope and set to a convenient side. *Use clean gloves when handling filters.*
10. Separate one filter from the protective sheets and grasp by opposite corners making sure that the labeled side is up.
11. Hold the filter so that the center falls like a *valley* running the length of the sheet from corner to corner, and align the distal corner with the corresponding corner of a filter plate.
12. Carefully lower the distal corner of the filter to the agar while simultaneously aligning the proximal corner with the corresponding corner of the filter plate.
13. Carefully lower the second corner to the agar and, slowly lay the rest of the now-aligned filter onto the agar surface.
14. Cover the filter in the tray with a lid and repeat from **step 9** until all the filters are set into the trays.

3.13.3. Gridding of Filters Using an Automatic Colony-Gridding Machine

1. Power up computers bearing the same numbers as those on the robots to be used.
2. Manually zero all axes of motion on each robot before powering up the robot (*see Note 46*).
3. Once manually zeroed, power up by actuating the switch at the rear portion of the right-side panel of the robot.
4. Ensure that the heater switch is also in the “on” or “red” position. It is located on the same side panel as the power switch but near the front.
5. Follow screen prompts and depress the Interrupt button to initiate the powered zero step (*see Note 47*).
6. Insert the plates with the bar code facing out beginning with BioBank #1. Place the lowest numbered 384-well plate in the topmost left position. Place the next

consecutively numbered plate in the topmost right position. Repeat these steps alternating left to right until the holder is filled (*see Note 48*).

7. Repeat for BioBank #2 following **steps 1–6**.
8. Double-check BioBanks #1 and #2 for proper placement and orientation of all plates.
9. Use BioBank #3 for the control clone plate. Insert this plate into the uppermost left position. *This plate will have no bar code but will be loaded in the same orientation as the library plates.*

3.13.4. Gridding

1. Click on the box to the right of “load previously saved parameters.”
2. Type “default” in the text box and hit enter.
3. Click on the box next to “Go.” All four filter trays will slide out.
4. Beginning with the top tray, uncover and load the filter with the lowest accession no. into the tray with labeling oriented to the left.
5. Slide the filter plate to the left until it stops, and lock into place using a thumb-screw at the front of the tray. Apply enough pressure to secure the filter plate without deforming it more than 1 mm.
6. Once the first filter is secure, press the red or yellow Interrupt button on the right side near the front of the robot. The filter, now in a tray, will slide back into the robot.
7. Repeat three more times until all four consecutively numbered filters are in closed position.
8. The screen will prompt for the placement of BioBank #1. Slide the BioBank into the rack and press the Interrupt button while supporting the majority of the BioBank’s weight until it stops in the lowered position.
9. The screen will prompt for installation of the 384-pin tool. Before attaching, invert and inspect the pins to ensure that they are uniform and unbent (*see Note 49*).
10. Join the tool to the transfer arm with a thumbscrew oriented to the right. Snug the tool up, and to the right, on the base of the transfer arm before tightening. Press the Interrupt button and follow the prompts on the screen.
11. Place a metal bath in the center position below the 384-pin tool. Add methanol to center up the bath to the level of the step. Do not exceed the height of the step although 1 or 2 mm below is sufficient.
12. Start the process by depressing the Interrupt button. The screen will prompt for library and filter information such as library copy number and filter number with letter range. This number is not the accession no. but, rather, the number that corresponds to the plate range used. The letter follows that number. The result is multiple, identical filters from the same plate range with unique accession nos. and letters. After each entry “enter” must be keyed. Follow when finished by clicking “OK” to begin gridding.
13. On completion of the first 24 plates, the screen will prompt for BioBank #2 containing the second set of 24 plates to complete the filter. Repeat **step 8** here.

3.13.5. Control

1. Once prompted that the run is complete, remove BioBank #2 and press the Interrupt button.
2. Press the Interrupt button four times to return all four trays to the closed position without removing the filter plates. Leave the tool attached to the transfer arm.
3. Repeat **steps 10–12 in Subheading 3.13.4.**, and enter “control” instead of “default.” Leave the tool as is and ensure that the level of methanol is sufficient. Press the Interrupt button to begin the control plate cycle.
4. When the screen prompts that the run is complete, remove the BioBank and press the Interrupt button.
5. Remove the filter plates and cover beginning from the top tray. Depress the Interrupt button as prompted. *This completes one cycle of four filters.*
6. Place the filters in a 37°C incubator inverted as they were in the refrigerator at the beginning of the process.
7. Repeat steps until the job is complete. Change methanol in the bath when prompted “fill to level of step.”

3.13.6. Shutting Down

1. Select “Main” from the menu.
2. Select “cancel” at the bottom of the screen.
3. Shut down Windows.
4. Power down the computer and Gridding robot.

3.13.7. Processing of Filters

1. Turn on a water bath set at 97°C. It takes at least an hour to heat up.
2. Take out a plate from the incubator and do a general examination of the growth of the BAC clones (*see Note 50*).
3. Because these criteria for filters of acceptable quality are met:
 - a. Control positions located at the four corners of each panel must have sufficient growth.
 - b. Missing clones, scratching, or smashing of the clones is minimal.
 - c. Irregular or bad gridding clones do not exist.
4. Prepare three large baking dishes with chromatography paper of precut fitting size on each bottom. Saturate the first two with denaturation solution. Saturate the third with neutralization solution. Drain excess solution for better results.
5. Pick up a filter by the corners from the LB agar tray using forceps, lay it flat onto the first baking dish without bubbles underneath, and incubate for 4 min at room temperature.
6. Pick up a filter from the first baking dish, lay it flat onto the second baking dish, and incubate for 4 min in a 97°C water bath. No splashing on the filter is allowed. The water bath lid must be wiped dry to reduce any condensation that may drip on the filter.

7. Take out the hot baking dish from the water bath, remove the filter, and lay flat onto a third baking dish.
8. Incubate the filter for 4 min at room temperature before taking it out to air-dry on large chromatography paper.
9. Continue to process all the remaining filters up to this point (*see Note 51*).
10. Dissolve 1.3 g of Pronase in 32.5 mL of deionized distilled water to produce a working stock solution of 40 mg/mL, which is kept on ice ready for use.
11. Measure out 197.5 mL of prewarmed ProPK buffer at 37°C, pour onto a 24.5 × 24.5 cm bioassay tray, and add 2.5 mL of Pronase stock solution to make a final concentration of 500 µg/mL.
12. Submerge filters one by one in each tray containing Pronase solution, placing meshes on top of each filter by pushing out any bubbles trapped under the filter or meshes; cover with a lid; and incubate at 37°C for 1 h.
13. Remove the filters from all the trays and air-dry on large chromatography paper overnight.
14. Turn on a UV crosslinker, select Program C3 (150 mJ), and crosslink each filter accordingly.
15. Sort out the filters into sets; seal them in hybridization bags; and store in a cool, dry place.

4. Notes

1. Store at -80°C in small aliquots (210 µL). It is important to maintain a 2 mM magnesium concentration in the reaction mixture. *EcoRI* Methylase retains only 50% activity in a 4 mM Mg⁺⁺ concentration. By contrast, *EcoRI* may not be active below a 2 mM Mg⁺⁺ concentration. The commercially supplied *EcoRI* buffer contains 10 mM Mg⁺⁺ and *EcoRI* Methylase buffer contains 10 mM EDTA. Thus, the reaction buffer should be prepared.
2. The initial BAC vector, pBAC108L, lacks a selection system of recombinant clones over nonrecombinant clones. It is therefore expected to contain a high level of noninsert clones in the first generation of human BAC library. Recombinant clones were therefore screened through hybridization using human repetitive DNA as a probe (3). The second generation of the BAC vector pBeloBAC11 permits screening by α -complementation to distinguish recombinant clones from noninsert clones (4). It is, however, sometimes difficult to identify white colonies over blue colonies using a robotic device. In addition, IPTG and X-gal are relatively expensive reagents, making their use costly for construction of a highly redundant library. PAC vectors (1,3) have been derived from the bacteriophage P1 vector. Unlike these two BAC vectors, the PAC vectors possess a positive selection system for clones that carry insert DNA. A self-circularized PAC vector molecule allows expression of the levansucrase gene (*sacB*), resulting in conversion of sucrose in the medium to levan, which is toxic to *E. coli*. In theory, only recombinant clones are able to grow on medium containing sucrose. It is thus anticipated that resulting libraries contain fewer noninsert clones in the library of this

positive selection. Therefore, it is feasible to construct a BAC library with a low level (<2%) of nonrecombinant clones, allowing random end sequencing and DNA fingerprinting from the library. In addition, a pUC-vector-derived fragment is inserted in between the cloning sites, thus inactivating *sacB* gene and permitting purification of large DNA quantities owing to the high copy number. This “pUC-link” fragment is removed during the library preparation, resulting in a BAC plasmid maintained at a single copy. However, the sizes (15 to 16 kb) of PAC vectors are twice as large as the original BAC vectors (7.4 kb). A genomic library based on a smaller vector is more economical with respect to the shotgun-sequencing approach because fewer reads are derived from the large-insert vector. A positive-selection BAC vector, pBACe3.6, has been constructed by transferring the *SacBII* and pUC-link sequences into the original pBAC108L vector. In recent years, there has been increased interest in cloning DNA from targeted regions without the expense of preparing and screening an entire library. Transformation-associated recombination (TAR) cloning utilizes yeast homologous recombination to allow rescue of a targeted DNA fragment by cotransforming genomic DNA and a linear vector containing small genomic DNA fragments from the target region at both vector termini. It is, however, laborious and costly to prepare tailored TAR-vectors if many target regions are considered and the corresponding genome sequence is not available. To facilitate downstream use of BAC clones for targeted recloning of the same genomic fragment, the yeast centromere (CEN 6) and a yeast-selectable marker (His 3) were inserted in the pBACe3.6 vector. This led eventually to two new BAC vectors: pTARBAC1.3 and pTARBAC2.1. BAC libraries derived from these vectors not only contain BAC clones but are easily converted through deletion of most of the insert into recloning vectors for targeted repair of the deletion in yeast, possibly using a different haplotype.

3. The protocol for vector preparation has been optimized to maximize for cloning efficiency and minimize the level of nonrecombinant clones. The integrity of cohesive cloning ends and the absence of undigested vector molecules are a critical requirement for preparing high-quality BAC/PAC libraries. Clones containing undigested vectors cannot be separated from recombinant clones because both grow on sucrose-containing agar medium. Great care should thus be taken to ensure complete digestion of the vector molecule with the first and second enzymes during vector preparation. It is necessary to optimize the conditions for each vector preparation and for different batches of restriction enzymes. The protocol provides a guideline to optimize these conditions. Note that BAC vectors such as pBeloBAC11 vector (4) do not contain this pUC-link and require other procedures.
4. Minimize the exposure to UV light. If the top band is visible, the lower band is the closed-circular plasmid. It is feasible to see a DNA band under white light, without exposing to UV light.
5. Prepare 10, 5, and 2 ng/ μ L of λ DNA concentration standards and store at 4°C. It is important to mix standard DNA, loading buffer, and 0.5 XTBE buffer thoroughly by repeating the cocktail up and down by pipet.

6. The number should not exceed 50 colonies on chloramphenicol plates. If the number is too high, it is expected to contain a high level of nonrecombinant clones in the library. A high fraction of noninsert clones will be troublesome if the library is used for random end sequencing and BAC DNA fingerprinting. These clones will not generate any useful information.
7. There should be no colony on chloramphenicol/ampicillin plates. If a colony is observed, it is advised that the vector be discarded and a new vector be prepared. It is anticipated that there will be a large number of nonrecombinant clones that contain intact vector in the library. This type of nonrecombinant clone can be troublesome when the library is screened via hybridization using a probe containing DNA sequence derived from pUC plasmid derivative, such as cDNA. The intact vector contains pUC19 sequence, thus replicating a high copy number of plasmid. On the other hand, recombinant clones that contain the true sequence are replicated as single-copy plasmid. Therefore, the nonrecombinant clones will give stronger hybridization signals than the true positives.
8. It is preferable to construct a library from a single animal to avoid sequence assembly problems owing to polymorphisms. It may be difficult to obtain a 50-mL blood sample at once from a small animal such as a monkey. It is possible to construct a library from a small number (two to four) of DNA blocks if the library constructor is well trained. In this case, a 15-mL blood sample should be sufficient. The blood sample is stable on ice for 24 h. It is also possible to count the total number of leukocytes on an automated hematology counter, if it is available. Be sure to perform all the following steps in a safety cabinet—since blood sample is a potential biohazardous material.
9. Carbonic anhydrase catalyzes an exchange reaction resulting in accumulation of ammonium chloride inside the erythrocytes. The accumulation increases the internal osmotic strength and causes swelling and bursting of erythrocytes. Leukocytes have much lower carbonic anhydrase activity and do not undergo lysis under this condition.
10. There are two chambers on a hemocytometer, each of which is divided into nine large (1-mm) squares. The 1-mm square in the middle is divided into 25 medium-size (0.2-mm) squares. The medium-size square is divided into 16 small squares. The number of cells will be in the range of 100–200 cells per five 0.2-mm squares if the blood sample is collected from a healthy animal.
11. Heart and lung are not desirable for preparation of DNA blocks. In the case of using an “inbred” small animal, such as a laboratory mouse, it is possible to mix the same tissue from several individuals to make it easy to prepare DNA blocks. However, if the animal is not inbred, it is desirable to construct a library from a single individual.
12. Approximately 4×10^6 cells are embedded in each block considering the volume of each block to be approx 80 μ L. Each block therefore should contain approx 24 μ g of DNA considering 6 pg of DNA/cell. Approximately 45 blocks are normally obtained from a 45-mL blood sample.
13. It is necessary to keep empty wells for PFG markers on both sides and at least one empty well between the markers and the preparative slot. This space is sufficient

to place 9 to 10 DNA blocks. In the case of a larger number of samples, prepare a larger gel that is feasible to create a larger preparative well.

14. Avoid contacting DNA blocks with EtBr. It is possible to observe whether DNA in an agarose block is maintained as high molecular weight by taking an image of the agarose gel. If DNA is degraded, a large smear is visible on the gel. Confirm that the DNA degradation level is minimal. If significant DNA degradation is observed, it is recommended that DNA blocks be prepared again. High-molecular-weight DNA does not move in the agarose block using the electrophoresis condition previously described. A significant amount of sheared DNA can be an inhibitor of the cloning process. Store the blocks in TE buffer (pH 8.0) at 4°C for a short period, and, if needed, change the solution to 0.5 M EDTA (pH 8.0) for long-term storage.
15. The sample from tube 1 is a negative control; no smearing pattern should be observed. If overdigestion is observed using all the conditions in **steps 4a–d**, reduce the amount of *Mbo*I. If underdigestion is observed, increase the amount of *Mbo*I or prolong the incubation time at **step 7**.
16. This is an opposite gel orientation compared to “normal” electrophoresis. This step allows the elimination of unnecessary small DNA fragments (<150 kb) from the gel. The small DNA fragments can be an inhibitor for construction of a large-insert library with narrow insert-size distribution. The voltage 4.7 V/cm is the standard condition for construction of a BAC library with average insert sizes of 150–200 kb. It is, however, observed that this condition is slightly different depending on the CHEF apparatus and DNA concentration per block. DNA concentration is supposed to be the same as long as it is prepared using the protocol as described in this chapter. However, it is difficult to prepare the same DNA concentration of DNA blocks from nonmammalian species, such as fruit fly and fish. In this case, it is necessary to optimize the electrophoresis condition varying the voltage from 4.5 to 5.7 V/cm.
17. The outer parts of the gel should contain a broad smear extending from 140 kb to megabase sizes. The image permits assessment of the success of partial digestion as well as size fractionation. If the broad smear band extends below 100 kb, it is recommended that the electrophoresis be run using the higher voltage at **step 8** for the next size fractionation. The marker fragments below 150 kb from the original markers should not be viewed, because these fragments should be removed during the electrophoresis at **step 8**. On the other hand, the second marker should be observed as intact.
18. Do not stain the gel portion containing DNA fragments between 150 and 300 kb. The lowest cutoff point should be the 150-kb position and is normally 4.5 cm below the well. Four to six slices are obtained and two to three will be the candidate for construction of a library. The size determination of this step does not correspond to the real size. It is often observed that the size range of 150–200 kb becomes 120–150 kb after confirmation at **steps 21** and **22**. It is therefore necessary to recover higher-molecular-weight DNA than is seen at this stage.
19. Boric acid is an inhibitor of ligation reaction, it is therefore important to dialyze the solution.

20. Do not place the filter completely flat on the cover. It is easier to recover the filter if one edge hangs over the wall of the dish, allowing the filter to be slanted toward the cover while keeping a space between the filter and the cover.
21. The ligation mixture is viscous and the volume of the solution is small. Thus, the solution is not dispersed during the filter transfer. If the volume is larger for large-scale ligation and transformation, the solution may be collected before the filter is transferred.
22. Test ligation and transformation is performed in duplicate. For large-scale ligation and transformation, increase the volume of ligation mixture and competent cells maintaining the same ratio.
23. Do not increase the volume per transformation. The electroporation cuvet can be used up to 10 times as long as transformation is repeated using the same mixture. For large-scale transformation, the number of transformations should be increased, because the volume of the mixture cannot be increased per transformation.
24. Alternatively, it is possible to analyze using a FIGE apparatus (cat. no. 170-3716; Bio-Rad). The FIGE apparatus is less expensive than CHEF, and permits obtaining good resolution for this purpose. The conditions for FIGE are as follows: 1% agarose, 0.5X TBE buffer, room temperature, 180-V forward voltage, 120-V reverse voltage, 16 h with a 0.1- to 14-s linear switch time.
25. It is possible to place the bottle in a bucket containing water to help cooling. Do not use ice-cold water, because the glassware may crack or break because of the immediate temperature change. It is important to cool the medium down below 55°C to avoid deterioration of antibiotics owing to the high temperature.
26. It is important to minimize the time and height of lifting the cover to avoid contamination from the air while pouring.
27. Flaming the medium facilitates eliminating bubbles from the surface. The bubbles may cause many empty wells or unpicked clones, because it is difficult to distinguish the bubbles from the true colonies through the image that was created by the colony-picking machine. The colony picker may count the bubbles as colonies or may not count the true colonies.
28. It is important to maintain a 7.5% glycerol concentration. Concentrations near or higher than 10% are excellent for cryopreservation but increasingly start to become a growth inhibitor. Growth inhibition is minimal at 7.5% (or below), but at lower than 7.5% the cryopreservation effect starts to disappear and the clones lose more viability during repeated freezing/thawing cycles.
29. Autoclaving liquid in the bottles requires loosening the cap to facilitate leakage of air pressure out of the bottle during the cycle. Failure to do this may cause the bottle to break because of the intense high pressure.
30. Plate filling is performed under a sterile flow hood and should be done 2 d before use.
31. The volume setting 0048 is supposed to fill 48 μ L of medium/well. However, we found that it is necessary to adjust the volume using a different machine. The volume of medium decreases to approx 10–20% during storage at room temperature, colony picking, and incubation at 37°C. Bacteria do not grow well if the volume of medium becomes too low during these processes. On the other hand,

excess medium causes well-to-well cross-contamination during colony picking and subsequent storage at -80°C . We found that the optimal level of medium is 1 mm below the top of the well.

32. Approximately 1500–2000 colonies per plate is the optimal number of colonies. If more than two colonies grow sticking together, these colonies are not picked by the colony picker. The colony picker can pick only singly isolated colonies. Hence, higher colony density per plate generates many unpicked colonies.
33. Plates should be ordered sequentially with the lowest number being the first on the left in the back row. Lay plates from left to right and begin again after 12 with a new row on the left until four rows have been arranged.
34. Light condensation will evaporate. Any excess moisture that may form drops and contact the tops of wells must be dried at this point. Use a clean Kimwipe to dry the lid only.
35. Moisture will accumulate between the wells. If the volume of accumulated moisture between the wells is sufficient to cross-contaminate adjacent wells go to **step 9**. If moisture is not sufficient to cross contaminate go to **step 10**.
36. Thawed plates must be handled differently from frozen plates. Although thawed plates can be tipped a little without danger of spilling, they cannot be subjected to shocklike bumping, dropping, or hitting. The medium is very subject to “sloshing” and “jumping” right out of the wells into adjacent wells. This is one form of contamination referred to as “cross contamination.” It is better not to remove lids completely owing to the potential for contamination from an outside source. This is another form of contamination. Some moisture may not create cross-contamination and will evaporate during the thawing process.
37. Once the tool has been passed over the flame, set it down on the side of the hood with the burner. Do not pass over the stainless steel dish with ethanol. Do this for each tool and leave to cool briefly together near the front of the hood. This process is called flaming. Do not double-dip the tools. The alcohol flame is difficult to see, especially when it is burning low. Dipping a flaming tool is dangerous because the entire dish of ethanol can ignite and a fire will result.
38. The stack of template plates should be of six different consecutively numbered plates. The stack of new copy plates should be all of the same number, matching the number of the plate in the top position of the template stack. Both stacks should be oriented such that the labels are facing out (toward your body), and the A1 corner position of all plates should be aligned.
39. Do this by first angling the tool so that the pins closest to you are closer to the wells than the pins farthest from you, and visualize the front two or three corner pins (from one side or the other) going into the wells. This will ensure that the rest of the tool is lined up. Then the tool can be lowered completely into the well and removed.
40. If replicating from a Replication Master copy, tools are sterilized after all copies of a single template have been inoculated. If replicating from an original (R0) copy, tools must be flamed between every inoculation of a new copy plate. Tools must not be used for more than one inoculation. In other words, tools must not

enter the template plate once they have touched anything else, including a new copy plate, before first being flamed.

41. The bath should hold at least two tools at a time. Always remove the least recently placed tool for flaming because this will allow the other tools to soak and dilute the glycerol medium on the pins.
42. After every two completed template stacks (six plates each) for R0 or four completed template stacks (six plates each) for Master, discard the ethanol and rinse the steel dish with water. Change the ethanol more often if it takes on a yellow or cloudy appearance. Visually inspect the pins from the side and look for a buildup of glycerol periodically. If this is found, change the ethanol more frequently.
43. Do not use any detergent to clean the pin tool.
44. The Square Bio Assay Dish is deeper (Corning) than a Q-tray (Genetix). It is our experience that the deeper dish works better for gridding filters using the BioGrid machine.
45. The lid is left more or less horizontal to the floor of the hood while the tray bottom is lifted free. The bottom of the tray should be righted and set to the side. The reason that the tray with lid is not simply righted and the lid easily removed is that the liquid on the lid may not be sterile and could contaminate the agar in the tray.
46. Manual zeroing means to physically move all the movable parts on the robot to their “home” positions. The entire process of gridding depends on the robot having found its “zero” or “home” positions before beginning. If manual zeroing is done incorrectly, a message will be displayed, once “powered zero” is attempted, telling the operator that zero failed owing to a sensor not being triggered. The sensors relay to the software that the components are in zero positions (direction given relative to the operator’s body as positioned directly in front of the robot). Following are the directions of movement for the components of the robot and the correct positions for zero. Most of the time, if the robot was shut down correctly, the components will be found in their zero positions (X direction is left to right, Y direction is back to front, and Z direction is up and down).
 - a. Transfer arm, X and Z directions, Push fully Right and Up.
 - b. Filter plate trays, Y direction, Push fully to back.
 - c. BioBank rack, Z direction, Push fully down.
 - d. Vacuum lid lifter, Z direction, Push fully down.
 - e. Source plate hooks, Y direction, Pull fully to front.
47. The Interrupt button is a large red or yellow button located near the heater switch. If only a Windows desktop is displayed, look for a minimized program window at the bottom of the screen or find the BioGrid icon and double-click on it to open the program if it is not already open. Now follow prompts to powered zero.
48. The holder will, on completion, contain 24 of the 48 necessary plates for one run. Observe that correctly oriented plates will result in a pattern of all odd-numbered plates being on the left side of the holder and all even-numbered plates being on the right side of the holder.
49. A good way to visualize the condition of the pins is to set the tool into an empty 384-well plate and invert. This will allow you to see how the 384 pins of the tool

align against a known regular interval of marks. The suspect pins can be straightened with a fingernail or forceps.

50. Normal BAC clones grow within the gridding pattern. Any missing major clones or any contaminated clones should be noted. Major gridding errors should be reported for immediate corrections. Fill out a daily filter-processing record, describing any unusual conditions, such as contaminated clones, irregular gridding, missing clones, or overgrowth.
51. Make sure that the chromatography paper on each baking dish is well saturated for each step, especially the hot denaturation step owing to high evaporation, and change the paper after processing every four filters.

References

1. Frengen, E., Zhao, B., Howe, S., Weichenhan, D., Osoegawa, K., Gjernes, E., Jessee, J., Prydz, H., Huxley, C., and de Jong, P. J. (2000) Modular bacterial artificial chromosome vectors for transfer of large inserts into mammalian cells. *Genomics* **68**, 118–126.
2. Zeng, C., Kouprina, N., Zhu, B., Wang, Y., Hoek, M., Cross, G., Osoegawa, K., Larinov, V., and de Jong, P. J. (2001) Large-insert BAC/YAC shuttle libraries for selective re-isolation of genomic regions by homologous recombination in yeast. *Genomics* **77**, 27–34.
3. Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **15**, 8794–8797.
4. Kim, U. J., Birren, B. W., Slepak, T., Mancino, V., Boysen, C., Kang, H. L., Simon, M. I., and Shizuya H. (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**, 213–218.
5. Ioannou, P., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A., and de Jong, P. J. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* **6**, 84–89.

Construction of Small Genome BAC Library for Functional and Genomic Applications

David Šmajs, Steven J. Norris, and George M. Weinstock

1. Introduction

The use of genetic approaches to study bacteria is limited by the following considerations: first, the bacteria must be easy to culture and nonpathogenic; and, second, a broad spectrum of genetic techniques for the particular bacteria being studied must be available. For other bacteria, such as the syphilis-causing spirochete, *Treponema pallidum*, genetic studies are highly limited because the bacterium cannot be continuously grown in the laboratory, because of its high virulence and a total lack of genetic tools to study this organism.

Construction of genomic libraries represents a powerful resource for genetic studies of bacteria including nonculturable organisms. For example, screening of a library for specific functions can be used for identification of genes coding for antigens, exported proteins, enzymes, receptors, regulators, and other activities.

The bacterial artificial chromosome (BAC) cloning system was invented for cloning of large fragments (80–300 kb) of eukaryotic DNA. However, there are several examples of using this technology for bacteria (1–4). In contrast to eukaryotic BAC libraries, gene expression from inserts of bacterial DNA in BACs was detectable and the functions of expressed genes can be studied (4,5). This expression is likely from bacterial transcription and translation signals in the insert and does not require special vector sequences. The major advantage of the BAC vector, derived from the F plasmid, is the strictly controlled copy number in *Escherichia coli*. The BAC plasmid is present in one to two copies per cell, which is important for genes that are toxic when overexpressed. Owing to the low BAC copy number, the insert length that can be recovered in BAC clones is usually much larger than for other cloning systems. BAC clones thus

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

can be used for construction of libraries covering the bacterial genome with a relatively small number of *E. coli* clones. Minimal sets of clones covering the bacterial chromosome can be subsequently used for strain comparisons, experiments in functional genomics, and genomic applications.

2. Materials

2.1. Preparation of Chromosomal DNA

1. Tris/EDTA (TE) buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.
2. Low-melting-point InCert agarose (FMC BioProducts, Rockland, ME).
3. TE buffer supplemented with 0.5% sodium dodecyl sulfate (SDS).
4. Proteinase K (Sigma, St. Louis, MO).

2.2. Partial Digestion of *T. pallidum* Chromosomal DNA

1. Triton X-100 (0.1%).
2. *Hind*III restriction endonuclease and *Hind*III buffer 2 (New England Biolabs, Beverly, MA).
3. 50 mM EDTA (pH 8.0).

2.3. Pulsed-Field Gel Electrophoresis of Digested DNA, Size Selection, and Electroelution

1. I.D.NA[®] agarose (BioWhittaker, Rockland, ME).
2. DR II apparatus (Bio-Rad, Hercules, CA), for electrophoresis by the contour-clamped homogeneous electric field (CHEF).
3. 0.5X Tris/acetate (TAE) buffer: 20 mM Tris-acetate, 0.5 mM EDTA, pH 8.3.
4. 1X TAE buffer: 40 mM Tris-acetate, 1 mM EDTA, pH 8.3.
5. λ DNA ladder (BMA, Rockland, ME).
6. 0.5 M EDTA (pH 8.0).
7. Dialysis tubing ($\frac{1}{4}$ - to $\frac{3}{4}$ -in. diameter) (Life Technologies, Gaithersburg, MD).

2.4. Preparation and Digestion of Vector DNA

1. *E. coli* VCS257 (Stratagene, La Jolla, CA) carrying pBeloBAC11 (6).
2. Luria Bertani (LB) plates containing 25 μ g/mL of chloramphenicol, 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG), and 40 μ g/mL of 5-bromo-4-chloro-3-indolyl- β -galactoside (X-Gal).
3. LB medium containing 25 μ g of chloramphenicol/mL.
4. Qiagen Plasmid Kit for isolation of BAC DNA (Qiagen, Valencia, CA) containing buffers P1, P2, P3, QBT, QC, QF, and Qiagen-tip 500 column.
5. P1 buffer: 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 μ g/mL of ribonuclease A (Sigma).
6. P2 buffer: 200 mM NaOH, 1% SDS.
7. P3 buffer: 3.0 M potassium acetate, pH 5.5.
8. Isopropanol and 70% ethanol.

9. *Hind*III endonuclease and calf intestinal alkaline phosphatase (CIP) (New England Biolabs).
10. Ultrapure agarose (Gibco-BRL, Life Technologies, Gaithersburg, MD).
11. TBE buffer: 89 mM Tris-borate, 2 mM EDTA, pH 8.3.
12. QIAquick Gel Extraction Kit (Qiagen).

2.5. Ligation, Dialysis, and Electroporation

1. T4 DNA ligase (New England Biolabs).
2. VSWP 0.025- μ m membranes (Millipore, Bedford, MA).
3. Electrocompetent ElectroMAX DH10B cells (Life Technologies).
4. LB medium.
5. Filter-sterilized glycerol (10%).
6. Gene Pulse Controller II apparatus (Bio-Rad) and 0.2-cm gap electrode cuvetts.
7. SOC medium: 2% bacto-tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose.
8. Chloramphenicol-containing LB plates (12.5 μ g/mL) supplemented with 0.5 mM IPTG and X-Gal (40 μ g/mL).

2.6. Isolation of DNA From Individual BAC Clones

1. Qiagen Plasmid Kit for isolation of BAC DNA (Qiagen), Qiagen-tip 20 columns.
2. LB medium supplemented with 12.5 μ g/mL of chloramphenicol.

2.7. End Sequencing and Restriction Mapping of BAC Clones

1. *Hind*III, *Eco*RI, *Xba*I, and *Xho*I restriction endonucleases with corresponding reaction buffers (New England Biolabs).
2. Model 377 DNA sequencing system (Applied Biosystems, Foster City, CA).
3. ABI Prism® BigDye™ Terminators v3.0 Cycle Sequencing Kit (Applied Biosystems).
4. Polymerase chain reaction (PCR) primers with target sites on pBeloBAC11 (GW386: 5'-TTGTAAAACGACGGCCAGTG-3'; GW387: 5'-TTACGCCAAGC TATTTAGGTGAC-3').

3. Methods

3.1. Preparation of Chromosomal DNA

1. Grow *T. pallidum* subspecies *pallidum* (Nichols) strain in rabbit testes, harvest, and purify the cells using sodium diatrizoate gradient centrifugation (7,8). Wash the *T. pallidum* cells four times, and resuspend in TE buffer to a concentration of 2×10^{10} cells/mL.
2. Mix an equal volume of *T. pallidum* cells with molten 1.6% low-melting-point InCert agarose, and apply 200 μ L of this mix into plug molds and allow solidification.
3. Gently remove the resulting $15 \times 9 \times 1.5$ mm plugs, and put into 30 mL of TE buffer supplemented with 0.5% SDS and incubate overnight at 37°C. Subse-

quently, add proteinase K to a final concentration of 100 µg/mL and incubate the plugs for an additional 48 h at 55°C. Then wash the plugs four times with TE buffer for 60 min each.

3.2. Partial Digestion of *T. pallidum* Chromosomal DNA

1. Partial digestion is performed according to Brosch et al. (1). Wash the chromosomal DNA-containing plugs three times with 0.1% Triton X-100, and then equilibrate three times in 10 mL of *Hind*III buffer 2 supplemented with 0.1% Triton X-100 for 60 min at 4°C.
2. The last equilibration step is done on ice. After removal of the buffer, transfer each plug to 1 mL of ice-cold buffer 2 containing *Hind*III restriction endonuclease (20 U/mL) and incubate for 2 h on ice. Then incubate the plugs for 30 min at 37°C (see Note 1).
3. Stop digestion by adding 0.5 mL of 50 mM EDTA (pH 8.0) to 1 mL of *Hind*III-containing buffer 2.
4. Use the plugs with digested DNA for fragment separation by pulsed-field gel electrophoresis (PFGE).

3.3. PFGE of Digested DNA, Size Selection, and Electroelution

1. Place the plugs with partially digested DNA into the wells of a 1% I.D.NA agarose gel, and subject to electrophoresis by the CHEF method using the DR II apparatus. Prepare and run the gels in 0.5X TAE buffer. Perform PFGE at 14°C and 6 V/cm for 16 h with a 5- to 45-s pulse time at a 120° angle. A λ DNA ladder is used for size markers from 48.5 kb to more than 1 Mb.
2. Remove the marker lanes from the gel and stain with ethidium bromide to visualize the positions of individual marker bands. Then cut out the lanes containing digested genomic DNA in the region corresponding to DNA fragment lengths of 40–200 kb, and divide into an additional four agarose blocks containing fragments with 40–80, 80–120, 120–160, and 160–200 kb, respectively. The size selection of digested *T. pallidum* DNA is shown in Fig. 1. For cloning of DNA fragments, digested DNA is exposed neither to ethidium bromide nor to ultraviolet light (see Note 2).
3. Directly use gel slices for electroelution or store in 0.5 M EDTA (pH 8.0) at 4°C.
4. Electroelution of digested genomic DNA from gel slices is performed according to Strong et al. (9). Equilibrate the gel slices with 1.0X TAE buffer for 3 h and subsequently transfer into dialysis tubing of 1/4- to 3/4-in. diameter (Life Technologies) with 200–400 µL of fresh 1.0X TAE buffer. Adjust the gel slice to one side of the dialysis tubing with the slice long axis perpendicular to the voltage gradient and put closer to the negative electrode. Elute the DNA from the gel at 2.5 V/cm for 2 h, and at the end of elution, reverse the polarity of current for 30 s to detach the DNA molecules from the bag.
5. Carefully remove the eluted DNA with a wide-bore pipet tip and either use directly for ligation or store at 4°C (see Note 3).

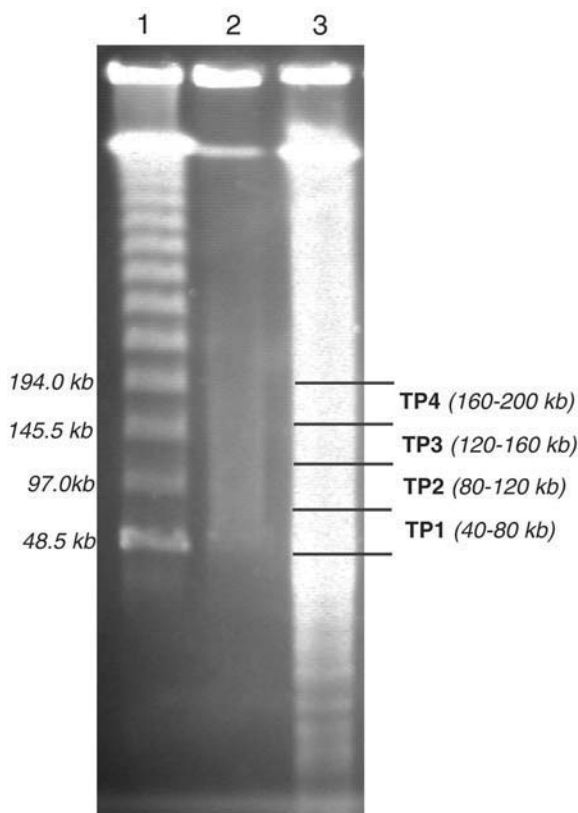


Fig. 1. PFGE of undigested and *Hind*III-treated *T. pallidum* chromosomal DNA. Lane 1, λ DNA ladder (48.5 kb to more than 1 Mb); lane 2, *T. pallidum* chromosomal DNA undigested; lane 3, *T. pallidum* chromosomal DNA digested for 30 min with *Hind*III restriction endonuclease. Note the four regions, TP1–TP4, of the gel used for excision of size-separated DNA fragments.

3.4. Preparation and Digestion of Vector DNA

1. The pBeloBAC11 vector (6) is used for construction of the library. Inoculate *E. coli* VCS257 carrying pBeloBAC11 on an LB plate containing 25 μ g/mL of chloramphenicol, 0.5 mM IPTG, and 40 μ g/mL of X-Gal and incubate overnight at 37°C. Use a single blue colony for inoculation of 1 L of LB culture containing 25 μ g of chloramphenicol/mL of medium.
2. Harvest the cells and isolate the plasmid DNA using a modified Qiagen Plasmid Protocol for isolation of BAC DNA (Qiagen). Carefully resuspend the cells to 100 mL of P1 buffer at 4°C until no cell clumps are visible. Add 100 mL of room

temperature P2 buffer, gently mix with the cells resuspended in P1, and incubate at room temperature for 5–10 min to lyse the cells. After visible lysis, add 100 mL of prechilled buffer P3, incubate the mixture on ice for 10 min, and centrifuge twice at 20,000g for 20 min at 4°C, discarding the sediment. Apply the clear supernatant to a QBT buffer–equilibrated Qiagen-tip 500 column, wash twice with QC buffer, and elute the vector DNA from the column with QF buffer prewarmed to 60°C according to the manufacturer’s recommendations.

3. Precipitate the eluted DNA with 0.7 vol of room temperature isopropanol, and immediately centrifuge at 15,000g for 30 min at 4°C. Discard the supernatants, wash the DNA with 70% ethanol, and recentrifuge the DNA for 15 min at 15,000g at 4°C. Remove the ethanol and dry the DNA pellet for 5 min in a vacuum.
4. Resuspend the DNA in 100 μ L of distilled water and measure the DNA concentration. Digest the pBeloBAC11 plasmid (100 ng) with 5 U of *Hind*III for 2 h at 37°C, dephosphorylate by adding 5 U of CIP, and incubate for 30 min at 37°C. Run the digested vector DNA on a 1% agarose gel in TBE buffer, and cut out the DNA band corresponding to the digested vector DNA, and extract from the gel using the QIAquick Gel Extraction Kit, and measure the DNA concentration (*see Note 4*).

3.5. Ligation, Dialysis, and Electroporation

1. Ligate 10 ng of size-selected *T. pallidum* DNA to 1 ng of *Hind*III-digested and -dephosphorylated pBeloBAC11 DNA overnight at 16°C with 200 U of T4 DNA ligase.
2. Inactivate T4 DNA ligase at 65°C for 10 min, and then drop-dialyze the ligation solution against TE buffer using VSWP 0.025- μ m membranes.
3. Prepare electrocompetent ElectroMAX DH10B cells after harvesting of an LB culture at $OD_{600} = 0.5$. Then extensively wash the cells five times with ice-cold water and resuspend in 10% ice-cold filter-sterilized glycerol to an OD_{600} of approx 100, and either directly use for electroporation or store as 50- μ L aliquots at –80°C.
4. Mix 50 μ L of competent cells with 1 μ L of ligation mixture in a 0.2-cm gap electrode cuvet on ice. Use a Gene Pulse Controller II apparatus set to 2.5 kV, 25 μ F, and 100 Ω .
5. Immediately after electroporation, add 0.6 mL of sterile SOC medium prewarmed to 37°C and grow the cells for 1 h at 37°C with shaking at 100 rpm.
6. Plate the cells on chloramphenicol-containing LB plates (12.5 μ g/mL) supplemented with 0.5 mM IPTG and X-Gal (40 μ g/mL). Incubate the plates for at least 24 h at 37°C or, for better results, for more than 48 h. Isolate the white colonies and use for further investigations (*see Note 5*).

3.6. Isolation of DNA From Individual BAC Clones

1. Use white colonies for isolation of clone DNA (*see Note 6*).
2. Use the same, but scaled down, procedure as for isolation of pBeloBAC11 DNA. Inoculate each white colony into 10 mL of LB medium with 12.5 μ g/mL of chlo-

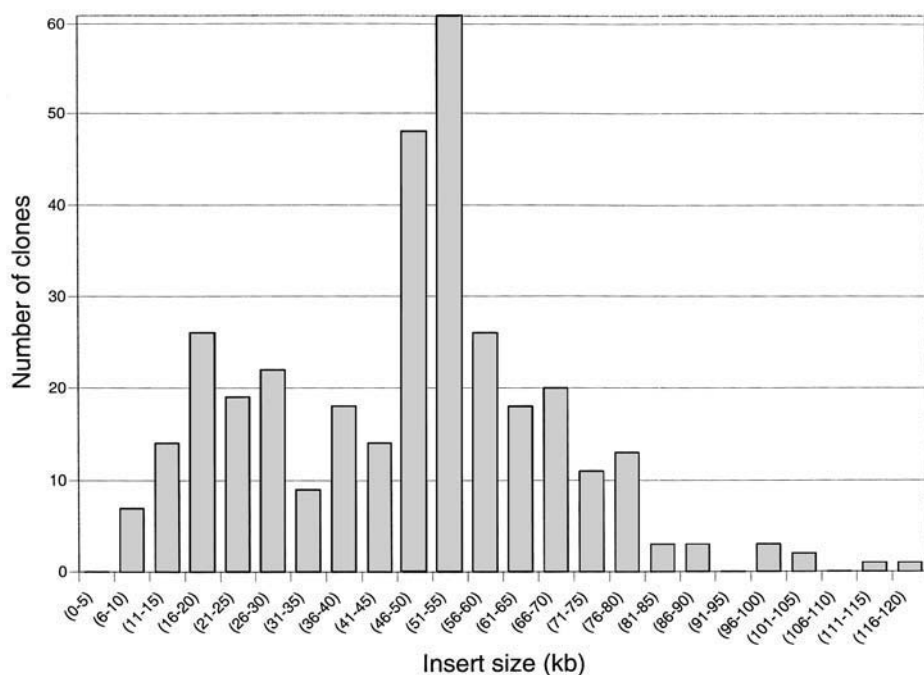


Fig. 2. Distribution of insert lengths of 339 clones in the library. Numbers of clones containing insert sizes in 5-kb increments are shown starting with 0–5 kb. The majority of the clones contained insert sizes of 45–50 and 50–55 kb. The largest insert obtained slightly exceeded 120 kb. A second peak for the insert lengths of 10–30 kb indicates preferential cloning of small inserts, probably as a result of higher cloning, ligation, and/or transformation efficiency.

ramphenicol, and after overnight incubation, extract the DNA using 2 mL of P1, P2, and P3 buffers and Qiagen-tip 20.

3. Resuspend the dried DNA pellet in 20 μ L of distilled water and store at -20°C .

3.7. End Sequencing and Restriction Mapping of BAC Clones

1. Use the isolated clone DNA (4 μ L) for an initial *Hind*III restriction digestion analysis to test for the presence of the insert (see **Note 7**).
2. Use the clone DNA (9 μ L) as a template for DNA sequencing reactions. Sequence the DNA using the *Taq* Dye-deoxy Terminator method and a model 377 DNA sequencing system. Use two PCR primers with target sites on pBeloBAC11 to sequence both insert termini, GW386 and GW387.
3. Align the DNA sequence obtained with the *T. pallidum* whole genome sequence (10) and map the position and length of each clone. No noncontiguous clone sequences were found. The distribution of insert lengths is shown in **Fig. 2**.

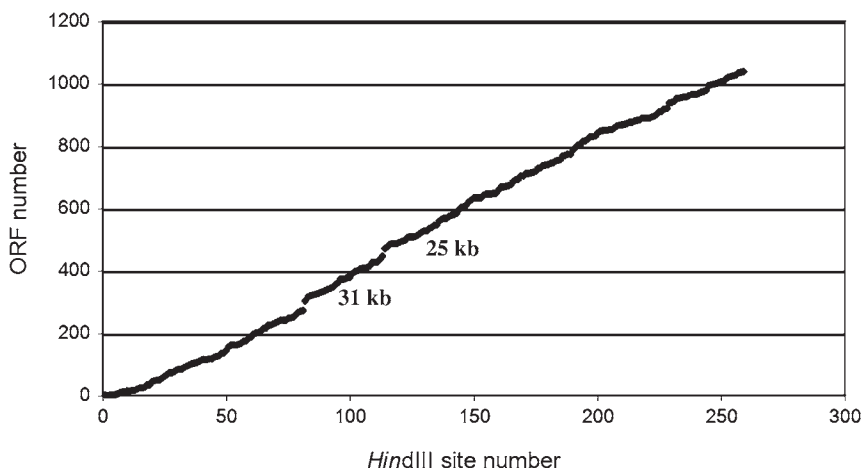


Fig. 3. Distribution of *Hind*III target sites along *T. pallidum* chromosome. For each individual *Hind*III site, the corresponding ORF containing the target site is plotted. The 259 *Hind*III sites are randomly distributed through 1040 predicted *T. pallidum* ORFs. The two largest regions without *Hind*III sites (31 and 25 kb) are indicated.

4. Additionally map selected clones with *Hind*III, *Eco*RI, *Xba*I, and *Xho*I restriction endonucleases, and compare the fragments obtained with those predicted (see Note 8).

4. Notes

1. The *Hind*III restriction endonuclease was chosen for construction of the *T. pallidum* library because of the unique cloning site on the pBeloBAC11 vector (6) and random distribution of 259 individual *Hind*III target sites throughout the *T. pallidum* chromosome (Fig. 3). Large regions without *Hind*III target sites would prevent their cloning and presence in the library. The two largest regions of the *T. pallidum* chromosome without *Hind*III target sites are a 31-kb region between bp 288,534 and 316,779 (open reading frame [ORF] TP0273–TP0304) and a 25-kb region between bp 474,463 and 499,924 (ORF TP00448–TP0471; Fig. 3). Both of these are small enough to be clonable in a large-insert (>30 kb) library.
2. Preelectrophoresis and elimination of small fragments can be used to achieve more homogeneous distribution of large inserts (11).
3. Alternatively, isolation of DNA fragments from agarose blocks can be performed by enzyme digestion of agarose with Gelase (Epicentre, Madison, WI) after agarose is melted for 10 min at 65°C (1). However, electroelution was shown to yield more intact DNA than isolation by the agarose digestion treatment (9). To our knowledge, electroelution is gentler and more efficient and thus more suitable for isolation of DNA fragments in which the source of DNA is limited.

4. A high-quality preparation of vector DNA can be used to achieve a reduced background of nonrecombinant clones (11).
5. Increased electroporation efficiency can be achieved by the method published previously (12).
6. Most of the white colonies were obtained for the DNA fraction with fragment lengths between 40 and 80 kb. Approximately 10% of all white colonies isolated were obtained with ligated DNA fragments 80–120 kb, and two fractions with an insert size of 120–200 kb repeatedly resulted in only clones with no inserts. A similar observation was made for construction of a *Mycobacterium tuberculosis* large-insert library in which human DNA was used as a positive control. The maximum size of the insert for prokaryotic DNA was shown to be considerably lower than for eukaryotic DNA (1).
7. Approximately 20% of the white colonies contained no insert clones and were discarded.
8. In 2 of 26 (7.7%) clones investigated, deletions of more than 30 kb inside the insert were observed. However, the clones used for restriction mapping were not selected randomly; that is, only the largest clones along the *T. pallidum* chromosome were used. These data, together with the finding that the cloned *T. pallidum* DNA fragments in the library were unequally distributed throughout the *T. pallidum* chromosome (not shown), suggest that the maximum size and stability of large inserts in BAC vectors depend not only on the prokaryotic vs eukaryotic source, but also on the specific gene content of each clone.

References

1. Brosch, R., Gordon, S. V., Billault, A., et al. (1998) Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infect. Immun.* **66**, 2221–2229.
2. Dewar, K., Sabbagh, L., Cardinal, G., Veilleux, F., Sanschagrin, F., Birren, B., and Levesque, R. C. (1998) *Pseudomonas aeruginosa* PAO1 bacterial artificial chromosomes: strategies for mapping, screening, and sequencing 100 kb loci of the 5.9 Mb genome. *Microb. Comp. Genomics* **3**, 105–117.
3. Tomkins, J. P., Miller-Smith, H., Sasinowski, M., et al. (1999) Physical map and gene survey of the *Ochrobactrum anthropi* genome using bacterial artificial chromosome contigs. *Microb. Comp. Genomics* **4**, 203–217.
4. Xu, Y., Murray, B. E., and Weinstock, G. M. (1998) A cluster of genes involved in polysaccharide biosynthesis from *Enterococcus faecalis* OG1RF. *Infect. Immun.* **66**, 4313–4323.
5. Rondon, M. R., Raffel, S. J., Goodman, R. M., and Handelsman, J. (1999) Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. *Proc. Natl. Acad. Sci. USA* **96**, 6451–6455.
6. Kim, U. J., Birren, B. W., Slepak, T., et al. (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**, 213–218.

7. Baseman, J. B. and Hayes, N. S. (1974) Protein synthesis by *Treponema pallidum* extracted from infected rabbit tissue. *Infect. Immun.* **10**, 1350–1355.
8. Hanff, P. A., Norris, S. J., Lovett, M. A., and Miller, J. N. (1984) Purification of *Treponema pallidum*, Nichols strain, by Percoll density gradient centrifugation. *Sex. Transm. Dis.* **11**, 275–286.
9. Strong, S. J., Ohta, Y., Litman, G. W., and Amemiya, C. T. (1997) Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. *Nucleic Acids Res.* **25**, 3959–3961.
10. Fraser, C. M., Norris, S. J., Weinstock, G. M., et al. (1998) Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**, 375–388.
11. Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52**, 1–8.
12. Zhu, H. and Dean, R. A. (1999) A novel method for increasing the transformation efficiency of *Escherichia coli*—application for bacterial artificial chromosome library construction. *Nucleic Acids Res.* **27**, 910, 911.

Preparation of BAC Libraries From Bacterial Genomes by In Vitro Packaging

Sangita Pal, Solida Mak, and George M. Weinstock

1. Introduction

Genomic libraries represent a powerful resource for genetic studies of bacteria. Large-insert libraries in bacterial artificial chromosome (BAC) vectors are particularly important not only for genome-sequencing projects, but for comparative and functional genomic studies once a complete sequence is known. Cloned DNA can be introduced into bacterial cells in two different ways. One way is to have cells take up naked DNA; this is known as transformation (1). The other way is to package the recombinant DNA inside bacteriophage particles in vitro using a phage such as λ . This process, known as in vitro packaging, allows DNA to be introduced by infection (2). Although transformation allows DNA of any size to be introduced, DNA uptake occurs at very low frequency. On the other hand, in vitro packaging limits the size of the DNA based on the phage head stability. However, in vitro package is the most efficient method for introducing large-insert clones.

In vitro packaging uses cosmid vectors such as a BAC vector to package DNA into the bacteriophage head. The cosmid vector has a cos site from phage λ , which is recognized by the phage-packaging apparatus and allows the DNA to be encapsulated in the phage head. One major constraint of cosmids is that the size of the cloned DNA is limited by the size of the phage head, about 50 kb for vector plus insert. If the size of DNA cloned into cosmid is too large or too small, the phage head will be unstable (3). Because of the size limitation of packaging protocols and the size of BAC vectors (8 to 9 kb), the average insert size is about 40 kb. For prokaryotic DNA, such large inserts can be unstable owing to expression of genes from the insert in the heterologous prokaryotic

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

host. However, BAC vectors, because of their low copy number, allow large-insert clones to be better tolerated.

In this chapter, we focus on the cloning of *Treponema denticola* genomic DNA by an in vitro packaging method. *T. denticola* is a spirochete that is part of the normal flora of the oral cavity and plays a role in periodontitis. Briefly, genomic DNA is prepared for insertion into a chosen cosmid vector. The vector and target DNA are then ligated together, and these concatemers are introduced into *Escherichia coli* through packaging by the infection process known as transduction.

2. Materials

All enzymes and reagents should be of molecular grade. Solutions should be made from double-distilled water. Reagents for in vitro packaging, DNA preparations, and sequencing are available as kits. Storage condition is usually at room temperature unless otherwise specified.

2.1. Preparation of Chromosomal DNA

1. Freshly prepared New Oral Spirochete medium (4).
2. 10% Rabbit serum (store at -20°C).
3. Anaerobic chamber (Coy, Grass Lake, MI).
4. TE buffer: 10 mM Tris-HCl; 1 mM EDTA, pH 8.0.
5. 10% sodium dodecyl sulfate (SDS).
6. Proteinase K (store at -20°C).
7. 5 M NaCl.
8. Hexadecyltrimethylammonium bromide (CTAB)/NaCl solution: 10% CTAB and 0.7 M NaCl.
9. Phenol/chloroform/isoamyl alcohol (25:24:1 [v:v:v] solution (Gibco-BRL, Gaithersburg, MD); store at -4°C . Phenol can cause severe burns to the skin and mucosal membranes.
10. Isopropanol.

2.2. Partial Digestion of Genomic DNA

1. Sau3A enzyme (store at -20°C) (see Note 1).
2. 1X TBE buffer: 89 mM Tris-borate and 2 mM EDTA, pH 8.3.
3. 1-kb Molecular weight marker (New England Biolabs).
4. HindIII-digested λ DNA molecular weight marker (Gibco-BRL).
5. 0.4–1.0% Agarose gel (Gibco-BRL, Life Technologies).
6. Ethidium bromide (EtBr) stock solution (50 $\mu\text{g}/\text{mL}$): EtBr is a mutagen; gloves should be worn when handling this reagent.
7. 10X Sau3A buffer (store at -20°C).
8. EDTA solution.
9. Phenol/chloroform/isoamyl alcohol (25:24:1) solution (Gibco-BRL); store at -4°C .
10. 3 M Na acetate.

11. Ethanol.
12. 70% Ethanol.

2.3. DNA Size Fractionation by Sucrose Gradient

1. pBeloBAC11OriT vector (pFT-2) (store at -20°C) (see **Note 2**).
2. Gradient former (Gibco-BRL).
3. 10% (w/v) Sucrose (Sigma, St. Louis, MO) in buffer containing 20 mM Tris-HCl (pH 8.0), 1 M NaCl, and 5 mM EDTA (pH 8.0) (see **Note 3**).
4. 40% (w/v) Sucrose (Sigma) in buffer containing 20 mM Tris-HCl (pH 8.0), 1 M NaCl, and 5 mM EDTA (pH 8.0).
5. Stirring bar.
6. Glass capillary tube.
7. 30-mL Ultracentrifuge tube (Beckman, Palo Alto, CA).
8. Ultracentrifuge (Sorvall).
9. Eppendorf tubes.

2.4. Agarose Gel Electrophoresis of Partially Digested DNA

1. 0.4–1.0% Agarose gel (FMC, Rockland, ME).
2. EtBr stock solution (50 $\mu\text{g}/\text{mL}$): EtBr is a mutagen; gloves should be worn when handling this reagent.
3. 1X TBE buffer: 89 mM Tris-borate and 2 mM EDTA, pH 8.3.
4. 1-kb molecular weight marker (New England Biolabs).
5. *Hind*III-digested λ DNA molecular weight marker (Gibco-BRL).
6. 10X loading buffer: 0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol in ddH_2O .
7. Ultraviolet (UV) transilluminator.

2.5. Purification of Digested Genomic DNA

1. Dialyzed tubing of $\frac{1}{4}$ -in. diameter (Gibco-BRL).
2. Stirring bar.
3. Magnetic stirrer.
4. Butanol (Sigma).
5. 3 M Na acetate.
6. Ethanol.
7. 70% Ethanol.
8. Drying vacuum.
9. 1X TE solution: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.
10. Fluorometer (Cambridge Technology, Watertown, MA).

2.6. Preparation of BAC Vector DNA

1. Chloramphenicol solution (store at -20°C).
2. Luria Bertani (LB) medium.
3. Qiagen MidiPrep kit (Qiagen, Valencia, CA).
4. Vacuum.

2.7. Digestion and Purification of Vector DNA

1. *Bam*HI (New England Biolabs) (store at -20°C).
2. Calf intestinal alkaline phosphatase (CIP) (New England Biolabs) (store at -20°C).
3. QIAquick Gel Extraction Kit (Qiagen, Valencia, CA).
4. 37°C Water bath.

2.8. Ligation for In Vitro Packaging

1. T4 DNA ligase (New England Biolabs) (store at -20°C).
2. 10X ligase buffer (store at -20°C).
3. 16°C Water bath.

2.9. Preparation of Bacterial Host Cells

1. *E. coli* (DH5 α).
2. 10 mM MgSO_4 .
3. 0.2% (w/v) Maltose (Sigma).
4. LB medium.

2.10. Phage-Packaging Reaction

1. Gigapack III Gold Packaging Kit (Stratagene, La Jolla, CA) (store at -80°C).
2. Ligation mixture (store at -20°C).
3. SM buffer: 0.01% gelatin, 0.1 M NaCl, 0.016 M MgSO_4 , and 0.05 M Tris-HCl.
4. *E. coli* (DH5 α) host cells.
5. LB medium.
6. LB agar plate containing 12.5 $\mu\text{g/mL}$ of chloramphenicol (store at 4°C).
7. Glycerol.

2.11. Isolation and Determination of DNA Insert Size From Individual BAC Clones

1. LB medium with chloramphenicol (12.5 $\mu\text{g/mL}$).
2. Qiagen-tip 20 (Qiagen).
3. *Not*I (New England Biolabs) (store at -20°C).
4. *Hind*III (New England Biolabs) (store at -20°C).
5. Pulsed-field gel electrophoresis (PFGE) contour-clamped homogeneous electric field DR II apparatus (Bio-Rad, Hercules, CA).
6. 0.5X TBE buffer.
7. I.D.NA[®] agarose (1.0%) (BioWhittaker, Rockland, ME).
8. λ DNA ladder (BioWhittaker).

2.12. End Sequencing of BAC Clones

1. Model 377 DNA-sequencing system (Applied Biosystems, Foster City, CA).
2. Polymerase chain reaction (PCR) primers (2 pmol/ μL) with target sites on pBeloBAC11 (GW570: 5'-TATAATGACCCCGAAGCAGG-3', GW387: 5'-TTACGCCAAGCTATTAGGTGAC-3').

3. Template DNA (400–500 ng).
4. Sequencing kit (Perkin-Elmer).
5. PCR (Perkin-Elmer).

3. Methods

3.1. Preparation of Chromosomal DNA

1. Grow *T. denticola* in 250 mL of New Oral Spirochete (NOS) medium (4) with 10% heat-inactivated rabbit serum at 37°C in an anaerobic chamber (Coy) under static condition.
2. After 6 to 7 d, harvest the cells by centrifugating at 4000g for 10 min.
3. Add 4.75 mL of TE into the cell pellet, digest the cells by adding 0.25 mL of 10% SDS and 25 µL of 20 mg/mL proteinase K and incubating for 1 h at 37°C.
4. Add 1.8 mL of 5 M NaCl and 1.5 mL of CTAB/NaCl solution (10% CTAB, 0.7 M NaCl) (5) to the mixture, gently mix the solution, and incubate for another 20 min at 65°C.
5. Extract the DNA by adding an equal volume (8.3 mL) of phenol/chloroform/isoamylalcohol (25:24:1) solution. Invert the tube for 15 s until the mixture becomes milky, and place it in a microfuge at 23,000g for 10 min.
6. Transfer the upper aqueous phase into a 10-mL Corex tube, and precipitate the DNA by adding 0.6 vol (5 mL) of room temperature isopropanol.
7. Place the tube containing the DNA mixture in a 4°C centrifuge and spin at 23,000g for 30 min.
8. Discard the solution; white DNA pellet is visible at the bottom of the tube.
9. Wash the DNA pellet twice with 70% alcohol to remove any residual CTAB.
10. Dry the DNA pellet and resuspend it in 1 mL of TE buffer.
11. Load 1–3 µg of DNA on a 0.4% agarose gel to check the quality of the DNA (see Note 4).

3.2. Partial Digestion of Genomic DNA

1. In a pilot experiment, set up two sets of seven microcentrifuge tubes. To each tube add 1 µg of genomic DNA and digest it with various amounts of Sau3A enzyme (1.0, 0.5, 0.25, 0.13, 0.06, 0.03, and 0.01 U). Incubate the first set of tubes at 37°C for 1 h and the second set at 37°C for 2.0 h.
2. Electrophorese the digested DNA in a 0.4% agarose gel containing 0.5 µg/mL of EtBr for overnight at 40 V. From the result of the pilot experiment, the correct time point and Sau3A enzyme concentration is chosen to be 0.025 U at 37°C for 1 h.
3. Partially digest 200 µg of high molecular weight genomic DNA with Sau3A.
4. Set up digestions in five serial dilutions to achieve a range of enzyme concentration 0.1, 0.05, 0.025, 0.013, and 0.006 U, respectively. This enzyme range should span around the optimal enzyme concentration with plus or minus twofold dilution.
5. Set up a serial dilution as follows: Divide 200 µg of DNA equally among five tubes (40 µg/tube) with the exception of the first tube having twice the amount of the DNA (80 µg) and the last tube being empty (0 µg) (see Note 5).

6. Add 5 μL of 10X Sau3A buffer to tubes 2–5 and 10 μL to tube 1.
7. Fill each tube with sterile water up to 100 μL for the first tube and 50 μL for the rest.
8. Add 8 U of Sau3A enzyme to the first tube containing 80 μg of DNA to achieve a 0.1 U enzyme concentration.
9. Mix gently and serially transfer 50 μL of that mixture into the second tube, then transfer 50 μL from the second tube to the third tube, and so on. The last tube should end up with a 100- μL volume.
10. Perform the enzyme digestion at 37°C for 1 h (3).
11. Stop the DNA digestion reaction by adding 6.5 μL of 100 mM EDTA solution into each tube and 13 μL to the last tube. The final concentration of EDTA solution should be 13 mM.
12. Check 1 μL of digested DNA from each tube on a 0.4% agarose gel.
13. Extract the DNA once with an equal volume of phenol/chloroform/isoamyl alcohol.
14. Precipitate with 10% 3M Na acetate and 3 vol of ethanol and wash with 70% ethanol (6).
15. Dry the DNA pellet under vacuum and resuspend each pellet in 500 μL of TE.
16. Pool the digested DNA mixture together and incubate at 65°C for 5 min to dissociate any DNA aggregation (see **Note 6**).

3.3. DNA Size Fractionation by Sucrose Gradient

1. To make a sucrose gradient, first set up a gradient apparatus on a stirring plate and turn the switch button to the “off” position to prevent leaking.
2. Pour 15 mL of 10% (w/v) sucrose (Sigma) in buffer containing 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, and 1 mM EDTA (pH 8.0) into the inside chamber of the apparatus (7).
3. Pour 15 mL of 40% sucrose in buffer containing 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, and 1 mM EDTA (pH 8.0) into the outside chamber of the apparatus (see **Note 7**).
4. Place an appropriate-size stirring bar into the inner chamber of the gradient former to mix the two sucrose solutions as they pass through a glass capillary tube.
5. Turn the switch to the “on” position, and carefully watch the two sucrose solutions passing through the capillary tube into a 30-mL ultracentrifuge collection tube.
6. Discard the initial few drops to get rid of the first few bubbles. Collect the rest of the sucrose gradient by inserting the tip of the glass capillary tube into the bottom of the collection tube until the last drop of solution has gone through.
7. Load the digested DNA pool carefully on top of freshly poured sucrose gradient in a dropwise manner without disturbing the gradient.
8. Place a gradient tube in an SW-27 rotor and centrifuge at 6,000g, 20°C for 16–20 h.
9. Remove the DNA gradient tube and carefully place it in a tube holder.
10. Collect fractions of DNA gradient by inserting a glass capillary tube into the bottom of the DNA gradient tube and allowing about 1000 μL (10 drops) of the DNA fraction to be collected in each microcentrifuge tube (see **Note 8**).

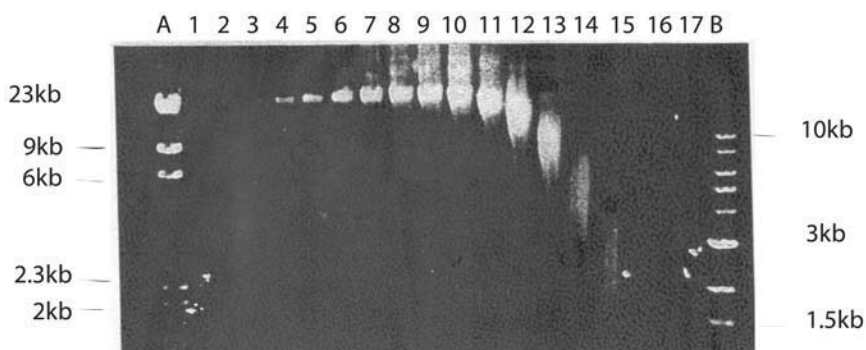


Fig. 1. DNA sucrose gradient fractionation. DNA is partially digested with *Sau3A* and fractionated by sucrose gradient. Lane A, high-molecular-weight marker; lanes 1–17, DNA gradient fractions collected in 1-mL aliquots each; lane B, 1-kb molecular weight marker. The first aliquot contained the densest part of the gradient and the DNA of highest molecular weight.

3.4. Agarose Gel Electrophoresis of Partially Digested DNA

1. Electrophorese 30 μ L of each digested sample in a 0.4% agarose gel containing EtBr.
2. Run the gel in 1X TBE at low voltage overnight, and photograph the gel to characterize the fractionated genomic DNA (**Fig. 1**).

3.5. Purification of Digested Genomic DNA

1. Dialyze fractions containing the desired DNA fragments of 39–47 kb in $\frac{1}{4}$ -in.-diameter dialyzed tubing against 0.5 L of 1X TE at 4°C (*see Note 9*).
2. Dialyze at 4°C overnight using a stirring bar on a magnetic stirrer.
3. The next day, extract each dialyzed DNA fraction once with an equal volume of butanol to reduce the water content. Mix the solution well and spin at 4000 rpm for 15 min.
4. Transfer the bottom DNA layer into a clean tube and precipitate with 10% 3 M Na acetate and 3 vol of ethanol.
5. Wash the DNA once with 70% ethanol and dry in a vacuum for 5–10 min. Resuspend each DNA pellet in 30 μ L of distilled water.
6. Check the quality and quantity of the DNA by absorption at OD₂₆₀ and OD₂₈₀ and by agarose gel electrophoresis.

3.6. Preparation of BAC Vector DNA

1. Streak *E. coli*, DH5 α -carrying pFT-2 vector on an LB agar plate containing 12.5 μ g/mL of chloramphenicol and incubate overnight at 37°C.
2. Select a single colony and inoculate in 5 mL of LB medium containing chloramphenicol. Incubate at 37°C overnight under constant shaking.

3. Isolate the vector DNA from the overnight 5 mL culture by using the rapid DNA isolation procedure according to a Qiagen Miniprep kit.
4. Digest $\frac{1}{4}$ vol of the DNA with *Bam*HI for 1 h at 37°C, and perform gel electrophoresis to confirm the presence of an appropriate-size vector.
5. Add 1% inoculum from the overnight culture to 100 mL of LB medium containing 12.5 µg/mL of chloramphenicol, and incubate at 37°C for 16–18 h.
6. The next day, harvest the cells at 1829g for 15 min at room temperature.
7. Isolate the pFT-2 vector DNA using a Qiagen MidiPrep kit.
8. Dry the DNA pellet in a vacuum and resuspend in 100 µL of distilled water.
9. Check the DNA quantity and quality by absorption at OD₂₆₀ and OD₂₈₀ and by agarose gel electrophoresis.

3.7. Digestion and Purification of Vector DNA

1. Digest 10 µg of pFT-2 vector DNA with 10 U of *Bam*HI at 37°C for 4 h.
2. Dephosphorylate the digested DNA by adding 2 µL of 5 U/µL CIP and incubate at 37°C for 30 min.
3. Inactivate the CIP reaction by incubating it at 65°C for 10 min.
4. Load the dephosphorylated DNA into a 1.0% agarose gel, and electrophorese in 1X TBE at 80–100 V for 3–5 h.
5. Place the gel on a UV illuminator, and using a clean razor, cut out the DNA band corresponding to the digested vector control, and transfer the gel slice into a 1.5-mL microfuge tube.
6. Purify the DNA gel slices using a QIAquick Gel Extraction Kit. Quantitate the purified vector DNA by taking the absorption at OD₂₆₀ and OD₂₈₀.

3.8. Ligation for In Vitro Packaging

1. Ligate 0.25 µg of purified vector DNA with three different molar ratios (1:1, 1:2, and 1:3) of purified 37 to 52-kb *Sau*3A-digested *T. denticola* DNA (see **Note 10**).
2. Perform ligation in a 5-µL vol with 5 U of T4 DNA ligase overnight at 16°C.
3. The next day, inactivate the T4 DNA ligase at 65°C for 10 min, and store it in a –20°C freezer before proceeding to the phage-packaging step.

3.9. Preparation of Bacterial Host Cells

1. Streak *E. coli* (DH5α) glycerol stock onto an LB agar plate and incubate overnight at 37°C.
2. Inoculate a single, isolated DH5α colony from an agar plate into 5 mL of LB medium supplemented with 10 mM MgSO₄ and 0.2% (w/v) maltose to maximize the expression of phage receptor protein.
3. Incubate the culture at 37°C, under shaking for 4–6 h. Monitor the bacterial growth by measuring the OD₆₀₀ until it reaches a range of 0.5–1.0. Do not let it grow past an OD₆₀₀ of 1.0.
4. Centrifuge the cells at 11000g for 10 min at room temperature, and wash twice with 2.5 mL of 10 mM MgSO₄ solution.

5. Dilute the cell pellet with sterile 10 mM MgSO₄ solution to an OD₆₀₀ of 0.5. Bacterial cells should be used in phage transduction immediately after dilution.

3.10. Phage-Packaging Reaction

1. Remove a packaging extracts tube from a Gigapack III Gold kit stored in a -80°C freezer, and thaw between the fingers until it melts.
2. Add 1 µL of ligation mixture to a packaging extracts tube.
3. Mix the tube gently and briefly spin down to ensure that all contents are at the bottom of the tube.
4. Incubate at room temperature for 2 h to allow the DNA to be packaged.
5. Stop the packaging reaction by adding 500 µL of SM buffer. Add 20 µL of chloroform to the reaction mixture and mix well.
6. Briefly spin the tube in a microcentrifuge to remove cellular debris.
7. Carefully transfer the supernatant into a clean 1.5-mL tube (the supernatant can be stored at 4°C for up to 1 mo).
8. Dilute the supernatant from the packaging solution with SM buffer to make 25 µL of 1;10 and 1;50 dilutions.
9. To each dilution, add 25 µL of freshly prepared DH5α cells, and incubate at room temperature for 30 min to allow the phages to infect bacterial cells.
10. Add 200 µL of LB medium to each reaction mixture, and incubate at 37°C for 1 h with gentle shaking every 15 min to allow the antibiotic gene to be expressed.
11. Spin the cells in a microfuge for 1 min at 11000g. Discard the supernatant and resuspend the cell pellet in 50 µL of LB medium.
12. Spread the cells on LB agar plates containing 12.5 µg/mL of chloramphenicol and incubate overnight at 37°C.
13. Select clones from overnight plates and inoculate in a 96-well plate containing 100 µL of LB medium with 12.5 µg/mL of chloramphenicol. Incubate the 96-well plate overnight at 37°C.
14. The next day, add glycerol solution to each well to a final concentration of 15% glycerol and store at -80°C (see **Note 11**).

3.11. Isolation and Determination of DNA Insert Size From Individual BAC Clones

1. Inoculate each colony into 10 mL of LB medium with 12.5 µg/mL of chloramphenicol, and incubate for 16 h at 37°C under constant shaking.
2. Extract the DNA using a Qiagen-tip 20 (Qiagen).
3. Resuspend the dried DNA pellet in 22 µL of distilled water and store at -20°C.
4. Use 4 µL of prepared DNA from the isolated clone for a *NotI* (New England Biolabs) restriction digestion analysis to test for the presence of insert (see **Note 12**).
5. Analyze the digested material using PFGE at a 5- to 45-s ramping over 16 h at 14°C and 6 V/cm (8).
6. Further confirm the size of each DNA insert with *HindIII* digestion (see **Note 13**). The distribution of insert lengths is shown in **Fig. 2**.

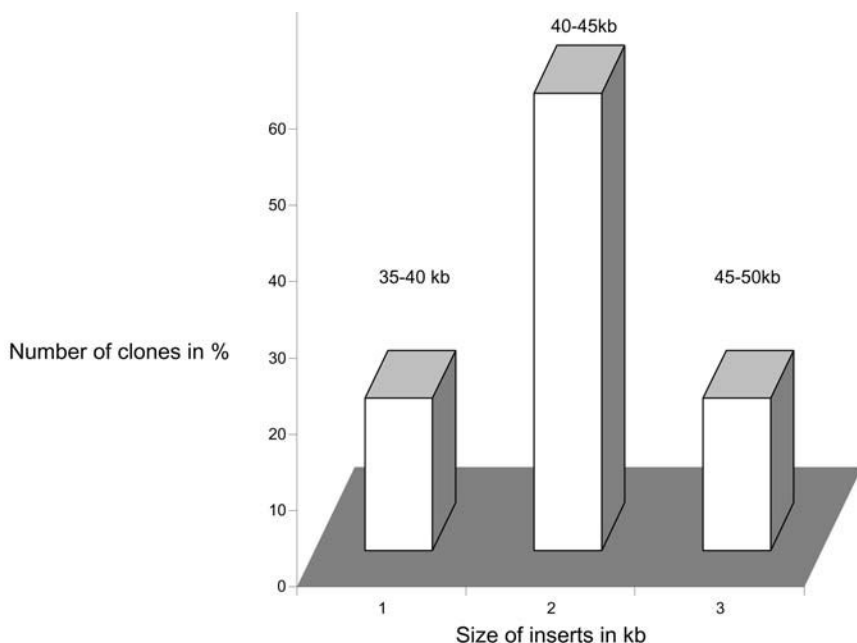


Fig. 2. Distribution of insert lengths of clones in *T. denticola* library. Approximately 60% of all colonies screened contained DNA fragments of insert sizes in the 35–38 kb range, 30% had insert sizes in the range of 39–40 kb, and the remaining 10% contained insert sizes in the range of 41–42 kb.

3.12. End Sequencing of BAC Clones

1. Use 9 μ L of clone DNA as a template in a PCR reaction for DNA sequencing.
2. Perform sequencing using the *Taq* Dye-deoxy Terminator method with a 377 DNA-sequencing system (Applied Biosystems). Two PCR primers (GW570 and GW387) with target sites on pBeloBAC11 are used to sequence the insert from both the 3'- and 5'-termini of each insert.
3. Align the results of the DNA sequences with the available *T. denticola* genome sequence to map the position and length of each clone.

4. Notes

1. The *Sau*3A restriction endonuclease was chosen for construction of the *T. denticola* library because of the unique, compatible *Bam*HI cloning site on the BAC vector pFT-2 (9) and the frequent occurrence of *Sau*3A target sites throughout the chromosome of *T. denticola*. Large regions without *Sau*3A target sites would prevent their cloning and presence in the library. *Sau*3A restriction endonuclease produces DNA ends that can be ligated to *Bam*HI-digested vector.

2. pFT-2 (9) vector is a modified pBeloBAC11 with an insertion of 0.6 kb OriT. The crucial element in the packaging reaction is the cos site, which governs the incorporation of phage DNA into infectious particles. Packaging requires two cos sites oriented in the same direction and separated by 37–55 kb. The first site is used to begin packaging, and the second is used to complete it and also to begin filling another particle (2).
3. Ten percent sucrose solution is prepared by adding 10 g of sucrose to sterile high-salt solution containing 20 mM Tris-HCl (pH 8.0) 1 M NaCl, and 5 mM EDTA (pH 8.0). In some protocols, this buffer contains a low amount of salt (10 mM NaCl). In our experience, using low-salt buffer can result in a great loss of DNA during the precipitation step after dialysis.
4. It is important to ensure that the genomic DNA obtained is of high molecular weight quality. Using low molecular weight DNA may yield poor results because DNA fractions obtained from degraded DNA contain fewer Sau3A sites and, therefore, are less efficient in ligating to vector.
5. To alleviate the problem with over- or underdigestion, the DNA digestions should be set up in five serial dilutions to achieve a wider range of enzyme concentrations (0.1, 0.05, 0.025, 0.013, and 0.006). This enzyme range should span around the optimal enzyme concentration (0.025 U) with a plus and minus twofold dilution of enzyme concentrations.
6. It is important to prevent the DNA pellet from becoming overly dry and to make sure that the DNA solution does not aggregate. DNA aggregation can significantly affect the result of the DNA gradient fractionation. To minimize this problem, DNA suspensions are pooled together and incubated at 65°C for 5 min to dissociate any DNA aggregation.
7. For beginners, pouring sucrose gradient can be a challenge. Therefore, it is better to practice pouring gradient containing dye first to ensure that the gradient is properly formed. To do this, add 50–100 μ L of 10X loading dye to a 40% sucrose solution (outer chamber only). Follow the instructions given in **Subheading 3.3**. As the gradient forms, it should be obvious that the blue loading dye is gradually fading, leaving the bottom layer with a darker blue and the upper layer with a lighter blue.
8. High-quality chromosomal DNA can be prepared following the sucrose density gradient protocol used in this method, which will exclude the possibility of contamination with different-sized DNA fractions.
9. During dialysis, TE buffer should be changed every 3 to 4 h to ensure complete removal of sucrose. This will be confirmed by expansion of the DNA solution by two- to threefold of the original volume.
10. For a more efficient phage packaging, DNA and vector ligation should be carried out at a DNA concentration of 0.2 μ g/ μ L or greater, in order to favor formation of concatemers and not circular DNA molecule (2). The ligation reaction is set up in three different molar ratios (1:1, 1:2, and 1:3) of vector to purified 37- to 52-kb Sau3A-digested *T. denticola* DNA.
11. It is comparatively easy to store libraries in the relatively stable λ phage head using this protocol.

12. So far, we have not obtained any clone from the library that does not have DNA insert. The chance of getting blank clones using this protocol is less likely. This is one of the major advantages of using the in vitro packaging protocol; by contrast, transfection or transformation methods may result in up to 10% empty clones (unpublished data from our laboratory).
13. The colonies analyzed were with insert DNA fragment lengths between 35 and 50 kb. Approximately 60% of all colonies screened contained DNA fragments of insert sizes in the range of 35–38 kb, 30% contained insert sizes in the range of 39–40 kb, and 10% contained insert sizes in the range of 41–42 kb (**Fig. 2**). Average insert size was 40.9 kb. Use of cosmids ensures that the pieces of DNA cloned into a vector will all be approximately of the same size.

References

1. Zhu, H. and Dean, R. A. (1999) A novel method for increasing the transformation efficiency of *Escherichia coli*—application for bacterial artificial chromosome library construction. *Nucleic Acid Res.* **27**, 910–911.
2. Champness, W. and Snyder, L. (1997) *Molecular Genetics of Bacteria*. American Society of Microbiology.
3. Weinstock, G. M., Maurer, R., and Berget, P. B. (1990) *Advanced Bacterial Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
4. Limberger, J. R., Silvenski, L. L., Izard, J., and Samsonoff, A. W. (1999) Insertional inactivation of *Treponema denticola* tap1 results in a nonmotile mutant with elongated flagellar hooks. *J. Bacteriol.* **181**(12), 3743–3750.
5. Murray, M. G. and Thompson, W. F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acid Res.* **8**, 4321–4325.
6. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
7. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Straw, K. (1989) *Current Protocols in Molecular Biology*. John Wiley, New York.
8. Murray, E. M., Singh, K. V., Ross, R. P., Heath, J. D., Dunny, G. M., and Weinstock, G. M. (1993) Generation of restriction map of *Enterococcus faecalis* OG1 and investigation of growth requirements and regions encoding biosynthetic function. *J. Bacteriol.* **175**(16), 5216–5222.
9. Teng, F., Murray, E. B., and Weinstock, G. M. (1998) Conjugational transfer of plasmid DNA from *E. coli* to *Enterococcus*: A method to make insertional mutation. *Plasmid* **39**, 182–186.

Exploring Transformation-Associated Recombination Cloning for Selective Isolation of Genomic Regions

Natalay Kouprina, Vladimir N. Noskov, Maxim Koriabine,
Sun-Hee Leem, and Vladimir Larionov

1. Introduction

Mammalian genome analysis has been advanced considerably by the development of yeast artificial chromosome (YAC) and bacterial artificial chromosome (BAC) cloning systems (1,2). These techniques have made the isolation of large, random DNA fragments possible, thereby greatly simplifying the physical mapping of chromosomes. However, isolation of entire genes and specific chromosomal regions for functional studies remained a laborious process because it relies on characterization of random clones in libraries comprising YACs or BACs. Even if a YAC or BAC contains an entire gene, the specific recovery of the gene was an arduous process. Often a gene is available as a contiguous set of fragments that must be pieced together.

To overcome this problem, a novel cloning system, transformation-associated recombination (TAR), has been developed that allows the specific isolation of genes and specific chromosomal regions directly from total genomic DNA (3–5). This system is based on the pioneering work of Ma et al. (6), who demonstrated that a double-strand DNA break in a vector can be repaired by cotransformation with a linear DNA fragment containing DNA sequence that flanks the double-strand DNA break. TAR cloning draws on several features of the yeast *Saccharomyces cerevisiae*. First, during transformation, yeast can take up several small and large DNA molecules. Second, intermolecular recombination between homologous DNAs is highly efficient during transformation.

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

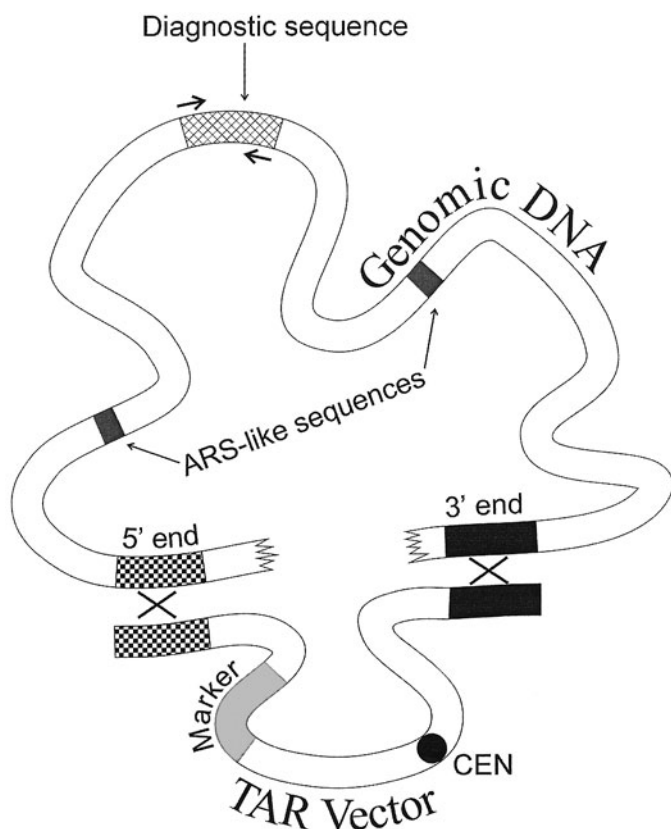


Fig. 1. Isolation of single-copy gene using TAR cloning. The DNA fragments involved when yeast spheroplasts are transformed are shown. The TAR cloning vector contains sequences from the 5' promoter region (checkered box) and from the 3' end (solid box) of a gene of interest (hooks). Hooks are cloned into the vector in such a way that after linearization of the vector, orientation of the hooks corresponds to their orientation in genomic DNA as shown. Recombination between sequences in the vector and the locus in genomic DNA creates a circular YAC. *CEN* corresponds to the yeast chromosome VI centromere and Marker is a marker for selection in yeast. The arrows indicate the positions of primers for an internal region that can be used for initial identification of YAC clones containing the gene of interest. (Adapted from **ref. 29**).

Third, genomic DNA of all mammals contains multiple copies of yeast-like *ARS* sequences that can function as origins of replication in yeast (7).

Two general schemes for the isolation of a specific gene as a circular YAC by TAR are shown in **Figs. 1** and **2**. When both 3'- and 5'-end sequence infor-

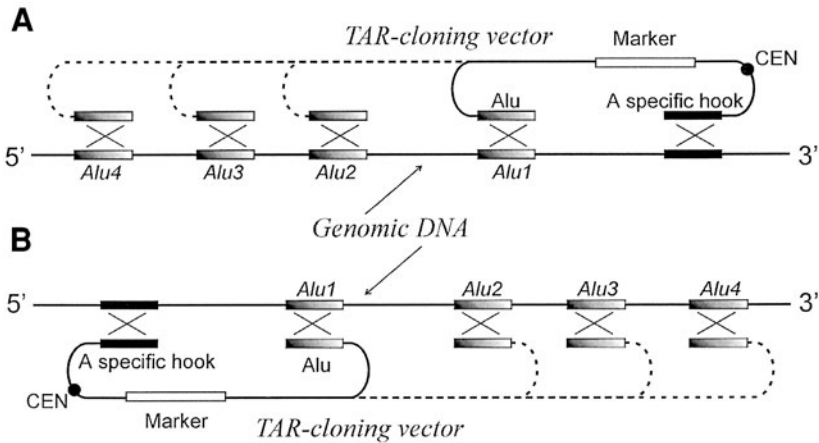


Fig. 2. Isolation of a single-copy gene by radial TAR cloning. The DNA fragments involved when yeast spheroplasts are transformed are shown. The TAR vector includes a gene-specific sequence (a specific hook) and a repeat sequence (i.e., *Alu* for cloning human DNA) at either end of the linearized vector. Two different relative orientations of the specific hook are possible, as shown. Recombination between the vector and genomic DNA can create circular YACs that extend from the unique sequence to upstream *Alu* sequences (A) or to downstream *Alu* sequences (B). Thus, a set of 3' and 5' YACs is generated. *CEN* corresponds to a yeast centromere and Marker is a marker for selection in yeast. (Adapted from ref. 9).

mation is available, a gene can be isolated by TAR using a vector containing two short, unique sequences flanking the gene (Fig. 1). These sequences (hooks; see Notes 1–3) are cloned in the vector in such a way that linearization of the vector releases the gene-targeting sequences. The size of the hooks can be as small as 60 bp (8). The vector also contains a yeast centromere (*CEN*), which is required for proper YAC segregation, and a yeast-selectable marker. TAR cloning is based on copenetration into yeast spheroplasts of gently isolated genomic DNA along with vector DNA, followed by recombination between the vector and the human DNA to establish a circular YAC. Propagation of the YAC can occur if the human DNA contains sequences that can function as an origin of replication (*ARS*) in yeast. These sequences are common in mammalian DNA with approximately one *ARS*-like sequence/40 kb ([7]; unpublished data), suggesting that most human genes can be isolated by TAR cloning using vectors with two specific hooks. TAR cloning with two specific hooks is highly selective and produces libraries in which nearly 1.0% of the transformants contain the desired gene (9,10). A clone containing a gene of interest can be easily identified in the libraries by polymerase chain reaction (PCR).

The use of TAR vectors with two specific targeting sequences has limitations. First, often only one sequence of information is available, such as a 3'-flanking expressed sequence tag. Second, some chromosomal regions may be unclonable because they lack yeast *ARS*-like sequences. Taking into account these limitations, a modified version of TAR cloning has been developed. This approach uses a vector that has one unique sequence hook and one repeated sequence hook (*Alu* or *B1* repeats for human or mouse DNA, respectively). The repeated element makes it possible to isolate a set of nested overlapping fragments that extend from the specific hook to different upstream or downstream *Alu* positions. Such an approach increases the likelihood that a genome region (a gene) will be isolated because it should include an *ARS*-like sequence. Because only one of the ends is fixed, this approach is called radial TAR cloning. It is important to note that by simply changing the arrangement of the same targeting hook, it is possible to clone in both directions along the chromosome from this specific targeting sequence (see **Fig. 2A,B**). The enrichment with gene-positive clones for radial TAR cloning is comparable with that observed for TAR cloning with vectors containing two gene-specific sequences. YAC isolates vary in size from 50 up to 600 kb (see **Note 4**).

Circular YACs have two main advantages compared with their linear counterparts: (1) they can be isolated as covalently closed circular DNA molecules directly from yeast cells (**11,12**); and (2) alternatively, they can be modified by homologous recombination into BACs and transferred into *Escherichia coli* cells for further DNA isolation (**13,14**).

During recent years, the TAR cloning method was successfully applied to the isolation of several single-copy genes from human and mouse genomes (**9,14–18**). Among other utilities of TAR cloning are the isolation of human DNA from radiation hybrids (**5,19,20**), closing of the gaps on physical maps (**21**), cloning of translocations, separation of haplotypes, and verification of genomic contigs. For some genomic regions that are toxic for *E. coli* cells (summarized in **ref. 22**), TAR cloning may be the only method to isolate clones as YACs. Based on our recent estimates, approx 5% of euchromatic regions cannot be cloned or stably propagated in *E. coli* cells (**9**).

This chapter presents four protocols. The first describes preparation of chromosome-size genomic DNA in solid agarose plugs for gene capture. The second details preparation of highly competent yeast spheroplasts and transformation of the spheroplasts by genomic DNA along with a TAR vector. The third describes identification of positive clones among primary yeast transformants using PCR. And, finally, the fourth provides a method for retrofitting of TAR-isolated YACs into BACs with a mammalian-selectable marker and transferring the YACs/BACs into *E. coli* cells.

2. Materials

2.1. Strains and Vectors

1. VL6-48N (*MAT alpha*, *his3Δ200*, *trp1Δ1*, *ura3Δ1*, *lys2*, *ade2-101*, *met14*), a highly transformable *S. cerevisiae* strain that has *HIS3*, *TRP1*, and *URA3* deleted, for use as a host for TAR cloning experiments. This strain is available by request from the Laboratory of Biosystems and Cancer, National Cancer Institute (National Institutes of Health [NIH]) (see **Note 5**).
2. Basic TAR cloning vector pVC604 (see **Fig. 3**) containing a yeast-selectable marker (*HIS3*) and a yeast centromeric sequence (*CEN6*), also available by request from the Laboratory of Biosystems and Cancer, National Cancer Institute (NIH). Before use, the vector is “activated” by insertion of gene-specific hooks into the polylinker.
3. Set of vectors for retrofitting circular YACs into BACs with different selectable mammalian markers, developed in the Laboratory of Biosystems and Cancer, National Cancer Institute (NIH). Physical maps of these vectors are presented in **Fig. 4**. **Figure 5** shows a schematic representation of retrofitting of a circular YAC into a BAC by homologous recombination in yeast.

2.2. Preparation of Chromosome-Size DNA Plugs for TAR Cloning in Solid Agarose Plugs

1. Mammalian cells or disaggregated tissue.
2. Low gelling/melting temperature agarose: 1% agarose gel prepared in 0.125 M EDTA, pH 7.5.
3. EDTA mix: 0.05 M EDTA, 0.01 M Tris-HCl, pH 7.5.
4. LET solution: 0.5 M EDTA, 0.01 M Tris-HCl, pH 7.5.
5. NDS cell lysis buffer: 0.39 M EDTA, 0.01 M Tris-HCl, pH 7.5, 1% *N*-lauroyl sarcosine; 2 mg/mL of proteinase K.
6. Phosphate-buffered saline (PBS) solution: 137 mM NaCl, 2.7 mM KCl, 8.0 mM Na₂HPO₄•7H₂O, 1.5 mM KH₂PO₄, pH 7.3.
7. Dimethylsulfoxide (DMSO) (J.T. Baker).

2.3. Preparation of Highly Competent Yeast Spheroplasts and Transformation of the Spheroplasts by Genomic DNA Along With a TAR Vector

1. 1 M sorbitol.
2. SPE solution: 1 M sorbitol; 10 mM Na₂EDTA, 0.01 M Na phosphate, pH 7.5.
3. SOS solution: 1 M sorbitol, 6.5 mM CaCl₂, 0.25% yeast extract (Difco, Detroit, MI), 0.5% Bactopeptone (Difco).
4. STC solution: 1 M sorbitol; 10 mM CaCl₂, 10 mM Tris-HCl, pH 7.5.
5. Zymolyase solution: 10 mg/mL of zymolyase 20T (ICN) in 20% glycerol (kept as frozen aliquots).
6. PEG 8000 solution: 20% (w/v) polyethylene glycol (PEG) 8000 (Sigma, St. Louis, MO); 10 mM CaCl₂, 10 mM Tris-HCl, pH 7.5.

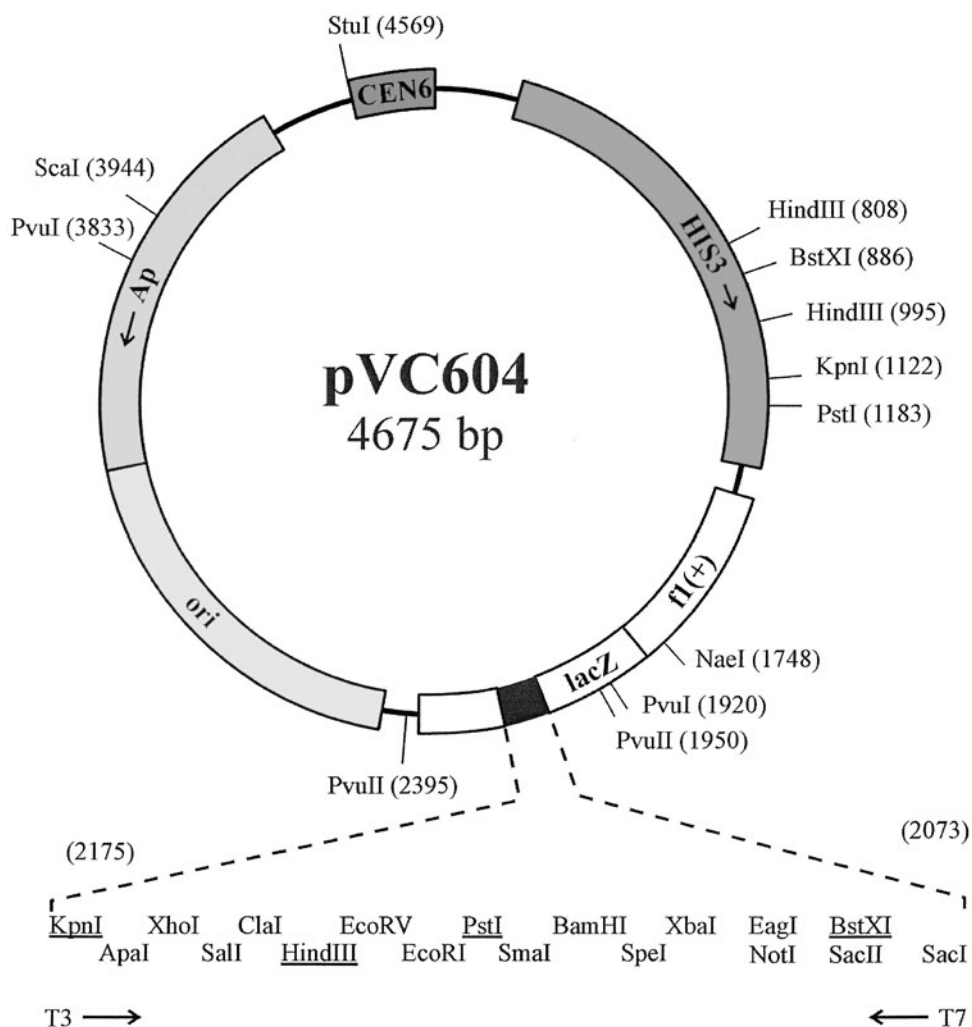


Fig. 3. Scheme of basic plasmid pVC604 for construction of TAR cloning vectors. pVC604 plasmid is a derivative of the Bluescript-based yeast-*E. coli* shuttle vector pRS313 (28). This plasmid was generated by deletion of a 295-bp fragment containing a yeast origin of replication (*ARSH4*) from pRS313. pVC604 has an extensive polylinker consisting of 14 restriction endonuclease 6- and 8-bp recognition sites for flexibility in cloning of particular fragments of interest. The functional DNA segments of the plasmid are indicated as follows: *CEN6* = a 196-bp fragment of the yeast centromere VI; *HIS3* = marker for yeast cells; *Ap* = ampicillin resistance gene. Construction of a TAR vector includes cloning of short, specific sequences (hooks) that flank a gene of interest into the plasmid. For TAR cloning experiments, a vector DNA is linearized by an endonuclease digestion to expose targeting sequences.

7. TE buffer: 0.1 mM EDTA; 0.01 M Tris-HCl, pH 7.5.
8. YPD medium: 2% D-glucose, 2% Bacto peptone (Difco), 1% yeast extract (Difco).
9. TOP agar-His: 1 M sorbitol, 2% D-glucose, 0.17% Yeast Nitrogen Base (Difco), 0.5% (NH₄)₂SO₄, 3% Bacto agar (Difco) containing the following supplements: 0.006% adenine sulfate, 0.006% uracil, 0.005% L-arginine•HCl, 0.008% L-aspartic acid, 0.01% L-glutamic acid, 0.005% L-isoleucine, 0.01% L-leucine, 0.012% L-lysine•HCl, 0.002% L-methionine, 0.005% L-phenylalanine, 0.0375% L-serine, 0.01% L-threonine, 0.005% L-tryptophan, 0.005% L-tyrosine, 0.015% L-valine.
10. SORB-His plates: 1 M sorbitol, 2% D-glucose, 0.17% Yeast Nitrogen Base (Difco), 0.5% (NH₄)₂SO₄, 2% Bacto agar (Difco) supplemented as described in **item 9**.
11. SD-His plates: 2% D-glucose, 0.17% Yeast Nitrogen Base (Difco), 0.5% (NH₄)₂SO₄, 2% Bacto agar (Difco) supplemented as described in **item 9**.
12. β-agarase (New England Biolabs) (store at -20°C).

2.4. Isolation of Total DNA From Pool of Yeast Transformants for PCR Analysis

1. 1 M sorbitol.
2. SP solution: 1.2 M sorbitol, 0.1 M Na phosphate, pH 7.5.
3. SPE solution: 1 M sorbitol, 10 mM Na₂EDTA, 0.01 M Na phosphate, pH 7.5.
4. Zymolyase solution: 10 mg/mL of zymolyase 20T (ICN) in 20% glycerol (kept as frozen aliquots).
5. 5 M potassium acetate (KAc).
6. Diethylpyrocarbonate (DEPC) (Sigma).
7. TE buffer: 0.1 mM EDTA, 0.01 M Tris-HCl, pH 7.5.
8. 100 and 70% Ethanol.
9. Appropriate PCR primers.

2.5. Retrofitting of Circular YACs into BACs With Mammalian-Selectable Marker and Transferring of YACs/BACs into E. coli Cells

1. BRV-based vector linearized at *Bam*HI and *Aat*II sites.
2. DH10B *E. coli* competent cells (Gibco-BRL, Gaithersburg, MD).
3. Lithium acetate solution: 100 mM lithium acetate; 10 mM Tris-HCl; 0.1 mM EDTA, pH 7.5.
4. SD-His synthetic liquid medium: 2% D-glucose, 0.17% Yeast Nitrogen Base (Difco), 0.5% (NH₄)₂SO₄, 0.006% adenine sulfate, 0.006% uracil, 0.005% L-arginine•HCl, 0.008% L-aspartic acid, 0.01% L-glutamic acid, 0.005% L-isoleucine, 0.01% L-leucine, 0.012% L-lysine•HCl, 0.002% L-methionine, 0.005% L-phenylalanine, 0.0375% L-serine, 0.01% L-threonine, 0.005% L-tryptophan, 0.005% L-tyrosine, 0.015% L-valine.
5. PEG 4000 solution: 40% (w/v) PEG 4000 (Fluka) aqueous solution.
6. SOC solution: 2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM D-glucose.

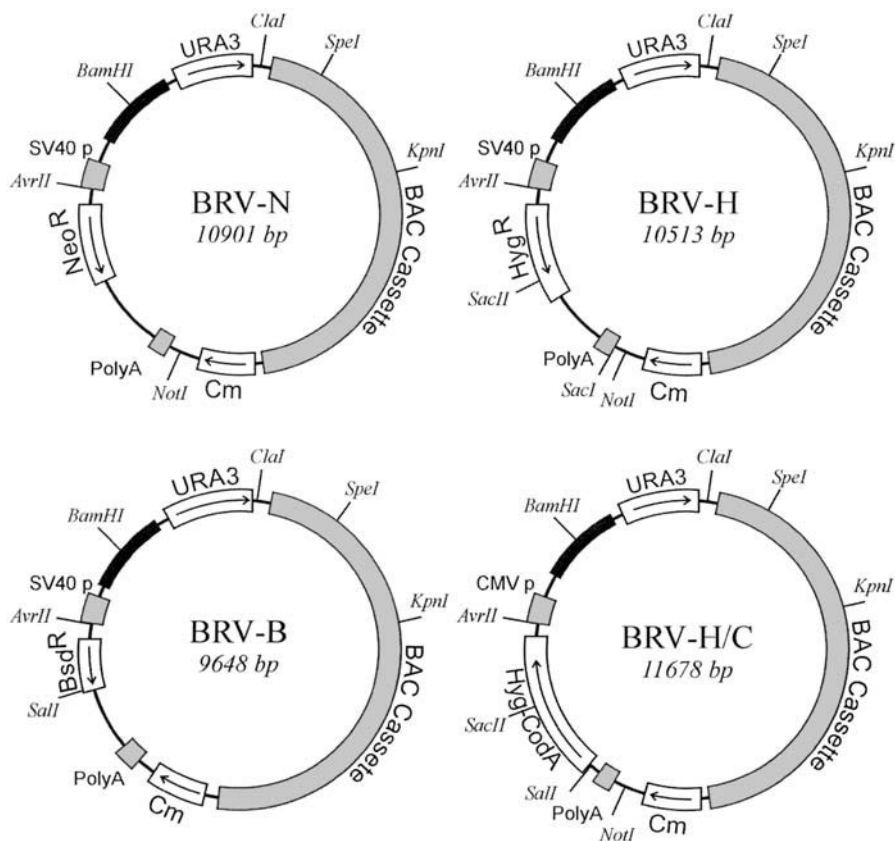


Fig. 4. Schemes of BRV vectors for retrofitting of a circular YAC into a BAC with a mammalian marker. Retrofitting vectors BRV-N, BRV-H, BRV-H/C, and BRV-B were constructed using a pBeloBAC11 BAC cloning vector (29); therefore, they contain F factor origin of replication and chloramphenicol acetyltransferase gene (*Cm*) providing resistance to chloramphenicol. The size of the BAC cassette in retrofitting vectors is 6.3 kb. Each retrofitting vector contains two short (approx 300 bp each) targeting sequences, A and B, flanking the *ColE1* origin of replication and the *Ap* gene in pVC604-based TAR cloning vectors (Fig. 3). These targeting sequences are separated by a unique *Bam*HI site. Each vector also contains the *URA3* yeast-selectable marker. The vectors differ by a mammalian-selectable marker. BRV-N retrofitting vector contains a *Neo*^R-selectable marker that provides resistance to G-418 Geneticin (neomycin). The *Neo*^R gene was PCR amplified as a 2.7-kb fragment from the pSV2neo vector (30). Expression of *Neo*^R is controlled by SV40 promoter and terminator. Construction and application of BRV-N (BRV-1) retrofitting vector was previously described (14). BRV-H retrofitting vector contains a Hygromycin-resistance cassette in which *Hyg*^R is expressed under the control of SV40 promoter and LTR Mo-MuL V terminator. This cassette was PCR amplified as a 2.3-kb fragment from the pLXSH vector (31). BRV-H/C retrofitting vector contains *Hyg*^R gene fused in frame with Cytidine deaminase gene (*CodA*), thus providing both positive selection for Hygromycin resistance and

7. LET solution: 0.5 M EDTA, 0.01 M Tris-HCl, pH 7.5.
8. Carrier salmon DNA: 10 mg/mL of sonicated salmon sperm DNA (Stratagene) denaturated by boiling for 10 min.
9. SD-Ura plates: 2% D-glucose, 0.17% Yeast Nitrogen Base (Difco), 0.5% $(\text{NH}_4)_2\text{SO}_4$, 0.006% adenine sulfate, 0.005% L-arginine•HCl, 0.008% L-aspartic acid, 0.01% L-glutamic acid, 0.004% L-histidine•HCl, 0.005% L-isoleucine, 0.01% L-leucine, 0.012% L-lysine•HCl, 0.002% L-methionine, 0.005% L-phenylalanine, 0.0375% L-serine, 0.01% L-threonine, 0.005% L-tryptophan, 0.005% L-tyrosine, 0.015% L-valine, 2% Bacto agar (Difco).
10. Standard LB plates supplemented with 12.5 µg/mL of chloramphenicol (Cm plates).

3. Methods

3.1. Preparation of Chromosome-Size Genomic DNA in Solid Agarose Plugs for TAR Cloning

This protocol describes the preparation of high molecular weight genomic DNA for TAR cloning. The procedure is similar to that described for analysis of chromosome-size DNA for pulsed-field gel electrophoresis (PFGE). Living cells are embedded in low-melting-point agarose plugs and lysed with detergent in the presence of proteinase K. Unwanted cellular components are removed by dialysis. The resulting agarose plugs are stable for several months when stored at 5°C in EDTA mix. This procedure will generate suitable DNA samples from all mammalian lines and can be adapted for cells from disaggregated tissue (*see Note 6*).

1. Melt an appropriate quantity of 1.0% low gelling/melting temperature agarose by placing a flask containing the agarose in a beaker half filled with water and heating just to boiling point. Quickly place in a 50°C water bath to cool. After 5 min, mix well using a magnetic stirrer and keep in a 50°C water bath before use. It is important not to lose a water volume from the agarose during melting.
2. Harvest the cells by centrifuging for 8 min at 150g at room temperature. Remove and discard the supernatant.
3. Resuspend the cell pellet in 10 mL of PBS and mix to produce a single cell suspension. Count the cells using a hemocytometer.

Fig. 4. (*continued*) negative selection in the presence of 5-Fluor Cytidin (5-FIC). Hyg^R-CodA fusion open reading frame was PCR amplified as a 2.8-Kb fragment from the pCMVHYGCODA vector (32). Its expression is controlled by cytomegalovirus (CMV) promoter and SV40 terminator. BRV-B retrofitting vector contains a Bsd^R gene providing resistance to a Blasticidin S, expressed under the control of SV40 and CMV promoters, respectively, and SV40 terminator. Bsd^R gene was PCR amplified as a 0.8-kb fragment from pcDNA6/V5-His vector (Invitrogen, Carlsbad, CA).

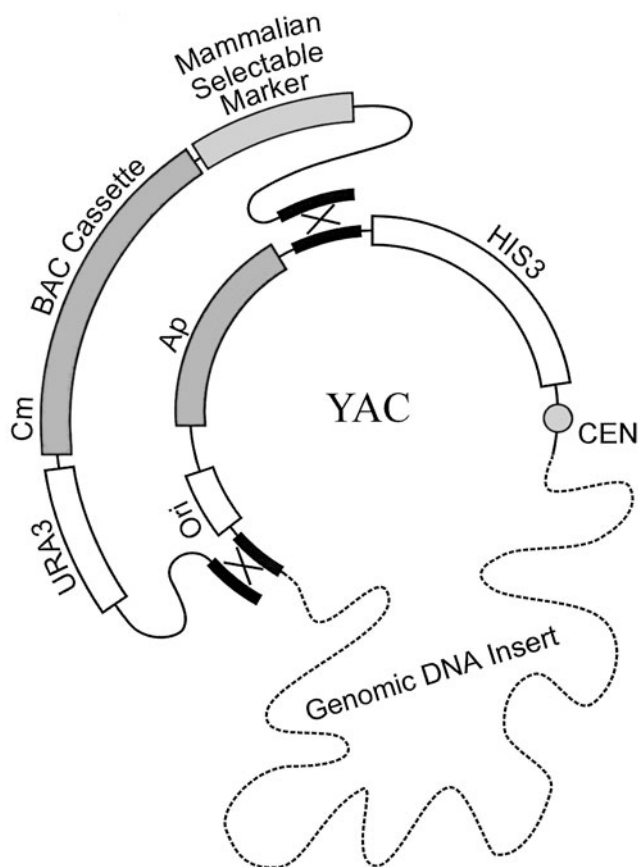


Fig. 5. Schematic representation of retrofitting a circular YAC into a BAC using BRV vectors. A linearized retrofitting vector is transformed into yeast cells carrying a circular YAC. Recombination between targeting sequences A and B in the vector and homologous regions in a YAC replaces the *ColE1* origin of replication and the *Ap* gene in the YAC by a cassette containing the *F*-factor origin of replication, the chloramphenicol acetyltransferase (*Cm*) gene, the *URA3* yeast-selectable marker, and a mammalian-selectable marker. (Adapted from ref. 14).

4. Wash the cells one more time by centrifuging with PBS. Calculate the final volume of suspension required to yield a cell concentration of 2×10^7 cells/mL. Resuspend the washed cells in PBS, measure the actual volume of suspension by carefully pipetting into a separate tube, and add PBS until the calculated volume is reached. The final cell concentration in agarose plugs will be 10^7 cells/mL, yielding a DNA concentration of approx 100 $\mu\text{g/mL}$ or approx 5 $\mu\text{g}/50\text{-}\mu\text{L}$ agarose plug.

5. Transfer the melted agarose from a 50°C water bath to a 42°C water bath (equilibration is not critical).
6. Place the tube with the cells in a 42°C water bath, add an equal volume of melted agarose (from the 42°C water bath), and mix well by vortexing. Keep the cell/agarose suspension at 42°C. It is important that the final concentration of agarose be equal to 0.5%. With a higher concentration of agarose, it is impossible to completely melt the plugs before TAR cloning.
7. Take approx 50-μL aliquots of the cell/agarose suspension and gently place into each Ultra Micro tip. Keep the tips horizontal for 10 min at 5°C until the agarose is completely solidified.
8. Transfer all the agarose plugs into a 50-mL polypropylene Corning tube. To do this, take up ~0.5 mL of NDS solution in a 6-cc syringe without a needle, place the tip of the Ultra Micro tip into the syringe lure, and gently apply pressure. The plug should slide out into the tube.
9. Adjust a volume of suspension in the tube up to 45 mL with NDS solution and incubate the plugs for 48 h at 50°C. Twenty milliliters of NDS is sufficient for agarose plugs prepared from 2×10^7 cells. When the cells are lysed, they will no longer be visible as light-scattering specks in the agarose.
10. Carefully remove the NDS solution and add 40 mL of EDTA mix into the tube. Incubate the plugs for 1 h at 50°C, and then for 20 min at room temperature, and carefully remove the solution.
11. Dialyze the agarose plugs 20 times against 40 mL of EDTA mix for 1 h each at room temperature and gently invert the tube several times. The plugs do not settle rapidly out of the EDTA mix. The solution should be removed by a 5.0-mL disposable pipet from the bottom.
12. Store the dialyzed plugs at 5°C in the EDTA mix. Before use, dialyze the plugs against 25 mM NaCl to completely remove the EDTA.
13. One day before the TAR cloning experiment, transfer each of 10 agarose plugs with genomic DNA into a 6-mL Falcon tube and dialyze against 1 mL of 25 mM NaCl solution containing 1% DMSO for 60 min at room temperature. Remove the DMSO by washing the plugs twice with 1 mL of 25 mM NaCl for 60 min at room temperature. Keep the plugs in 1 mL of 25 mM NaCl at room temperature overnight before use for yeast spheroplast transformation.

3.2. Preparation of Highly Competent Yeast Spheroplasts and Transformation of the Spheroplasts by Genomic DNA Along With a TAR Vector

1. One day before the TAR cloning experiment, inoculate 100-mL aliquots of YPD medium in a 500-mL Erlenmeyer flask with one, two, and three individual colonies of the host yeast strain VL6-48N from a YPD plate, and grow the cultures overnight at 30°C with a vigorous shaking to ensure good aeration.
2. In the morning, measure optical densities (ODs) of the cultures at 1- to 2-h intervals until an OD₆₆₀ of approx 2–4 is achieved in one of the flasks. (The actual

measurement is 0.14 after diluting 1/10 in water.) The culture with such an OD is ready for the preparation of highly competent spheroplasts. This OD corresponds to approx 2×10^7 cells/mL. (Cell density can be determined directly using a hemocytometer.)

3. Transfer the yeast culture into two 50-mL Falcon conical tubes, and pellet the cells by centrifuging 5 min at 1000g and 5°C. Remove and discard the supernatant.
4. Resuspend each cell pellet in 30 mL of sterile water by vortexing, and centrifuge for 5 min at 3000g and 5°C. Remove and discard the supernatant.
5. Resuspend each cell pellet in 20 mL of 1 M sorbitol by vortexing, and centrifuge for 5 min at 3000g and 5°C. Remove and discard the supernatant.
6. Resuspend each cell pellet in 20 mL of SPE solution. Add to a tube 20 μ L of 10 mg/mL zymolyase 20T and 40 μ L of 14 M β -mercaptoethanol, mix well, and incubate at 30°C for approx 20 min with slow shaking.
7. Check the level of spheroplasting by comparing the ODs of the cell suspension in 1 M sorbitol vs 2% sodium dodecyl sulfate (SDS). The spheroplasts are determined to be ready when the difference between the two OD₆₆₀ readings is three- to fivefold. (The OD reading for the sorbitol sample should not drop lower than 1.5 times compared with the reading before zymolyase treatment. If this value is lower, the spheroplasts have been overexposed to enzyme and cannot be used). To measure the OD₆₆₀ difference, aliquots of the zymolyase-treated cell suspension are diluted 10-fold by 1 M sorbitol or 2% SDS. Note that the treatment time varies depending on the zymolyase stock. With a new stock, the OD₆₆₀ readings should be done every 5 min. Both under- and overexposure to zymolyase greatly affects transformation efficiency. From this point on, extreme care must be taken to avoid lysing the delicate spheroplasts; very slow, gentle resuspensions are necessary.
8. Centrifuge the spheroplasts for 10 min at 570g and 5°C. Decant the supernatant, add 20 mL of 1.0 M sorbitol, and then rock very gently to resuspend the pellet. Pellet the spheroplasts again by centrifuging for 10 min at 300–600g and 5°C.
9. Repeat the wash with 1 M sorbitol two more times, and gently resuspend the final pellets in 2.0 mL of STC solution. The spheroplasts are stable at room temperature for at least 1 h.
10. Transfer each of the 10 agarose plugs with genomic DNA (conditions for preparation of the plugs for TAR cloning are described in the previous section) to a 6-mL Falcon tube, remove all the dialysis solution, and add 1 μ g of the linearized TAR vector in a volume of 1–5 μ L to each tube. Place the tubes in a 70°C temp-block, and incubate for 5 min until the agarose is melted completely. Failure of the agarose to melt under these conditions would indicate that the agarose plugs are not appropriately prepared (either the concentration of agarose is >0.5% or the plugs contain >5 μ g of genomic DNA). Although the plugs with an increased concentration of agarose cannot be used, the plugs with two to three higher concentrations of genomic DNA can be melted if an equal volume of 25 mM NaCl is added to the plugs before melting at 70°C.
11. Transfer the tubes to a 42°C tempblock for 10 min.

12. Add 1 U of β -agarase to each tube and mix gently by stirring with finger flicking. Incubate for 10–20 min at 42°C until the agarose is completely digested. Test for completion by placing the tube for 5 min on ice and examining for solid agarose. If solid agarose remains, remelt at 70°C, cool to 42°C, add an additional 1 U of β -agarase, and incubate for 10–20 min more at 42°C.
13. Add 450 μ L of spheroplast suspension to each DNA mixture, mix gently, and incubate for 10 min at room temperature. Using a cut 1-mL (blue) tip, transfer the spheroplasts into 15-mL Falcon tubes.
14. Add 4.5 mL of PEG solution, gently mix by inverting the tubes, and incubate for 10 min at room temperature.
15. Pellet the spheroplasts by centrifuging for 10 min at 300–500g and 5°C. Remove the supernatant, and gently resuspend the spheroplasts with a pipet tip in 1.0 mL of SOS solution.
16. Incubate the spheroplasts for 40 min at 30°C without shaking.
17. Transfer the spheroplasts into a 35-mL borosilicate glass test tube with a plain end containing 8.0 mL of melted TOP agar (equilibrated at 50°C), gently mix, and quickly pour agar onto a SORB plate with selective medium containing 1 M sorbitol.
18. Keep the plates at 30°C for 4–7 d until all the transformants become visible.

For the transformation conditions described (i.e., with 1 μ g of a vector, 5 μ g of genomic DNA, and approx 5×10^8 spheroplasts), the yield of transformants varies from 10 to 300 colonies per plate depending on the hooks used. Typically, the higher yield of transformants is observed with the hooks containing nonunique sequences. Most of these transformants result from recombination between the vector and genomic DNA.

3.3. Identification of Positive Clones by PCR

Typically, one among 100–300 primary His⁺ transformant colonies contains a gene of interest (*see* **Notes 6 and 7**). To identify positive colonies, primary transformants are combined into pools and examined for the presence of the gene by PCR using a pair of primers specific for its internal sequence (*see* **Figs. 1 and 2**). Individual clones from each positive pool are screened by a second round of PCR. While in reconstruction experiments, a gene-positive clone can be detected in a pool containing 1000 transformants, we recommend using pools containing not more than 30 transformants if DNAs are isolated by the fast protocol described next.

1. Transfer 1500–2000 primary transformants by toothpicks on SD-His plates with a synthetic medium lacking histidine. Individually streak 30 colonies onto each master plate.
2. Incubate the plates with pools of transformants at 30°C overnight and replica plate on new plates with the SD-His selective medium. Master plates should be sealed

with parafilm® and kept at 5°C; replica plates are used for detection of positive pools by PCR.

3. Wash the yeast cells from the replica plates containing 30–40 His⁺ transformants with 5 mL of water into 12-mL Falcon conical tubes, and pellet the cells by centrifuging for 5 min at 1000g and 5°C. Remove and discard the supernatant.
4. Resuspend each cell pellet in 1 mL of 1 M sorbitol by vortex, transfer the suspension to a 1.5-mL Eppendorf microfuge tube, and spin for 30 s. Remove and discard the supernatant.
5. Resuspend the cells in 0.5 mL of SPE solution containing 14 mM β -mercapto-ethanol, add into each tube 20 μ L of zymolyase 20T (10 mg/mL), and incubate for 2 h at 30°C.
6. Harvest the spheroplasts by centrifuging for 5 min at 1000g on the Eppendorf microfuge, and resuspend the pellets in 0.5 mL of 50 mM EDTA solution containing 0.2% SDS.
7. Add 1 μ L of DEPC at room temperature and vortex well.
8. Completely lyse the spheroplasts by incubating at 70°C for 15 min.
9. Add 50 μ L of 5 M KAc to the lysate and let the tubes sit on ice for 30 min. This step precipitates proteins and dodecyl sulfate as a potassium salt.
10. Pellet the precipitate by centrifuging for 15 min at maximum minifuge speed (2500g).
11. Transfer the supernatant to fresh microfuge tubes, fill the tubes with room temperature ethanol, mix, and pellet the DNA by centrifuging for 5 min. Remove the supernatant as much as possible, and dry the tubes by inverting on blotting paper.
12. Resuspend each damp DNA pellet in 0.4 mL of TE buffer, leave the tubes at room temperature for 30 min, and then vortex until the DNA is dissolved. (Samples can be incubated at 4°C overnight. In this case, the DNA dissolves more thoroughly.)
13. Remove all nondissolved material by centrifuging for 1 min, transfer the supernatant to a new tube, and add 1 mL of isopropanol. Mix well and immediately pellet the DNA precipitate by centrifuging for 5 min at room temperature.
14. Remove the supernatant as much as possible and dry the tubes well.
15. Wash the DNA pellet with 1 mL of 70% ethanol and dry at room temperature.
16. Dissolve the final pellet of DNA in 0.3 mL of water.
17. Use 1 μ L of the DNA solution in a 50- μ L PCR reaction to identify positive pools.
18. Screen individual clones from each positive pool by a second round of PCR to identify colonies containing a gene of interest. Conditions for analysis of a large number of individual yeast clones using lysed spheroplasts as template for PCR reaction are described by Ling et al. (23). Touch a streak of each transformant from a master plate with “a positive pool” with a sterile disposable pipet tip, and then thoroughly rinse the tip with 10 μ L of the SP solution containing 2.5 mg/mL of zymolyase 20T by pipeting the solution up and down three to five times.
19. Incubate the resulting suspension for 5 min at 37°C.
20. Use 1–5 μ L of the suspension for each 100- μ L PCR reaction to identify positive clones. The remaining samples can be stored at –20°C for repeated use. The final

concentration of Mg^{2+} should be increased up to 2.5 mM when lysed spheroplasts or preparations of yeast DNA are used as templates for PCR reactions.

3.4. Retrofitting of Circular YACs into BACs With Mammalian-Selectable Marker and Transferring of YACs/BACs into *E. coli* Cells

Although circular YACs generated by TAR cloning can be separated from host strain linear chromosomes by PFGE (5) or alkaline extraction (12), there is still no procedure for the isolation of quantitative amounts of YAC DNA from yeast cells for physical analysis and for transfection of the cloned material into mammalian cells. This protocol describes an efficient and accurate procedure for retrofitting YACs into BACs with different selectable markers using a set of yeast-bacteria-mammalian shuttle BRV vectors (Fig. 4) (see Notes 8 and 9). The retrofitted YACs can be moved to *E. coli* by electroporation for a standard large circular DNA isolation.

1. Inoculate 5 mL of SD-His synthetic medium without histidine with one individual colony containing a YAC, and grow overnight at 30°C with vigorous shaking to ensure good aeration.
2. Transfer the yeast culture into 50 mL of YPD medium, and grow for an additional 4 to 5 h at 30°C with vigorous shaking.
3. Pellet 5 mL of the culture by centrifuging for 5 min at 1000g and 5°C in a 12-mL Falcon conical tube. Remove and discard the supernatant.
4. Resuspend the cell pellet in 1 mL of sterile water by vortex, transfer into an Eppendorf tube, and pellet the cells by centrifuging for 1 min at maximum speed. Remove and discard the supernatant.
5. Resuspend the cells in 1 mL of 0.1 M LiAc solution. Incubate at 30°C for 1 h with slow shaking. Alternatively, cells can be stored at 5°C for 2 to 3 d with no effect on transformation efficiency.
6. Collect the cells by centrifugation.
7. Decant the supernatant and resuspend the cells in 50 µL of 0.1 M LiAc using a pipet.
8. Add 1 µg of a *Bam*HI/*Aat*II-linearized BRV vector DNA (in 5–10 µL) and 5 µL of carrier DNA (10 mg/mL) to the cells and mix well.
9. Add 0.45 mL of 40% PEG 4000, mix by vortexing or repeated inversion, and incubate for 1 h at 30°C.
10. Heat-shock the cells in a 42°C tempblock for 15 min.
11. Top off the tube with sterile, distilled water and mix by inversion.
12. Collect the cells by centrifuging at high speed for 1 min.
13. Decant the supernatant and resuspend the cells in 1 mL of water using a sterile toothpick.
14. Collect the cells by centrifuging for 1 min.
15. Decant the supernatant and resuspend the cells in 100 µL of water, and spread the suspension on a 100-mm SD-Ura plate lacking uracil.

16. Incubate the plates at 30°C. Colonies of Ura⁺ transformants should be visible in 2 to 3 d. With 1 µg of vector, the yield of Ura⁺ transformants varies from 50 to 200 colonies. More than 90% of the transformants should derive from recombination between a BRV vector and a circular YAC.
17. To transfer the retrofitted YACs/BACs into *E. coli* cells, inoculate 5 mL of YPD medium in a 20-mL flask with two individual Ura⁺His⁺ colonies, and grow overnight at 30°C with vigorous shaking.
18. Pellet the cells in a 15-mL Falcon tube. Remove and discard the supernatant.
19. Resuspend the cells in 100 µL of EDTA mix (vortex well) and transfer into 1.5-mL Eppendorf tubes. Add 50 µL of 10 mg/mL zymolyase 20T, vortex the cells for 4 s, and incubate the suspension for 30 min at 37°C.
20. Melt an appropriate quantity of 1% low gelling/melting temperature agarose and place it in a 50°C water bath to cool.
21. Transfer the melted agarose and resuspended cells into a 42°C tempblock and equilibrate for 15 min.
22. Add to the cell suspension an equal volume of the melted agarose and mix well by vortexing. Keep the cell/agarose suspension at 42°C. It is important that the final concentration of agarose be equal to 0.5%. With a higher concentration of agarose, it is impossible to completely melt the plugs for electroporation.
23. Take 50-µL aliquots of the cell/agarose suspension and gently place each into Ultra Micro tips. Keep the tips horizontal for 10 min at 5°C until the agarose is completely solidified.
24. Transfer the agarose plugs into Eppendorf tubes. To do this, take up LET solution in a 6-cc syringe without a needle, place the tip of the Ultra Micro tip into the syringe lumen, and gently apply pressure. The plug should slide out into the tube. Make three to four 50-µL agarose plugs and incubate them for 1 h at 37°C.
25. Remove the LET and add enough NDS solution to cover the plugs. Incubate the plugs for 1 h at 55°C.
26. Remove the NDS solution carefully and wash the plugs three times with EDTA mix (20 min each time at room temperature). Dialyzed plugs may be stored at 5°C in EDTA mix.
27. Incubate the plugs overnight at room temperature in water before melting, and use for electroporation.
28. To electroporate YACs/BACs into *E. coli*, melt the plugs at 68°C for 15 min, cool to 42°C for 10 min, treat with 1.5 U of agarase for 1 h at 42°C, and chill on ice for 10 min.
29. Dilute the treated plug twofold with sterile water.
30. Use 1 µL of the mixture to electroporate 20 µL of the *E. coli* DH10B competent cells using a Bio-Rad Gene Pulser with the settings 2.5 kV, 200 Ω, and 25 µF.
31. After electroporation, add 1 mL of SOC into a cuvet, mix well with a Pipetman, and transfer into a microfuge tube.
32. Incubate the cells for 1 h at 37°C.

33. Spread 30, 100, and 300 μ L of the cell suspension onto LB-Cm plates supplemented with 12.5 μ g/mL of chloramphenicol.
34. Incubate the plates at 37°C overnight.

4. Notes

1. Another important step is the selection of specific hook(s) for a TAR vector. Hooks should be unique sequences; no repeated sequences should be present in the hooks. For human and mouse genomes, the uniqueness of hooks now can be easily checked by blasting against draft sequences. We demonstrated that the size of a hook could be as small as 60 bp (8). A further increase in the length of a targeting sequence had no effect on selectivity of gene isolation. Hooks should also be free of yeast *ARS*-like sequences. Potential *ARS*-like sequences in hooks can be identified based on the presence of a 17-bp *ARS* core consensus, WWWTTT TAYRTTTWGTT, in which W = A or T, Y = T or C, and R = A or G (24). The final conclusion about the absence of the yeast origin of replication in a hook(s) can be obtained only by yeast transformation assay. No or only a few His⁺ transformants should appear when the TAR cloning vector (with its hooks) is transformed into LiAc-treated yeast cells deficient in *HIS3*.
2. One of the potential mistakes that can mislead selective gene isolation by recombination in yeast is incorrect orientation of the hooks in the TAR vector. Hooks should be cloned into the vector in such a way that after linearization of the vector, the orientation of the hooks should correspond to that illustrated in Figs. 1 and 2. The quality of the vector DNA can also affect the yield of transformants and selectivity of gene isolation. For TAR cloning experiments, the vector DNA should not be contaminated by chromosomal DNA, and the completeness of the vector linearization by endonuclease digestion should be carefully checked by electrophoresis. Nonlinearized vector molecules will be inactive for homologous targeting of chromosomal DNA. In addition, they can induce circularization of linear vector molecules through a gap repair mechanism when the molecules enter the same cell.
3. Because yeast *ARS*-like sequences are located predominately in intragenic regions and introns, some genes smaller than 100 kb may be unclonable by a vector with two specific hooks. For such regions, the radial TAR cloning approach should be exploited. A TAR vector with a common repeat as a second hook can target a region that is up to 600 kb away from a specific sequence, increasing the probability of *ARS* capture.
4. It should also be mentioned that TAR cloning has only been applied for the isolation of average-size mammalian genes (from 50 to ~280 kb). Isolation of megabase-size genes may require more careful manipulation of the genomic DNA or modification of the TAR cloning protocol itself.
5. Isolation of specific genes by TAR cloning can be routinely carried out in any laboratory by adhering to a few guidelines. A prerequisite for cloning a single-copy gene by TAR is a high efficiency of yeast transformation. To satisfy this prerequisite, the yeast strain VL6-48N, which exhibits abnormally high transfor-

mation efficiency, was identified. Under standard conditions, this strain yields approx 50 times more transformants than a strain such as AB1380, which is routinely used for the construction of YAC libraries. Three genes, *HIS3*, *TRP1*, and *URA3*, are deleted in VL6-48N, allowing their use as markers in TAR cloning and/or in YAC/BAC retrofitting vectors. In addition, VL6-48N contains a nonreverting mutation in *LYS2*, allowing additional modifications and retrofitting of the cloned material.

6. Quality of genomic DNA is also critical for TAR cloning. DNA agarose plugs should be carefully washed out from EDTA and from traces of proteinase K that can lyse yeast spheroplasts. The size of genomic DNA should be checked by PFGE before use. Typically, >90% DNA fragments are >1000 kb when prepared in agarose plugs. Recently, we have shown that DNA gently prepared in aqueous solutions can also be used for TAR cloning when a targeted gene is smaller than approx 100 kb. The yield of transformants with DNA prepared in aqueous solutions is about 10–20 times higher compared with that observed with DNA prepared in agarose plugs, because agarose fragments inhibit yeast transformation. This means that much less genomic DNA (~100 ng) is required for gene isolation by TAR cloning. Two average-size human genes (60 and 80 kb) were successfully cloned by TAR in our laboratory using genomic DNA prepared in aqueous solutions. For these experiments, human genomic DNA was prepared by a protocol described in a manual for construction of genomic P1-derived artificial chromosome (PAC) libraries (25). The average size of genomic DNA prepared by this method is approx 150 kb. Approximately the same size of human and mouse DNA can be purchased from Promega (Madison, WI).
7. The basic protocol given here yields 50–300 transformants/μg of vector DNA with 5 μg of genomic DNA prepared in agarose plugs. Most of these transformants contain mammalian DNA inserts. With both standard and radial TAR cloning, the yield of positive clones varies from 1 to 10 per 1000 primary transformants for a single-copy gene. This variation is determined by the nature of the hooks selected for gene isolation. For a gene family, the yield of positive clones is much higher (15).
8. Retrofitted YACs/BACs with sizes up to about 250 kb can be efficiently and faithfully transferred from yeast cells into *E. coli* cells by electroporation. Larger circular DNAs cannot be moved into bacterial cells intact but still can be purified from yeast by alkaline lysis preparation (12) or by making use of their differential mobility in circular and relinearized form (5,26). Approximately 5% of human DNA fragments cloned in YAC/BAC vectors exhibit an abnormally low transformation efficiency during electroporation into *E. coli* cells. The bacterial colonies that can be obtained with these YACs/BACs contain deletions (9). The nature of these toxic regions (including several functional genes) is not yet clear. Clones with such inserts should be analyzed in yeast.
9. Because YACs/BACs can be deleted during electroporation, it is necessary to compare the size of inserts in yeast and in *E. coli* cells. To estimate the size of circular

YACs or YACs/BACs in yeast, they should be linearized either by endonuclease digestion (a unique *NotI* site is present in pVC604 vector) or by irradiation with a low dose of γ -rays (5 krad), separation by PFGE, and blot-hybridization with a TAR vector-specific probe or with total genomic DNA as previously described (4,5).

References

1. Burke, D. T., Carle, G. F., and Olson, M. V. (1987) Cloning of large segments of DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806–812.
2. Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepax, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300-kilo-base-pair fragments of human DNA in *E. coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794–8797.
3. Ketner, G., Spencer, F., Tugendreich, S., Connelly, C., and Hieter, P. (1994) Efficient manipulation of the human adenovirus genome as an infectious yeast artificial chromosome clone. *Proc. Natl. Acad. Sci. USA* **91**, 6186–6190.
4. Larionov, V., Kouprina, N., Graves, J., and Resnick, M. A. (1996) Specific cloning of human DNA as YACs by transformation-associated recombination. *Proc. Natl. Acad. Sci. USA* **93**, 491–496.
5. Larionov, V., Kouprina, N., Graves, J., and Resnick, M. A. (1996) Highly selective isolation of human DNAs from rodent-human hybrid cells as circular YACs by TAR cloning. *Proc. Natl. Acad. Sci. USA* **93**, 13,925–13,930.
6. Ma, H., Kunes, S., Schatz, P. J., and Botstein, D. (1987) Plasmid construction by homologous recombination in yeast. *Gene* **58**, 201–216.
7. Stinchomb, D. T., Thomas, M., Kelly, I., Selker, E., and Davis, R. W. (1980) Eukaryotic DNA segments capable of autonomous replication in yeast. *Proc. Natl. Acad. Sci. USA* **77**, 4559–4563.
8. Noskov, V., Koriabine, M., Solomon, G., Randolph, M., Barrett, J. C., Leem, S.-H., Stubbs, L., Kouprina, N., and Larionov, V. (2001) Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. *Nucleic Acids Res.* **29**, E62.
9. Kouprina, N. and Larionov V. (2003) Exploiting the yeast *Saccharomyces cerevisiae* for the study of the organization of complex genomes. *FEMS Microbiol Rev.* **27**, 1–21.
10. Kouprina, N. and Larionov, V. (1999) Selective isolation of mammalian genes by TAR cloning, in: *Current Protocols in Human Genetics*, vol. I (Dracopoli, N. C., Haines, J. L., Korf, B. R., et al., eds.), John Wiley & Sons, New York, pp. 5.17.1–5.17.21.
11. Strathern, J. N., Newlon, C. S., Herskowitz, I., and Hicks, J. B. (1979) Isolation of a circular derivative of yeast chromosome III: implications for the mechanism of mating type interconversion. *Cell* **2**, 309–319.
12. Devenish, R. J. and Newlon, C. S. (1982) Isolation and characterization of yeast ring chromosome III by a method applicable to other circular DNAs. *Gene* **18**, 277–288.

13. Bradshaw, M. S., Bollekens, J. A., and Ruddle, F. H. (1995) A new vector for recombination-based cloning of large DNA fragments from yeast artificial chromosomes. *Nucleic Acids Res.* **23**, 4850–4856.
14. Kouprina, N., Annab, L., Graves, J., Afshari, C., Barrett, J. C., Resnick, M. A., and Larionov V. (1998) Functional copies of a human gene can be directly isolated by TAR cloning with a small 3' end target sequence. *Proc. Natl. Acad. Sci. USA* **95**, 4469–4474.
15. Kouprina, N., Graves, J., Resnick, M. A., and Larionov, V. (1997) Specific isolation of human rDNA genes by TAR cloning. *Gene* **197**, 269–276.
16. Cancilla, M., Tainton, K., Barry, A., Larionov, V., Kouprina, N., Resnick, M., Du Sart, D., and Choo, A. (1998) Direct cloning of human 10q25 neocentromere DNA transformation-associated recombination (TAR) in yeast. *Genomics* **47**, 399–404.
17. Annab, L., Kouprina, N., Solomon, G., Cable, L., Hill, D., Barrett, J. C., Larionov, V., and Afshari, C. (2000) Isolation of functional copy of the human BRCA1 gene by TAR cloning in yeast. *Gene* **250**, 201–208.
18. Humble, M., Kouprina, N., Noskov, V., Graves, J., Garner, E., Tennant, R., Resnick, M. A., Larionov, V., and Cannon, R. E. (2000) Radial TAR cloning from the TgAC mouse. *Genomics* **70**, 292–299.
19. Cancilla, M., Graves, J., Matesic, L., Reeves, R., Tainton, K., Choo, K., Larionov, V., and Kouprina, N. (1998) Rapid cloning of mouse DNA as yeast artificial chromosomes by transformation-associated recombination (TAR). *Mamm. Genome* **9**, 157–159.
20. Kouprina, N., Campbell, M., Graves, J., Campbell, E., Meincke, L., Tesmer, J., Grady, D., Doggett, N., Moyzis, R., Deaven, L., and Larionov, V. (1998) Construction of human chromosome 16- and 5-specific YAC/BAC libraries by *in vivo* recombination in yeast (TAR cloning). *Genomics* **53**, 21–28.
21. Kim, J., Noskov, V. N., Lu, X., Bergmann, A., Ren, X., Warth, T., Richardson, P., Kouprina, N., and Stubbs, L. (2000) Discovery of a novel, paternally expressed ubiquitin-specific processing protease gene through comparative analysis of an imprinted region of mouse chromosome 7 and human chromosome 19q13.4. *Genome Res.* **10**, 1138–1147.
22. Razin, S. V., Ioudinkova, E. S., Trifonov, E. N., and Scherer, K. (2001) Non-clonability correlates with genomic instability: a case study of a unique DNA region. *J. Mol. Biol.* **307**, 481–486.
23. Ling, M., Merante, F., and Robinson, B. H. (1995) A rapid and reliable DNA preparation method for screening a large number of yeast clones by polymerase chain reaction. *Nucleic Acids Res.* **23**, 4294–4295.
24. Theis, J. F. and Newlon, C. S. (1997) The *ARS309* chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional *ARS* consensus sequence. *Proc. Natl. Acad. Sci. USA* **94**, 10,786–10,791.
25. Shepherd, N. S. (1999) Construction of bacteriophage P1 libraries with large inserts, in: *Current Protocols in Human Genetics*, vol. I (Dracopoli, N. C., Haines, J. L., Korf, B. R., et al., eds.), John Wiley & Sons, New York, pp. 5.3.1–5.3.26.

26. Cocchia, M., Kouprina, N., Kim, S.-J., Larionov, V., Schlessinger, D., and Nagaraja, R. (2000) Recovery and potential utility of YACs as circular YACs/BACs. *Nucleic Acids Res.* **28**, E81.
27. Larionov, V., Kouprina, N., Solomon, G, Barrett, J. C., and Resnick, M. A. (1997) Direct isolation of human BRCA2 gene by transformation-associated recombination in yeast. *Proc. Natl. Acad. Sci. USA* **94**, 7384–7387.
28. Sikorski, R. S. and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27.
29. Kim, U.-J., Birren, B. W., Slepak, T., Mancino, V., Boysen, C., Kang, H.-L., Simon, M. I., and Shizuya, H. (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**, 213–218.
30. Southern, P. J. and Berg, P. (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Genet.* **1**, 327–341.
31. Miller, A. D., Miller, D. G., Garcia, J. V., and Lynch, C. M. (1993) Use of retroviral vectors for gene transfer and expression. *Methods Enzymol.* **217**, 581–599.
32. Karreman, C. (1998) A new set of positive/negative selectable markers for mammalian cells. *Gene* **218**, 57–61.

Purification of BAC DNA

Tim S. Poulsen

1. Introduction

The introduction of bacterial artificial chromosome (BAC) libraries has made it easy for the scientific world to gain access to an unlimited amount of DNA from different species. These BAC libraries are constructed by insertion of DNA fragments from different species into a vector, which can be replicated in a bacterial host. Choosing *Escherichia coli* as the model host has many advantages: rapid growth of the host, high stability of the DNA fragment when inside the host, few chimeric clones, easy and rapid purification of the BAC DNA, and large amounts of sequenced BAC clones. The easy and rapid isolation of the BAC DNA from *E. coli* is facilitated by using an alkaline method for purification of plasmids (1). To ensure that the BAC DNA is pure enough to be used for downstream application, additional protocols have been elaborated to be combined with the protocol for alkaline purification of BAC DNA to achieve RNA-free BAC DNA (1,2), protein-free BAC DNA (1,3), genomic DNA-free BAC DNA (3), or endotoxin-free BAC DNA (4). Cloning, library construction, DNA labeling, sequencing, fingerprinting, transfection, and DNA microarray are examples of downstream applications, which are described in the following chapters.

1.1. Principles of DNA Purification

BAC is a single-copy plasmid up to 350 kb in size. However, an average BAC is approx 150 kb (5). A low-copy number of BAC requires a higher number of host cells to obtain enough BAC DNA for the downstream applications than high-copy number plasmids. This is solved by using a larger amount of standard Luria Bertani (LB) medium or by using rich media such as Terrific

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

broth (TB) or 2X yeast tryptone (YT). The rich media have the advantage of producing two to five times more bacterial cells per volume (6). However, using rich media also increases the amount of cellular proteins and RNA in the BAC DNA preparation, producing less satisfactory results in the downstream applications (7). Therefore, the use of rich media when purifying BAC DNA is not advised. After harvesting the cells, the BAC DNA is released from the host by using sodium dodecyl sulfate (SDS) to solubilize the lipid layers of *E. coli* and to bind the proteins. NaOH is added to denature the DNA and the proteins. Adding potassium acetate neutralizes the solution and precipitates the SDS-salt complexes, including the denatured proteins, the chromosomal DNA, and the cellular debris. The BAC DNA is renatured and remains in the solution. The precipitated debris is removed by centrifugation. The BAC DNA is desalted and concentrated by precipitation using isopropanol and washed with ethanol. The BAC DNA pellet is finally dissolved in a suitable solvent, and the DNA concentration is measured.

If the downstream application requires more pure BAC DNA, additional protocols can be combined with the protocol for alkaline purification of BAC DNA. These additional protocols can be divided into five separate steps; these steps are described in more detail in the following sections.

1. RNase A digestion to ensure removal of cellular RNA.
2. Adenosine triphosphate (ATP)-dependent exonuclease digestion to ensure the removal of contaminating genomic DNA, as well as nicked or damaged DNA.
3. Phenol:chloroform:isoamyl alcohol extraction to remove proteins.
4. Column anion exchange using Qiagen Resin under appropriate conditions to remove RNA, proteins, dyes, and low-molecular-weight impurities.
5. Removal of lipopolysaccharides using a specific Qiagen endotoxin removal kit.

As a simple rule, the yield of BAC DNA is typically 1 µg from a 5-mL LB culture using the protocol for alkaline BAC DNA purification, 0.8 µg from a 1.3-mL TB culture using Qiagen R.E.A.L, 4 µg from a 20-mL LB culture using Qiagen-tip 20, and 100 µg from a 500-mL LB culture using Qiagen-tip 500.

2. Materials

1. Buffer P1: 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 µg/mL of RNase A (store at 4°C).
2. Buffer P2: 200 mM NaOH, 1% SDS (store at room temperature).
3. Buffer P3: 3.0 M potassium acetate, pH 5.5 (store at 4°C).
4. Buffer QBT: 750 mM NaCl; 50 mM (3-[N-Morpholino]propanesulfonic acid (MOPS), 15% isopropanol, 0.15% Triton X-100, pH 7.0 (store at room temperature).
5. Buffer QC: 1.0 M NaCl; 50 mM MOPS, 15% isopropanol, pH 7.0 (store at room temperature).

6. Buffer QS: 1.5 M NaCl, 100 mM MOPS, 15% isopropanol, pH 7.0 (store at room temperature).
7. Buffer QF: 1.25 M NaCl, 50 mM MOPS, 15% isopropanol, pH 8.5 (store at room temperature).
8. Buffer EX: 50 mM Tris-HCl, 10 mM MgCl₂, pH 8.5 (store at room temperature).
9. Buffer ES: 20 mM KCl, 20 mM potassium phosphate, pH 8.5 (store at room temperature).
10. Buffer TE: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 (store at 4°C).
11. 5X TBE buffer: 445 mM Tris base, 445 mM boric acid, 10 mM EDTA (store at 4°C).
12. 6X Loading buffer: 0.25% (w/v) bromophenol blue, 30% glycerol in H₂O (store at 4°C).
13. ATP: 100 mM ATP, pH 7.5 (store at -20°C).
14. H₂O: ddH₂O (store at 4°C).
15. Chloramphenicol: 12.5 mg/mL in 96% ethanol (stock), working concentration of 12.5 µg/mL (store at -20°C).
16. Kanamycin: 25 mg/mL in H₂O (stock), working concentration of 25 µg/mL (store at -20°C).
17. Phenol:chloroform:isoamyl alcohol (25:24:1), saturated with TE (store at 4°C).
18. Chloroform (store at room temperature).
19. Isopropanol: 2-propanol (store at room temperature).
20. 70% EtOH: Diluted from absolute ethanol (store at room temperature).
21. Qiagen-tips: 20–10,000 (store at room temperature).
22. LB (1 L): 10 g of tryptone, 5 g of yeast extract, 10 g of NaCl, pH 7.0 (store at 4°C).
23. TB (1 L): 12 g of tryptone, 24 g of yeast extract, 4 g of glycerol, 12.54 g of K₂HPO₄, 2.31 g of KH₂PO₄ (store at 4°C).
24. 2X YT (1 L): 16 g of tryptone, 10 g of yeast extract, 5 g of NaCl, pH 7.0 (store at 4°C).
25. RNase A: 100 mg/mL in TE (store at room temperature).
26. ATP-dependent exonuclease: 350 µg/mL in ES (store at 4°C).
27. Ethidium bromide (EtBr): 10 mg/mL (store at room temperature in the dark).

3. Methods

3.1. Alkaline Purification of BAC DNA (see Note 1)

1. Streak a small amount of *E. coli* onto a selective plate and incubate overnight at 37°C to obtain single colonies.
2. Pick a single colony and inoculate into 5 mL of LB medium (see Note 2) containing the selective agent (see Note 3), and grow overnight at 37°C with shaking (see Note 4).
3. Harvest the bacterial cells by centrifuging at 20,000g for 2 min at 4°C (see Note 5).
4. Resuspend the bacterial pellet in 400 µL of buffer P1 until no cell clumps remain (see Note 6).

5. Add 400 μL of buffer P2, mix by inverting six times, and incubate at RT for 5 min (*see Note 7*).
6. Add 400 μL of buffer P3, mix by inverting six times, and incubate on ice for 10 min (*see Note 8*).
7. Centrifuge at 20,000g for 15 min at 4°C (*see Note 9*).
8. Decant the supernatant to a new centrifuge tube and centrifuge at 20,000g for 15 min at 4°C (*see Note 10*).
9. Decant the supernatant to a new centrifuge tube containing 700 μL of isopropanol, mix by inverting six times, and centrifuge at 20,000g for 30 min at 4°C (*see Note 11*).
10. Discard the supernatant and wash the DNA pellet with 1 mL of 70% ethanol, and centrifuge at 20,000g for 15 min at 4°C (*see Note 12*).
11. Discard the supernatant and air-dry the pellet for 10 min (*see Note 13*).
12. Dissolve the DNA in a suitable volume of buffer (*see Note 14*).

3.2. Additional Protocols That Can Be Used in Combination With Protocol for Alkaline Purification of BAC DNA

3.2.1. RNA-Free BAC DNA

Omitting the RNase A digestion from the protocol for alkaline purification of BAC DNA results in contamination of the BAC DNA with cellular RNA. Including the RNase A treatment in **step 3** of the protocol for alkaline purification of BAC DNA (*see Subheading 3.1.*) does not remove the cellular RNA completely, but it does lower the concentration significantly. An RNase A treatment can be included after **step 10** to ensure complete removal of the cellular RNA. If RNase A is included at **step 10**, a phenol extraction of the proteins should also be included to remove the RNase A from the BAC DNA (*1*).

1. Dissolve the DNA in 250 μL of buffer P1 and incubate at 37°C for 15 min.
2. Add 250 μL of phenol:chloroform:isoamyl alcohol and mix by inverting for 2 min. Centrifuge at 20,000g for 5 min at room temperature (*see Note 15*).
3. Transfer the aqueous layer to a new centrifuge tube, add 250 μL of phenol:chloroform:isoamyl alcohol, mix by inverting for 2 min, and centrifuge at 20,000g for 5 min at room temperature.
4. Transfer the aqueous layer to a new centrifuge tube and add 250 μL of chloroform. Mix by inverting for 2 min, and centrifuge at 20,000g for 5 min at room temperature (*see Note 16*).
5. Transfer the aqueous layer to a new centrifuge tube, and add 150 μL of H_2O , 40 μL of buffer P3, and 280 μL of isopropanol. Mix by inverting six times and centrifuge at 20,000g for 30 min at 4°C.
6. Discard the supernatant, wash the DNA pellet with 1 mL of 70% ethanol, and centrifuge at 20,000g for 15 min at 4°C.
7. Discard the supernatant and air-dry the pellet for 10 min.
8. Dissolve the DNA in a suitable volume of buffer (*see Note 14*).

3.2.2. Protein-Free BAC DNA

The protocol for alkaline purification of BAC DNA does not remove all of the proteins. The consequence is that the DNA is not particularly stable, presumably owing to nucleases present in the sample. Proteins can be removed by using either phenol or a column. An acid-phenol extraction that removes proteins, genomic DNA, and nicked BAC DNA has previously been described (3). This method can be used instead of the method described. The use of an anion-exchange Qiagen resin column has the advantages that RNA, dyes, and low-molecular-weight impurities are also removed, and that harmful phenol is avoided. The additional steps are inserted between **steps 7** and **9** of the protocol for alkaline purification of BAC DNA (*see Subheading 3.1.*).

1. Equilibrate a Qiagen-tip 20 by applying 1 mL of buffer QBT (*see Note 17*).
2. Apply the supernatant from **step 7** of **Subheading 3.1.** to the Qiagen-tip 20.
3. Wash the Qiagen-tip 20 four times with 1 mL of buffer QC each time (*see Note 18*).
4. Elute the DNA twice with 400 μ L of 65°C buffer QF each time (*see Note 19*).
5. Add 500 μ L of isopropanol, mix by inverting six times, and centrifuge at 20,000g for 30 min at 4°C.
6. Continue at **step 10** of **Subheading 3.1.**

3.2.3. Genomic and Damaged DNA-Free BAC DNA

When using the protocol for alkaline purification of BAC DNA, the BAC DNA may be contaminated with chromosomal DNA. Chromosomal DNA can represent up to 30% of the yield. Nicked and damaged BAC DNA is also present and may disturb the downstream application. Using ATP-dependent exonuclease digestion ensures removal of contaminating chromosomal DNA, nicked DNA, and damaged DNA. The additional steps are inserted after **step 11** of the protocol for alkaline purification of BAC DNA (*see Subheading 3.1.*).

1. Dissolve the DNA in 380 μ L of buffer EX.
2. Add 8 μ L of ATP-dependent exonuclease and 12 μ L of ATP solution and incubate at 37°C for 45 min (*see Note 20*).
3. Add 500 μ L of buffer QS.
4. Equilibrate a Qiagen-tip 20 by applying 1 mL of buffer QBT.
5. Apply the supernatant from **step 11** to the Qiagen-tip 20.
6. Wash the Qiagen-tip 20 four times with 1 mL of buffer QC each time.
7. Elute the DNA twice with 400 μ L of 65°C buffer QF each time.
8. Add 500 μ L of isopropanol, mix by inverting six times, and centrifuge at 20,000g for 30 min at 4°C.
9. Discard the supernatant and wash the DNA with 1 mL of 70% ethanol, and centrifuge at 20,000g for 15 min at 4°C (*see Note 12*). Continue at **step 12** of **Subheading 3.1.**

3.2.4. Endotoxin-Free BAC DNA

The outer layer of the outer membrane of *E. coli* is composed of lipopolysaccharides (endotoxins). When using the protocol for alkaline purification of BAC DNA (see **Subheading 3.1.**), the BAC DNA is contaminated with endotoxins. Contamination of BAC DNA with endotoxins affects the transfection efficiencies in a negative manner (**4**). Endotoxins can be removed either by using two rounds of CsCl gradient ultracentrifugation or by using a Qiagen EndoFree plasmid kit (**8**). Removal of the endotoxins using two rounds of CsCl gradient ultracentrifugation is a very time-consuming procedure. Using the EndoFree plasmid kit from Qiagen together with endotoxin-free tubes and endotoxin-free buffers to ensure removal of lipopolysaccharides is less time-consuming (**8**). A protocol for CsCl ultracentrifugation has been published (**1**). The protocol for the EndoFree plasmid kit can be obtained by contacting the local Qiagen supplier.

3.3. Assessment of BAC DNA Quality

BAC DNA prepared by the protocol for alkaline purification of BAC DNA (see **Subheading 3.1.**) typically contains genomic bacterial DNA, RNA, proteins, dyes, and low-molecular-weight impurities. This leads to significant overestimation of the actual DNA yield when measured spectrophotometrically (**1**). Quantification of the DNA yield is therefore difficult, and at least two different approaches should be used to calculate the BAC DNA concentration as described in the following sections.

3.3.1. Spectrophotometric Measurement of BAC DNA

Spectrophotometric measurement of BAC DNA concentration using ultraviolet (UV) absorption is simple and accurate if the sample is not too contaminated with proteins, phenol, RNA, or genomic DNA (see **Note 21**).

1. Dilute the DNA 10X in TE buffer.
2. Transfer to a quartz cuvet.
3. Measure the Δ_{260} and Δ_{280} of the diluted DNA. Use TE as a reference. Δ_{260} should be between 0.1 and 1.0 for a reliable estimation of the DNA concentration.
4. Calculate the DNA concentration: $\mu\text{g}/\mu\text{L} = [\Delta_{260} \times 10 \times 50 \mu\text{g}/(\text{mL} \times \text{OD} \times \text{cm})]/(1 \text{ cm} \times 1000 \mu\text{L}/\text{mL})$ (see **Note 22**). The ratio $\Delta_{260}/\Delta_{280}$ should be between 1.8 and 2.0 (see **Note 23**).

3.3.2. Spot Test

A spot test can be performed if the sample is not too contaminated with RNA. This technique provides a rapid way to make a rough, but useful estimate of the DNA concentration in a given sample. Although it is not a highly accu-

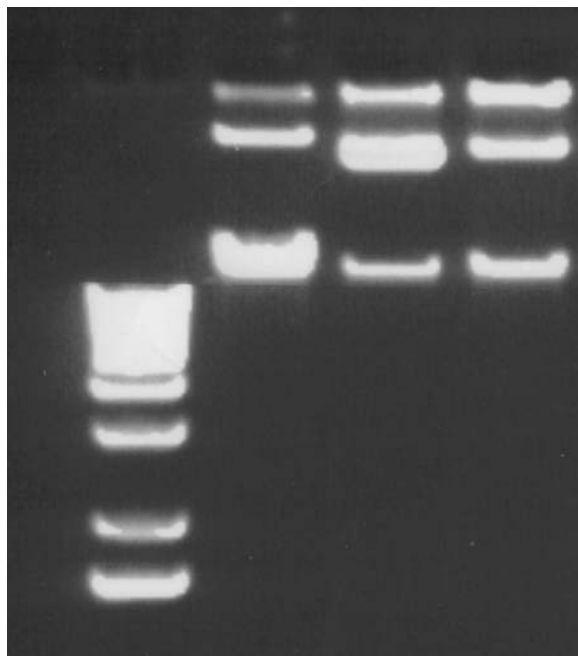


Fig. 1. Gel electrophoresis of BAC DNA. Lane 1, 1-kb ladder; lanes 2–4, 250 ng of BAC DNA. The lower band is supercoiled BAC DNA, the middle band is relaxed BAC DNA, and the upper band contains nicked BAC DNA and bacterial genomic DNA.

rate method, it is still useful when measuring the concentration of BAC DNA used for cloning and probe labeling (*see Note 24*).

1. Mix 5 μL of EtBr (1 $\mu\text{g}/\text{mL}$) and 5 μL of DNA solution in an Eppendorf tube.
2. Prepare control samples containing 100, 200, 400, and 800 ng.
3. Place a drop from each control sample and the sample to be measured on a UV transilluminator that is covered with Vita wrap and adjusted to 260 nm.
4. Estimate the approximate BAC DNA concentration by comparing the fluorescence intensity of the control samples with the sample to be measured.

3.3.3. Agarose Gel Electrophoresis

Agarose gel electrophoresis is the most reliable technique to estimate the BAC DNA concentration after purification if the sample is contaminated with genomic DNA and/or RNA. During agarose gel electrophoresis, the RNA will migrate faster than the BAC DNA and thereby be dissociated from the BAC DNA (*see Note 25*). An example of DNA purified from a BAC clone is shown in **Fig. 1**.

1. Prepare a 0.7% agarose gel in 1X TBE.
2. Load the gel with 250 ng of BAC DNA (typically 5 μ L) dissolved in 1X loading buffer.
3. Load control samples containing 100, 200, 300, and 400 ng of BAC DNA.
4. Run the agarose gel at a constant voltage (6 V/cm) for 1 h using a horizontal electrophoresis apparatus.
5. Stain the agarose gel in 1 μ g/mL of EtBr solution for 20 min.
6. Destain in H₂O for 10 min.
7. Take a photograph while exposing the agarose gel with UV light.
8. Estimate the approximate BAC DNA concentration by comparing the fluorescence intensity of the control samples with the sample to be measured.

4. Notes

1. The protocol can easily be modified to large-volume cultures by scaling. A protocol for multipurification of BAC DNA has previously been described (7).
2. Rich media such as TB or 2X YT produce more bacteria per volume (two to five times). The culture volume should then be reduced to match the amount of cells produced in an LB medium culture. An outgrown overnight culture grown in standard LB medium has a cell density of approx 3×10^9 cells/mL (OD₆₀₀) or 3×10^8 cells/mL (OD₄₃₆). This corresponds to a pellet wet wt of approx 3 g/L. It is not normally recommended that rich media be used when using Qiagen tips for preparation of BAC DNA.
3. The selective agent is used to ensure a selection pressure on *E. coli* containing the plasmid. The selective agent can be different, depending on the vector that has been used for creating the DNA libraries. The most widely used antibiotic for the BAC libraries is chloramphenicol. The working concentrations of the most commonly used antibiotics are listed in **Subheading 2**.
4. If a larger amount of *E. coli* cells is required, grow a 2-mL culture for 6 h and dilute 1/1000 into a vessel containing the medium with the appropriate selective agent. Then grow this overnight at 37°C with vigorous shaking (250–300 rpm). The vessel should have a volume that is at least four times greater than the volume of the medium.
5. The culture can also be harvested by centrifuging at 4000g for 30 min or at 20,000g for 2 min at 4°C. All of the culturing medium should be carefully removed.
6. RNase A can be omitted from buffer P1. It is crucial that the pellet be completely resuspended in buffer P1.
7. Check buffer P2 for SDS precipitation. The lysate should appear viscous. Avoid shaking, as this will result in shearing of the genomic and BAC DNA.
8. A white precipitate consisting of SDS-salt complexes should appear. Avoid shaking, as this will result in shearing of the genomic and BAC DNA, as well as trapping of the BAC DNA in the SDS-salt complexes.
9. Avoid disturbing the white precipitate consisting of SDS-salt complexes.

10. **Step 7** is included to avoid remaining white precipitate that can disturb downstream application.
11. When precipitating the DNA, using isopropanol at room temperature instead of ethanol minimizes salt precipitation.
12. Avoid disturbing the DNA pellet.
13. The pellet can be air-dried under a lamp for 10 min.
14. Twenty microliters of TE (pH 8.0), or Tris-HCl (pH 8.5), or H₂O can be used depending on the downstream application (e.g., H₂O is used for DNA sequencing, since the EDTA inhibits the polymerase).
15. RNase A is more efficiently removed when using a (25:24:1) mixture of phenol:chloroform:isoamyl alcohol compared with using phenol alone.
16. Chloroform is added to remove traces of phenol in the aqueous layer.
17. When the Qiagen column is equilibrated, the resin is stable for 6 h. The column can be reused by reequilibrating with QBT buffer.
18. This elutes all nonbinding impurities.
19. Eluting at 65°C eases the release of BAC DNA from the column.
20. Digestion with ATP-dependent exonuclease.
21. The DNA absorbs UV light at 260 nm. RNA and some amino acids within proteins also absorb UV light at 260 nm and may cause a significant overestimation of the DNA concentration. Genomic DNA may represent up to 30% of the total DNA concentration.
22. An OD₂₆₀ of 1 corresponds to approx 50 µg/mL for double-stranded DNA and 40 µg/mL for single-stranded DNA and RNA.
23. If there is contamination with protein, the OD₂₆₀/OD₂₈₀ will be significantly less than 1.8 and accurate quantification of the DNA is not possible.
24. Contamination with small amounts of RNA can easily cause an overestimation of the DNA concentration. Contamination with proteins and dyes does not lead to a significant overestimation.
25. Normally two bands are present in the agarose gel. The upper band corresponds to genomic DNA while the lower band corresponds to BAC DNA. Use the BAC DNA band to estimate the approximate concentration by comparing the fluorescence intensity of the control samples with the sample to be measured. The genomic DNA may represent up to 30% of the total DNA concentration.

References

1. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
2. Birnboim, H. C. (1983) A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods Enzymol.* **100**, 243–255.
3. Azad, A. K., Coote, J. G., and Parton, R. (1992) An improved method for rapid purification of covalently closed circular plasmid DNA over a wide size range. *Lett. Appl. Microbiol.* **14**, 250–254.

4. Weber, M., Möller, K., Welzeck, M., and Schorr, J. (1995) Effect of lipopoly-saccharide on transfection efficiency in eukaryotic cells. *BioTechniques* **19**, 930–940.
5. Zhao, S., Malek, J., Mahairas, G., Fu, L., Nierman, W., Venter, J. C., and Adams, M. D. (2000) Human BAC ends quality assessment and sequence analyses. *Genomics* **63**, 321–332.
6. Tartof, K. D. and Hobbs, C. A. (1987) Improved media for growing plasmid and cosmid clones. *Bethesda Res. Lab. Focus* **9**, 12.
7. Kelly, J. M., Field, C. E., Craven, M. B., Bocskai, D., Kim, U.-J., Rounsley, S. D., and Adams, M. D. (1999) High throughput direct end sequencing of BAC clones. *Nucleic Acids Res.* **6**, 1539–1546.
8. Ehlert, F., Bierbaum, P., and Schorr, J. (1993) Importance of DNA quality for transfection efficiency. *BioTechniques* **14**, 546.

Hybridization-Based Selection of BAC Clones

Chang-Su Lim and Ung-Jin Kim

1. Introduction

Studying large genomic regions at the molecular level requires access to the DNA representing such sites. The availability of large and stable clones derived from target genomic regions is essential for detailed analysis such as sequencing. Since its introduction in 1992, the bacterial artificial chromosome (BAC) library system has been widely employed as a standard cloning system for mapping and sequencing the genomes of human and model organisms (*1*), as well as in a variety of other research areas where large DNA insert-containing clones are needed (*2–24*). With the end of the Human Genome Project drawing near (*25*), the genome-sequencing community is now directing its efforts at sequencing commercially important organisms. BAC library construction and high-throughput screening has become an indispensable tool for the isolation of large chromosomal DNA fragments necessary for such projects.

The BAC vectors allow *lacZ*-based positive color selection of the BAC clones that carry insert DNA in the cloning sites at the time of library construction (*26*). BAC clones are arrayed into microtiter plates, gridded onto hybridization filters at high density. Clones carrying desired human DNA fragments are identified by colony hybridization with labeled DNA probes derived from cDNAs or oligonucleotides (*26*). In this chapter, we describe protocols that are useful for the identification of BAC clones from BAC libraries using colony hybridization. The protocols are pertinent to both small- and large-scale library screenings, using either single probes or pools of more than 100.

2. Materials

Chemicals should be of molecular biology grade or higher. Solutions should be prepared using deionized and distilled water. Some reagents are available as commercial kits, and in this case, vendor-provided instructions were followed unless otherwise mentioned.

2.1. Isolation and Purification of cDNA Inserts Using Polymerase Chain Reaction

1. Polymerase chain reaction (PCR) reaction kits were obtained from Qiagen (Valencia, CA). Any standard PCR reaction kit should suffice (*see Note 1*). PCR profile: 94°C for 3 min; 94°C for 30 s, 55°C for 30 s, and 72°C for 50 s for 30 cycles; 72°C for 5 min; and a 4°C hold or ready to purify.
2. Low-melting-point Sea Plaque GTG agarose was obtained from Fisher (Pittsburgh, PA). It seems that better-quality DNA probes can be obtained using low-melting-temperature agarose gels because of their greater resolving properties relative to standard-melting-temperature agarose (*see Note 2*).
3. 1X TAE gel electrophoresis buffer: 0.04 M Tris-acetate, 0.001 M EDTA. A 50X TAE stock can be stored at room temperature (*see Note 3*).
4. Ethidium bromide (EtBr) (10 mg/mL); store at room temperature (*see Note 4*).
5. 10X Gel-loading buffer: 0.25% xylene cyanol FF, 0.25% bromophenol blue, 30% glycerol (*see Note 5*).
6. GELase™ Agarose Gel-Digesting Preparation (Epicentre Technologies, Madison, WI). Normally 1 µL of enzyme GELase (1 U/µL) is used (*see Note 6*).
7. 1X TE: 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0.
8. 3 M Na acetate, pH 7.0.
9. Gel electrophoresis apparatus.
10. Ultraviolet transilluminator.
11. Scalpel razor blade, to excise agarose gel slices for the subsequent isolation of DNA fragments using Gelase.
12. Multichannel pipets.
13. 96-Well thin-walled PCR tubes.
14. PCR apparatus.

2.2. DNA Probe Labeling

1. Hexanucleotide mix (Roche, Indianapolis, IN).
2. Nucleotide mix: 0.5 mM dGTP, 0.5 mM dTTP, 0.5 mM dCTP except dATP. This can be made by mixing each dNTP except dATP.
3. Labeling-grade Klenow enzyme (Roche).
4. Sephadex G50 spin column; Quick Spin Columns (TE) for radiolabeled DNA purification (Roche).
5. ³²P-dATP (3000 Ci/mmol).
6. Scintillation counter, if needed (*see Note 7*).

2.3. Southern Hybridization to High-Density Filters

1. Human placental DNA; this can be purchased from Sigma (St. Louis, MO) (*see Note 8*).
2. Hybridization solution: 1 M NaCl; 0.05 M Tris-HCl, pH 8.0, 5 mM EDTA, 1% sodium dodecyl sulfate (SDS), 10% dextran sulfate (*see Note 9*).
3. Hybridization apparatus: It is convenient to use a commercially available hybridization oven; we obtained hybridization ovens from Robbins.
4. Hybridization bottles; these were purchased from Robbins.
5. Nylon filters harboring BAC clones gridded by a robotic system (*see Note 10*).

2.4. Washing and Autoradiography

1. 1X low-stringent washing solution: 1X saline sodium citrate (SSC), 0.5% SDS. 20X SSC is made up of 3.0 M NaCl and 0.3 M Na citrate, pH 7.0 (*see Note 11*).
2. 1X high-stringent washing solution: 0.1X SSC, 0.5%.
3. Saran Wrap®.
4. Kodak BioMax MS autoradiography film (New York, NY).
5. Intensifying screen.
6. X-ray film cassettes.
7. Film processor or phosphoimager.

2.5. OVERGO Hybridization

1. Solution Q: 1.25 M Tris-HCl, pH 8.0, 125 mM MgCl₂.
2. Solution A: 1 mL of solution Q, 18 μ L of 2-mercaptoethanol, 5 μ L 0.1 M dTTP, 5 μ L of 0.1 M dGTP.
3. Solution B: 2 M HEPES-NaOH, pH 6.6.
4. Solution C: 3 mM Tris-HCl, pH 7.4, 0.2 mM EDTA, pH 8.0.
5. ABC mixture: solution A:solution B:solution C (1:2.5:1.5). Make up by mixing 1 vol of solution A with 2.5 vol of solution B and 1.5 vol of solution C.
6. Hybridization solution (*see Note 9*): 7% SDS, 1 mM EDTA, pH 8.0, 1% bovine serum albumin (BSA), 0.5 M Na phosphate buffer.
7. ³²P-dATP (3000 Ci/mmol) and ³²P-dCTP (3000 Ci/mmol).
8. Sephadex-G25 spin columns.
9. Klenow DNA polymerase.

3. Methods

3.1. Isolation and Purification of cDNA Inserts Using PCR

1. Thaw glycerol stocks of cDNA clones (384-well plates, one cDNA clone per well) or prepare any template DNA such as plasmid, cosmid, or genomic DNA for PCR.
2. Using a 96-well thin-walled PCR plate, add 1 μ L of Unigene glycerol stock culture to each well containing 99 μ L of sterile H₂O (*see Note 12*).

3. Run the BOIL program (100°C for 5 min) on a PCR machine (MJ Research, Waltham, MA) (*see Note 13*).
4. Immediately put the PCR plate on ice.
5. Transfer 1 μL of boiled culture diluent into the wells of a new 96-well PCR plate (*see Note 14*).
6. Aliquot 24 μL of PCR reaction mixture per sample well using multichannel pipets (*see Note 15*). The PCR mixture per sample (25- μL reaction) is as follows: 2.5 μL of 10X reaction buffer, 2.5 μL of MgCl_2 (25 mM), 2.0 μL of dNTPs (2.5 mM), 1.0 μL of forward primer (5 μM), 1.0 μL of reverse primer (5 μM), 0.5 μL of *Taq* polymerase, and 14.5 μL of H_2O .
7. Set up the PCR reaction. The reaction conditions (30 cycles) are as follows: 94°C for 3 min; 94°C for 30 s, 55°C for 30 s, 72°C for 50 s; 72°C for 5 min; and a 4°C hold or ready to do the next step.
8. Stop the reaction by adding 2 μL of 10X gel-loading buffer.
9. Load 10 μL of PCR samples onto each well of a 0.8% low-melting-temperature Sea Plaque GTG agarose made in 1X TAE.
10. Run at 3 V/cm for up to a few hours depending on the size of the PCR products. Aim for the condition under which the DNA bands are best separated to facilitate the quality of purified DNAs, especially when there are contaminating nonspecific PCR bands (*see Note 16*).
11. Cut out PCR bands using a fresh razor blade, and place each into a separate 1.8-mL microcentrifuge tube (i.e., an individual probe DNA/tube).
12. Incubate for 2 h in 0.5 mL of 1X Gelase buffer to equilibrate.
13. Remove all liquid but the gel slice from the microcentrifuge tubes.
14. Incubate the tubes containing the gel slices at 70°C for 20 min.
15. Transfer the tubes directly into 45°C.
16. Equilibrate for 10 min.
17. Add 1 μL of Gelase (1 U/ μL ; *see Note 17*).
18. Incubate at 45°C for 2 h or longer.
19. Add 1 vol of 3 M Na acetate.
20. Add 2.5 vol of absolute ethanol kept at -20°C and mix well.
21. Incubate at -20°C for 1 h.
22. Centrifuge at 12,000g and 4°C for 30 min.
23. Pour off the supernatant.
24. Wash once with 0.5 mL of 70% ethanol kept at -20°C.
25. Centrifuge at 12,000g and 4°C for 5 min.
26. Decant the supernatant and air-dry the pellets.
27. Resuspend the pellets in 20 μL of 1X TE.

3.2. DNA Probe Labeling

1. Prepare a single probe or a mixture of probes by combining 3 μL from each of 10 probes to a total volume of 30 μL (*see Note 18*).
2. Denature by boiling at 100°C for 5 min.
3. Immediately place on ice.

4. Add 3 μL of hexanucleotide mix.
5. Add 10 μL of dNTPs mix.
6. Add 5 μL of 3000 Ci/mmol ^{32}P -dATP.
7. Add 3 μL of labeling-grade Klenow (2 U/ μL).
8. Incubate at room temperature overnight or at 37°C for 2 h.
9. Pass through a Sephadex-G50 spin column to remove unincorporated ^{32}P -dATPs.
10. Obtain specific activities of labeled probes using a scintillation counter, if desired.

3.3. Southern Hybridization to High-Density Filters

1. Proceed to the next step if a single probe is prepared or if many probes are labeled. Then combine eluates from 10 Sephadex-G50 spin columns into one 1.8-mL Eppendorf tube.
2. Add 150 μL of human placental DNA (10.9 mg/mL).
3. Add 250 μL of modified hybridization solution.
4. Incubate at 100°C for 10 min.
5. Immediately transfer the probes to 65°C for 2 h to prehybridize.

3.3.1. High-Density Filter Preparation

1. Prewet high-density nylon filters in a tray large enough to fit the filters with 1 L of distilled H_2O , then 1X SSC.
2. Roll no more than four high-density filter blots together and place them into a roller hybridization bottle.
3. Add 25–50 mL of modified hybridization solution to each roller bottle.
4. Prehybridize in a rotating oven at 65°C for 2 h or more.
5. Add probe to the hybridization bottle.
6. Rotate at 100–150 rpm at 65°C for 6 h or overnight (preferred).

3.4. Washing and Autoradiography

1. Carefully pour off hybridization solution into designated radioactive waste.
2. Wash with 200 mL of low-stringent washing solution prewarmed at 65°C with fast rotation for 1 h (see **Note 19**).
3. Repeat **step 2** once.
4. Wash with 200 mL of high-stringent washing solution prewarmed at 65°C with fast rotation for 1 h.
5. Repeat **step 4** once.
6. Rinse washed filters with H_2O once and put them on used films as a support.
7. Wrap each filter blot with Saran Wrap.
8. Expose to X-ray films.
9. Incubate at -80°C overnight.
10. The next day warm the cassettes at room temperature for no longer than 20–30 min (see **Note 20**).
11. Develop the films.
12. Address positive clones using a lightbox and grid overlay (refer to the source of high-density filters, which will accompany the calculation formula).

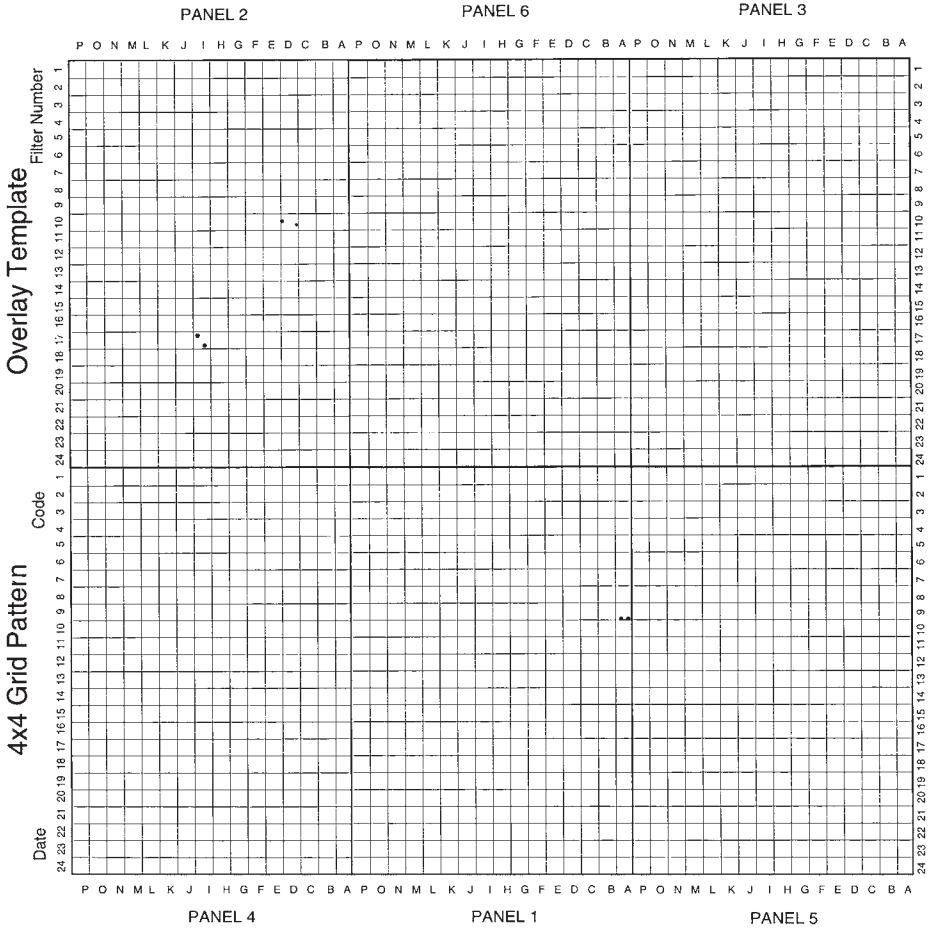


Fig. 1. Representative image of high-density filter hybridized with OVERGO probes. There are 55,296 spots gridded in duplicate by a robot (Q-bot) representing 27,648 individual BAC clones, which cover an entire human genome. A high-density nylon filter was hybridized with five OVERGO probes as a pool corresponding to five different genes with the conditions described in **Subheading 3.5**.

3.5. OVERGO Hybridization (see Fig. 1)

1. Add 50 mL of prewarmed hybridization solution to a roller bottle.
2. Soak filters in 2X SSC and then roll and insert into the bottle up to eight filters per bottle (see **Note 21**); prehybridize new filters for at least 2 h and reused filters for 1 h.
3. Add ^{32}P -labeled oligos (from 1 to 100 different oligos) after denaturation at 90°C for 5 min to the bottle.

3.5.1. OVERGO Probe Labeling (see **Note 22**)

1. Heat two overlapping oligos (10 pmol/ μ L) at 80°C for 5 min.
2. Incubate at 37°C for 10 min, and store on ice until use.
3. To make 10 μ L of labeling reaction, two oligos (1 μ L of each oligo) + H₂O = 5.5 μ L:
 - a. BSA (2 mg/mL): 0.5 μ L.
 - b. ABC mixture: 2.0 μ L.
 - c. ³²P-dATP: 0.5 μ L.
 - d. ³²P-dCTP: 0.5 μ L.
 - e. Klenow fragment: 1 μ L (2 U/ μ L).
4. Incubate at room temperature for 1 h.
5. Pass the reaction mixture through a Sephadex-G25 spin column.
6. Heat the probes at 95°C for 3 min before adding to the hybridization bottle.
7. Hybridize for 12–48 h (see **Note 23**).
8. Wash once with 1.5X SSC, 0.1% SDS at 58°C for 30 min. Then wash once with 0.5X SSC, 0.1% SDS at 58°C for 30 min.
9. Carry out autoradiography (see **Note 24**).

4. Notes

1. We found that Qiagen *Taq* is very robust and can tolerate moderate changes in salt concentration. Other *Taq* polymerases can be employed as long as PCR conditions are optimized to give clear PCR products without many other background bands. In addition, PCR conditions need to be adjusted depending on the source of PCR templates and the size of PCR products. Our PCR templates are directly from bacterial glycerol stocks unless otherwise stated.
2. We use a low percentage (typically 0.8%) of low-melting-temperature agarose gel to increase the quality of our probe DNAs and also to facilitate the probe DNA preparation from the gel pieces by digestion with Gelase (Epicentre Technologies). It is especially beneficial to use low-melting-temperature agarose gels when one has to reamplify a PCR product after gel purification because of poor yields in the first PCR reaction.
3. 1X TAE can be reused up to five times only when DNAs on the gel do not need to be gel purified. Otherwise, 1X TAE buffer should be changed when different probes need to be prepared to avoid any contamination, which will subsequently lead to false positives.
4. EtBr is potentially carcinogenic. Therefore, caution should be taken when handling this reagent. It is always better to wear gloves when handling stained gels with EtBr. It is also suggested that EtBr-stained gels be disposed of safely.
5. When desired PCR products are between 400–600 bp, we suggest that 10X loading buffer be used at a final 0.5X concentration or less to reduce interference of DNA intensities on the gel. Sometimes yields of PCR reactions can be underestimated.
6. GELase™ Agarose Gel-Digesting Preparation is a unique enzyme solution for simple, quantitative recovery of intact DNA and RNA from low-melting-point agarose gels following electrophoresis. GELase digests the carbohydrate back-

bone of molten agarose into small, soluble oligosaccharides. The purified nucleic acid can be rapidly recovered from the digested gel solution by ammonium acetate and ethanol precipitation.

7. Usually we do not measure specific activities of labeled probes. However, it should be done at least once to optimize probe labeling.
8. It is important not to skip this step. Nonspecific hybridization to highly repetitive Alu sequences can be suppressed during incubation with small fragments of human placental DNAs.
9. An alternative hybridization solution (7% SDS, 0.5 M Na phosphate, 1 mM EDTA, pH 8.0, 1% BSA) can also be used. A 1 M Na phosphate stock can be prepared as follows: Add 134 g of $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, then 4 mL of 85% H_3PO_4 , and adjust the volume to 1 L. Hybridization solution comprises a half volume of 1 M Na phosphate solution (such that the final solution will be 0.5 M with respect to Na^+ ions), 7% SDS, 1 mM EDTA (pH 8.0), and 1% BSA.
10. We used high-density BAC arrayed nylon filters (22 × 22 cm) by a robotic system. Each filter has 55,296 BAC colony spots, giving rise to 27,648 BAC total clones in duplicates.
11. 20X SSC can be used either concentrated or diluted. To prepare a less concentrated solution, dilute the 20X SSC using autoclaved or sterile-filtered H_2O . Diluted SSC is stable for at least 1 yr at room temperature.
12. When several different probes (10–10,000) are to be used to isolate corresponding positive BAC clones, it is convenient to use 96-well thin-walled PCR plates to facilitate processing time.
13. This is to prepare PCR templates from bacterial glycerol stock. If bacterial colonies are to be used, the same protocol applies. Too much template is always problematic in generating nonspecific PCR products.
14. As mentioned in **Note 13**, nonspecific PCR reactions arise if too much template is used. Therefore, 1 μL or less of template diluent should be used.
15. This PCR program is for plasmid templates from glycerol stock ranging from 300 bp up to 4 kb. However, PCR conditions need to be adjusted depending on the templates such as cosmid, YAC and genomic DNAs, and the expected size of PCR products (at 72°C *Taq* polymerase has an extension time of 1 min/kb).
16. It is best to run PCR products to get the sharpest bands to reduce false positives owing to contamination of probes with other nonspecific DNAs. If genomic DNAs are to be used as templates for probe preparation by PCR, more care should be taken.
17. Other commercial DNA extraction kits can be used (e.g., Gel Extraction Kit II, Qiagen). However, when many different DNA probes need to be gel purified it will not be practical, in a temporal sense, to process each probe using such kits. For our needs, Gelase was found to be the most time efficient.
18. To label several different DNA probes (between 10 and 100), it is efficient to label them as pools (10 probes/pool/tube) and then mix 10 pools after labeling with ^{32}P -dATP. These 100 probes can then be used in a hybridization bottle containing nylon membranes to perform hybridizations.

19. This step can be reduced to 30 min depending on the number of probes added, how many times filters are used, and so on. We obtained the most consistent results when we washed the filters under the conditions provided in the protocol.
20. It is important that cassettes be kept at room temperature no more than 20–30 min. Otherwise, films tend to stick to Saran wrap, leaving traces of the wrap, which makes analysis difficult owing to the interference with positive signals.
21. Care needs to be taken not to trap bubbles between filters when rolling several filters together and placing them into the bottle. The bottle can be rotated to allow the filters to unroll slowly. Be sure that all the filters are rolled in the same direction to reduce damaging the filter sets. We have reused nylon filters 20 times unless they have gotten physically damaged. It is not necessary to add blocking DNAs such as salmon sperm DNA.
22. Two 24-base oligonucleotides that overlap by 10 bases are first annealed. When added to a labeling reaction, both 5'-overhanging ends will be filled with ^{32}P -dATP and ^{32}P -dCTP by Klenow DNA polymerase, resulting in 38 bases of double-strand oligonucleotides labeled with ^{32}P .
23. Hybridizations are done in an oven at 58°C. OVERGOS (GC content between 40 and 60%) work best at this temperature. Therefore, when using AT-rich OVERGOS, lower hybridization temperatures ranging from 37 to 58°C should be employed. The best conditions need to be obtained empirically. Hybridization can be done for 12 h up to 72 h to obtain somewhat stronger signals. Longer hybridizations are particularly useful for older filters.
24. After washing, rinse the filters with distilled water. Filters are placed onto used films and then wrapped with Saran Wrap. Expose for 12 h up to 72 h at -70°C.

Acknowledgments

We thank Dr. Shane Rea for critical reading of the manuscript. This work was supported in part by a US Department of Energy grant on human and mouse BAC-EST mapping (DEFC03-96ER62242).

References

1. Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794–8797.
2. Woo, S. S., Jiang, J., Gill, B. S., Paterson, A. H., and Wing, R. A. (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic Acids Res.* **22**, 4922–4931.
3. Wang, M., Chen, X. N., Shouse, S., et al. (1994) Construction and characterization of a human chromosome 2-specific BAC library. *Genomics* **24**, 527–534.
4. Wang, G. L., Holsten, T. E., Song, W. Y., Wang, H. P., and Ronald, P. C. (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the Xa-21 disease resistance locus. *Plant J.* **7**, 525–533.

5. Cai, L., Taylor, J. F., Wing, R. A., Gallagher, D. S., Woo, S. S., and Davis, S. K. (1995) Construction and characterization of a bovine bacterial artificial chromosome library. *Genomics* **29**, 413–425.
6. Kim, U. J., Shizuya, H., Chen, X. N., Deaven, L., Speicher, S., Solomon, J., Korenberg, J., and Simon, M. I. (1995) Characterization of a human chromosome 22 enriched bacterial artificial chromosome sublibrary. *Genet. Anal.* **12**, 73–79.
7. Stone, N. E., Fan, J. B., Willour, V., Pennacchio, L. A., Warrington, J. A., Hu, A., de la Chapelle, A., Lehesjoki, A. E., Cox, D. R., and Myers, R. M. (1996) Construction of a 750-kb bacterial clone contig and restriction map in the region of human chromosome 21 containing the progressive myoclonus epilepsy gene. *Genome Res.* **6**, 218–225.
8. Schmitt, H., Kim, U. J., Slepak, T., Blin, N., Simon, M. I., and Shizuya, H. (1996) Framework for a physical map of the human 22q13 region using bacterial artificial chromosomes (BACs). *Genomics* **33**, 9–20.
9. Diaz-Perez, S. V., Crouch, V. W., and Orbach, M. J. (1996) Construction and characterization of a *Magnaporthe grisea* bacterial artificial chromosome library. *Fungal Genet. Biol.* **20**, 280–288.
10. Hubert, R. S., Mitchell, S., Chen, X. N., et al. (1997) BAC and PAC contigs covering 3.5 Mb of the Down syndrome congenital heart disease region between D21S55 and MX1 on chromosome 21. *Genomics* **41**, 218–226.
11. Zimmer, R. and Verrinder Gibbins, A. M. (1997) Construction and characterization of a large-fragment chicken bacterial artificial chromosome library. *Genomics* **42**, 217–226.
12. Schibler, L., Vaiman, D., Oustry, A., Guinec, N., Dangy-Caye, A. L., Billault, A., and Cribiu, E. P. (1998) Construction and extensive characterization of a goat bacterial artificial chromosome library with threefold genome coverage. *Mamm. Genome* **9**, 119–124.
13. Mozo, T., Fischer, S., Shizuya, H., and Altmann, T. (1998) Construction and characterization of the IGF Arabidopsis BAC library. *Mol. Gen. Genet.* **258**, 562–570.
14. Kouprina, N., Campbell, M., Graves, J., Campbell, E., Meincke, L., Tesmer, J., Grady, D. L., Doggett, N. A., Moyzis, R. K., Deaven, L. L., and Larionov, V. (1998) Construction of human chromosome 16- and 5-specific circular YAC/BAC libraries by *in vivo* recombination in yeast (TAR cloning). *Genomics* **53**, 21–28.
15. Li, R., Mignot, E., Faraco, J., Kadotani, H., Cantanese, J., Zhao, B., Lin, X., Hinton, L., Ostrander, E. A., Patterson, D. F., and de Jong, P. J. (1999) Construction and characterization of an eightfold redundant dog genomic bacterial artificial chromosome library. *Genomics* **58**, 9–17.
16. Vaiman, D., Billault, A., Tabet-Aoul, K., Schibler, L., Vilette, D., Oustry-Vaiman, A., Soravito, C., and Cribiu, E. P. (1999) Construction and characterization of a sheep BAC library of three genome equivalents. *Mamm. Genome* **10**, 585–587.
17. Wu, C., Asakawa, S., Shimizu, N., Kawasaki, S., and Yasukochi, Y. (1999) Construction and characterization of bacterial artificial chromosome libraries from the silkworm, *Bombyx mori*. *Mol. Gen. Genet.* **261**, 698–706.

18. Capela, D., Barloy-Hubler, F., Gatiús, M. T., Gouzy, J., and Galibert, F. (1999) A high-density physical map of *Sinorhizobium meliloti* 1021 chromosome derived from bacterial artificial chromosome library. *Proc. Natl. Acad. Sci. USA* **96**, 9357–9362.
19. Suzuki, K., Asakawa, S., Iida, M., Shimanuki, S., Fujishima, N., Hiraiwa, H., Murakami, Y., Shimizu, N., and Yasue, H. (2000) Construction and evaluation of a porcine bacterial artificial chromosome library. *Anim. Genet.* **31**, 8–12.
20. Song, J., Dong, F., and Jiang, J. (2000) Construction of a bacterial artificial chromosome (BAC) library for potato molecular cytogenetics research. *Genome* **43**, 199–204.
21. Han, C. S., Sutherland, R. D., Jewett, P. B., et al. (2000) Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. *Genome Res.* **10**, 714–721.
22. Fu, H. and Dooner, H. K. (2000) A gene-enriched BAC library for cloning large allele-specific fragments from maize: isolation of a 240-kb contig of the bronze region. *Genome Res.* **10**, 866–873.
23. Buitkamp, J., Kollers, S., Durstewitz, G., Fries, R., Welzel, K., Schafer, K., Kellermann, A., and Lehrach, H. (2000) Construction and characterization of a gridded cattle BAC library. *Anim. Genet.* **31**, 347–351.
24. Rogel-Gaillard, C., Piumi, F., Billault, A., Bourgeaux, N., Save, J. C., Urien, C., Salmon, J., and Chardon, P. (2000) Construction of a rabbit bacterial artificial chromosome (BAC) library: application to the mapping of the major histocompatibility complex to position 12q.1.1. *Mamm. Genome* **12**, 253–255.
25. Venter, J. C., Adams, M. D., Myers, E. W., et al. (2001) The sequence of the human genome. *Science* **291**, 1304–1351.
26. Kim, U.-J., Birren, B. W., Slepak, T., Mancino, V., Boysen, C., Kang, H. L., Simon, M. I., and Shizuya, H. (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**, 213–218.

Applications of Interspersed Repeat Sequence Polymerase Chain Reaction

**Heike Zimdahl, Claudia Gösele, Thomas Kreitler,
and Margit Knoblauch**

1. Introduction

Analysis of complex genomes includes characterization of complete large-insert genomic libraries comprising several hundreds of thousands of clones. Conventional methods to screen large-insert clone libraries for specific clones within a defined chromosomal interval are polymerase chain reaction (PCR) based using microsatellite markers. This strategy is labor- and cost-intensive and requires the PCR amplification of several thousands of DNA samples and verification of the PCR products via agarose gel electrophoresis. The number of PCR reactions is significantly reduced by the pooling of library clones in a three-dimensional (3D) pooling system. Nevertheless, several hundred PCR reactions are necessary to screen a P1-derived artificial chromosome (PAC) or bacterial artificial chromosome (BAC) library for one individual microsatellite marker. The application of high-density clone arrays, spotted robotically on nylon filters, offers the possibility of screening several tens of thousands of clones in a single working step. A new hybridization-based marker system has been established in the department of Hans Lehrach at the Max-Planck Institute for Molecular Genetics (Berlin, Germany). The interspersed repetitive sequence (IRS) markers are generated by amplification of genomic sequences that are located between two repetitive short interspersed repetitive elements (SINE) elements and are evenly distributed over the whole genome (*1*). IRS-PCR strategies have been applied to various species using the SINE sequences in human (Alu repeat) (*2,3*), mouse (B1 repeat) (*4,5*), rat (ID repeat) (*6*), and

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

zebrafish (DANA/mermaid repeat) (7). The application of these IRS marker systems enables high-throughput screening of large numbers of clones in a time- and cost-efficient way by hybridization.

The average size of amplified IRS-PCR products ranges from 300 to 2000 bp. The majority of IRS probes (about 99% for the rat genome) consist of unique sequences and are useful as genetic markers (data unpublished). The IRS marker strategy has already been successfully used for the construction of a genomewide integrated genetic and physical YAC and BAC map for the mouse genome (8). A related strategy is currently used by our group for the construction of an integrated genetic and physical map, for the rat genome. To construct a whole genome physical map, IRS-PCR products derived from random BAC, PAC, or YAC clones are assigned to the genetic framework map (by radiation hybrid mapping) and in parallel to YAC clones by hybridization against high-density YAC pool filters (9).

An extended IRS-PCR-based method (IRS-PCR walking) is used to construct clone contigs for defined chromosomal intervals (10) (e.g., disease regions) to facilitate positional cloning. The first step involves the identification of “anchor” yeast artificial chromosome (YAC) or PAC/BAC clones by PCR screening using defined markers (microsatellite and IRS markers) located in the region of interest (ROI). IRS-PCR probes generated from these “anchor” clones are used for the identification of flanking (overlapping) YAC clones by hybridization to YAC pool filter arrays. These steps are repeated for the generation of complete clone contigs. The YAC clone contigs can be converted into high-resolution PAC/BAC clone contigs by hybridization of purified YAC clones to high-density spotted colony PAC/BAC filter arrays. The IRS-PCR walking strategy is a cost-effective, rapid, and efficient strategy to construct contigs covering chromosomal ROIs (complete disease gene regions/QTL regions), even for large intervals spanning several centiMorgans.

2. Materials

All solutions should be made with double-distilled or deionized water. All chemicals should be of molecular biology grade.

2.1. IRS-PCR

2.1.1. Buffer and Primer

1. 10X PCR buffer: 50 mM KCl, 1.5 mM MgCl₂, 35 mM Tris base, 15 mM Tris-HCl, 0.1% Tween-20, 15 μ M cresol red. Store at -20°C (see **Notes 1** and **2**).
2. Rat IDR primer: 5'CCACTGAGCTAAATCCCCAACCCC 3'.

2.1.2. IRS-PCR Mix for Amplification of BAC/PAC Clones

1. Mix 6 μL of 10X PCR buffer, 0.6 μL of dNTPs (25 mM), 8 μL of MgCl_2 (20 mM), 0.7 μL of IDR primer (2 $\mu\text{g}/\mu\text{L}$), 0.25 μL of *Taq* polymerase (5 U/ μL), and sterile water up to 60 μL .
2. Add 1 μL of BAC culture to 60 μL of PCR mix (*see Note 3*).

2.1.3. IRS-PCR Mix for Amplification of Radiation Hybrid Panel DNA

1. Mix 6 μL of 10X PCR buffer, 0.6 μL of dNTPs (25 mM), 9 μL of MgCl_2 (20 mM), 0.5 μL of IDR primer (2 $\mu\text{g}/\mu\text{L}$), 0.2 μL of *Taq* polymerase (5 U/ μL), and sterile water up to 55 μL .
2. Add 5 μL of radiation hybrid (RH) DNA (2 ng/ μL) to 55 μL of PCR mixture (*see Note 4*). The DNA samples of the 106 rat/hamster RH cell line panel T55 are available from Research Genetics (Huntsville, AL).

2.1.4. Agarose Gel Electrophoresis

1. 1.2% Agarose gel in 1X TAE buffer, 1 $\mu\text{g}/\text{mL}$ of ethidium bromide (EtBr) (*see Notes 5 and 6*).
2. 50X TAE (1 L): 242 g of Tris base, 57.1 g of glacial acetic acid, 100 mL of 0.5 M EDTA (pH 8.0).

2.1.5. Southern Blot

1. Hybond N⁺ nylon membranes (Amersham): IRS-PCR products of the RH panel DNA are transferred onto Hybond N⁺ nylon membranes via conventional Southern blotting (*see Subheading 3.2.1*).
2. Denaturing solution: 0.5 M NaOH, 1.5 M NaCl. Store at room temperature.
3. Neutralization solution: 0.05 M Na_2HPO_4 . Store at room temperature.

2.2. Probe Labeling

2.2.1. Solutions

1. 1X TE: 10 mM Tris-HCL; 1 mM EDTA, pH 8.0.
2. TM buffer: 250 mM Tris-Cl, pH 8.0, 25 mM MgCl_2 , 50 mM β -mercaptoethanol. Store at -20°C .
3. Label solution (LS): 50 μL of HEPES buffer (1 M HEPES, pH 6.6), 50 μL of TM buffer, 14 μL of oligo hexanucleotides (Pharmacia, Uppsala, Sweden), 1:8 dilution of the stock (*see Note 7*).

2.2.2. Labeling of IRS-PCR Products

1. Make up label mix as follows: 18 μL of LS, 1.5 μL of bovine serum albumin (BSA) (10 mg/mL), 3 μL dNTP (100 μM mix of dATP, dGTP, dTTP each).
2. Add 24 μL of label mix to 16 μL of denatured IRS-PCR probes (*see Note 8*).

2.3. Filter Hybridization

1. Hybridization buffer (Church buffer): 0.5 M Na₂HPO₄, pH 7.2 (adjust with phosphoric acid); 5% sodium dodecyl sulfate (SDS); 2.5 mM EDTA, pH 8.0. Store at room temperature.
2. Washing solutions:
 - a. Washing solution I: 2X saline sodium citrate (SSC), 0.1% SDS. Store at room temperature.
 - b. Washing solution II: 0.5X SSC, 0.1% SDS. Store at room temperature.
 - c. 20X SSC (5 L): 3.0 M NaCl; 0.3 M sodium citrate, pH 7.0. Store at room temperature.
3. Stripping solution: 0.1% SDS, 2 mM EDTA. Store at room temperature.

2.4. PCR Screening of Pooled Genomic Libraries

1. PCR mix (30- μ L final volume) (for PCR buffer recipe, see **Subheading 2.1.1.**): 3 μ L of 10X PCR buffer, 0.24 μ L of dNTP mix (25 mM), 1 μ L of forward primer (6 μ M), 1 μ L of reverse primer (6 μ M), 0.2 μ L of *Taq* polymerase (5 U/ μ L), 5 μ L of DNA (lyophilized sample is diluted 1:50), and sterile water up to 30 μ L.
2. DNA of primary and secondary pools of genomic libraries is supplied by RZPD (German Ressource Center, www.rzpd.de).

2.4.1. Agarose Gel Electrophoresis (1.5% Agarose Gel)

1. 1X TAE buffer: 40 mM Tris-acetate, 1 mM EDTA.
2. EtBr (1 μ g/mL) (see **Note 6**).

2.5. IRS-PCR Walking

2.5.1. IRS-PCR

IRS-PCR conditions and solutions are described in **Subheadings 2.1.** and **2.2.** IRS-PCR products are denatured 2 min at 95°C and labeled by random priming using α -³²P-dCTP.

2.5.2. Labeling

1. Mix 16 μ L of denatured PCR product, 18 μ L of label solution, 1 μ L of BSA (10 mg/mL), 3 μ L of nonradioactive dNTP (at 100 μ M up to 1 mM), 3 μ L of α -³²P-dCTP, and 1 μ L of Klenow DNA polymerase.
2. Incubate at 37°C for 3 h or at room temperature overnight (see **Notes 7** and **8**).

2.5.3. Competition Hybridization of PCR Products

1. Competition mix (60 μ L): 100 μ g of genomic sonicated DNA, 25 mM EDTA, 0.25 mM sodium phosphate buffer (0.5 M NaH₂PO₄/0.5 M Na₂HPO₄). Labeled probes are denatured at 95°C for 10 min and incubated with 60 μ L of the competition mix for 2 h at 65°C.

2. Church buffer: 0.5 M Na₂HPO₄, pH 7.2, adjusted with phosphoric acid; 5% SDS; 2.5 mM EDTA, pH 8.0.
3. Washing solution I: 2X SSC, 0.1% SDS.
4. Washing solution II: 0.1X SSC, 0.1% SDS.

2.6. Conversion of YAC Clones into Corresponding PAC/BAC Clones

2.6.1. Purification of YAC Clones Via Pulsed-Field Electrophoresis

1. YAC broth agar (50 µg/mL of ampicillin): 20 g of glucose, 14 g of casamino acids, 0.055 g of tyrosine, 0.1 g of adenine, 100 mL of 6.7% yeast nitrogen base up to 1 L.
2. High-molecular-weight DNA is prepared in agarose plugs:
 - a. Lyticase mix: Dissolve lyticase (50,000 U) in 500 µL of SCE, and add 10 µL of lyticase (100 U/µL) to 10 mL of 1X TE and 100 µL of 1 M dithiothreitol (DTT).
 - b. SCE: 50 mL of 2 M sorbitol, 10 mL of 1 M sodium citrate, 2 mL of 0.5 M EDTA; add sterile water to 100 mL.
 - c. NDS (1 L): 30 mL of 30% sarcosyl, 10 mL of 1 M Tris-HCl, 0.5 M EDTA up to 1 L.
 - d. Proteinase K: Add 500 µL of proteinase K (100 mg/mL) to 1 L of NDS (*see Note 9*).
 - e. Phenylmethylsulfonyl fluoride (PMSF): Add 100 µL of 100 mM PMSF (in isopropanol) solution to 10 mL of 1X TE (*see Note 10*).
 - f. 1X TE: 10 mM Tris-HCl, 1 mM EDTA, pH 7.5.
3. Pulsed-field gel electrophoresis (PFGE) (Bio-Rad, Hercules, CA) for YAC DNA-agarose plugs:
 - a. 10X TBE buffer (1 L): 108 g of Tris base, 55 g of boric acid, 40 mL of 0.5 M EDTA, pH 8.0.
 - b. Conditions: 1% low-melting-point (LMP) agarose gel in 0.5X TBE buffer, 120° field angle, 6 V/cm, 40- to 100-s switch time, 24 h.

2.6.2. Labeling of YAC Clones

Purified YAC clones are labeled via random priming using Klenow enzyme and competed with sonicated genomic DNA. For label solutions, *see Subheading 2.2.1.* and for competition mix, *see Subheading 2.5.3.*

2.6.3. Sizing of PAC/BAC Clones

1. CHEF (Bio-Rad).
2. Conditions: 1% LMP agarose gel in 0.5X TBE buffer, 12 V/cm: 5- to 15-s switch time, 14 h.

3. Methods

3.1. IRS-PCR

Ten nanograms of RH DNA or 1 μ L of BAC culture (RPCI-32 Male, Brown Norway Rat BAC library, constructed by P. de Jong, <http://bacpac.chori.org>) (*II*) is used for IRS-PCR. For amplification of the DNA samples of the RH panel, a 120-fold PCR mix is recommended (*see Note 4*). PCR is performed in 96-well plates (Thermofast, AB-0600; Advanced Biotechnologies).

1. Chill PCR solutions on ice during preparation of the PCR mix.
2. Fill 96-well plates with 60 μ L of PCR mix for amplification of BAC clones and 110 μ L for amplification of the RH panel DNA using a multichannel pipet.
3. Add 1 μ L of BAC culture or 10 μ L of RH panel DNA (diluted to 2 ng/ μ L), and seal the PCR plates with plastic foil using a plate sealer (Genetix).
4. Perform a thermocycle using the following protocol: initial denaturation at 94°C for 2 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 60 s, extension at 72°C for 5 min, and final extension at 72°C for 5 min.

3.2. Filter Construction

3.2.1. RH Filter

Filters for RH mapping are obtained by alkaline transfer (Southern blot) of the IRS-PCR products of the 106 cell lines of the rat T55 RH panel from agarose gel to nylon membranes (Hybond N⁺, 20 $\times$ 4.5 cm). IRS-PCR products of rat genomic DNA of two different strains as positive controls and a “non-template control” (sterile water) as negative control are recommended.

1. Load 5 μ L of IRS-PCR products of the RH panel and controls onto a 1.2% agarose gel (20 $\times$ 20 cm). Run the gel for 10 min in 1X TAE buffer at a constant voltage of 135 V.
2. Prior to transfer denature the DNA. Place the gel in a dish immersed in denaturing solution and gently shake for 20 min. Drain off the solution and repeat **step 2**.
3. Slide the gel onto a glass plate covered with a wet sheet of Whatman paper (0.4 M NaOH) without trapping air bubbles, and put the plate on a tray containing 0.4 M NaOH.
4. Carefully place a nylon membrane that is cut to the size of the gel and rinsed in 0.4 M NaOH on top of the gel to avoid air bubbles.
5. Place a sheet of Whatman paper, a stack of paper towels, and a glass plate on top. Carry out Southern transfer onto nylon membranes overnight in denaturing solution.
6. Remove the towels and filter paper and mark the gel's position with a pen. Neutralize the filter by gentle shaking in 0.05 M Na₂HPO₄ for 5 min and air-dry.
7. Fix the DNA by exposing to ultraviolet light for 5 min (UV Stratalinker 2400; Stratagene).

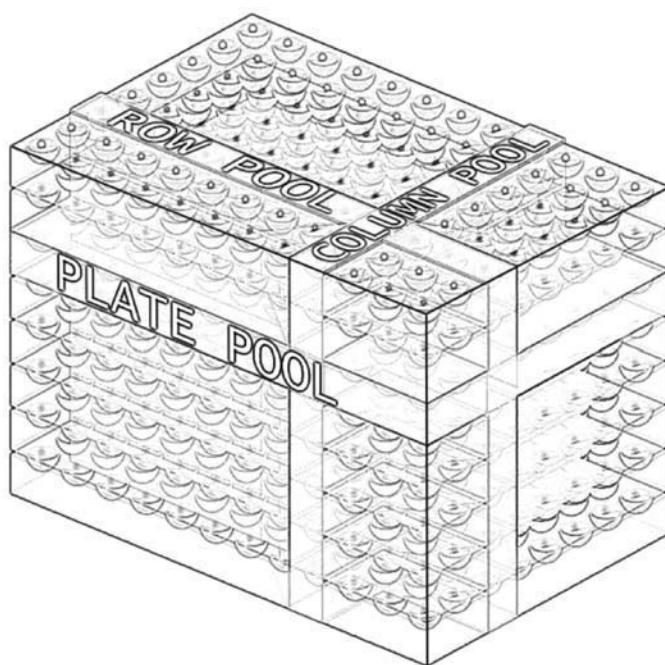


Fig. 1. 3D pooling of clone libraries. For PAC/BAC libraries stored in 384-well plates, one stack consists of 8 plate pools, 16 row pools, and 24 column pools. For YAC library stored in 96-well plates, one stack consists of 8 plate pools, 8 row pools, and 12 column pools.

3.2.2. YAC Pool Filter

High-density gridded arrays (11×7.5 cm) representing the IRS-PCR products of all 3D pools from two rat YAC libraries (**12,13**) are produced at the RZPD, Berlin (www.rzpd.de). The two YAC libraries MPIMGy916 and WIBRy933 (92,600 clones in total) represent 20-fold coverage of the rat genome. 3D pools are obtained by pooling all YAC clones from stacks of eight 96-well plates in three dimensions (plate, row, and column; see **Fig. 1**). IRS-PCR products of the pooled YAC libraries are spotted in duplicate onto nylon membranes. According to the 3D pooling system, an individual YAC clone is represented by three corresponding duplicated signals (one in each dimension: plate, row, and column). A filter example is given in **Fig. 2**.

3.3. Probe Labeling

Individual IRS-PCR products from random BAC clones (96-well plate) are labeled with α - ^{32}P -dCTP by random oligonucleotide priming (**14**). The following is a detailed working protocol for 96 BAC IRS-PCR samples.

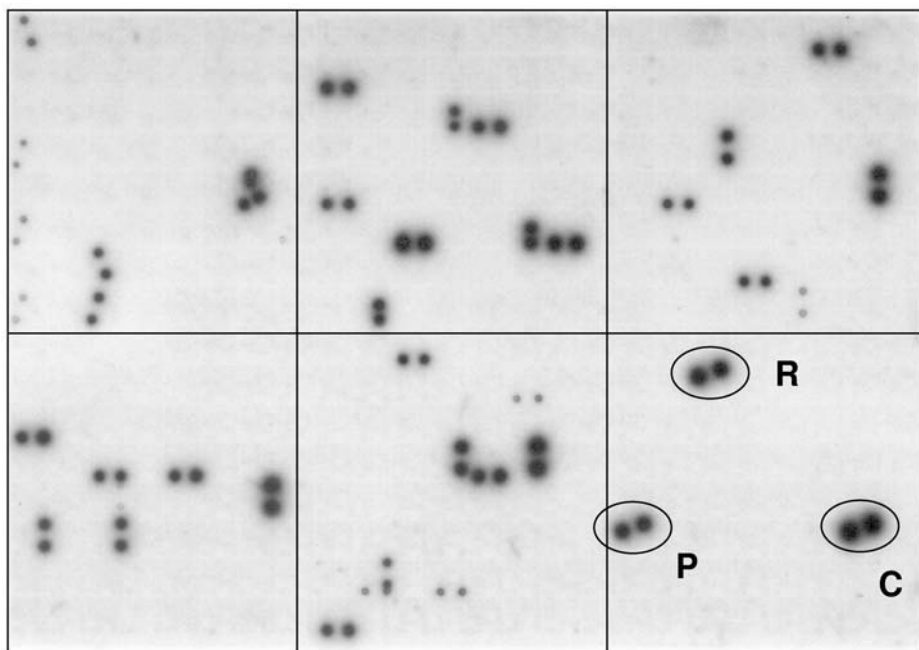


Fig. 2. A high-density robotically spotted filter carrying IRS-PCR products of 3D pooled YAC libraries is hybridized with an IRS-PCR product of a single clone to detect overlapping clones.

1. Check IRS-PCR products on a 1.2% agarose gel.
2. Load 15–20 μL of positive IRS-PCR probes onto a 1% low-melting-point agarose gel. For an 18×8.5 cm gel, 100 mL of LMP agarose is needed, 4 combs with 18 slots each for 32 samples (see **Note 5**). To estimate the fragment size, load 50 ng of Φ 174 DNA/*Bsu*RI (*Hae*III) DNA marker in the first slot.
3. Cut all DNA fragments larger than 200 bp out of the gel and transfer to 96-well plates. Add 30 μL of 1X TE to each agarose slice (see **Note 11**).
4. Melt the agarose slices at 75°C in a PCR cycler, and transfer 16 μL of each sample (see **Note 8**) into another 96-well plate for labeling.
5. Prepare the label mix (for 96 samples a 110-fold mix is recommended) in a 5-mL tube and chill on ice. A 110-fold label mix is prepared as follows:
 - a. 1-fold: 16 μL of LS, 1.5 μL of BSA (10 mg/mL), 3 μL of ATG mix (100 nM), 0.5 μL of Klenow polymerase, and 1 μL of α - ^{32}P -dCTP.
 - b. 100-fold: 1980 μL of LS, 165 μL of BSA (10 mg/mL), 330 μL of ATG mix (100 nM), 55 μL of Klenow polymerase, and 100 μL ($=1$ mCi) of α - ^{32}P -dCTP.
6. Denature the PCR probes in 96-well plates for 2 min at 94°C in a PCR cycler and chill on ice immediately.

7. Dispense 24 μL of the labeling mix to each denatured PCR probe. Seal the 96-well plates with foil (Biostat), and incubate at room temperature overnight kept in a safety box.

3.4. Filter Hybridization

The hybridization process of YAC pool filters and RH filters (in one tube) can be divided into three steps: prehybridization, hybridization with radioactive probe, and washing. For a large-scale hybridization, each of the 96 PCR probes is transferred individually into 15-mL plastic tubes (Falcon) containing 10 mL of hybridization buffer.

3.4.1. Hybridization

1. Fill Falcon tubes with 10 mL of hybridization buffer. Roll wet YAC pool filters (e.g., using a 10-mL glass pipet) and immerse in the hybridization buffer in a 15-mL tube. The spotted side of the filter carrying the DNA is outside and adheres to the inside wall of the Falcon tubes.
2. Place the wet rolled RH filters in the same tube in the center of the YAC pool filter (*see* **Notes 12** and **13**). Place the Falcon tubes in a 96-tube rack (*see* **Note 14**).
3. Add 40 μL of 100 mM NaOH to denature the labeled probes.
4. Incubate the denatured probes in a PCR cycler at 80°C for 5 min to melt the agarose slices, and add to the prehybridized filters.
5. Carefully mix the tubes by inverting them and place in a rack.
6. Put the 96-tube rack into a gently shaking water bath at 65°C. Perform hybridization overnight (*see* **Note 15**).

3.4.2. Washing

Washing is done in dishes (25 $\times$ 25 $\times$ 6 cm plastic boxes) placed in a water bath preheated to 65°C.

1. Fill dishes with 1.5 L of washing solution I.
2. Use a pair of tweezers to take the filters out of the Falcon tube, separate the YAC and RH filters (*see* **Note 16**), and put into a plastic box filled with washing solution I. A maximum of 24 filters can be washed in one plastic box (*see* **Note 16**).
3. Put the plastic boxes into a water bath and wash the filters (by shaking) for 30 min at 65°C.
4. Decant the washing solution I and add 1.5 L of prewarmed washing solution II. Wash the RH filters for 30 min and the YAC filters for 10–15 min in washing solution II.
5. Monitor the filters after washing with a Geiger counter. Monitoring between washing steps is recommended.
6. Wash filters above 50 cpm again. Drain the other filters, place on solid plastic pads, and wrap with Saran foil (*see* **Note 13**).

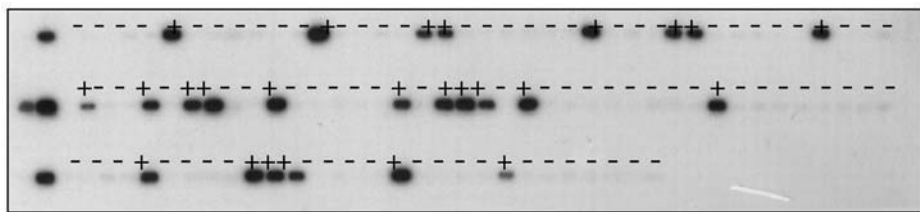


Fig. 3. RH mapping of IRS-PCR products by hybridization. The RH filter carries IRS-PCR products of genomic DNA (positive control, second sample in each row) and of the DNA of the RH panel. The third sample in each row contains sterile water as negative control.

7. Expose the filters to X-ray films for at least 24 h (*see Note 17*). Scoring of the images and evaluation of the data are described in **Subheading 3.8**.
8. Store used filters in washing solution I for at least 14 d at 4°C.
9. Dehybridize the filters in stripping solution: use 1 L for 15 filters in a pizza box. Gently shake the filters 30 min at 65°C (*see Note 18*). It is advisable to check that the probe has been efficiently removed by monitoring the filters or even autoradiographing them overnight.
10. Store the filters in washing solution II at 4°C (*see Note 19*). Filters can be used up to 10 times (*see Note 13*).

Examples of hybridized YAC pool and RH filters are given in **Figs. 2 and 3**, respectively.

3.5. PCR Screening of Genomic Libraries

The first step is to identify YAC/PAC or BAC clones within the relevant region (“anchor” clones) by PCR screening of large-insert clone libraries using microsatellite-specific primer pairs. To reduce the number of required PCR reactions, the clone libraries are pooled in a 3D system (*see Fig. 1*). The primary and secondary pools are distributed via the RZPD (www.rzpd.de).

1. Start the screening by testing primary pools consisting of purified DNA from the vector constructs of all clones from eight microtiter plates of the genomic library. About 50 ng of the pooled DNA are used in one PCR reaction.
2. Generate PCR products using microsatellite primers located in the defined region. Perform cycling according to the following conditions: initial denaturation at 95°C for 3 min, 30 cycles of denaturation at 94°C for 45 s, annealing at 55°C for 45 s, extension at 72°C for 1 min, and final extension at 72°C for 5 min.
3. Check the PCR results via 1.5% agarose gel (1 μ m/mL of EtBr) electrophoresis (*see Notes 2 and 6*). Once a positive signal within a primary pool is obtained, the screening of the corresponding secondary pools follows. The secondary pools

consist of individually pooled microtiter plates from the positive primary pool (plate pools) as well as pooled rows and columns from all eight microtiter plates of the positive primary pool. Each secondary pool consists of 28 pools (8 plate pools, 8 pooled rows, and 12 pooled columns) for YAC libraries, cultured in 96-well microtiter plates, and 48 pools for the other genomic libraries (e.g., BAC and PAC libraries), respectively which are maintained in 384-well microtiter plates.

4. Separate the PCR products via conventional 1.5% agarose gel electrophoresis (*see Notes 2 and 6*). PCR screening of each secondary pool should result in three corresponding positive signals—the plate, the row, and the column—which can be combined for determination of the exact clone name consisting of RZPD library number, plate number, and plate coordinates.

3.6. IRS-PCR Walking

The clones identified in the PCR screening are used as “starting points” for the IRS-PCR walking technique (*see Fig. 4*), which allows detection of overlapping individual IRS-PCR products by hybridization to 3D IRS-PCR product pool filters spotted robotically on nylon filters. In parallel, all probes are mapped by RH mapping to confirm the chromosomal localization by hybridization (*see Fig. 3*). For generation of the IRS-PCR YAC pool filters and RH filters used for the clone contig building, *see Subheadings 3.1–3.4*.

1. Generate IRS-PCR products using primer (IDR) complementary to the 5' sequence of the rat ID-consensus sequence. Perform cycling according to the following conditions: initial denaturation at 95°C for 3 min, 30 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 60 s, extension at 72°C for 3 min, and final extension at 72°C for 5 min.
2. Verify PCR products by loading 7 μ L of PCR product on a 1.2% agarose gel (1 μ m/mL of EtBr). Run the gel in 1X TAE buffer for 20 min at 120 V.
3. Cut positive PCR products out of a LMP agarose gel and transfer into 96-well plates (*see Subheading 3.3.*) (*see Note 5*).
4. Add 20 μ L of 1X TE to the PCR products, which can be stored at 4°C for 3 to 4 wk or at –20°C for several months.
5. Pipet 16 μ L of molten (70°C) PCR products in eight-well stripes or 96-well plates, denature at 95°C for 2 min, and label by random priming using α -³²P-dCTP (*see Subheading 2.5.*) (*see Note 7*).
6. Compete labeled probes with sonicated genomic DNA to block repetitive sequences. Add the competition mix to the labeled probes, and incubate for 10 min at 95°C followed by 2 h at 65°C.
7. Prewet high-density IRS-PCR YAC pool filters in hybridization buffer, and roll into 15-mL Falcon tubes. Place the RH filter in the center of the YAC pool filter in the same tube (*see Notes 12 and 13*).
8. Prehybridize the filters at 65°C for 30–60 min in 10 mL of hybridization buffer (*see Note 14*).

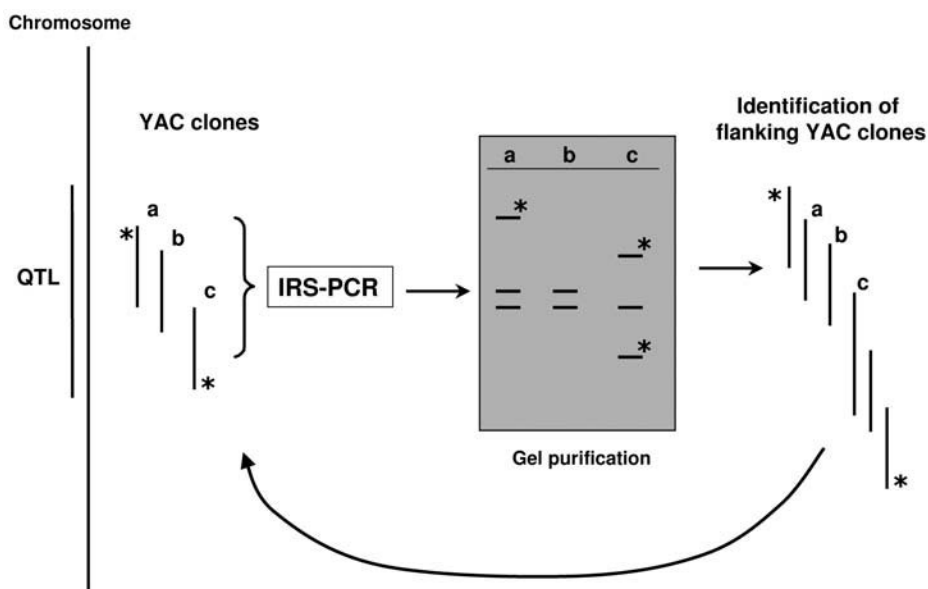


Fig. 4. IRS-PCR walking strategy for the detection of overlapping clones.

9. Add the competed probes to 10 mL of hybridization buffer. Hybridize the filters for 16 h at 65°C. Immediately transfer the competed probes into the prewarmed hybridization buffer. (Do not denature competed probes!)
10. Wash the filters in washing solution I for 30 min and in washing solution II for 15–30 min (*see Note 15*).
11. Expose the wrapped filters to X-ray films for at least 24 h (*see Notes 13 and 17*). Scoring of the images and evaluation of the data are described in **Subheading 3.8.** (*see Notes 20–23*).

3.7. Conversion of YAC Clones into PAC/BAC Clones

Physical maps and clone contigs are established in multiple steps starting with YAC clones, which are capable of carrying very large fragments (>1 Mb) of exogenous DNA in a single clone. However, working with YAC clones has disadvantages: high rates of instability and chimerism, which restrict the reliability of YAC clones for mapping and sequencing purposes. Therefore, YAC contigs have to be converted into PAC/BAC contigs. The following is a useful clone-clone hybridization method to detect corresponding PAC/BAC clones.

3.7.1. YAC DNA Preparation in Agarose Plugs

Optimal growing conditions for YAC clones are 2 to 3 d at 30°C in YAC broth medium/50 µg/mL of ampicillin.

1. Pick single colonies from the plate and grow for 48 h in 5 mL of YAC broth medium in the presence of 50 µg/mL of ampicillin.
2. Centrifuge the grown cultures at 2000 rpm for 5 min and wash twice with 1X TE buffer.
3. Treat the cells with lyticase solution (1X TE buffer, 100 mM DTT, 100 U of lyticase) for 1 h at room temperature, and embed in agarose blocks (2% LMP agarose in SCE).
4. Carefully remove agarose plugs from the bottom of the Falcon tube by adding NDS solution containing proteinase K (*see Note 9*). An inoculating loop helps to achieve this.
5. Fix the tubes on a rocker with sticky tape, and gently shake the agarose plugs for 36–48 h at 55°C.
6. Wash the agarose plugs three times on a rocker at 55°C in 10 mL of 1X TE buffer, 1X TE buffer/100 mM PMSF (*see Note 10*), and 1X TE buffer. Store the agarose plugs in 1X TE buffer/10 mM EDTA at 4°C (*see Note 24*).

3.7.2. Purification and Hybridization of YAC Clones to BAC or PAC Colony Filter

1. Use a blue inoculating loop can be used to break off an agarose chunk from one of the agarose blocks (*see Note 25*). Load the chunk onto a horizontal comb toward the end of the tooth, and seal to it using a couple of drops of molten agarose from a loop dipped into cooling 1% agarose in 0.5X TBE (*see Note 26*).
2. Pour the gel at 40–50°C. Upright the comb vertically into a gel tray bracket, and allow the gel to set around the loaded samples.
3. Precool the running buffer (0.5X TBE) to 15°C in the tank. The running time is 24 h up to 36 h at 15°C (*see Note 27*).
4. Immerse the gel in 1 µg/mL of EtBr solution for 30 min, and, if necessary, destain in water (*see Note 6*).
5. Visually determine the position of YAC bands in the gel, and confirm by conventional Southern blotting of the gel and hybridization with 10 ng of radioactively labeled rat genomic DNA (for labeling conditions, *see Subheading 3.2.*).
6. Cut identified YAC bands out of the gel, and label overnight at room temperature with 30 µCi of α -³²P-dCTP each using random priming as described. Compete the denatured probes with sonicated rat genomic DNA (final concentration of 0.1 mg/mL) for 3 h at 65°C, and hybridize at 65°C overnight against high-density rat PAC filters of library RPCI-31 Rat PAC (**15**) in 15 mL of Church buffer.
7. Manually store the coordinates of positive signals and deconvolute into plate, row, and column positions by the RZPD method (www.rzpd.de). Recheck the positive clones by RH mapping (*see Fig. 3*).

3.7.3. Sizing of PAC/BAC Clones and Detection of Overlapping Clones Using CHEF PFGE

1. Digest 0.5 µg of liquid BAC/PAC DNA with *NotI* for 3 h at 37°C, and fractionate the resulting fragments by PFGE using a CHEF system (Bio-Rad).

2. Perform electrophoresis using 1.0% LMP agarose gel in 0.5% TBE buffer for 14 h at 12 V/cm with the initial switch time ramped from 5 s to a final switch time of 15 s.
3. Use high-molecular-weight and λ HindIII ladders can be used to estimate the fragment size. Overlapping clones should share some bands.

3.8. Scoring and Evaluation of Data

Hybridizations for RH filters and YAC pool arrays are manually scored by two persons. For determination of individual clone addresses, an in-house-developed deconvolution program is used RH and YAC data are added into our database using in-house-developed scoring editor functions. The deconvolution program and C-source are available at www.molgen.mpg.de/~ratgenome/rh_maps/software. RH signals are scored as 1 for safe positive result, 0 for safe negative result, and 2 for questionable. IRS markers are placed within both available RH framework maps (for mcw framework: <http://rgd.mcw.edu/pub/maps/rhframework/v.2>; for Oxford framework: www.well.ox.ac.uk/rat_mapping_resources) in a two-step process using the RHMPPER1.22 program (Stein, 1996–1998, <http://www-genome.wi.mit.edu/ftp/pub/software/rhmapper>) (16). In a first step, chromosomal assignments for each marker are determined by two-point analysis with a threshold of logarithmic odds ratio (LOD) > 10 and the three highest linkage results within a contiguous region of the same chromosome. Mapping results are excluded as ambiguous by the following criteria: (1) >40, (2) <9, or (3) >10 unsure positive RH signals. After chromosomal assignment, markers are placed within the RH framework (17,18) of the individual chromosome with a threshold of LOD 6 and a maximal distance of 25 cR. Chromosome maps are created as PDF documents by an in-house-developed PHP script. The maps can be viewed and printed by using the freely available Acrobat Reader (www.adobe.com).

4. Notes

1. We routinely prepare 10X PCR buffer and autoclave before 750 μ L of 0.1 M sterile filtered cresol red is added. 10X PCR buffer can be aliquoted in 2-mL tubes and stored at -20°C up to 1 yr.
2. Cresol red acts as loading dye; therefore, no loading dye/buffer needs to be added before loading probes on an agarose gel.
3. BAC clones are arrayed in 384-well plates using LB medium containing 5% sucrose, and chloramphenicol (20 $\mu\text{g/mL}$). Plates are stored at -80°C . Library plates from the RPCI-32 Rat BAC (<http://bacpac.chori.org>) are obtained from the Resource Center/Primary Database with RZPD number 657 (RZPD, Berlin, Germany; for detailed information, see www.rzpd.de). To amplify a 384-well plate, a 420-fold PCR mix is recommended.

4. The volume of the IRS-PCR for RH filter can be increased to 120 μ L (110 μ L of PCR mix plus 10 μ L of RH DNA).
5. For optimal cutting out of PCR bands, slots are separated by an empty slot in between (only every second slot is loaded).
6. EtBr stock solution should be made up in a fume hood and stored in a light-protected tube or glass. Wear gloves and a laboratory coat.
7. LS aliquots can be stored at -20°C . LS is stable for at least 1 yr at -20°C , but do not thaw and refreeze too often; β -mercaptoethanol can disaggregate, which can result in poor hybridization results.
8. Up to 20 μ L of PCR product can be used in the labeling reaction if the DNA concentration is low. Safety instructions should be followed when working with radioactivity.
9. Dissolve proteinase K in NDS just before use. Pronase works just as well as proteinase K and is much cheaper.
10. PMSF is extremely toxic and should be handled with care.
11. Sealed PCR products and cut-out bands can be stored at 4°C up to 6 mo if necessary.
12. Do not roll the YAC pool and the RH filter together in one step (filters may stick together); this may result in bad hybridization results. Wear gloves.
13. Do not allow the filters to dry out, and avoid creases.
14. New filters have to be prehybridized overnight (for at least 4 h) at 65°C in hybridization buffer. For reused filter, 1 h of prehybridization is sufficient.
15. Hybridization can be performed up to 40 h if required.
16. It is recommended that in one dish (plastic box) 20 RH filters or 15 YAC pool filters be washed.
17. Problems of high background may be owing to dirt, agarose, air bubbles trapped on the filter, insufficient prehybridization, or insufficient washing.
18. Do not strip the filter longer than 1 h; this can decrease filter quality.
19. Store filters in a box or sealed in plastic bags in strip or washing solution II at 4°C .
20. Positive duplicated signals indicate overlapping clones (*see Fig. 2*) and are scored as described in **Subheading 3.8**. Because of the 3D pooling of the IRS-PCR products spotted on the filters, a clone is identified by three corresponding signals: row, column, plate (*see Fig. 2*, IRS-PCR product hybridized to a YAC pool filter). The identified clones contain sequences overlapping with the hybridization probe. In a next step, they can be used for IRS-PCR again. Because of the occurrence of overlapping sequences, these clones share some PCR products. For the next hybridization round, only the unique PCR products corresponding to nonoverlapping regions of the clones are used for hybridization against the IRS-PCR pool filters to identify additional overlapping clones. The IRS-PCR walking procedure is summarized in **Fig. 4**.
21. By repeating these steps a clone contig for a certain chromosomal region can be constructed. The physical size of the overlapping clones and the whole contig is

determined by PFGE on 1% agarose gels, Southern blotting, and hybridization using the radioactively labeled genomic DNA as hybridization probe. IRS-PCR walking has several advantages over the conventional sequence-tagged site (STS) content mapping: the simultaneous screening of multiple libraries, the analysis of a large number of clones at once, and the identification of multiple independent clones in a single step. By eliminating sequencing steps, STS generation, and primer synthesis, the efficiency is further increased and the costs of contig assembly are reduced significantly.

22. The identified large-insert clones can be used next to identify expressed sequences for the construction of a transcript map by hybridization against cDNA filters.
23. A clone-clone hybridization with YAC clones is technically difficult because of the large amount of repetitive DNA in the inserts. This can be overcome by blocking repetitive sequences in the probe by competition, but this will result in weak hybridization signals. Clone-clone hybridization is feasible using BAC or PAC clones. How YAC clones can be translated into BAC or PAC clones by hybridization is described in **Subheading 3.7**.
24. Agarose plugs in 1X TE buffer/10 mM EDTA can be stored at 4°C for several months. If DNA in agarose plugs is used for PCR, wash agarose plugs twice in 1X TE buffer to remove the EDTA.
25. Use “clean” DNA, such as agarose blocks for PFGE, not crude lysates. Liquid DNA should be loaded carefully onto gel to avoid shearing (use cutoff tips).
26. If agarose plugs are cut too big for the slot size they may touch their neighbor. This can be avoided by leaving a blank slot in between. Loaded samples should not be too long (vertically) to ensure that all the DNA has roughly the same start point.
27. The ramp is set depending on the range of YAC/PAC sizes to be resolved. Pulses of 100 s are required for a 1.5-Mb YAC. A YAC may typically lie in the range of 200 kb to 1.5 Mb. An initial pulse of 40 s can be chosen if resolution of smaller fragments is not critical.

Acknowledgments

We wish to thank Dr. Kathrin Meissner for helpful discussions and suggestions. This work was supported by European Community Grant BIO4 TC960372 and by German Human Initiative Grant 01KW9607/0.

References

1. Ledbetter, S. A., Nelson, D. L., Warren, S. T., and Ledbetter, D. H. (1990) Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. *Genomics* **6**(3), 475–481.
2. Cayanis, E., Russo, J. J., Kalachikov, S., et al. (1998) High-resolution YAC-cosmid-STS map of human chromosome 13. *Genomics* **47**(1), 26–43.
3. Kass, D. H. and Batzer, M. A. (1995) Inter-Alu polymerase chain reaction: advancements and applications. *Anal. Biochem.* **228**(2), 185–193.
4. McCarthy, L., Hunter, K., Schalkwyk, L., et al. (1995) Efficient high-resolution genetic mapping of mouse interspersed repetitive sequence PCR products, toward

- integrated genetic and physical mapping of the mouse genome. *Proc. Natl. Acad. Sci. USA* **92**(12), 5302–5306.
5. Hunter, K. W., Riba, L., Schalkwyk, L., Clark, M., Resenchuk, S., Beeghly, A., Su, J., Tinkov, F., Lee, P., Ramu, E., Lehrach, H., and Housman, D. (1996) Toward the construction of integrated physical and genetic maps of the mouse genome using interspersed repetitive sequence PCR (IRS-PCR) genomics. *Genome Res.* **6**, 290–299.
 6. Kim, J. and Deininger, P. L. (1996) Recent amplification of rat ID sequences. *J. Mol. Biol.* **261**, 322–327.
 7. Shimoda, N., Chevrette, M., Ekker, M., Kikuchi, Y., Hotta, Y., and Okamoto, H. (1996) Mermaid, a family of short interspersed repetitive elements, is useful for zebrafish genome mapping. *Biochem. Biophys. Res. Commun.* **220**(1), 233–237.
 8. Himmelbauer, H., Schalkwyk, L. C., and Lehrach, H. (2000) Interspersed repetitive sequence (IRS)-PCR for typing of whole genome radiation hybrid panels. *Nucleic Acids Res.* **28**(2), e7.
 9. Goesele, C., Hong, L., Kreitler, T., et al. (2000) High-throughput scanning of the rat genome using interspersed repetitive sequence-PCR markers. *Genomics* **69**(3), 287–294.
 10. Hunter, K. W., Ontiveros, S. D., Watson, M. L., et al. (1994) Rapid and efficient construction of yeast artificial chromosome contigs in the mouse genome with interspersed repetitive sequence PCR (IRS-PCR): generation of a 5-cM, >5 megabase contig on mouse chromosome 1. *Mamm Genome* **5**(10), 597–607.
 11. Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52**(1), 1–8.
 12. Cai, L., Schalkwyk, L. C., Schoeberlein-Stehli, A., Zee, R. Y., Smith, A., Haaf, T., Georges, M., Lehrach, H., and Lindpaintner, K. (1997) Construction and characterization of a 10-genome equivalent yeast artificial chromosome library for the laboratory rat, *Rattus norvegicus*. *Genomics* **39**, 385–392.
 13. Haldi, M. L., Lim, P., Kaphingst, K., Akella, U., Whang, J., and Lander, E. S. (1997) Construction of a large-insert yeast artificial chromosome library of the rat genome. *Mamm. Genome* **4**, 284.
 14. Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13.
 15. Woon, P. Y., Osoegawa, K., Kaisaki, P. J., Zhao, B. H., Catanese, J. J., Gauguier, D., Cox, R., Levy, E. R., Lathrop, G. M., Monaco, A. P., and DeJong, P. J. (1998) Construction and characterization of a 10-fold genome equivalent rat P1-derived artificial chromosome library. *Genomics* **50**(3), 306–316.
 16. Stein, L. (1996) RHMAPPY, Installation and user's guide. <http://www-genome.wi.mit.edu/ftp/pub/software/rhmapper>.
 17. Steen, R. G., Kwitek-Black, A. E., Glenn, C., et al. (1999) A high density integrated genetic linkage and radiation hybrid map of the laboratory rat. *Genome Res.* **9**, AP1–AP8.
 18. Watanabe, T. K., Bihoreau, M.-T., McCarthy, L. C., et al. (1999) A radiation hybrid map of the rat genome containing 5,255 markers. *Nat. Genet.* **22**, 27–36.

BAC Mapping Using Fluorescence *In Situ* Hybridization

Xiao-Ning Chen and Julie R. Korenberg

1. Introduction

The ultimate goal of the Human Genome Project is to establish the DNA sequence of human and model organism genomes as the critical first step in understanding disease, development and evolution. To accomplish this goal and a broad spectrum of applications requires integration of genome sequence information to genetic markers (expressed sequence tag/cDNA/gene transcripts content) and to reagents that can be seen through a microscope and linked to cytogenetic landmarks. Such linkage/integration should be dense large fragments and for reagents well characterized with respect to low-copy repeats that are present at multiple other points in the genome. Therefore, ideally, the same templates should be used as an integrating framework of entry points for sequencing and then applied to gene isolation and mapping, studies of genome organization and evolution, and a myriad of clinical applications (1).

Fluorescence *in situ* hybridization (FISH) is a powerful technique for detecting and mapping the position of DNA or RNA sequences in cells, tissues, and tumors. This molecular cytogenetic technique enables the localization of specific DNA sequences within interphase chromatin and metaphase chromosomes and the identification of both structural and numerical chromosome changes. FISH is quickly becoming one of the most extensively used cytochemical staining techniques owing to its sensitivity and versatility, and with the improvement of current technology, its cost-effectiveness (2).

Bacterial artificial chromosomes (BACs) are ideal for the purpose of integrating sequence to genetic markers. They have been the major vectors used in

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

genome sequencing. BACs are also well suited for FISH, because they represent a stable and easily manipulated form of cloned DNA that is readily sequenced and produce bright, well-defined signals on metaphase and interphase chromosome preparations (**1**). BACs have been ideal reagents for linking cytogenetic marks with DNA sequence.

We have described a detailed method of obtaining fluorescent signals on single bands of human metaphase chromosomes using BAC DNA probes imaged with either a conventional or charge-coupled device (CCD) camera linked to a fluorescence microscope (**3**). This procedure includes the steps described in the protocol provided herein (*see* **Note 1**).

2. Materials

2.1. Equipment and Supplies

1. Fluorescence microscope (Zeiss Axiovert 135) with seven fluorescence filter sets (Chroma Technology).
2. Cooled-CCD camera (Photometrics CH250).
3. Phase microscope (Zeiss Axiophot 20).
4. Slide warmer (Precision).
5. Spectrophotometer (Beckman) or fluorometer (Turner Designs).
6. Shaking water bath (Precision).
7. Incubator oven (Fisher) (set to 37°C).
8. S/P Brand superfrost slides (25 × 75 mm) (Allegiance).
9. S/P Brand cover glass (22 × 50 mm, 22 × 40 mm, 22 × 22 mm) (Allegiance).

2.2. Chemicals

1. RPMI-1640 medium (Life Technologies, Gaithersburg, MD). Store at 4°C.
2. L-Glutamine (Sigma, St. Louis, MO). Store at -20°C.
3. Fetal bovine serum (Sigma). Store at -20°C.
4. Penicillin/streptomycin (Life Technologies). Store at -20°C.
5. Phytohemagglutinin (10 mg/mL) (Life Technologies). Store at 4°C.
6. 5-Bromodeoxyuridine (Sigma).
7. 10⁻⁵ M Thymidine (Life Technologies).
8. Colcemid (10 mg/mL) (Life Technologies).
9. RNase (Life Technologies), DNase-free: 20 mg/mL in sterile water; boil at 100°C for 10 min. Aliquot and store at -20°C.
10. DNA extraction kit (NucleoBond Nucleic Acid Kit [ClonTech]; Qiagen Plasmid Midi/Maxi kit [Qiagen]).
11. Agarose (Life Technologies).
12. Chloramphenicol (Sigma).
13. Potassium acetate (KAc) (Sigma).
14. Phenol:chloroform (Life Technologies).
15. Sodium acetate (Sigma).
16. Nicktranslation kit and Bionicktranslation kit (Life Technologies). Store at -20°C.

17. 1 mM Digoxigenin-11-dUTP (Boehringer Mannheim). Store at -20°C .
18. Fluorolink Cy3-dCTP (Amersham Pharmacia Biotech). Store at -20°C .
19. Fluorolink Cy5-dCTP (Amersham Pharmacia Biotech). Store at -20°C .
20. DNase I (Gibco-BRL, Gaithersburg, MD).
21. G50 Sephadex Quick spin column (Life Technologies).
22. 1-kb DNA marker (Gibco-BRL).
23. Formamide for denaturation (EM Science). Store at -20°C .
24. Dextran sulfate (Sigma).
25. Salmon sperm DNA (3'-5') Store at -20°C .
26. Human COT1TM DNA (1 mg/mL) (Life Technologies).
27. Tween-20 (Sigma).
28. Formamide for posthybridization washing step (Fisher). Store at 4°C .
29. Bovine serum albumin (BSA) (Sigma). Store at 4°C .
30. Avidin-conjugated fluorescein isothiocyanate (FITC) (Vector).
31. Sheep-antidigoxigenin-Rhodamine (Boehringer Mannheim).
32. Chromomycin A3 (0.5 mg/mL, in 50% McIlvaine's buffer) (Sigma).
33. Distamycin A (0.1 mg/mL in 50% McIlvaine's buffer) (Sigma).
34. 0.5 M EDTA, pH 8.0 (Life Technologies).

2.3. Buffers and Other Solutions

1. Hank's balanced salt solution (HBSS) (Life Technologies).
2. Hypotonic solution (0.075 M KCl): 5.6 g of KCl/L of purified water.
3. Fixative: acetic acid/methanol (1/3 [v/v]). Make fresh right before use and keep on ice.
4. Terrific broth (Life Technologies): Add 48.2 g/L of purified water. Mix well to dissolve, add 8 mL of glycerol, and then autoclave at 121°C for 15 min. Cool to 50°C , and add chloramphenicol to 12.5 $\mu\text{g/mL}$.
5. Denaturing solution: Add 35 mL of formamide; 10 mL of distilled water; and 5 mL of 20X saline sodium citrate (SSC), pH 7.0. Store at 4°C . Prepare fresh every 2–4 wk.
6. Ethanol series: prepare 70, 80, and 100% ethanol in distilled water and keep on ice.
7. Hybridization master mix: 10% (v/v) dextran sulfate, 50% formamide, 2X SSC, pH 7.0.
8. 50% Formamide/2X SSC wash solution: Add 15 mL of 20X SSC; 60 mL of distilled water; and 75 mL of formamide, pH 7.0. Store at 4°C .
9. SSC: 20X SSC consists of 3 M NaCl, 0.3 M Na_3 citrate, pH 7.0.
10. 4X SSC/0.1% Tween-20: Add 100 mL of 20X SSC, 400 mL of distilled water, and 500 μL of Tween-20. Mix well.
11. TE (Tris-EDTA) buffer: 1X TE consists of 10 mM Tris-HCl, 1 mM EDTA, pH 7.4, 7.5, or 8.0.
12. Tris-HCl: 1 M, pH 7.5 or 8.0.
13. McIlvane's buffer (pH 9.0): 0.63 of citric acid, 6.19 g of sodium phosphate dibasic, and 500 mL of purified water.

14. Antifade solution (4): Dissolve 100 mg of *p*-phenylenediamine dihydrochloride (Sigma) in 10 mL of phosphate-buffered saline (PBS). Adjust to pH 8.0 with 0.5 M bicarbonate buffer (0.42 g of NaHCO₃ in 10 mL of water, pH 9.0 with NaOH). Add to 90 mL of glycerol. Store in aliquots at -20°C. The solution darkens with time but remains effective even after a few years.

3. Methods (3)

3.1. Metaphase Preparation and Slide Selection

1. Grow human peripheral lymphocytes for 72 h at 37°C in RPMI-1640 supplemented with L-glutamine (2 mM), 15% fetal calf serum, penicillin (100 IU/mL), streptomycin (0.05 mg/mL), and 0.02% phytohemagglutinin.
2. Block the cells in S-phase by adding 5-bromodeoxyuridine (0.8 mg/mL) for 16 h.
3. Wash the cells once with HBSS to remove the synchronizing agent, and release the cells by additional incubation for 6 h in supplemented medium (*see step 1*) with 2.5 µg/mL of thymidine.
4. Harvest cultures by the addition of 0.1 µg/mL of colcemid for 10 min followed by treatment with 10 mL of 0.075 M KCl for 15 min at 37°C and fixation four times in a freshly made solution of methanol and glacial acid (3:1 [v/v]).
5. To obtain high-quality chromosome preparations, prepare metaphase spreads by letting one drop of cell suspension fall 15 in. onto alcohol-cleaned slides, then placing above a container filled with heated (close to boiling) water for 20–40 s. This time varies with the ambient humidity and with the individual cell preparations. It must be determined for each cell preparation and checked using a phase contrast microscope. If cytoplasmic residue is visible around the metaphase spread, wash the remaining cells with fixative several times before beginning dropping. Ideally, the chromosomes in metaphase spreads should appear dark black. If chromosomes appear glassy or refractile, this suggests that the cells have dried too slowly. The slides are then kept in the dark for at least 2 to 3 wk at room temperature and stored at -70°C until use (*see Note 2*).
6. Review slides before using them for FISH. Look at the slide under phase contrast (×10), select a region of interest containing at least five metaphase spreads per field, and mark the area at the edges of the cell spreads with a diamond-tip pen.
7. The RNase treatment step is usually omitted if the slides are aged more than 2 wk. In general, it does not affect the signal-to-noise ratio. If RNase treatment is implemented, use 100 µg/mL of RNase for 30 min at 37°C followed by dehydration through an ethanol series (70, 90, and 100%).

3.2. Extraction of BAC DNA and Probe Labeling

3.2.1. Isolation of BAC DNA

The following protocol from Qiagen is good for preparing 20–50 µg of DNA (Qiagen Plasmid Midi/Maxi kit). Alternative protocols employ the NucleoBond BAC Miniprep kit or the NucleoBond BAC Maxi kit (Clontech) according to the manufacturer's directions (*see Note 3*).

1. Streak the bacterial culture on a Luria Bertani (LB) agar plate containing chloramphenicol (12.5 $\mu\text{g/mL}$) for overnight growth at 37°C.
2. Pick up single colonies and grow in 100 mL of LB containing chloramphenicol (12.5 $\mu\text{g/mL}$) overnight at 37°C. Try to achieve an $\text{OD}_{600\text{nm}}$ of 1.4–1.6.
3. Spin the culture in 250-mL buckets using a GSA rotor/Sorval at 5000 rpm for 15 min.
4. Resuspend the bacterial pellet in 10 mL of 50 mM Tris-HCl, pH 8.0, with 10 mM EDTA and RNase A (100 $\mu\text{g/mL}$). Pipet up and down using a 10-mL pipet and resuspend completely. Transfer the cells to Oak Ridge tubes (Allegiance).
5. Add 10 mL of 0.2 M NaOH and 1% sodium dodecyl sulfate (SDS). Mix the tubes gently by inverting them repeatedly but slowly 10 times. Incubate at room temperature for 5–10 min, and make sure the solution turns from turbid to translucent.
6. Add 8 mL of chilled 3 M KAc, pH 5.5. Mix the tubes gently by inverting slowly 10 times. Excessive violent mixing will increase contamination with *Escherichia coli* genomic DNA. Allow to stand on ice for 15 min.
7. Centrifuge at 4°C in a Sorvall RC 28S centrifuge using an SS34 rotor for 8 tubes or an SA-600 (12 tubes). Spin at ~33,000g or more (16,000 rpm for the SA-600 rotor) for 30 min and promptly remove the supernatant.
8. Carefully remove the supernatant using a 25-mL pipet while avoiding debris. Place the supernatant in a clean autoclaved Oak Ridge tube.
9. Centrifuge again as in **step 7**. Even though the supernatant may appear clear, you must centrifuge again because skipping this step will cause the column to clog. Transfer the supernatant to a fresh 50-mL tube and keep on ice.
10. Equilibrate QIAGEN-tip 100 columns with 4 mL of QBT buffer.
11. Apply the supernatant from **step 9** to a QIAGEN-tip 100 column.
12. Wash the QIAGEN-tip 100 with 10 mL of QC buffer twice.
13. Elute the DNA from the column with 5 mL of 60°C QF Buffer. Collect the eluent in a 14-mL culture tube containing 5 mL of isopropanol. Use VWR cat. no. 60818-725 culture tubes because these can withstand high g-forces; thinner-walled tubes can break. Oak Ridge tubes can be substituted but the pellet is more difficult to observe.
14. Centrifuge the tubes at 11,700g for 30 min at 4°C. For the Sorval SA-600 rotor, use 9000 rpm.
15. Wash the pellet with 5 mL of 70% EtOH at room temperature. Spin again for 10 min. Carefully pour off the supernatant. Air-dry for 10 min. This step removes excess salt from the BAC DNA preparation.
16. Resuspend the pellets in 400 μL of TE buffer using pipet tips with the ends cut off. Because BAC DNAs are large, it may take some time to totally resuspend the DNA. It may be necessary to let the tubes sit overnight at room temperature until the pellets are dissolved into the TE buffer. Transfer the DNA to a clean 1.5-mL Eppendorf tube.
17. The following phenol-chloroform extraction results in higher-quality DNA than the DNAs extracted from the steps described above. Add 400 μL of phenol/chloroform (1:1) and vortex. Spin for 5 min at a maximum speed of any micro-centrifuge.

- a. Remove and transfer the aqueous phase (top) to clean 1.5-mL tubes.
 - b. To precipitate the DNAs, add 40 μ L of 3 M NaOAc and 1100 μ L of cold EtOH and keep at -20°C for 2 h.
 - c. Centrifuge at maximum speed for 30 min, and wash the pellets with 750 μ L of cold 70% EtOH twice.
 - d. Dissolve the pellets in 50 μ L of TE buffer, and leave the tube at 4°C overnight or at 37°C overnight if the pellets are big.
18. Expect a total yield of 20–50 μ g for concentrations of about 0.5–1 μ g/ μ L when resuspending in 50 μ L.
 19. Determine the BAC DNA yield by ultraviolet spectrophotometry or fluorometry. Confirm BAC integrity by agarose gel electrophoresis.

3.2.2. Probe DNA Labeling

The following procedures use the protocol provided in the Nicktranslation kits (Gibco-BRL, Life Technologies) (*see Note 4*).

1. Select the nucleotide to be used from one of the five mixes that contain all dNTPs except those to be used for tagging with biotin or digoxigenin, and thaw it on ice. Keep the DNase I/Polymerase (POL) I mix on ice or at -20°C before use.
2. Add the following reagents to a 1.5-mL microcentrifuge tube placed on ice and mix briefly: 5 μ L of dNTP mix, X μ L of solution containing 1 μ g of test DNA, 1 μ L of digoxigenin-11-dUTP (for digoxigenin labeling), and Y μ L of distilled water, for a 42- μ L total volume. (The volumes of X and Y may vary depending on the concentration of DNA.)
3. Add 5 μ L of DNase I/Pol mix and 3 μ L of a 1/1000 dilution of DNase I (3 mg/mL). Mix gently but thoroughly. Centrifuge for 5 s in a microcentrifuge to bring down the solution from the cap.
4. Incubate at 15°C for 60 min.
5. Take 5 μ L of the labeling solution and run on a 1.2% nondenaturing agarose gel to check the fragment size by comparing with a 1-kb DNA marker.
6. If the fragment size is within the size range of 100–500 bp, add 5 μ L of Stop Buffer. If the size is >500 bp, add 1/100 diluted DNase I (3 mg/mL) to the tube and incubate for 15–40 min (the time length will depend on the measured size of the DNA fragments).
7. Separate unincorporated nucleotides by chromatography (Sephadex G-50) or by using ethanol precipitation.

3.3. Fluorescence In Situ Hybridization

3.3.1. Hybridization and Detection (*see Notes 5 and 6*)

1. Denature chromosome slides at 67 – 70°C in 70% formamide/2X SSC (0.15 M NaCl, 0.015 M sodium citrate) for 10 s to 2 min. (Note that fresh chromosome preparations need a shorter time.)

2. Make up hybridization solution as follows: 100–200 ng of probe DNA, 3 μg of Cot 1 DNA, and 7 μg of sonicated salmon sperm DNA/10 μL of hybridization mixture (70% formamide, 10% dextran sulfate, and 2X SSC).
3. Denature the hybridization solution at 75°C in a water bath for 5 min, and transfer to a 37°C water bath for 30 min for preannealing.
4. Apply 10 μL of denatured and preannealed solution from **step 3** to denatured chromosome slides. Place a cover slip on top of the solution, squeeze bubbles out, and seal with rubber cement.
5. Incubate the slides in a humidified chamber at 37°C overnight.

3.3.2. Posthybridization Washes

1. Wash slides four times in Coplin jars placed in a buffer containing 2X SSC and 50% formamide for 5 min each at 44°C. For directly labeled probes, go to the chromosome counterstaining step in **Subheading 3.4**.
2. Wash slides three times at 50°C in a shaking water bath for 5 min each in 0.1X SSC with gentle shaking.
3. Block sites of specific hybridization on the slides with 100 μL of 4X SSC, 3% BSA, and 0.1% Tween-20 covered with 22 $\times$ 50 cm cover slips and incubate at 37°C for 20 min.
4. Remove the cover slips and drain the blocking solution very briefly. Then replace the blocking solution with a detection solution. For biotinylated probes, avidin-conjugated FITC is used (5 $\mu\text{g}/\text{mL}$ in 0.4 $\mu\text{g}/\text{mL}$ of 4X SSC, 1% BSA, and 0.1% Tween-20). For digoxigenin-labeled probes, sheep-antidigoxigenin antibody is used (0.4 $\mu\text{g}/\text{mL}$ in 4X SSC, 1% BSA, and 0.1% Tween-20).
5. Cover the chromosomal area with cover slips and incubate the slides in a chamber (or a large Petri dish) at 37°C for 30 min.
6. Remove the cover slips and wash the slides three times in 2X SSC, 0.1% Tween-20 at 42°C for 5 min each.

3.4. Chromosome Counterstain (R Banding)

To view chromosome bands and fluorescence signals simultaneously, chromomycin A3 and distamycin A are used as a counterstain (**3**) (*see Note 7*). This reverse banding pattern generates a more reproducible and higher-resolution banding pattern than the Q-type pattern revealed by 4'-diamidino-2-phenylindole (DAPI) (**5**). Although the emission spectrum of chromomycin overlaps that of FITC, it can be separated by using the appropriate filter combination.

1. Rinse the slides briefly in 50% McIlvaine's buffer (pH 9.0) prior to staining, and shake off the excess fluid.
2. Place 100 μL of chromomycin A3 (0.5 mg/mL in 50% McIlvaine's buffer, pH 9.0) onto the slides for 40–60 min at room temperature.
3. Rinse the slides in 50% McIlvaine's buffer for 1 min at room temperature and shake off the excess fluid.

4. Place 50 μ L of 0.1 mg/mL distamycin A on the slide for 1 to 2 min at room temperature.
5. Rinse the slides in 50% McIlvaine's buffer very briefly (10–20 s).
6. Place a very thin layer of antifade solution on the slides and cover with cover slips (20 $\times$ 50 cm).

3.5. Microscopy, Photography, Image Capture, and Analysis

3.5.1. Microscopy

Analysis of *in situ* hybridization preparations may be performed by visual inspection, photography, or electronic image capture combined with digital image processing, using two different types of Zeiss fluorescence microscopes. The Zeiss Axiophot 100 microscope was used for generating black-and-white photographs from experiments using single-color FISH. Kodak Technical pan ASA100 black-and-white films were used. For capturing color images from single, dual, or multicolor FISH experiments, the Zeiss Axiovert 135 microscope was equipped with a 200-W mercury lamp and combined with a Photometrics Cooled-CCD camera employing BDS (Biological Detection System) image software (*see Note 8*).

3.5.2. Image Capture and Analysis

To view multiple labeled probes that have been simultaneously hybridized to chromosome slides, images are viewed sequentially with single-bandpass or with multiple-bandpass filter sets. In our hands, the images were acquired using a Plan-APO 63X/1.40 oil. objective and filter sets for excitation and observation of Texas Red/rhodamine and FITC or Cy3 and Cy5 fluorescence, respectively (ChromaTechnology, Brattleboro, VT). Chromomycin A3 and distamycin A reverse-banded chromosomes are captured by using Quinacrine filter sets (excitation: 440 nm; emission: 495 nm) (*see Note 9*).

To map a single BAC, a total of 20 metaphase cells are chosen in the best area (this area should have the highest signal-to-noise ratio, best metaphase spreads) of the slide to be evaluated. Signals per cell are counted, and images from two to four cells are acquired and stored in an image database. Only the raw images are saved. No enhancement, correction, or modification should be performed at this stage. A BAC carrying a human-specific sequence should be localized to a unique chromosomal band in >50% of the cells viewed. If multiple locations are seen, a second BAC colony should be picked and all of the steps just described repeated (*see Note 10*).

4. Notes

1. FISH, combined with a well-characterized reagent resource, is one of the most rapid and accurate methods for mapping human genes and novel chromosomal

break points. Moreover, the production of a reproducible high-resolution banding pattern has been highly useful for the FISH assignment of DNA fragments to human metaphase chromosomal bands essential for the analysis and identification of chromosome rearrangements. Mapped and sequence-linked BACs provide an ideal vector-insert combination for human disease genetic analyses as well as for applications including gene isolation, large-scale sequencing, and molecular cytogenetic diagnosis and prognosis. We have described here a detailed protocol for performing FISH analyses with BAC DNAs that may be applied to a broad spectrum of molecular cytogenetic analyses. To produce consistent FISH mapping results, the points given in **Notes 2–10** should be taken into consideration. These involve each step, from chromosome preparation to image capture and analysis. One of the most important elements is to begin with high-quality chromosome preparations. The DNA denaturation, hybridization, and probe detection parameters appear to be less important. We describe these issues in sequential order as we go through the protocol.

2. The quality and pretreatment of metaphase preparations are both critical to the success of signal production. A high-quality chromosome preparation should meet the following criteria: (1) metaphase spreads that are evenly distributed on the slides with as little residual cytoplasm as possible; (2) few overlapping chromosomes; (3) metaphase spreads that appear dark black. To achieve this quality, the following parameters must be adjusted: hypotonic treatment time and fixation of cell pellets; dropping of cell suspension and steaming of slide during preparation; chromosome slide baking; and, finally, length of chromosome denaturation. The appropriate adjustments should result in chromosomes that are dark and not refractile. These should provide enhanced DNA target accessibility to the probe without undesirable loss of DNA from the target.
3. The best preparation of BAC DNAs is performed with kits that can be obtained from several companies. The DNA should be free of contaminants that might inhibit DNA polymerases such as metals and detergents. In many cases, an alkaline lysis DNA isolation protocol followed by phenol-chloroform extraction and isopropanol precipitation is sufficient.
4. Several different labeling methods can be used to map BACs using FISH. These include nick translation or random priming, each employing either direct or indirect labeling reagents. Probes labeled with Nicktranslation give much higher hybridization signal-to-background noise ratio than those labeled with random priming; however, Nicktranslation uses 10 times more DNAs. The commercially available reagents used for direct labeling (i.e., fluorochrome-labeled nucleotides) are more expensive than the reagents used for indirect labeling (i.e., biotin or digoxigenin). However, both labeling methods give equally bright signals when using BACs as probes. In our hands, Nicktranslation was often used in the presence of either biotin-14-dATP or digoxigenin-11-dUTP (Boehringer Mannheim) for indirect labeling, and fluorolink Cy5-dCTP and fluorolink Cy3-dCTP (Amersham Pharmacia Biotech) was used for direct labeling employing a Nicktranslation kit (Gibco-BRL). The procedures are performed according

to the manufacturer's instructions, except for DNase concentration and length of incubation of nicktranslation (as described in this chapter). The fragment size after labeling is important and should be in the range of 300–500 bp. A fragment size of 1000 bp or above results in high background and less intense signals, contributing to failure of experiments.

5. The blocking and detection steps are important, but straightforward as long as the correct solutions and procedures are used. However, a word of caution is in order for the step involving immunocytochemical signal detection. The slides should *never* be allowed to dry, not even part of them. Although it is important to drain the liquid from the slides to prevent overdiluting the antibodies, the next solution such as the blocking solution or antibodies should be applied immediately. If the slides are allowed to dry, the signal-to-noise ratio can decrease significantly owing to highly increased background.
6. The antibodies should be kept at -20°C in small aliquots. If stored for more than 1 yr, the antibody concentration should be increased twofold. If it yields a high background, a Sephadex G-50 column can be used to purify the antibody. Usually this will restore the signal-to-noise ratio, but if it does not give sufficient signal, a fresh antibody should be obtained.
7. As we described previously (3), the best banding is obtained by using the most aged chromosomes, namely from months to as many years as 10 yr. However, aged chromosomes do not produce the strongest signals. Therefore, to obtain optimal signals with clear banding, freshly prepared chromosomes need to be baked at $50\text{--}55^{\circ}\text{C}$ for about 4 h (they can be left for 15 h with no change in signal). Aging chromosomes at room temperature ($20\text{--}25^{\circ}\text{C}$) for 2–4 wk will yield the best banding results plus reasonable bright hybridization signals. After 1 mo, the slides can be stored at -70°C until use. For fresh slides, the duration of staining with chromomycin A3 should be extended to 1.5 h, and for aged slides, the staining time may be as little as 30 min. The antifade solution is crucial to prevent fading, but only a thin layer should be used.
8. The metaphase selected for image capture and storage should meet the following criteria: (1) strong but homogeneous hybridization signals, (2) low background, (3) little or no overlap of chromosomes, and (4) relatively even condensation along the length of the chromosomes.
9. Most fluorochromes fade quickly, including the reverse banding obtained with chromomycin. Thus, expose the slides to any excitation light only for the minimum amount of time.
10. To map each BAC DNA, count at least 10–20 cells and save a minimum of two images for analysis and databasing. If needed, signals should be enhanced to identify additional sites that may have been missed owing to low signal intensity. If BAC DNAs are used for cytogenetic analysis, a standard quality control protocol should be developed. This is crucial to the success of experiments.

References

1. Korenberg, J. R., Chen, X. N., Sun, Z., et al. (1999) Human genome anatomy: BACs integrating the genetic and cytogenetic maps for bridging genome and biomedicine. *Genome Res.* **9**, 994–1001.
2. Nath, J. and Johnson, K. L. (2000) A review of fluorescence in situ hybridization (FISH): current status and future prospects. *Biotech. Histochem.* **75**, 54–78.
3. Korenberg, J. R. and Chen, X. N. (1995) Human cDNA mapping using a high-resolution R-banding technique and fluorescence in situ hybridization. *Cytogenet. Cell Genet.* **69**, 196–200.
4. Johnson, G. D. and Nogueira Araujo, G. M. (1981) A simple method of reducing the fading of immunofluorescence during microscopy. *J. Immunol. Methods* **43**, 349–350.
5. Mitelman, E. (1994) *ISCN 1995 An International System for Human Cytogenetic Nomenclature*. Chap. 2.2.2 p. 7, Fig. 1.

High-Throughput BAC Fingerprinting

**Jacqueline Schein, Tamara Kucaba, Mandeep Sekhon,
Duane Smailus, Robert Waterston, and Marco Marra**

1. Introduction

This chapter describes a nonradioactive, agarose gel-based, high-throughput DNA restriction digest fingerprinting methodology first described by Marra et al. (1) for use in the construction of high-resolution physical maps from low-copy-number, large-insert clones. The procedure is robust and allows for the recovery of clone insert size information. Initially used to construct sequence tag site (STS)-based contigs (1), the methodology has also been applied to whole-genome, random-clone strategies that have resulted in the construction of high-resolution, sequence-ready physical maps of the genomes of *Arabidopsis thaliana* (2,3), human (4,5), *Caenorhabditis briggsae* (6), and *Cryptococcus neoformans* (7). The methodology is currently being employed in the construction of physical maps for several other large, mammalian genomes, such as those of mouse (8), rat and bovine.

The basic approach used in the construction of whole-genome fingerprint maps is to fingerprint a set of randomly selected clones that together represent in a redundant fashion the genome of interest, computationally identify overlapping clones based on shared restriction fragments, and assemble them into ordered arrays representing contiguous stretches of DNA ("contigs"). By contrast, STS-based contig construction employs a directed approach wherein specific clones are fingerprinted based on their determined STS content. Depending on the depth of STS markers in the genomic region of interest, several iterative rounds of clone identification and fingerprinting may be required to achieve contiguous coverage of the entire region.

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

There are several key factors to consider when working with large-insert, low-copy-number clones for fingerprinting purposes. DNA quantity and quality are of primary importance. The single-copy, large-insert characteristics of bacterial artificial chromosome (BAC) clones that make them attractive vehicles for genomic cloning can lead to difficulties in isolating reproducibly sufficient quantities of BAC DNA with negligible relative bacterial chromosomal DNA contamination, particularly when the use of small-volume preparations that lend themselves to high-throughput applications is desired. For large-scale fingerprinting projects, it is important to carefully control electrophoresis conditions in order to achieve adequate restriction fragment separation and to minimize variations in inter- and intragel fragment mobility, which adversely affect the downstream restriction fragment-based clone comparisons required for physical map construction. It is also necessary to optimize the quantity of DNA loaded onto the gel; too much DNA can result in band distortion that will reduce the accuracy of the restriction fragments identified, while too little DNA will result in failure to reliably detect the fragments at all. A high-sensitivity DNA detection and gel-imaging method is therefore required to achieve accurate and reliable restriction fragment identification.

In this chapter, we cover all the steps required for fingerprint generation, from clone growth to fingerprint gel imaging. The procedures as described are for high-throughput data acquisition in a 96-well format, without the use of specialized kits, and reflect modifications made to the original procedure that have improved efficiency, throughput, and data consistency. The major steps involved are inoculation of overnight cultures, isolation of BAC DNA using an alkaline lysis procedure, restriction enzyme digestion of BAC DNA, agarose gel preparation, agarose gel electrophoresis, and fingerprint gel imaging.

While downstream analysis of the fingerprint data is beyond the scope of this chapter, we mention here briefly the software used. The Sanger Centre program Image (9) is used for manipulation and interactive analysis of the gel images for purposes of gel lane identification, restriction fragment identification, and calculation of normalized fragment mobility and size information. Clone-by-clone comparisons of restriction fragment data and contig assemblies are performed using the program FPC (10,11). Image software and documentation may be obtained at www.sanger.ac.uk/Software/Image. FPC software and documentation may be obtained at www.genome.arizona.edu/software/fpc.

2. Materials

Suggested suppliers of reagents and equipment are provided; however, other suppliers and equipment may be adequate but have not been tested. The heavy-duty aluminum foil sealing tape used is Scotch No. 425. Where clear plastic

sealing tape is required a precut, 96-well adhesive plate sealer is used. The use of a 96-tip liquid-handling device is recommended for efficiency and accuracy of plate-to-plate sample transfers, and we typically use a Hydra 96 instrument (Matrix Technologies) for these purposes. All solutions should be prepared with double-distilled or deionized water.

2.1. Cell Culturing

1. 2X YT medium: We routinely use a premixed powder (Becton Dickinson) prepared following the manufacturer's directions. If preparing using individual reagents, combine 16 g of tryptone, 10 g of yeast extract, 5 g of NaCl, and water to 1 L. Adjust the pH to 7.0 with dilute NaOH if necessary. Sterilize by autoclaving. Store at room temperature.
2. 96-Pin, slotted pin replicator: The pins should be a minimum of 22 mm in length, with 5- μ L slots (V&P Scientific). Sterilize according to the manufacturer's recommendations.
3. 96-Well, square-well growth blocks with 2-mL well volume (Beckman Coulter). The blocks must be capable of withstanding centrifugation at 5250g. Sterilize by autoclaving.
4. Microporous tape sheets for sealing growth blocks while permitting gas exchange (AirPore Tape; Qiagen) (*see Note 1*).
5. Incubator shaker with a platform capable of holding 96-well blocks.

2.2. Preparation of BAC DNA

1. GET buffer: 50 mM glucose, 10 mM EDTA, pH 8.0, 25 mM Tris-HCl, pH 8.0. Filter sterilize and store at 4°C.
2. DNase-free RNase A (Sigma, St. Louis, MO): 10 mg/mL in TE (10;0.1) (*see item 9*). Store at -20°C.
3. 10% Sodium dodecyl sulfate (SDS) (w/v) in water. Store at room temperature.
4. 10 N NaOH. Store in a plastic container at room temperature.
5. 3 M Potassium acetate (KAc), pH 5.5: 2400 mL of 5 M KAc (1472.25 g of KAc, water to 3 L), 460 mL of glacial acetic acid, and water to 4 L. Confirm pH, filter sterilize, and store at 4°C (*see Note 2*).
6. Isopropanol. Store at room temperature.
7. 95% Ethanol. Store at room temperature (*see Note 3*).
8. 80% Ethanol. Prepare fresh from 95% ethanol.
9. TE (10;0.1): 10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, pH 8.0. Sterilize by autoclaving and store at room temperature.
10. 96-Well collection plates with a minimum 800- μ L well capacity (Uniplate; Whatman Polyfiltronics).
11. 96-Well, non-tissue culture-treated microtiter plates (Corning).
12. Multitube vortexer (VWR).
13. Microtiter plate shaker with 4 mm orbital diameter (IKA Vibrax-VXR, with dish attachment modified to hold four 96-well, deep-well blocks).

2.3. Restriction Digest of BAC DNA

1. Restriction enzyme: For processing large numbers of BAC DNA samples, we recommend obtaining the highest enzyme concentration available from the supplier (*see Note 4*).
2. Appropriate 10X enzyme reaction buffer, supplied with the restriction enzyme (*see Note 5*).
3. Sterile water, stored at 4°C.
4. Digest brew: Just prior to use (*see Subheading 3.3.*), prepare sufficient digest brew to digest all the BAC DNA samples. Digest brew for a restriction digest of a single 5- μ L DNA sample contains 3.8 μ L of water, 1.0 μ L of 10X enzyme buffer, and 0.2 μ L (20 U) of restriction enzyme (at a stock concentration of 100 U/ μ L), for a total reaction volume of 10.0 μ L (*see Note 6*). Combine the water and 10X reaction buffer first, and then add the restriction enzyme and mix thoroughly. Store on ice until ready for use.
5. 5X Loading buffer: 0.21% (w/v) bromophenol blue, 12.5% (w/v) Ficoll Type 400 (Sigma) in water. Store at room temperature.
6. 96-Well, thin-walled cycle plates.

2.4. Preparation of Agarose Gel

1. SeaKem LE agarose (BioWhittaker). This is a standard melting temperature agarose with low electroendosmosis (EEO).
2. 1X TAE buffer freshly prepared from 50X TAE stock (242 g/L of Trizma Base [Sigma]; 100 mL/L of 0.5 M EDTA, pH 8.0; 57.1 mL/L of glacial acetic acid).
3. Gel-casting trays for the electrophoresis apparatus (*see Subheading 2.5., item 2*): The casting trays used with the Gator A3-1 apparatus measure 23 cm wide $\times$ 42.5 cm long. With the use of a spacer comb, these trays allow two 23 $\times$ 21 cm gels to be run in tandem. For each 23 $\times$ 21 cm gel, the distance from the wells to the bottom edge of the gel measures approx 19 cm.
4. Gel combs and gel spacers: Two gel combs and one spacer are required to form two gels in each casting tray. The format of the combs will determine the number of samples that may be loaded at one time. We use a custom manufactured comb with 121 teeth (unpublished) that allows 96 samples and 25 marker lanes to be loaded onto each gel. The comb has a 9-mm center-to-center distance for every five teeth, which allows the use of a multichannel Hamilton syringe gel loader for loading samples and markers (*see Subheading 3.5., steps 3 and 4*). The spacer has a solid insert (no teeth) that effectively divides the casting tray into two, allowing easy separation of the two gels following electrophoresis. Store the combs and spacers submerged in distilled water.

2.5. Agarose Gel Electrophoresis

1. Molecular weight marker mix: 5.0 ng/ μ L of Analytical Marker DNA Wide Range (Promega, Madison, WI), 0.36 ng/ μ L of Marker V (Roche), and 20% (v/v) 5X loading buffer (*see Subheading 2.3., item 5*) in TE (*see Subheading 2.2., item 9*). Immediately prior to use, prepare sufficient marker mix for all gels to be

loaded. This mixture of commercially available molecular weight markers provides regularly spaced fragments of known size spanning an effective range of approx 30,000–400 bp.

2. Large-format, horizontal gel electrophoresis apparatus (Gator A3-1; Owl Separations) with buffer recirculation ports.
3. Power supply, preferably with built-in timer.
4. 1X TAE electrophoresis buffer diluted from 50X stock (*see Subheading 2.4., item 2*), and stored at 4°C.
5. Peristaltic pumps (Master Flex Easy Load L/S; Cole Parmer) and tubing (Tygon LFL L/S 17; Cole Parmer) for buffer recirculation during electrophoresis (*see Note 7*).
6. Recirculating, refrigerated chiller (Model 1173; VWR) and large water bath. Peristaltic tubing from the electrophoresis chambers is submerged in the chilled water bath in order to cool the recirculating electrophoresis buffer. Tubing from multiple electrophoresis units may be placed, evenly distributed, in the same chilled water bath. The volume of water in the bath, the length of tubing that is submerged, and the capacity of the recirculating chiller determine the number of electrophoresis units that can be efficiently cooled by a single bath. We use 25 ft of peristaltic tubing per electrophoresis chamber, 15–20 ft of which is loosely coiled and submerged, with the use of lead ring flask weights, in a chilled, 55-L water bath. We typically place tubing from seven electrophoresis units in a single bath.

2.6. Gel Staining and Imaging

1. SYBR Green I DNA stain (Molecular Probes), stored at –20°C. Thaw just prior to use (*see Subheading 3.6.*), and then dilute 1/10,000 in 1X TAE at room temperature (*see Note 8*). (**Caution:** The stain contains dimethylsulfoxide. Wear gloves and protective eyewear at all times.)
2. Plastic containers for gel staining. We use custom trays manufactured from 3-mm acrylic, 22 cm wide × 28 cm long × 2.5 cm high.
3. Light tight cabinet or chamber, capable of holding multiple staining trays and providing gentle agitation.
4. Fluorimager compatible with SYBR Green I excitation and emission spectra (Fluorimager 595; Molecular Dynamics).

3. Methods

3.1. Cell Culturing

The starting materials are bacterial stocks of BAC clones arrayed into 384-well plates. A 96-pin replicator is used to inoculate from the 384-well plates (*see Note 9*). The spacing of the replicator pins is twice that of the wells in the 384-well plate, such that the replicator samples from every other well in each row and each column. It is therefore necessary to generate four sets of 96 clones (“quadrants”) from each 384-well plate in order to sample from all wells. The quadrants can be distinguished by the well position of the top, left

pin of the replicator when it is placed into the 384-well plate; well A1, A2, B1, or B2.

1. Sterilize the work area.
2. Fill each well of the sterile growth blocks with 1.2 mL of 2X YT medium containing appropriate antibiotic.
3. If the bacterial stocks from which inoculation is to be performed are frozen, allow them to thaw at room temperature. Monitor the thawing process periodically because it is important to minimize the time the cultures are thawed in order to maintain viability of the bacterial stocks. When processing large numbers of plates, it is advisable to stagger the removal of library plates from the freezer. Multiple freeze/thaw cycles will also decrease the viability of the bacterial stocks.
4. For each growth block, sterilize the 96-pin replicator according to the manufacturer's directions (*see Note 10*). Carefully place the replicator pins into the appropriate quadrant of the 384-well plate and allow the pins to rest on the bottom of the wells for several seconds to allow liquid to fill the slots in the pins. Transfer the replicator to the appropriate growth block, gently swirl the pins in the medium to wash the inoculums from the slots, and then seal the block with a sheet of Air-Pore tape.
5. Place the blocks in a 37°C incubator shaker for 16 h at 290 rpm (*see Note 11*).
6. Pellet the bacterial cells by centrifuging at 1400g for 20 min.
7. Decant the supernatant from the blocks by inverting them and shaking out the liquid. Be somewhat vigorous to ensure that the medium is removed, but do not dislodge the pellet. Invert the blocks onto paper towels and tap several times to dislodge excess medium from the sides of the wells, then let them drain, inverted, for 10 min. Lightly tap the blocks a second time (do not dislodge the pellet), and place inverted on fresh paper towels for an additional 10 min.
8. Seal the blocks with foil tape and store at -80°C for a minimum of 2 h prior to DNA preparation (*see Note 12*).

3.2. Preparation of BAC DNA

The methods described in this chapter have been customized to provide a consistent product provided the procedures are followed closely. We therefore do not include a DNA quantitation step following preparation of the BAC DNA. The yield of DNA recovered under the conditions described is roughly on the order of 1 µg.

1. Freshly prepare the required volume of GET buffer containing 150 µg/mL of RNase A. Keep on ice until ready to use.
2. Freshly prepare the required volume of lysis solution: 1% SDS, 0.2 N NaOH. Keep at room temperature.
3. Thaw the frozen 96-well blocks containing pelleted bacterial cells on the bench-top for 30 min.

4. Add 200 μL of GET/RNase A to each well. Seal with clear tape and resuspend the bacterial pellets by mixing on a multitube vortexer for 5 min at top motor speed (2400 cpm). Ensure that all pellets are completely resuspended before proceeding to the next step.
5. Arrange the blocks in sequential order on the benchtop and peel back the clear tape. Add 200 μL of lysis solution to each well and allow the cells to lyse for 5 min (*see Note 13*). Monitor the wells to ensure that lysis is progressing.
6. Add 200 μL of cold 3 *M* KAc to each well in the same order that the lysis solution was added. Best results are obtained if the liquid is directed straight down into the well at a velocity adequate to effect some mixing of the solution. Reseal the blocks with clear tape, and immediately place them onto an IKA microtiter plate shaker. Mix at 1100 rpm for 3 min (*see Note 14*).
7. Centrifuge at 4°C for 45 min at 5250*g* to pellet the precipitate. After centrifugation, inspect the blocks to ensure that there is a compact pellet with little particulate matter in the supernatant.
8. For each block, use a 96-tip liquid-handling instrument to aspirate 400 μL of the cleared lysate from each well. Ensure that none of the pelleted debris is aspirated. Dispense the lysate into a 96-well collection plate containing 300 μL of isopropanol in each well. Use a mixing cycle to thoroughly mix the two liquid phases (*see Note 15*).
9. Place a clear plastic sealer on the collection plates and centrifuge at 4°C for 15 min at 2830*g* to precipitate the DNA (*see Note 16*).
10. Decant the isopropanol by inverting the plates and gently shaking to dislodge the liquid. Blot the inverted plates on paper towels. Leave each plate inverted until ready to perform the next step.
11. To each plate in succession, wash the DNA pellets by adding 200 μL of 80% ethanol to the wells. Dispense the fluid slowly and direct it to the side of the wells so as not to disturb the DNA pellet. Decant the ethanol immediately by inverting the plate and shaking gently to dislodge the fluid from the wells. Blot the inverted plate briefly on a paper towel, and then move it to dry paper toweling and allow it to drain inverted for 5 min.
12. Gently tap the plates to dislodge excess ethanol, move them to dry paper toweling, and air-dry inverted for an additional 20 min.
13. Remove residual ethanol by drying the pellets under vacuum for 6 min with moderate heat (*see Note 17*). Ensure that the wells and pellets are dry because residual ethanol can inhibit restriction enzyme activity.
14. Resuspend the DNA by adding 50 μL of TE to each well. Seal the plates with clear plastic tape sealer. Lightly tap the plates to deposit the TE to the bottom of the wells. Incubate in a 37°C air incubator for 10 min. Following incubation, vortex the plates on a microtiter plate shaker for 5 min, and then briefly centrifuge to deposit the liquid to the bottom of the wells.
15. Transfer the DNA to microtiter plates using a 96-tip liquid-handling instrument. Check both the collection plates and the microtiter plates for efficient transfer.

Seal the microtiter plates tightly with foil tape, and store at 4°C until ready to perform the restriction digest.

3.3. Restriction Digest of BAC DNA

1. Add 5 μL of digest brew to each well of the required number of 96-well digest plates. Seal lightly with foil tape and centrifuge briefly to deposit the brew to the bottom of the wells. Visually inspect the plates for the correct volume in all wells. Store the plates on ice.
2. Briefly centrifuge the microtiter plates containing the prepared BAC DNA to deposit the DNA to the bottom of the wells. Store the plates on ice.
3. For each microtiter plate containing BAC DNA, use a 96-tip liquid-handling instrument to transfer 5 μL of DNA from each well into one of the prepared 96-well digest plates.
4. Centrifuge the digest plates briefly to deposit all the liquid to the bottom of the wells. Visually inspect the plates to ensure that all wells have had the correct volume of DNA added.
5. Seal the digest trays tightly with foil tape to prevent evaporation during incubation. Incubate the plates in a 37°C incubator for 2 h, and then briefly centrifuge to collect the liquid at the bottom of the wells.
6. Add 2.4 μL of 5X bromophenol blue loading buffer to each well, reseal the plates, vortex to mix, and briefly centrifuge to deposit the samples to the bottom of the wells. Tightly seal the plates with foil tape and store at 4°C until ready to load onto gels (*see Note 18*).

3.4. Preparation of Agarose Gel

The quality of the agarose gels used for electrophoresis of the digested BAC DNA directly affects the quality of the fingerprint data obtained. Even small imperfections in the gels can adversely affect the quality of the data. It is therefore of utmost importance that great care be taken during the preparation and pouring of agarose gels. The primary considerations are that the gels be of uniform thickness and agarose composition, be free of contaminating particulates, and have properly formed and undamaged wells. Careful technique must be employed to achieve consistency both within a gel and between gels.

1. Prepare the gel-pouring surface as follows: Wipe down the surface of the bench to be used for pouring the gels. Cover the area where each gel will be cast (e.g., with a piece of aluminum foil) to keep the area clean until ready for pouring (*see Note 19*). The same cover will later be placed over the casting tray while the agarose cools, so it must be large enough to loosely but completely cover the casting tray without contacting the gel surface.
2. Prepare the gel-casting trays as follows: Wear powder-free gloves that have been washed with soap and thoroughly rinsed with water to remove any particulates. Using gloved hands, thoroughly wash the gel-casting trays and rinse well with

distilled water. Using gentle pressure, wipe the trays with lint-free tissues to remove most of the water. Leave the trays just slightly damp; wiping dry trays with dry tissue will cause the tissue to deteriorate and leave lint on the casting tray surface. Invert the gel trays and allow them to air-dry completely. When dry, seal the ends of the casting trays with tape.

3. Prepare the gel combs and spacers as follows: Clean the combs and spacers using a soft-bristled brush and a small amount of dilute soap. Rinse thoroughly and allow them to air-dry completely. Placing the combs in an air stream such as that provided by a small fan will reduce the drying time. Before use, check the combs to ensure that the space between the teeth is completely clean and dry.
4. Measure 4.8 g of agarose and 400 mL of 1X TAE buffer into a 1-L Erlenmeyer screw-top flask (the resulting 1.2% agarose gel, when cast, will be approx 4 mm thick). Loosely fasten the screw cap and record the weight of the flask. (**Caution:** It is extremely important that the cap is not tightly fastened in order to allow steam to escape during heating and avoid a buildup of pressure inside the flask.)
5. Prepare the molten agarose by performing the following for each flask. Wear protective eyewear and appropriate safety protection. Heating times are based on use of an 1100-W microwave.
 - a. Microwave the flask on high for 4 min.
 - b. Keeping the flask inside the microwave, use thermal gloves to carefully mix the solution by gentle swirling to release agarose from the bottom of the flask and disperse it.
 - c. Add water until the flask weighs approx 10 g more than the initial weight. The extra volume will compensate for the next round of heating. Swirl to mix.
 - d. Loosely fasten the screw cap and microwave on high for 2 min. Monitor the flask to ensure that it does not boil over.
 - e. Keeping the flask inside the microwave, use thermal gloves to gently and carefully swirl the flask to release the gas trapped in the gel solution and allow the steam to escape. (**Caution:** The gel will boil when disturbed and steam will be forced out under the lid. If the gel is mixed too vigorously, the trapped gas will be released explosively, forcing steam and molten agarose out under the lid.)
 - f. Add water until the flask weighs approx 8 g more than the initial weight. Swirl to mix.
 - g. Loosely fasten the screw cap and microwave for 1 min. Monitor the flask to ensure that it does not boil over.
 - h. Keeping the flask inside the microwave, use thermal gloves to gently and carefully swirl the flask to release the gas trapped in the gel solution and allow the steam to escape. (**Caution:** Observe the cautionary measures outlined in **step 5.**)
 - i. Add water to bring the flask up to the initial weight. Swirl gently to mix but do not introduce air bubbles. Place a nonmercury thermometer into the flask and place the cap loosely on top.
 - j. Place the flask in a 55°C water bath. The water level in the bath should be equal to or slightly higher than the level of the agarose. Use the thermometer to gently stir the agarose every few minutes to prevent differential cooling and

to ensure that the agarose solution is uniform. Stir gently to avoid introducing air bubbles. Allow the liquid to cool to 55°C.

6. Fairly quickly, but carefully, pour all of the molten agarose into one corner of the casting tray. Rotate the flask while pouring to pick up condensation on the sides of the flask. Be careful not to introduce air bubbles. When all of the liquid has been poured from the flask, gently rock the casting tray once to the back and once to the side to ensure that the agarose solution is evenly distributed. Use the wide end of 200- μ L pipet tips to remove any bubbles or lint near the areas where the gel combs and spacer are to be placed.
7. Carefully place the two combs and the spacer into the designated slots in the casting tray. Gently blow on the agarose solution around the comb teeth to ensure that the liquid is properly dispersed between all the teeth.
8. Remove lint and bubbles from the gel using the wide end of 200- μ L pipet tips. Be as thorough as possible, but stop the moment the gel visibly starts to set, as evidenced by slight indentations left in the gel surface by the tips. Loosely cover the casting tray in order to protect the cooling liquid from dust and air currents, but allow steam to escape. Allow the agarose to set for at least 1 hr. Any movement of the casting tray prior to the agarose being completely set will disturb gel uniformity and adversely affect data quality.
9. Once the gel is set, apply water along the comb/gel and spacer/gel interfaces. Very slowly and carefully pull each of the combs and the spacer straight up until the vacuum is broken, and then smoothly pull up to remove them completely.
10. Remove the tape from the ends of the casting tray and pour off any excess water from the gel surface. Wrap the tray tightly with plastic wrap and store at 4°C. Best results are obtained when the gels are used within 1 or 2 d, but if carefully wrapped they may be stored for 3 d before use.

3.5. Agarose Gel Electrophoresis

1. Turn on the recirculating chiller and ensure that all tubing in the water bath is submerged and that there are no kinks restricting buffer flow. In our hands, a setting of 16°C on the chiller maintains the buffer in the electrophoresis chamber at a temperature of 19°C.
2. Check that the electrophoresis chamber is level. Add sufficient cold 1X TAE buffer to the chamber to submerge the recirculation ports and allow proper buffer recirculation (the Gator A3-1 chambers will require approx 3.8 L). Insert the peristaltic tubing from one of the recirculation ports into the pump head, ensuring that the tubing is not pinched. Run the peristaltic pump until the buffer has completely circulated through the tubing and all the air is displaced. Turn the pump off and place a gel in the buffer chamber. The buffer level should be sufficient to just cover the wells.
3. Load 1.5 μ L of marker mix into every fifth well, including the first and last wells, ensuring that the liquid is deposited at the bottom of the wells.
4. Load 1.5 μ L of each digested sample, ensuring that the liquid is deposited at the bottom of the wells. The marker spacing on the gel allows four samples to be

loaded between marker wells. Any sample will therefore be at most two wells away from a known size standard.

5. Check that the buffer level is minimally 5 mm above all areas of the gel, adding additional 1X TAE if necessary. Allow the samples and markers to settle for 10 min. Electrophorese the samples at 140 V (3 V/cm) for 8 h with buffer recirculation at approx 420 mL/min. After 8 h, the bromophenol blue in the loading dye will have typically migrated approx 15 cm from the wells.

3.6. Gel Staining and Imaging

1. Immediately following electrophoresis, remove the gel tray from the electrophoresis chamber, carefully separate the two gels, and gently slide each into a staining tray (*see Note 20*). Add a sufficient volume of prepared SYBR Green I stain to cover the gels. Allow the gels to stain, with gentle agitation and protected from light, for a minimum of 30 min (if the stain is being used for the second time, allow the gels to stain for a minimum of 45 min).
2. Carefully transfer the stained gel onto the surface of a clean scanning plate. This is best achieved with a sheet of thin plastic, because using one's hands to manipulate the gels should be avoided. Using a wash bottle, thoroughly rinse the gel with distilled water to remove excess stain and to wash any particulates off the gel surface. Ensure that there are no air bubbles or particles trapped underneath the gel. These may be more easily identified by placing the plate over a dark surface (*see Note 21*).
3. Image the gel on a fluorimager. The following parameters are used for scanning with the Molecular Dynamics Fluorimager 595: 200 μm pixel size, 16-bit digital resolution, 530df30 filter, single-label dye(s), 488-nm excitation filter. An example of the data collected is shown in **Fig. 1**.

4. Notes

1. Plastic lids can be used in place of AirPore tape but they may prohibit uniform aeration of the wells, resulting in uneven cell growth and nonuniform BAC DNA yields.
2. The BAC DNA preparation is sensitive to the quality of this solution. New batches of 3 M KAc should always be assessed with a set of test samples prior to being put into general use.
3. We do not recommend the use of 100% ethanol owing to possible contaminants that may affect DNA recovery (*12*).
4. The enzyme should be robust and produce complex fingerprints, with fragments widely distributed over the resolvable area of the gel. It may be necessary to digest the DNA with two enzymes to achieve the desired number and distribution of restriction fragments. Cost and availability of enzymes is an important consideration when processing large numbers of samples. We have found *HindIII* (New England Biolabs) to be well suited for most applications of this methodology.
5. If digesting the DNA with more than one enzyme, it is necessary to select a reaction buffer in which the enzyme activities are compatible.

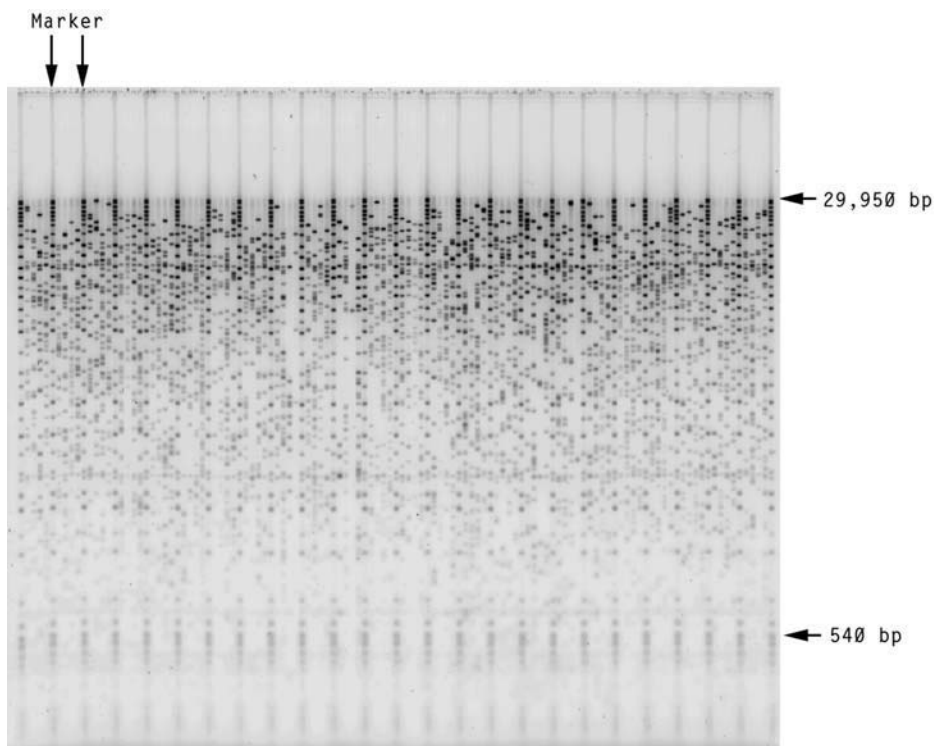


Fig. 1. Typical result of a fingerprinting gel using protocol described. Restriction fragments are resolvable over a size range of approx 30 kb to 500 bp. Internal vector fragments of 6.5, 1.5, and 0.64 kb are visible as common restriction fragments in all lanes. Every fifth lane contains marker DNA.

6. Although 1 U of enzyme is defined as the amount required for digestion of 1 μ g of DNA in 1 h at 37°C, addition of excess enzyme is recommended to ensure complete digestion of the DNA. Care must be taken, however, to avoid star activity.
7. The tubing clamped in the pump heads wears quickly from heavy daily usage and should be examined at the end of each run. If excessive wear or cracking is noted, the tubing should be replaced. We recommend using tubing connectors to splice a short length of tubing (approx 30 cm) into the section of the tubing that is placed into the peristaltic pump heads. When worn, this small section is easily replaced without disturbing the rest of the tubing.
8. SYBR Green I is light sensitive, particularly once it is diluted. In our experience, the diluted stain can be used a second time within 24 h, with a slight decrease in activity, if care is taken to limit light exposure and the diluted stain is stored in a light tight container at 4°C.

9. The clones are not single-colony purified in this procedure. Prior to embarking on a full-scale fingerprinting effort on a library, clones from a sample set of 384-well plates in the library should be fingerprinted and the fingerprints assessed. If the 384-well source plates have significant cross-contamination (more than one species of BAC clone per well), then a colony purification step will be required prior to inoculation of overnight cultures.
10. We typically dip and swirl the pins, in series, in detergent, sterile water, and 95% ethanol, followed by drying with a small fan or a hot-air drier. The pins must be cool and dry before they are dipped into the bacterial stocks. Increased efficiency of the inoculation process can be realized by the alternate use of two replicators.
11. It is advisable to initially test a range of growth times in order to determine which provides the best BAC DNA yields for a particular library. In our hands, 16-h overnight growth typically produces the best results.
12. We routinely store frozen, pelleted cells for up to 1 mo without noticeable effect.
13. A physical mixing step is not recommended because it can result in increased bacterial genomic DNA contamination in the BAC DNA preparation, particularly when working with BAC clones containing inserts of approx 200 kb or greater.
14. The white precipitate on the surface of the wells should somewhat resemble a layer of lily pads covering the majority of the liquid surface area in each of the wells. Insufficient mixing, as evidenced by a limited amount of white precipitate, will result in low DNA yield. If mixing is too vigorous or allowed to proceed for an extended period of time, the precipitate will be very fragmented and the solution will foam. This will result in an increase in contaminating proteins and bacterial genomic DNA in the BAC DNA preparation.
15. The alcohol and aqueous phases will remain separated if not adequately mixed, with the result that the DNA will not be efficiently precipitated.
16. The polystyrene Uniplates from Whatman Polyfiltronics will break if they are not padded beneath the wells during centrifugation.
17. We typically use a Savant Speed Vac Model SC210A (with rotor removed) with GP110 vacuum pump for this purpose. We have also had good success placing the plates in a 37°C air incubator for 30 min. The DNA will be difficult to resuspend if dried too long.
18. Because of desiccation, best results are obtained if the digested samples are loaded within 24 h. However, if the plates are tightly sealed and wrapped in plastic, they may be stored for up to 3 d with minimal loss of volume.
19. To produce a gel of uniform thickness, the molten agarose must be distributed evenly within the casting tray. It is therefore important that each casting tray be placed on an area of the benchtop where it is known that even liquid distribution will be achieved.
20. Clean up of the electrophoresis chamber is simplified by the use of a wet/dry vacuum to remove the buffer.
21. Air bubbles, lint, and other particulates are visible on the captured images and can interfere with fragment identification and subsequent analysis.

Acknowledgments

We wish to thank Michael Smith, Steven Jones, Letticia Hsiao, Ian Bosdet, Martin Krzywinski, Readman Chiu, John McPherson, the staff of the B.C. Cancer Agency Genome Sciences Centre, and members of the mapping group at the Washington University Genome Sequencing Center for contributions to this work.

References

1. Marra, M. A., Kucaba, T. A., Dietrich, N. L., Green, E. D., Brownstein, B., Wilson, R. K., McDonald, K. M., Hillier, L. W., McPherson, J. D., and Waterston, R. H. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Res.* **7**, 1072–1084.
2. Marra, M., Kucaba, T., Sekhon, M., et al. (1999) A map for sequence analysis of the *Arabidopsis thaliana* genome. *Nat. Genet.* **22**, 265–270.
3. Mozo, T., Dewar, K., Dunn, P., Ecker, J. R., Fischer, S., Kloska, S., Lehrach, H., Marra, M., Martienssen, R., Meier-Ewert, S., and Altmann, T. (1999) A complete BAC-based physical map of the *Arabidopsis thaliana* genome. *Nat. Genet.* **22**, 271–275.
4. McPherson, J. D., Marra, M., Hiller, L., et al. (2001) A physical map of the human genome. *Nature* **409**, 934–941.
5. Lander, E. S., Linton, L. M., Birren, B., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
6. Stein, L. D., Bao, Z., Blasiar, D., et al. The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. PLOS Biology, in press.
7. Schein, J. E., Tangen, K. L., Chiu, R., et al. (2002) Physical maps for genome analysis of serotype A and D strains of the fungal pathogen *Cryptococcus neoformans*. *Genome Res.* **12**, 1445–1453.
8. Gregory S. G., Sekhon, M., Schein, J., et al. (2002) A physical map of the mouse genome. *Nature* **418**, 743–750.
9. Sulston, J., Mallett, F., Staden, R., Durbin, R., Horsnell, T., and Coulson, A. (1988) Software for genome mapping by fingerprinting techniques. *CABIOS* **4**, 125–132.
10. Soderlund, C., Humphray, S., Dunham, I., and French, L. (2000) Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res.* **11**, 934–941.
11. Soderlund, C., Longden, I., and Mott, R. (1997) FPC: a system for building contigs from restriction fingerprinted clones. *CABIOS* **13**, 523–535.
12. Ito, K. (1992) Nearly complete loss of nucleic acids by commercially available highly purified ethanol. *Biotechniques* **12**, 69–70.

BAC End Sequencing

Tim S. Poulsen and Hans E. Johnsen

1. Introduction

Large-insert genomic DNA libraries are based on the *Escherichia coli* F factor, a low-copy plasmid that exists in a supercoiled circular form in the host cells. These libraries are used to provide a way to divide complex genomes into DNA segments, thereby reducing the complexity (**1**). The libraries are arrayed in microtiter dishes, providing the opportunity for many researchers to accumulate and use information regarding particular clones. The information about the end sequence of the bacterial artificial chromosome (BAC) clones is currently used in a wide array of applications from genome sequencing to gene discovery. The BAC end sequences of individual clones with insert DNA are collected in the large databases that can be accessed by the researcher. The information regarding the BAC end sequence stored in these databases makes it easier to identify the minimally overlapping clones that can be used as a source for shotgun-sequencing projects (**2**), to find clones for restriction fingerprints for building overlapping clone sets (**3**), to find appropriate clones for fluorescence *in situ* hybridization mapping (**4**), or to select a BAC clone that contains genes of interest (**5**). Some important BAC end sequence databases are as follows:

1. Human: www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html, 470,000 clones.
2. Rice: www.genome.clemson.edu/projects/rice/rice_bac_end/index.html, 92,000 clones.
3. Mouse: www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html, 300,000 clones.
4. Rat: www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html, 200,000 clones.
5. Sea urchin: <http://sugp.caltech.edu:7000>, 25,000 clones.

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

6. *Arabidopsis thaliana*: http://ftp.tigr.org/tdb/at/abe/bac_end_search.html, >40,000 clones.
7. *Trypanosoma brucei*: http://ftp.tigr.org/tdb/mdb/tbdb/bac_end_search.html, 10,000 clones.

1.1. Principles of BAC End Sequencing

Most DNA-sequencing methods presently used are variations of the chain-termination method developed by Sanger et al. in 1977 (6). In principle, the DNA to be sequenced acts as a template for enzymatic elongation from a defined primer-binding site. The DNA polymerase enzyme incorporates radioactive isotopes or fluorescent dye detection molecules into the synthesized DNA single strand. The synthesized DNA is then separated by length using electrophoresis and detected by using either X-ray films or a laser and a set of excitation/emission filters. To ensure the health of the users and to save time, many laboratories now use fluorescent dyes. There are three major methods for sequencing with fluorescent dyes: (1) using fluorescent dye conjugated to the primers, (2) using fluorescent dye conjugated to the deoxynucleotides, and (3) using fluorescent dye conjugated to the terminator dideoxynucleotides. Each has its own advantages and disadvantages; however, the first two require a single reaction for each of the four nucleotides, while the last method only requires one reaction for all four nucleotides. Dye terminator chemistry sequencing with primers specific for the BAC vector T7 and SP6 promotor region has been advised for end sequencing of BAC clones (7).

An important key factor in BAC end sequencing is the quality of the bacterial culture from which the DNA is extracted. Using *E. coli* strain DH10B as the host has many advantages, including the *endA* mutation (lowers the amount of DNA nucleases); elimination of *mcrA*, *mcrB*, *mcrC*, and *mrr* (prevents the strain from methylating cytosine and adenine residues); *recA1* and *deoR* mutations (ensure stability of large plasmid insert DNA); $\Delta lacX74$ deletion (deletion of the *lac* operon); and $\Phi 80dlacZ\Delta M15$ insertion (insertion of a part of the *lac* gene that can be used for α complementation). Another key factor to obtain high-template-quality DNA from *E. coli* relies on the effective removal of RNA, salt, and proteins. The RNA can give a higher signal background during end sequencing (see **Note 1**). The salt contamination significantly reduces the fluorescence intensity, the accuracy, and the reading length of the end sequencing (see **Note 2**). Phenol contamination after purification of DNA template from proteins may reduce the signal intensity, read length, and accuracy (see **Note 3**). The DNA template itself may also be a key factor. GC-rich templates often result in a strong artifact stop band. This stop is owing to a secondary structure resembling hairpins. This structure may not melt at the temperature of sequencing reactions (see **Note 4**). The template may also be contaminated

with ethanol after ethanol precipitation, which significantly reduces the peak height of the analyzed end sequence (*see Note 5*). Solvent containing EDTA may significantly reduce the fluorescence intensity, the accuracy, and the reading length of the end sequencing owing to inhibition of the DNA polymerase (*see Note 6*). The protocol described herein is optimized for BAC end sequencing using ABI PRISM BigDye Terminator Sequencing Ready Reaction Kits, an ABI PRISM 310 Genetic Analyzer, a GeneAmp PCR system 2400 Perkin Elmer Biosystems, and precipitating of the DNA pellet with sodium acetate/ethanol. A typical BAC end sequence run gives 400–500 nt of good sequence.

2. Materials

1. 3.0 M Sodium acetate, pH 4.6 (store at room temperature).
2. 95% Ethanol, diluted from absolute ethanol (store at room temperature).
3. 70% Ethanol, diluted from absolute ethanol (store at room temperature).
4. BAC DNA dissolved in ddH₂O (store at –20°C) (*see Note 6*).
5. Terminator Ready Reaction Mix (store at –20°C).
6. Template suppression reagent (TSR) (store at –20°C).
7. 10 μ M Primer stock (store at –20°C). Useful primers are as follows:
 - a. Vector pBeloBAC 11, promotor T7, primer 5' TAATACGACTCACTATAGGG (20mer), promotor, 5' GTTTTTTGCGATCTGCCGTTTC (22mer).
 - b. Vector pBACe3.6 promotor T7, primer 5' CGGTCGAGCTTGACATTGTAG (21mer), promotor SP6, primer 5' GATCCTCCCGAATTGACTAGTG (22mer).

3. Methods

Other protocols may be required for systems other than the ABI PRISM 310 Genetic Analyzer, such as the ABI PRISM 3700 DNA Analyzer, ABI PRISM 377 DNA sequencers, or ABI PRISM 373 DNA sequencers. Application of the sample and operation of the ABI PRISM 310 Genetic Analyzer should be performed as described by the manufacturer (Perkin Elmer Biosystems). A part of a typical BAC end sequence is presented in **Fig. 1**.

1. Mix reaction mixture on ice in a polymerase chain reaction (PCR) tube (*see Note 7*) as follows:
 - a. BAC DNA template (2 μ g): X μ L.
 - b. Primer (3 pmol): 0.3 μ L.
 - c. Terminator Ready Reaction Mix: 8 μ L.
 - d. ddH₂O: 11.7–X μ L.
 - e. Total volume: 20 μ L.
2. Mix well and centrifuge briefly (*see Note 8*).
3. Place the tube in a thermocycler and set the volume to 20 μ L.
4. Run the cycle-sequencing program on a GeneAmp PCR system 2400 Perkin Elmer Biosystems (*see Note 9*) as follows:

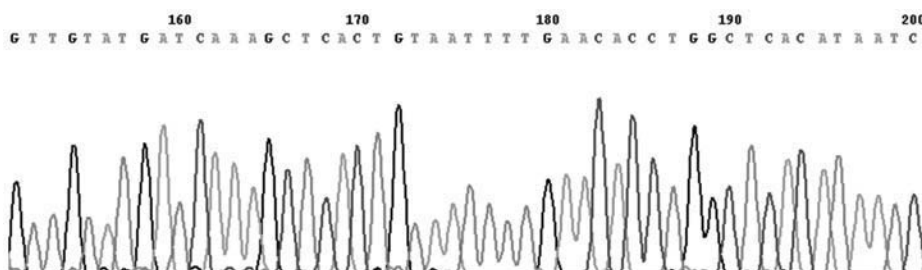


Fig. 1. Part of end sequence of RPCI-11 BAC clone 47P24 when using protocol as described and SP6 primer for pBACe3.6.

- a. 96°C: 3 min.
- b. 96°C: 10 s.
- c. 50°C: 10 s, 99 cycles.
- d. 60°C: 4 min.
- e. 4°C: ∞.
5. Transfer the DNA sample to an Eppendorf tube and precipitate by adding 2 μ L of 3 M sodium acetate and 50 μ L of 95% ethanol (*see Note 10*).
6. Vortex briefly.
7. Incubate for 10 min at room temperature.
8. Centrifuge at 20,000g for 30 min.
9. Discard the supernatant and remove remaining ethanol drops (*see Note 11*).
10. Wash the pellet with 250 μ L of 70% ethanol and centrifuge at 20,000g for 5 min.
11. Discard the supernatant and remove remaining ethanol drops (*see Note 11*).
12. Dry the pellet in a vacuum centrifuge for 5 min (*see Note 12*).
13. Resuspend the pellet in 16 μ L of TSR (*see Note 13*).
14. Vortex and centrifuge briefly (*see Note 14*).
15. Heat the samples at 95°C for 5 min and then chill on ice.
16. Vortex and centrifuge briefly (*see Note 8*).
17. Place the sample on ice.

4. Notes

1. RNA can be removed by treating the DNA with RNases.
2. Salt contamination in template DNA can result from coprecipitation of salt in alcohol when incubating at low temperatures, by insufficient removal of supernatant, or by an insufficient wash with 70% ethanol. If traces of salt are suspected, careful precipitation of the template at room temperature followed by a 70% ethanol wash at room temperature can solve the problem.
3. Avoid using phenol, which is harmful and may contaminate the DNA template. Instead, use other purification methods such as a gel column or a resin column.
4. The addition of dimethylsulfoxide or formamide to the reaction mixture further reduces the melting temperature of GC-rich regions, and thereby prevents strong stop bands without affecting the sequencing reaction.

5. Alcohol contamination in the template can arise from insufficient drying of the DNA pellet after precipitation. Alcohol contamination can be eliminated by evaporation.
6. Dissolve the BAC DNA in ddH₂O instead of TE. For long-term storage dissolve in TE.
7. Add the Terminator Ready Reaction Mix as the last component to avoid exposing the fluorescent dye to light.
8. Vortexing ensures a good mix of the components, and a brief centrifugation ensures that all of the reaction mix is located at the bottom of the PCR tube.
9. Using 99 cycles has been shown to increase the success of end sequencing of BAC clones. If another PCR thermal cycler is used, it might be necessary to optimize the thermal cycling conditions.
10. Instead of ethanol/sodium acetate precipitation, the DNA can be purified with a gel spin column or a resin spin column. If purifying is done with a column, dry the pellet in a vacuum centrifuge for 15 min and continue at **step 13**.
11. The remaining ethanol drops can be removed with a piece of 3M paper or by using a vacuum centrifuge. Avoid disturbing the DNA pellet.
12. This step ensures that all remaining ethanol is effectively removed.
13. From this point on the work should be performed without exposing the DNA to light.
14. The sample can be frozen for several weeks before running on the DNA analyzer.

References

1. Shizuya, H., Birren, B., Kim, U., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. L. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **92**, 10,831–10,835.
2. Venter, J. C., Smith, H. O., and Hood, L. (1996) A new strategy for genome sequencing. *Nature* **381**, 364–366.
3. Park, J. H., Dixit, M. P., Onuchic, L. F., et al. (1999) A 1-Mb BAC/PAC-based physical map of the autosomal recessive polycystic kidney disease gene (*PKHD1*) region on chromosome 6. *Genomics* **57**, 249–255.
4. Poulsen, T. S., Silahtaroglu, A. N., Gisselø, C. G., Gaarsdal, E., Rasmussen, T., Tommerup, N., and Johnsen, H. E. (2001) Detection of illegitimate rearrangements within the immunoglobulin locus on 14q32.3 in B-cell malignancies using end sequenced probes. *Genes Chromosomes Cancer*, **32**, 265–274.
5. Boysen, C., Simon, M., and Hood, L. (1997) Analysis of the 1.1-Mb human α/δ T-cell receptor locus with bacterial artificial chromosome clones. *Genome Res.* **7**, 330–338.
6. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463.
7. Kelly, J. M., Field, C. E., Craven, M. B., Bocskai, D., Kim, U.-J., Rounsley, S. D., and Adams, M. D. (1999) High throughput direct end sequencing of BAC clones. *Nucleic Acids Res.* **27**, 1539–1546.

Radiation Hybrid Mapping With BAC Ends

Michael Olivier, Shannon Brady, and David R. Cox

1. Introduction

Recent advances in the Human Genome Project have opened the door to new approaches in biologic research. The availability of the human draft sequence (*1*) now offers the tool for sequence-based genetic analyses on a genomewide level. However, owing to the fragmented and incomplete nature of the draft version currently available, work is focusing on ordering and orienting the individual sequence segments relative to each other to unambiguously place them within the sequence of a single, unique chromosome.

One of the methods that can be used to assign sequences to unique positions in the human genome is radiation hybrid (RH) mapping (*2*). In short, human donor DNA is irradiated to induce DNA strand breaks. The resulting DNA fragments are fused with hamster cells, and a random proportion of the human DNA fragments is retained in each hamster cell. By isolating a set of independent cell lines, the entire human genome is retained in these hybrid cell lines. RH cell lines have been used in building maps of numerous species, including several RH maps of the human genome. The number of radiation breaks induced in DNA is dependent on the dose of X-ray radiation used, with higher doses resulting in smaller fragments. Thus, several human RH mapping panels have been generated using different doses of radiation (*3,4*). When more DNA breaks are induced, sequences in the genome that are close to each other can then be mapped relative to each other since occasionally breaks will be induced between them. Consequently, RH maps constructed with higher doses of radiation allow ordering of close markers but also require larger numbers of sequences to be mapped in order to cover the entire genome. The most recent

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

RH map was constructed at the Stanford Human Genome Center using 50,000 rad of X-rays (5). The resulting panel of hybrid cell lines (TNG) contains 90 independent hybrid cell lines and allows the unambiguous placement of sequence-tagged sites (STSs) at a resolution of 100 kb. In all, the current map contains 36,678 ordered STSs (6).

To map a sequence of interest, RH cell lines are analyzed individually for the presence or absence of the specific sequence. Commonly, STSs are amplified using polymerase chain reaction (PCR) (7). For this, PCR primers are designed for the specific sequences in the human genome, and subsequently each hybrid cell line DNA is used as template in individual PCR reactions. An amplification product of the expected size indicates that the genomic segment containing the sequence of interest is present in a specific hybrid cell line. By comparing the presence or absence of amplification products for an unknown sequence with the pattern for known sequences in the human genome, new STSs can be assigned to a unique genome location.

Bacterial artificial chromosomes (BACs) are commonly used in genetic analyses because they contain continuous segments of the human genome. As a first step in characterizing individual BAC clones, both ends of a BAC are sequenced, resulting in sequences of approx 500 bp on either end of the BAC clone. By designing STSs in both end sequences and mapping the STSs to the TNG RH map, BAC clones can be placed and oriented on the TNG map.

In this chapter, we describe the methods used routinely at the Stanford Human Genome Center to design and map STSs derived from BAC end sequences to the TNG hybrid panel. We also describe how positional information about the newly designed STSs can be obtained using computational tools provided through the Stanford Human Genome Center and marker information from the existing TNG RH map.

2. Materials

All solutions should be made with ddH₂O.

2.1. Polymerase Chain Reaction

1. RH DNA (5 ng/μL) (Research Genetics/Invitrogen, Huntsville, AL) (*see Note 1*).
2. 10 μM (for each individual oligonucleotide primer) Oligonucleotide primer pair for BAC of interest.
3. 2.5 mM DNTPs (*see Note 2*).
4. 5X Buffer: Mix equal volumes of 150 mM Tris-HCl, pH 8.0; 500 mM KCl; and 25 mM MgCl₂ in water. 10X Buffer and 25 mM MgCl₂ solutions are available with AmpliTaq Gold from Perkin Elmer (Foster City, CA) (*see Note 3*).
5. AmpliTaq Gold (5 U/μL) (Perkin Elmer).
6. ddH₂O, autoclaved.

2.2. Agarose Gel Electrophoresis

1. Ultrapure agarose (Gibco-BRL, Life Technologies, Gaithersburg, MD).
2. 1X TBE buffer: 54 g/L of Tris, 27.5 g/L of boric acid, 3.72 g/L of EDTA. Stock solutions can be made and stored at room temperature (*see Note 4*).
3. Ethidium bromide (EtBr) solution: EtBr is a mutagen, so adequate safety precautions should be used (*see Note 5*).
4. 3X Loading buffer: 10% (w/v) Ficoll 400; 0.1 M EDTA, pH 8.0; 0.025% (w/v) bromophenol blue in ddH₂O.
5. Gel-imaging system with an ultraviolet transilluminator.

3. Methods

3.1. Primer Selection

Primers are designed according to a protocol described by Beasley et al. (8) using the program Primer3. The program is available at www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi/. The following modified conditions are used:

1. The initial start sequence (in this case the sequence obtained from sequencing the ends of a BAC) is modified using a repeat masker program. Primer3 offers a mispriming library (repeat library) for this task as well.
2. The modified sequence is loaded into the program, and primers are designed using the default parameters except for the following:
 - a. Product size: min, 90; opt, 220; max, 350.
 - b. Max 3' end stability: 8.0.
 - c. Primer size: min, 21; opt, 23; max, 26.
 - d. Primer *Tm*: min, 59; opt, 62; max, 65.
 - e. Primer GC%: max, 50.
3. Designed primers are ordered and diluted to a concentration of 20 μ M. Diluted primers are mixed with equal volumes of forward and reverse primer. This mix is referred to as 10 μ M primer pair.

3.2. Polymerase Chain Reaction

1. Add 5 μ L of 5 ng/ μ L RH DNA for each hybrid into individual wells of a 96-well PCR plate; include human, hamster, and water control wells (*see Note 6*).
2. Prepare PCR mixture (per reaction): 2.0 μ L of 5X buffer, 0.8 μ L of 2.5 mM dNTPs, 0.8 μ L of 10 μ M primer pair, 0.07 μ L of AmpliTaq Gold (5 U/ μ L), 1.33 μ L of ddH₂O. We strongly suggest typing in duplicate (*see Note 7*).
3. Vortex the PCR mixture and add 5 μ L to each well, bringing the total PCR volume to 10 μ L (*see Note 8*).
4. Cover with heat-sealing tape and run in a 96-well PCR thermocycler using the following conditions: 95°C for 10 min, followed by 30 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 23 s, followed by 72°C for 3 min, 30 s (*see Note 9*).

3.3. Agarose Gel Electrophoresis

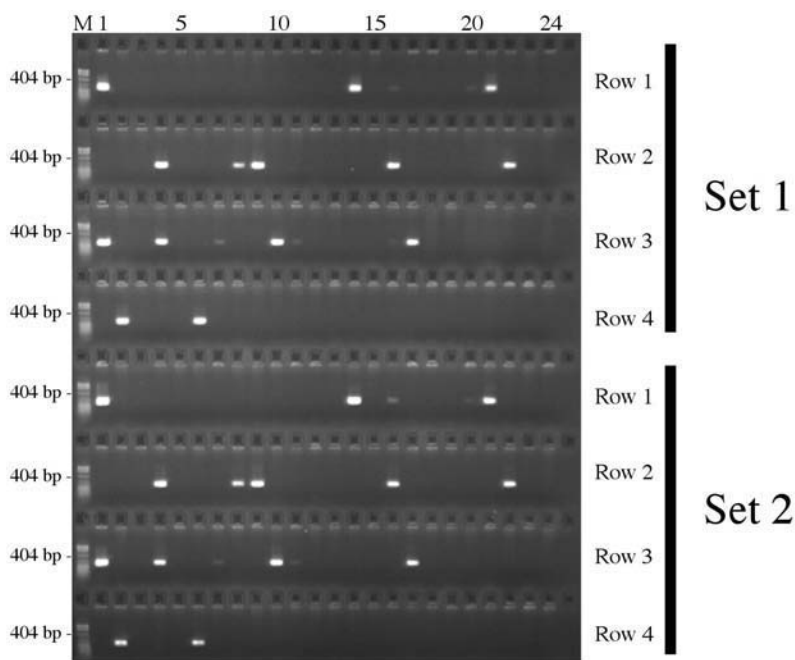
1. Pour a 3% agarose gel with 1X TBE buffer. Use a comb that will create wells to hold a 15- μ L vol (*see Note 10*).
2. Add 5 μ L of agarose gel loading buffer to the PCR reaction and load 13–15 μ L into the gel wells. Add a size standard to each row on the gel to confirm PCR product size (*see Note 11*).
3. Run electrophoresis in a horizontal gel chamber using 1X TBE buffer at 120 V for 20–45 min, depending on the desired resolution (*see Note 12*).
4. Stain in 0.9 mg/L of EtBr, 1X TBE for 15 min, and destain in 1X TBE for 30 min (*see Note 13*).
5. Take an image of the gel using a transilluminator and charge-coupled (CCD) camera imaging system (*see Note 14*). A representative image of one STS can be seen in **Fig. 1**.

3.4. RH Server Analysis

1. Determine raw score data by assigning scores based on the results from the gel image:
 - a. A score of 1 is assigned to a hybrid if a PCR band is present at the expected size in the lane for that hybrid.
 - b. A score of 0 is assigned if a band is not present in the lane for that hybrid.
 - c. A score of R is assigned if the result is ambiguous (*see Notes 15 and 16*).The raw score data for the STS gel image in **Fig. 1** are depicted below the gel image.
2. Access the SHGC RH server website at www-shgc.stanford.edu/RH/index.html and enter the raw scores to position the STS from the BAC of interest relative to markers on the TNG RH map. The RH server will send the results of a two-point statistical analysis via e-mail and will provide the following data:
 - a. The name(s) of SHGC markers linked to your BAC.
 - b. The chromosome on which your BAC is located.
 - c. The LOD score indicating the confidence of the position.
 - d. The distance in centiRay units between your BAC and the linked marker(s) (*see Note 17*).

4. Notes

1. The TNG RH DNA panel of 90 hybrids is available from Research Genetics/Invitrogen. Panels are provided at a concentration of 25 ng/ μ L in TE buffer and include positive and negative control DNAs.
2. The dNTPs are prepared by mixing the following: 10% DNA polymerization mix at 25 mM/dNTP, 1% 10X Perkin Elmer PCR Buffer II, and 89% ddH₂O. These stocks are stored at –20°C.
3. We have made 5X solution in bulk and store in 50-mL aliquots at –20°C.
4. We have made our TBE in 10X stock, which is diluted to 1X TBE in carboys for direct use.



STS vector: 0000000000010100011000000100011000000100000100100100110000010000000001000100000000000000000

Fig. 1. Representative data from agarose gel electrophoresis of STS on TNG RH panel. PCR products of STS with all 90 hybrid DNAs are separated on an agarose gel. The reaction is run in duplicate (sets 1 and 2). Each row contains 24 PCR reactions (1–24) and a size marker (M). Wells 1–3 of rows 1 and 3 contain positive (well 1), negative (well 2), and water (well 3) controls. The resulting raw data to be used in the RH server for this STS are shown below the image (STS vector).

5. Stock solutions of 1% Biotech-grade EtBr can be stored at room temperature. We make gel staining solutions of 0.9 mg of EtBr/L of 1X TBE in light-protected containers to use for up to 5 d, depending on the number of gels stained.
6. Human and hamster controls are provided with RH panels from Research Genetics/Invitrogen. We reassay plates that show no PCR product in our human controls or contamination in our water controls. Product in the hamster controls that is of the same size as the sequence of interest indicates the presence of that sequence in the hamster genome; therefore, mapping of this STS on these hybrids is impossible. A redesign of the STS is suggested.
7. We make enough mixture for 110% of the number of samples in our assay. Since ambiguities in typing can limit the ability to localize sequences of interest, we type in duplicate by preparing two identical 96-well PCR plates for each marker.

8. For ease in adding the PCR mixture to the PCR wells, the use of a multichannel pipettor, electronic multipipettor, or pipetting robot is recommended. In our high-throughput setup, we use an 8-channel Hamilton MicroLab 2200 robot.
9. Our conditions have been standardized for high-throughput mapping and are optimized for use with a Perkin Elmer 9700 thermocycler. We recommend using Perkin Elmer 96-well plasticware designed for use with the 9700 and Perkin Elmer MicroAmp clear adhesive films for sealing. You may choose to use your own PCR method that is optimized for your thermocycler and plasticware designs.
10. We recommend a gel-well setup that allows for multiple combs in a single gel. If typing in duplicate, it is ideal to load both PCR plates into one gel to avoid differences in staining or gel background that might interfere with gel data analysis.
11. We use *MspI* digested pBR322 marker as our size standard. If the PCR product does not match the known size of our STS, we will not map that sequence; if multiple PCR products are present, we only use the data of the STS if the correct fragment size is distinct from other products.
12. Our gels are made with eight rows of 26-well combs to accommodate two duplicate PCR plates per assayed marker, and agarose gel electrophoresis requires 20–25 min. We decrease the running time slightly for markers that are 125 bp or shorter, and we increase the running time if markers are longer than 350 bp. For a gel setup with four rows of combs, electrophoresis time of 45 min may be required.
13. Staining and destaining can be done in ddH₂O rather than TBE if desired. We use TBE to maintain the buffer concentration in the agarose gels since we melt and repour the gels once without a significant decrease in gel quality.
14. Our gel-imaging system is based on a UV transilluminator and a 640 × 480 pixel CCD camera, and the images are printed on thermal paper.
15. Data are considered ambiguous if the gel images from duplicate PCR plates show conflicting results. That is, if a PCR product is present in only one of the two duplicate PCR assays for a particular hybrid, we assign that hybrid a score of R; the RH server does not include hybrids with an R assignment when it maps the STS. There are several possible explanations for these conflicting results:
 - a. There could be false positives, meaning the PCR product is present in only one of the two assays, owing to human or PCR error.
 - b. There could be false negatives owing to human error in PCR setup or gel loading.
 - c. The desired sequence is present in such low quantities, owing to loss of DNA during the culturing of the hybrids, that PCR results are not always reproducible in that hybrid.
16. If the gel background, owing to either imperfections in the gel or the presence of PCR artifacts, is significant enough to obscure PCR product bands, we do not attempt to score despite the background. We have standardized our PCR conditions for our high-throughput mapping setup, so we will typically reassay that sequence in duplicate and score only if the gel background has decreased.

17. The RH server will compare the raw-score data for the STS with the scores of 36,678 markers mapped on the TNG RH panel. The average retention frequency of markers successfully mapped at SHGC is 18.8%, calculated as the percentage of 1 scores among the 90 hybrids per marker. However, the average maximum retention frequency per chromosome is 56.3% and the minimum average is 3.8%, so unusually high or low retention numbers do not necessarily prevent successful linkage. The probability of successful mapping of an STS is higher if the chromosome the sequence is on is already known. In addition, there is a greater chance of success when there are few ambiguities in the raw-score data. We do not map a marker when there are ambiguous results for eight or more hybrids, and we consider reassaying a marker that shows four or more ambiguities if the retention frequency for that marker is <20%.

References

1. International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
2. Cox, D. R., Burmeister, M., Price, E. R., Kim, S., and Myers, R. M. (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**, 245–250.
3. Stewart, E. A., McKusick, K. B., Aggarwal, A., et al. (1997) An STS-based radiation hybrid map of the human genome. *Genome Res.* **7**, 422–433.
4. Goodfellow, P. J., Povey, S., Nevanlinna, H. A., and Goodfellow, P. N. (1990) Generation of a panel of somatic cell hybrids containing unselected fragments of human chromosome 10 by X-ray irradiation and cell fusion: application to isolating the MEN2A region in hybrid cells. *Somat. Cell. Mol. Genet.* **16**, 163–171.
5. Lunetta KL, Boehnke M, Lange K, and Cox DR (1996) Selected locus and multiple panel models for radiation hybrid mapping. *Am. J. Hum. Genet.* **59**, 717–725.
6. Olivier, M., Aggarwal, A., Allen, J. R., et al. (2001) A high resolution radiation hybrid map of the human genome draft sequence. *Science* **291**, 1298–1302.
7. Olsen, M. L., Hood, L., Cantor, C., and Botstein, D. (1989) A common language for physical mapping of the human genome. *Science* **245**, 1434–1435.
8. Beasley, E. M., Myers, R. M., Cox, D. R., and Lazzaroni, L. C. (1999) Statistical refinement of primer design parameters, in *PCR Applications* (Innis, M. A., Gelfand, D., and Sninsky, J. J., eds.), Academic, San Diego, pp. 55–71.

Shotgun Library Construction for DNA Sequencing

Bruce A. Roe

1. Introduction

Shotgun cloning is a method to generate the templates needed for DNA sequencing. This process entails breaking a large target DNA randomly into smaller fragments; end sequencing these smaller fragments; and from the overlapping sequences of the randomly generated fragments, reassembling the initial target sequence. Although this random strategy initially was described more than two decades ago (*1–3*), only a few years following the original reports describing the dideoxynucleotide method for DNA sequencing (*4–7*), it immediately did not gain wide-scale acceptance. It was not until the introduction of instrumentation (*8–10*) that was capable of collecting the required large quantity of data required for successfully implementing the shotgun method, that this approach began to be widely accepted. The shotgun cloning strategy presently is the method of choice for generating the major portion of the sequence data for sequencing projects. This holds true for target DNAs as small as a 4-kbp restriction digest fragment (*1*) or as large as an entire 3-Gbp mammalian or plant genome (*11–13*) since many of the methods used have been automated (*14,15*). These partially or fully automated methods include the shotgun DNA template clone isolation; DNA sequence reaction pipetting; DNA sequence data collection; DNA sequencing chemistry; and computer-based shotgun sequence data assembly, visualization, and editing (*16–18*). Therefore, it is reasonable to initially obtain from 6- to 10-fold shotgun sequence coverage and then proceed with more directed closure and finishing methods. In this chapter, the methods I describe are those presently used in my laboratory's genome center that have evolved over the years into reliable and robust techniques for shotgun library construction and subsequent DNA sequencing.

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

2. Materials

Chemicals should be American Chemical Society grade or higher, and solutions should be prepared in deionized distilled water. For those reagents for which commercial kits are available, the vendor-specified instructions were followed unless otherwise stated.

2.1. Bacterial Artificial Chromosome DNA Isolation From 200-mL Cultures by a Cleared Lysate Method Followed by Double Acetate Precipitation

1. Luria Bertani (LB) medium: 10 g of Bacto Tryptone, 5 g of Bacto yeast extract, and 10 g of NaCl in 1 L of distilled water, autoclaved.
2. 0.5 M EDTA: Dissolve 186.1 g of Na₂EDTA in 400 mL of sterile distilled water, adjust the pH to 8.0 with 10 M NaOH, and then adjust to 1-L final volume with sterile distilled water.
3. 10 mM EDTA, pH 8.0, prepared by diluting 2 mL of 0.5 M EDTA.
4. GET: 50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0.
5. Lysis solution: 0.2 N NaOH, 1% sodium dodecyl sulfate (SDS).
6. 3 M KOAc: Mix 50 mL of 7.5 M potassium acetate (KAc) without pH adjustment with 23 mL of acetic acid and 127 mL of ddH₂O.
7. 10:50 TE: 10 mM Tris-HCl, pH 7.6, 50 mM EDTA, pH 8.0.
8. DNase-free RNase A: 20 mg/mL of RNase A in 1 mM NaOAc (pH 4.5) prepared by adding 200 mg of RNase A (R-5500; Sigma, St. Louis, MO) to 3.3 μ L of 3 M sodium acetate (pH 4.5) and ddH₂O to a 10-mL final volume. After incubating for 10 min in a boiling water bath and aliquoting into 1-mL portions, store this DNase-free RNase frozen at -20°C.
9. RNase T1: 100 U/ μ L in 50 mM Tris-HCl, pH 7.6, 500- μ L final volume stored at -20°C in 50- μ L aliquots:
 - a. 100 μ L of RNase T1 (cat. no. R-8251; Sigma) (100,000 U/0.2 mL).
 - b. 25 μ L of 1 M Tris-HCl, pH 7.6.
 - c. 375 μ L of ddH₂O.
10. ITE, pH 8.0: 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0.

2.2. Physical Shearing in a Nebulizer or Hydroshear

1. Nebulizer (part number 4101, 4101, or UO 4207; IPI Medical Products, Chicago, IL) or Hydroshear (Gene Machines).
2. Centrifuge tubes (1.5 mL) (part no. 72-690; Startstedt).
3. Tabletop centrifuge (Marathon 13k/M)
4. 10X TM buffer: 500 mM Tris-HCl, pH 8.0, 150 mM MgCl₂ in sterile ddH₂O.

2.3. Random Fragment End Repair, Size Selection, and Phosphorylation

1. T4 polynucleotide kinase and 10X kinase buffer (cat. no. 70031; United States Biochemicals).
2. Klenow DNA polymerase (cat. no. 210L; New England Biochemicals).

3. T4 DNA polymerase (cat. no. 203L; New England Biochemicals).
4. 10 mM rATP: Diluted from a 100 mM solution of dipotassium adenosine triphosphate (ATP) made by dissolving 619 mg of dipotassium ATP (cat. no. 27-1006-01; Amersham Pharmacia Biotech) in 10 mL of sterile distilled water, aliquot, and store at -20°C .
5. 0.25 mM dNTPs: Dilute from 100 mM stock solutions (cat. no. 27-2035-01; Amersham Pharmacia Biotech).

2.4. DNA Ligation

1. T4 DNA ligase (cat. no. 202L; New England Biolabs).
2. *Sma*I-linearized, CIAP-dephosphorylated pUC vector (cat. no. 27-4860-01; Amersham Pharmacia Biotech).

2.5. Preparation of Competent Cells

1. 50 mM calcium chloride: Dissolve 0.74 g of $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ in sterile distilled water to a final volume of 100 mL, autoclave to sterilize, and store at 4°C .
2. XL1-Blue *Escherichia coli* cells: recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac (F'proAB lacIqZDM15 Tn10 [Tetr]) (cat. no. 200228; Stratagene).
3. YENB: Bring 7.5 g of Bacto yeast extract and 8 g of Bacto nutrient broth to 1 L with distilled water and autoclave.
4. Tetracycline (10 mg/mL) in 50% ethanol–sterile distilled water: Prepare by dissolving 1 g of tetracycline (cat. no. T-3383; Sigma) in 50 mL of 100% ethanol and add sterile distilled water to 100 mL. Store at 4°C in the dark.
5. 5-Bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal) (20 mg/mL) (cat. no. B-4252; Sigma): Dissolve in dimethylformamide (DMF), aliquot, and store protected from light at -20°C .
6. Isopropyl- β -D-thiogalactopyranoside (IPTG) (30 μL of 25 mg/mL) (cat. no. I5502; Sigma): Dissolve in sterile distilled water.
7. LB-Amp Petri dishes: 10 g of Bacto Tryptone (cat. no. 0123-01-1; Difco, Detroit, MI) plus 5 g yeast extract (cat. no. 0127-05-3; Difco), 10 g of NaCl, and 15 g agar (cat. no. 0140-01; Difco): Dissolve in 1 L of sterile distilled water, autoclave to sterilize, and cool to 55°C . Add antibiotic to a final concentration of 100 $\mu\text{g}/\text{mL}$ and then pour the mixture into sterile Petri dishes (approx 20 mL/plate).
8. Tetracycline plate: 20 μg of tetracycline/mL of LB agar.
9. GPR centrifuge (Beckman) or RC5-B centrifuge (DuPont) equipped with an SS-34 rotor.

2.6. 384-Well DNA-Sequencing Template Picking, Growth, and Isolation Using a Flexys Colony Picker, HiGro Oxygenated Shaker Incubator, and Hydra 96 Automated Pipettor With Moving Stage

1. 384-Well flat-bottomed microtiter plate (cat. no. 242757; Nunc).
2. 10X TB: 12 g of Bacto Tryptone plus 24 g of yeast extract and 4 mL of glycerol brought to 900 mL with distilled water. After sterilization by autoclaving, cool

the solution. Then add 100 mL of 10X TB salts and adjust the final volume to 1 L with sterile distilled water.

3. 10X TB salts: Dissolve 2.31 g of potassium dihydrogen phosphate and 12.54 g of potassium phosphate, dibasic in 100 mL of distilled water and autoclave.
4. TE-RNase: 50 mM Tris-HCl, pH 7.6, 0.5 M EDTA-Na, 40 $\mu\text{g}/\mu\text{L}$ of RNase A, 0.04 U/ μL of T1 RNase.
5. Lysis buffer: 1% SDS, 0.2 M NaOH.
6. 3 M NaOAc: 408.24 g of sodium acetate $\cdot 3\text{H}_2\text{O}$ in a total volume of 1 L adjusted to pH 4.5 with acetic acid.
7. 3 M KOAc: 294.45 g of potassium in a total volume of 1 L adjusted to pH 4.5 with acetic acid.
8. BigDye v3.0 (cat. no. 4390253; Applied Biosystems) diluted 1 : 16 with 5X reaction buffer (400 mM Tris-HCl, pH 9.0, containing 10 mM magnesium chloride).
9. Robins plates: part no. 1047-20-1 for the red 96-well Robins plates and part no. 1047-20-4 for the green 96-well Robins plates.
10. Dimethylsulfoxide (DMSO) (cat. no. D8779; Sigma).
11. 5X TM buffer: 20 mL of Tris-HCl, pH 9.0, plus 0.5 mL of 1 M magnesium chloride and 29.5 mL of sterile distilled water.
12. HiGro incubator (Gene Machines).

3. Methods

3.1. Bacterial Artificial Chromosome DNA Isolation From 200-mL Cultures by a Cleared Lysate Method (19) Followed by Double Acetate Precipitation

1. Pick a smear (rather than a single colony) of bacterial artificial chromosome (BAC) colonies and transfer into a 12 $\times$ 75 mm Falcon tube containing 3 mL of LB medium with appropriate antibiotic. After incubating at 37°C for 8–10 h with 250 rpm shaking, transfer the culture to a 500-mL flask containing 200 mL of the same medium and incubate for 8–10 h under the same conditions. After harvesting the cells by centrifuging at 4000g for 15 min in a 250-mL bottle in the RC5-B using the GSA rotor, freeze the cell pellets and store at -70°C .
2. Prior to use, thaw the cells from the 200-mL growth, and resuspend in 8 mL of 10 mM EDTA, pH 8.0, by gently pipetting up and down with a 10-mL pipet but do not vortex. Cells should be resuspended completely. The cells suspend more efficiently with 10 mM EDTA alone (rather than with GET [50 mM glucose; 25 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0]). After mixing gently, incubate the solution at room temperature for 5 min.
3. To the resuspended cells, add 16 mL of alkaline lysis solution, and after very gently swirling until the solution is homogeneous, incubate for 5 min at room temperature. The whole step should be finished within 10 min.
4. Immediately add 12 mL of cold, approx 3 M KOAc and mix very gently by swirling the bottle several times. Then place the bottle in a freezer (either -20 or -80°C) and store frozen at -20 or -80°C overnight. Freezing at this stage has recently been shown to result in a much cleaner final DNA preparation because

the resulting pellet is much more tightly packed if the samples are frozen for several hours prior to clearing the lysate by the subsequent centrifugation step.

5. Clear the lysate from the precipitated SDS, proteins, membranes, and chromosomal DNA by centrifuging at 16,000g for 15 min at 4°C in the RC5-B using the GSA rotor. Prior to recentrifugation, the cleared lysate supernatant can be filtered through a double layer of cheesecloth to remove any floating material, although this typically is not necessary. Then, an additional centrifugation is performed to ensure that all insolubles are removed.
6. Transfer the supernatant into one 250-mL bottle for each 200 mL of original cell growth, precipitate the DNA by adding an equal volume of isopropanol, and mix by swirling. After centrifuging at 4000g for 15 min in the RC5-B at 4°C using the GSA rotor, decant the supernatant and drain the pellet.
7. To perform a second acetate precipitation step, quickly and gently dissolve the DNA pellet in 3.6 mL of 10:ITE. Then add 1.8 mL of 7.5 M KOAc (this 7.5 M KOAc solution is used without pH adjustment) to each bottle and transfer to a 50-mL Sorval centrifuge tube.
8. After mixing, freeze the tubes at -70°C for at least 30 min (until frozen solid), but typically they can be frozen overnight. This freezing step ensures that the pellet produced in the subsequent centrifugation step forms a tight layer.
9. After thawing, centrifuge at 16,000g for 10 min in the SS34 rotor.
10. Transfer the supernatant of each tube into a 50-mL Corning centrifuge tube. Then add DNase-free RNase A to a final concentration of 100 µg/mL from a 20 mg/mL stock solution and RNase T1 to a final concentration of 40 µL/100 mL from a 100 U/µL stock solution followed by incubation in a 37°C water bath for 1 h. This ribonuclease treatment step reduces the amount of RNA present in the final BAC preparation.
11. Precipitate the DNA by adding 30 mL of cold 95% ethanol to each tube. After mixing, by inverting, incubate the tubes in an ice-water bath for 15 min followed by pelleting the precipitated DNA by centrifuging at 2000g for 25–30 min in a Beckman GS-R centrifuge.
12. Wash each pellet with 30 mL of 70% ethanol, and dry in a vacuum.
13. Dissolve each pellet in 100 µL of 10/0.1 TE by pipetting up and down and incubating at 37°C. Store this solution 4°C overnight to ensure that all the high-molecular-weight large-insert clone DNA is dissolved completely (*see Notes 1 and 2*).

3.2. Physical Shearing in a Nebulizer or Hydroshear

3.2.1. Nebulizer

1. To modify a nebulizer, remove the plastic cylinder drip ring, cut off the outer rim of the cylinder, invert it, and place it back into the nebulizer. Then seal the large hole in the top cover (where the mouthpiece was attached) with a plastic stopper, and connect one end of a 0.25-in-id length of Tygon tubing (the other end of which eventually will be connected to a compressed air source) to the smaller hole (*see Notes 2 and 3*).

2. Prepare the following DNA sample and place in the nebulizer cup: 50 μg of DNA, 200 μL of 10X TM buffer, 0.5–1 mL of sterile glycerol, and sterile ddH_2O , to a final volume of 2 mL.
3. Place the nebulizer in an acetone–dry ice bath, and nebulize the DNA samples by applying 30 psi for 2.5 min for plasmids, or 8–10 psi for 2.5 min for BACs, P1-derived artificial chromosomes, fosmids, or cosmids (*see Note 4*).
4. Place the entire nebulizer unit in the rotor bucket of a tabletop centrifuge (Beckman GPR tabletop centrifuge) fitted with pieces of styrofoam to cushion the plastic nebulizer, and centrifuge at 2000g to collect the sample at the bottom of the nebulizer unit.
5. Distribute the sample into four 1.5-mL microcentrifuge tubes, and ethanol precipitate with 2.5 vol of ethanol acetate. After collecting the DNA pellet by centrifugation, dissolve in 35 μL of 1X TM buffer prior to proceeding with fragment end repair (*see Note 5*).

3.2.2. Hydroshear (20)

The Hydroshear has a volume range of 40–300 μL , and the DNA concentration and initial fragment length do not dramatically affect the postshearing fragment length.

1. Before shearing the DNA, transfer 40–300 μL of approx 2 $\mu\text{g}/\mu\text{L}$ DNA solution to a 1.5-mL centrifuge tube, and centrifuge on a tabletop centrifuge at 12,000g for 30 min at room temperature to remove any undissolved particulate matter. In addition, create a file on the Hydroshear-attached computer, where the shearing parameters are set, i.e., number of shearing cycles and syringe rate needed to obtain fragments of the desired length.
2. After transferring the supernatant to a new tube and discarding the undissolved material in the pellet, incubate the DNA solution on ice for 10 min.
3. Turn on the Hydroshear device using the switch on the bottom of the back.
4. Using the mouse attached to the computer attached to the Hydroshear, click on the Hydroshear icon on the screen.
5. Click on the “Read From File” button located on the Hydroshear Device Main Panel and choose the file called HS.Sp10. (This is a file in which the parameters that produce an average length of DNA fragment of between 2000 and 4000 bp were previously specified.)
6. Change the volume setting on the main panel to reflect the actual volume of the DNA sample to be sheared.
7. Click the Start button on the main panel and a dialog box will appear displaying “Do you want to wash the shearing device before starting the shearing process?”
8. Click OK unless the Hydroshear Device has just been cleaned, and the wash scheme specified will be performed. On completion of this wash step, a dialog box appears displaying, “Washing complete” and an OK button.
9. Click OK and a dialog box will appear displaying, “Proceed with loading the sample?” along with two buttons: Load and Bypass. If the sample is already in the

syringe, click Bypass; however, usually the sample will not yet have been loaded so Load should be clicked. A dialog box then will appear displaying "Prepare to load the sample" and an OK button.

10. Check to ensure that the DNA sample does not contain any particulate material and that it is thoroughly in solution. If any particulate material is noticed, repeat **steps 1 and 2**.
11. Click OK and a dialog box will appear displaying "Turn valve to input, bring sample to input tube, and click OK."
12. Turn the valve handle counterclockwise to the input position.
13. Hold the 1.5-mL tube containing the sample into the input tube on the Hydroshear, and make sure that the input tube reaches the bottom of the 1.5-mL sample tube.
14. Click OK and the specified volume will be drawn from the sample tube into the input tube. When the dialog box appears displaying "Remove sample from the input tube and click OK," remove the input tube from the sample tube, place the output tube into the empty sample tube, and secure it there using tape. Then place the sample tube back into the ice bath.
15. Click OK to draw the sample into the syringe from the input tube. When the dialog box appears displaying "Turn valve to output and click OK," turn the valve halfway so that the valve handle is vertical, i.e., pointing downward (do not turn more than halfway), and again click OK.
16. Watch the bubble rise as air is pushed out of the syringe, and then turn the output valve handle entirely to its output position. A dialog box then will appear displaying "Sample loading complete" and an OK button. Click the OK button (*see Note 6*).
17. When the first message appears asking "Is there an air gap near the bottom of the syringe?" click No. Also, click No in response to the second and third Clear Air Gap messages.
18. When the "Clear Air Gap" dialog box disappears and a dialog box appears displaying "Click OK to begin shearing cycles," click OK. The hydroshear then will perform 20 cycles of shearing passes as specified in the Shearing Parameters area of the Hydroshear Device Main Panel. On completion of the shearing cycles, a dialog box appears displaying "Shearing cycles complete" and an OK button. Click OK.
19. A dialog box then appears displaying "Proceed to eject the sample from the device?" and a Yes button. Click Yes and a dialog box will appear displaying "Turn the valve to input and click OK."
20. Click OK to draw air into the input tube, and a dialog box will appear displaying "Turn valve to output to eject sample" and an OK button.
21. Click OK and the sheared sample will be pushed through the Hydroshear DNA Shearing Device's output tube into the sample tube.
22. When the dialog box appears displaying "Do you want to wash the device?" and a Yes and No button, Click Yes to wash the Hydroshear Device as specified in the Wash Scheme window.
23. Once this wash is completed, exit the Hydroshear file on the computer and turn off the Hydroshear device. Then the sheared DNA may be dried in a vacuum and

brought to the volume desired for the end repair and fill-in kinase reaction, or it may be ethanol precipitated, collected by centrifugation, and subsequently dissolved in the desired volume of water.

3.3. Random Fragment End Repair, Size Selection, and Phosphorylation

Since both sonicated and nebulized DNA fragments usually contain single-stranded ends, the samples must be end repaired prior to ligation into blunt-ended vectors (*10,11*). A combination of T4 DNA polymerase and Klenow DNA polymerase are used to “fill in” the DNA fragments by catalyzing the 3′–5′ incorporation of complementary nucleotides into resultant double-stranded fragments with a 5′ overhang. Additionally, the single-stranded 3′–5′ exonuclease activity of T4 DNA polymerase is used to degrade 3′ overhangs. These reactions, which include the two enzymes, buffer, and deoxynucleotides, typically are incubated at 37°C.

Following fragment end repair, the DNA samples are electrophoresed on a preparative low-melting-temperature agarose gel vs the ϕ -X 174 marker, and after appropriate separation, the fragments in the size ranges of 1 to 2 and 2–4 Kbp are excised and eluted separately from the gel. Here, the band of interest is excised with a sterile razor blade, placed in a microcentrifuge tube, frozen at –70°C, and then melted. Next, TE-saturated phenol is added to the melted gel slice, and the mixture again is frozen and then thawed. After this second thawing, the tube is centrifuged and the aqueous layer removed to a new tube. Residual phenol is removed with two ether extractions, and the DNA is concentrated by ethanol precipitation. Alternatively, the fragments can be purified by fractionation on a Sephacryl S-500 spin column as also discussed. In both instances, the purified fragments are concentrated by ethanol precipitation followed by resuspension in kinase buffer, and phosphorylation using T4 polynucleotide kinase and rATP. The polynucleotide kinase is removed by phenol extraction and the DNA fragments are concentrated by ethanol precipitation, dried, resuspended in buffer, and ligated into blunt-ended cloning vectors. Note that because a significant portion of nebulized DNA fragments are easily cloned without end repair or kinase treatment, these two steps can be combined without significantly affecting the overall number of resulting transformed. The detailed protocol is as follows:

1. Resuspend the DNA in 27 μ L of 1X TM buffer. Add the following: 5 μ L of 10X kinase buffer, 5 μ L of 10 mM rATP, 7 μ L of 0.25 mM dNTPs, 1 μ L (3 U/ μ L) of T4 polynucleotide kinase, 2 μ L (5 U/ μ L) of Klenow DNA polymerase, and 3 μ L (3 U/ μ L) of T4 DNA polymerase, to a final volume of 50 μ L.
2. Incubate at 37°C for 30 min.

3. Add 5 μL of agarose gel-loading dye, apply to a separate well of a 1% low-melting-temperature agarose gel, and electrophorese for 30–60 min at 100–120 mA.
4. Elute the DNA from each sample lane by placing the excised DNA-containing agarose gel slice in a 1.5-mL microcentrifuge tube, and freeze at -70°C for at least 15 min, or until frozen. It is possible to pause at this stage in the elution procedure and leave the gel slice frozen at -70°C .
5. Melt the slice by incubating the tube at 65°C .
6. Add 1 vol of TE-saturated phenol, vortex for 30 s, and freeze the sample at -70°C for 15 min.
7. Thaw the sample, and centrifuge in a microcentrifuge at 12,000g for 5 min at room temperature to separate the phases. Then remove the aqueous phase to a clean tube, extract twice with an equal volume of ether, and ethanol precipitate by adding 2.5 vol of ethanol acetate. Collect the DNA by centrifugation, rinse once with 70% ethanol, and dry. Dissolve the resulting dried pellet in 10 μL of 10:0.1 TE buffer.

3.4. DNA Ligation

DNA ligations are performed by incubating DNA fragments with appropriately linearized cloning vector in the presence of buffer, rATP, and T4 DNA ligase (21,22). For random shotgun cloning, nebulized or hydrosheared fragments are ligated to either *Sma*I-linearized, dephosphorylated double-stranded M13 replicative form or pUC vector by incubation at 4°C overnight. A practical range of concentrations is determined based on the amount of initial DNA, and several different ligations, each with an amount of insert DNA within that range, are used to determine the appropriate insert-to-vector ratio for the ligation reaction. In addition, several control ligations are performed to test the efficiency of the blunt-ending process, the ligation reaction, and the quality of the vector (21,22). These usually include parallel ligations in the absence of insert DNA to determine the background clones arising from self-ligation of inefficiently phosphatased vector. Parallel ligations also are performed with a known blunt-ended insert or insert library, typically an *Alu*I digest of a cosmid, to ensure that the blunt-ended ligation reaction will yield sufficient insert-containing clones, independent of the repair process. The protocol is as follows:

1. Combine the following reagents in a microcentrifuge tube, and incubate overnight at 4° (see Note 7): 100–1000 ng of DNA fragments, 2 μL (10 ng/ μL) of cloning vector, 1 μL of 10X ligation buffer, 1 μL (400 U/ μL) of T4 DNA ligase, and sterile ddH₂O, to a total volume of 10 μL .
2. Include control ligation reactions with no insert DNA and with a known blunt-ended insert (such as an *Alu*I-digested cosmid).

3.5. Preparation of Competent Cells

There are two main methods for preparation of competent bacterial cells for transformation: the calcium chloride (23) and the electroporation (24,25) methods. For the calcium chloride method, a glycerol cell culture stock of the respective *E. coli* strain is thawed and added to 50 mL of liquid medium. This culture is then preincubated at 37°C for 1 h, transferred to an incubator-shaker, and incubated further for 2 to 3 h. The cells are pelleted by centrifugation, resuspended in calcium chloride solution, and incubated in an ice-water bath. After another centrifugation step, the resulting cell pellet is again resuspended in calcium chloride to yield the final competent cell suspension. Competent cells are stored at 4°C for up to several days.

3.5.1. Calcium Chloride Protocol

1. Thaw a frozen glycerol stock of the appropriate strain of *E. coli*, add it to an Erlenmeyer flask containing 50 mL of prewarmed 2X TY medium, and preincubate in a 37°C water bath for 1 h with no shaking. Further incubate for 2 to 3 h at 37°C with shaking at 250 rpm.
2. Transfer 40 mL of the cells to a sterile 50-mL polypropylene centrifuge tube, and collect the cells by centrifuging at 2000g for 8 min at 4°C in a Beckman GPR centrifuge or 4000g for 8 min at 4°C in a DuPont RC5-B centrifuge equipped with an SS-34 rotor.
3. Decant the supernatant and resuspend the cell pellet in ½ vol (20 mL) of cold, sterile 50 mM calcium chloride; incubate in an ice-water bath for 20 min; and centrifuge as in **step 2**.
4. Decant the supernatant and gently resuspend the cell pellet in ⅓ vol (4 mL) of cold, sterile 50 mM calcium chloride to yield the final competent cell suspension.

3.5.2. Preparation of Calcium Chloride Competent Cells for Frozen Storage

1. Transfer 166 µL of the competent cell suspension to sterile Falcon culture tubes.
2. Add 34 µL of sterile 100% glycerol to the 166-µL aliquots of the final competent cell suspension prepared in **step 1**, giving a final concentration of 17% glycerol.
3. Place the competent cells at -70°C; they can be stored indefinitely.
4. To use competent cells for transformation, remove from the freezer and thaw for a few minutes at 37°C. Place on ice, add plasmid DNA, and incubate for 1 h as in the standard transformation procedure. Then heat shock at 42°C for 2 min, cool briefly, add 1 mL of 2X TY, and incubate for 1 h at 37°C before spreading on plates.

3.5.3. Preparation of Electrocompetent Cells

1. Grow XL1-Blue cells on a tetracycline plate (20 µg of tetracycline/mL of LB agar).
2. Inoculate 3 mL of YENB and grow overnight at 37°C with shaking at 250 rpm in a New Brunswick incubator shaker.

3. Inoculate the 3 mL of overnight growth into 1 L of YENB and grow to an A_{600} of 0.5 (typically requires 3 to 4 h of shaking at 250 rpm in the New Brunswick incubator shaker at 37°C).
4. Distribute the 1 L of cells into four 500-mL Sorval (GS-3) centrifuge bottles, and centrifuge at 4000g at 4°C for 10 min.
5. Resuspend each pellet in 100 mL of ice-cold, sterile ddH₂O, and combine the resuspended pellets into two Sorval centrifuge bottles (i.e., each bottle will then contain 200 mL of resuspended pellet) (see **Note 8**).
6. Centrifuge at 4000g at 4°C for 10 min in the Sorval GS-3 rotor.
7. Resuspend each of the two pellets in 100 mL of ice-cold, sterile ddH₂O, and combine the resuspended pellets into one Sorval centrifuge bottle and centrifuge at 4000g at 4°C for 10 min in the Sorval GS-3 rotor once more (see **Note 9**).
8. Resuspend the pellet in 100 mL of 10% ice-cold, sterile glycerol; centrifuge as above, and finally resuspend the pellet in 2 mL of 10% ice-cold, sterile glycerol to give salt-free, concentrated electrocompetent cells.
9. Aliquot 40 μ L of these electrocompetent cells into small snap-cap tubes, immediately freeze by placing in crushed dry ice, and then store at -70°C until needed.

3.5.4 Electroporation Protocol for Transformations

Using Double-Stranded Plasmids

1. Thaw the electrocompetent cells on ice for about 1 min.
2. Add 2 to 3 μ L of the ligation mix to the cells.
3. Transfer 40 μ L of the cells into a BTX Electroporation cuvet and make sure that the cells cover the bottom of the cuvet.
4. Turn on the Bio-Rad *E. coli* Pulser and set the current to 2.5 kV by pushing the “Lower” and “Raise” bottoms simultaneously twice.
5. Place the cuvet in the holder and slide it into position.
6. Charge by pressing the “Charge” bottom until you hear the beep.
7. Immediately suspend the cells in 1 mL of YENB and transfer into a Falcon tube.
8. Incubate the cells in a shaker at 37°C for 30 min at 250 rpm.
9. Spin the cells in a Beckman tabletop centrifuge for 8 min at 1100g.
10. Resuspend the cells in 200 μ L of fresh YENB, and add 30 μ L of 20 mg/mL X-gal and 30 μ L of 25 mg/mL IPTG.
11. Plate approx 130 μ L of the cells on prewarmed LB-Amp plates.

3.5.5. Calcium Chloride-Treated Bacterial Cell Transformation

For DNA transformation (23,26), the entire DNA ligation reaction is added to an aliquot of competent cells, which is mixed gently, and incubated in an ice-water bath. This mixture is then heat shocked briefly in a 42°C water bath for 2–5 min.

For pUC-based transformation (26), an aliquot of liquid medium is added to the heat-shocked mixture, which is then incubated in a 37°C water bath for 15–20 min. After recovery, the cell suspension is concentrated by centrifugation

and then gently resuspended in a smaller volume of fresh liquid medium. IPTG and X-gal are added to the cell mixture, which is spread onto the surface of an ampicillin-containing agar plate. After the cell mixture has diffused into the agar medium, the plates are inverted and incubated overnight at 37°C.

1. Add the entire ligation reaction to a 12 × 75 Falcon tube containing 0.2 to 0.3 mL of competent cells, mix gently, and incubate in an ice-water bath for 40–60 min. (For retransformation of recombinant DNA, add approx 10–100 ng of DNA directly to the competent cells.)
2. Heat shock the cells by incubating at 42°C for 2–5 min.
3. Add 1 mL of fresh 2X TY to the heat-shocked transformation mixture, and incubate in a 37°C water bath for 15–30 min.
4. Collect the cells by centrifuging at 2000g for 5 min, decant the supernatant, and gently resuspend in 0.2 mL of fresh 2X TY.
5. Add 25 µL/IPTG (25 mg/mL of water) and 25 µL/X-gal (20 mg/mL of DMF), mix, and pour onto the surface of a prewarmed LB-Amp Petri dish. Spread over the agar surface using a sterile bent glass rod or sterile inoculating loop.
6. Allow 10–20 min for the liquid to diffuse into the agar, and then invert and incubate overnight at 37°C (*see Note 10*).

3.6. 384-Well DNA-Sequencing Template Picking, Growth, and Isolation Using a Flexys Colony Picker, HiGro Oxygenated Shaker Incubator, and Hydra 96 Automated Pipettor With Moving Stage

1. Pick the colonies into a 384-well flat-bottomed microtiter plate containing 70 µL of TB + salt supplemented with 100 µg/mL of ampicillin using the Flexys colony picker. Also pick a few dozen colonies into an extra 96-well plate containing 200 µL of medium for replacing the contents of wells with no cell growth.
2. Incubate the plates in a HiGro incubator for 22 h at 37°C with shaking at 520 rpm. The oxygenated flow is set to begin 3.5 h after shaking begins and a full open flow rate with the HiGro Oxygen Flow Setting at 0.5 s on and 0.5 min off.
3. Examine the wells and replace the contents of the wells with no cell growth with culture from positive growth cells obtained from the extra plate containing additional colonies picked into a partial plate.
4. Centrifuge the 384-well plates at 3000 rpm in a Beckman CS-6R tabletop microtiter plate centrifuge for 10 min.
5. Decant the supernatant by inverting the plates onto three to four layers of paper towels and gently tapping the inverted bottom of the plate.
6. Freeze the plates for at least 2 to 3 h or overnight at –20°C.
7. Remove the plates from the freezer, and using the Hydra add 25 µL of TE-RNase solution and shake on a benchtop shaker for 30 min at a setting of 10.
8. Add 25 µL of lysis buffer and shake on the benchtop shaker for 30 min at a setting of 8.

9. Add 25 μL of 3 *M* NaOAc or 3 *M* KOAc and shake in the HiGro at 37°C at a setting of 520 rpm for 30 min.
10. Freeze overnight at -80°C.
11. Thaw the plates for approx 30 min and centrifuge at 2000g for 45 min at 4°C in the Beckman CS-6R tabletop microtiter plate centrifuge.
12. Transfer 40 μL of the supernatant into a new 384-well plate.
13. Add 40 μL of 100% isopropanol using the Hydra.
14. Let stand at room temperature for 3–5 min, and then centrifuge at 2000g for 30 min at 4°C in the Beckman CS-6R tabletop microtiter plate centrifuge.
15. Decant the supernatant by inverting the plates onto three to four layers of paper towels and gently tapping the inverted bottom of the plate. Then, immediately wash by adding 50 μL of room temperature or 4°C 70% ethanol using the Hydra, and centrifuge at 3000 rpm for 10 min at 4°C in the Beckman CS-6R tabletop microtiter plate centrifuge.
16. Decant the final ethanol wash, and dry the pelleted DNA for 10 min in a vacuum. Then dissolve the final, dried DNA in 20 μL of water and shake on a benchtop shaker for 30 min at a setting of 8 (see Note 10).

3.7. 384-Well DNA Sequence Reaction Pipetting From 96-Well Predispensed BigDye Mix Plates to 384-Well Viper Plates for BigDye Version 3

Because of the large number of sequencing reactions typically performed in a shotgun sequencing approach, it is useful to aliquot the BigDye Version 3 premix into 96-well microtiter plates and store these diluted mixes frozen. Typically the Robins plates (part no. 1047-20-1 for the red 96-well Robins plates and part no. 1047-20-4 for the green 96-well Robins plates) are convenient to use for this purpose. The diluted reaction mix includes BigDye premix (1:16 dilution), either forward or reverse primer, as well as DMSO and 5X TM buffer.

1. To make the premix, combine the following: 400 μL of forward or reverse primer (6.5 μM stock primer solution), 200 μL of ABI BigDye (Version 3), 50 μL of DMSO, 150 μL of 5X TM buffer, and 200 μL of ddH₂O, to a final volume of 1000 μL .
2. After mixing, dispense 10 μL of this premix into each well of the 96-well colored Robins plates using the Hydra, and store frozen at -20°C. The green 96-well Robins plates contain the forward primer BigDye premix. The red 96-well Robins plates contain the reverse primer BigDye premix. Once all the premix has been dispensed from the colored plates, they can be reused.
3. Prior to use, remove a 96-well premix-containing plate from the freezer and thaw for a few seconds. Then, after centrifuging at 500g for 2 s to concentrate the BigDye mix to the bottom of the wells in the 96-well plate, place the plate in the source position on the Hydra, and place an empty 384-well viper plate in the target position on the Hydra. A program called Transfer Big Dye v3.0 Mix was

written on the Hydra-associated computer that initially transfers 2 μL of BigDye mix to each well of the viper plate.

4. Remove the viper plates and centrifuge at 500g for 2 s to concentrate the reaction mixture at the bottom of the wells in the 384-well viper plate.
5. After placing the centrifuged 384-well viper plates back on the Hydra, run a program in which the Hydra transfers 4 μL of DNA-sequencing template that was isolated as described in the 384-well DNA sequencing template isolation protocol and that was dissolved in 20 μL of ddH₂O into each of 384 wells of the viper plate containing the already diluted BigDye Version 3.0 premix.
6. Following centrifugation of the viper plates containing the entire reaction mixture, centrifuge at 500g for 2 s to concentrate the reaction mixture at the bottom of the wells in the 384-well viper plate.
7. Transfer the 384-well reaction plate to the Viper PCR instrument, and begin the cycle-sequencing protocol as recommended by ABI but modified to run for 60 cycles.
8. Once the cycle sequencing is completed, remove excess unreacted dye from the reactions by ethanol precipitation of the sequencing reaction products with 25 μL of ethanol acetate, followed by centrifugation, washing with 25 μL of 70% ethanol, centrifugation, removal of ethanol by inverting onto paper towels, and subsequently storing the dried reactions at -20°C until they are dissolved in 20 μL of sterile distilled water loaded into the first 96 wells in the 384-well plate, 24 μL loaded into the second 96 wells, 28 μL loaded into the third 96 wells, and 32 μL loaded into the fourth 96 wells and placed in the loading position in the ABI 3700 fluorescent DNA sequencers without covering. If the plates are to be foil covered and use the ABI piercing mechanism, then add only 20 μL of sterile distilled water to each well of the 384-well plates to dissolve the washed DNA-sequencing reactions.

Acknowledgments

I gratefully acknowledge the outstanding undergraduate, graduate, and post-doctoral students who have worked in my laboratory over the past 35 yr for their contributions to these protocols and the science that resulted from their successful implementation. This work was supported, in part, by US Public Health Services grant HG02153.

4. Notes

1. If the DNA is to be stored for more than a few days, it should be frozen, and then when needed, it should be thawed and redissolved by incubating overnight at 4°C . If the BAC is to be end sequenced, then a portion of the BAC-TE solution should be reprecipitated with ethanol and dissolved in ddH₂O instead of TE because the EDTA present in the TE inhibits *Taq* polymerase.
2. The DNA concentration typically can be estimated by agarose gel electrophoresis in parallel with size and known amounts of pUC or other small plasmids as stan-

dards and/or by measuring the A_{260} in the spectrometer. Typical BAC yields per 200 mL of original cell growth are approx 150 μ g.

3. Nebulizers, no. 4101 or 4101UO, can be purchased from IPI Medical Products, 3217 North Kilpatrick, Chicago, IL 60641; phone: (773) 777-0900. The hole where the mouthpiece is normally attached should be covered with a cap QS-T from Isolab (Drawer 4350, Akron, OH 44303; 100 caps for \$9.50). The nebulizer should be attached to a nitrogen source using Nalgene tubing (VI grade 3/16 in. id) because it makes a better seal than the tubing that comes with the nebulizer.
4. Nebulizing DNA at lower temperatures (*[21]*; S. J. Surzycki, personal communication) insures the generation of evenly distributed DNA fragments. During the nebulization process, unavoidable leaks are minimized by securely tightening the lid to the nebulizer chamber and sealing the larger hole in the top piece with a plastic cap.
5. To prepare for fragment end repair, the nebulized DNA typically is divided into four tubes and concentrated by ethanol precipitation.
6. A series of three messages will appear in a dialog box titled "Clear Air Gap." Three buttons also will appear in the dialog box each time: Yes, No, and Abort Protocol.
7. The cloning vector typically is *Sma*I-linearized, CIAP-dephosphorylated pUC vector (27-4860-01; Pharmacia); several years ago we switched from M13 to pUC-based shotgun cloning. The advantage of obtaining two sequence reads off one isolated shotgun subclone seems to outweigh the disadvantage of a few bases less in double-stranded vs single-stranded read lengths. In some instances, including 5% polyethylene glycol in the ligation reactions also seems to slightly improve ligation efficiency.
8. **Steps 5–9** should be performed in the cold room and typically approx 600 mL of ice-cold sterile water and 150 mL of ice-cold sterile 10% glycerol are required for manipulating the cells from a 1-L growth.
9. The purpose of these centrifugation/resuspension/centrifugation steps is to ensure that the cells are essentially "salt free" since salt causes arching during the electroporation step.
10. An alternate procedure uses "lasagna dishes" for pUC-based transformations. After transformation of competent *E. coli* host cells, follow these steps:
 - a. Add 1 mL of fresh 2X TY (or YENB) to each sample, and recover the transformed XL1-Blue (MRF') cells by incubating at 37°C for 15–30 min.
 - b. Centrifuge at 2000g in the Beckman CS-6R tabletop microtiter plate centrifuge for 5 min and decant the supernatant.
 - c. Resuspend the cells in 1.0 mL of YENB or 2X TY.
 - d. Add 130 μ L of 20 mg/mL IPTG (in water) and 130 μ L of 24 mg/mL X-Gal (in DMF).
 - e. Spread one-half of the cell suspension over each of two prewarmed "lasagna" dishes using a sterile inoculating loop; the "lasagna" dishes are #240853 Nunc Bio-Assay dishes (243- $\times$ 243- $\times$ 18 mm) (#25384-002; VWR) prepared by

- mixing 10 g of Bacto Tryptone (#0123-01-1; Difco), 5 g of Bacto yeast extract (#0127-05-03; Difco), 10 g of NaCl, and 18 g of Bacto agar (#0140-01; Difco) and brought to 1 L with ddH₂O. Sterilize this mixture and cool to 55°C. Then add 10 mL of ampicillin (10 mg/mL in sterile ddH₂O (#A-9518; Sigma), and pour 310 mL of this LB + Amp medium onto each “lasagna” plate in the sterile hood. Allow the plates to cool to room temperature before storing in a cold room.
- f. Allow 10–20 min for the cell suspension to diffuse into the agar, and then invert and incubate for at least 20 h at 37°C. Finally incubate the plates in a cold room (4°C) for an additional 3 to 4 h to intensify the blue color.
11. Typically, 3 to 4 μ L (200 ng) of the resulting 20 μ L of isolated DNA-sequencing template solution are used for sequencing with a 1:16 dilution of BigDye v3.0 mix and incubated following the cycle-sequencing protocol as recommended by the manufacturer (Applied Biosystems).

References

1. Anderson, S. (1981) Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res.* **9**, 3015–3027.
2. Messing, J., Crea, R., and Seeburg, P. H. (1981) A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9**, 309–321.
3. Deininger, P. L. (1983) Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal. Biochem.* **129**, 216–223.
4. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
5. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**, 161–178.
6. Bankier, A. T., Weston, K. M., and Barrell, B. G. (1987) Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods Enzymol.* **155**, 51–93.
7. Bankier, A. T. and Barrell, B. G. (1989) Sequencing single-stranded DNA using the chain-termination method in *Nucleic Acids Sequencing: A Practical Approach* (Howe, C. J. and Ward, E. S., eds.), IRL, Oxford, UK, pp. 37–78.
8. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., and Hood, L. E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679.
9. Ansorge, W., Sproat, B., Stegemann, J., Schwager, C., and Zenke, M. (1987). Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.* **15**, 4593–4602.
10. Brumbaugh, J. A., Middendorf, L. R., Grone, D. L., and Ruth, J. L. (1988) Continuous, on-line DNA sequencing using oligodeoxynucleotide primers with multiple fluorophores. *Proc. Natl. Acad. Sci. USA* **85**, 5610–5614.
11. Venter, C. J., Adams, M. D., Myers, E. W., et al. (2001) The sequence of the human genome. *Science* **291**, 1304–1351.

12. Yu, J., Hu, S., Wang, J., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* **296**, 79–92.
13. Goff, S. A., Ricke, D., Lan, T. H., et. al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* **296**, 92–100.
14. Mardis, E. R. and Roe, B. A. (1989) Automated methods for single-stranded DNA isolation and dideoxynucleotide DNA sequencing reactions on a robotic workstation. *Biotechniques* **7**, 840–850.
15. Bodenteich, A., Chissoe, S., Wang, Y. F., and Roe, B. A. (1993) Shotgun cloning as the strategy of choice to generate templates for high-throughput dideoxynucleotide sequencing, in *Automated DNA Sequencing and Analysis Techniques* (Venter, J. C., ed.), Academic, London, pp. 42–50.
16. Ewing, B., Hillier, L., Wendl, M., and Green, P. (1998) Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185.
17. Ewing, B. and Green, P. (1998) Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.
18. Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202.
19. Birnboim, H. C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**(6), 1513–1523.
20. Oefner, P. J., Hunnicke-Smith, S. P. Chiang, L., Dietrich, F., Mulligan, J., and Davis, R. W. (1996) Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic Acids Res.* **24**, 3879–3886.
21. Pan, H., Chissoe, S. L., Bodenteich, A., Wang, Z., Iyer, K., Clifton, S. W., Crabtree, J. S., and Roe, B. A. (1994) The complete nucleotide sequences of the SacBII Kan domain of the P1 and pAD10-SacBII cloning vector and three cosmid vectors: pTCF, svPHEP, and LAWRIST16. *GATA* **11**(5–6), 181–186.
22. Bankier, A. T., Weston, K. M., and Barrell, B. G. (1987) Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods Enzymol.* **155**, 51–93.
23. Mandel, M. and Higa, A. (1970) Calcium dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**, 154–159.
24. Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127–6145.
25. Sharma, R. C. and Schimke, R. T. (1996) Preparation of electro-competent *E. coli* using salt-free growth medium. *Biotechniques* **20**, 42–44.
26. Cohen, S. N., Chang, A. C. Y., and Hsu, L. (1992) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc. Natl. Acad. Sci. USA* **69**, 2110–2114.

Rolling Circle Amplification for Sequencing Templates

**Paul F. Predki, Chris Elkin, Hitesh Kapur, Jamie Jett,
Susan Lucas, Tijana Glavina, and Trevor Hawkins**

1. Introduction

Robust and reproducible isolation of high-quality templates is a requirement for successful DNA sequencing. To date, approaches for template generation have been limited to purification of biologically propagated M13 or plasmid-based templates, or in vitro amplification of such templates by polymerase chain reaction (PCR). In this chapter, we describe a protocol for a new approach to template generation: rolling circle amplification (RCA). We have found that templates produced through RCA yield more consistent and higher-quality sequence than identical templates generated from plasmid-prep methods. The protocol is simple, amenable to high throughput, and currently in use at the DOE Joint Genome Institute (Walnut Creek, CA) for the daily production of 30,000 sequencing templates.

RCA has long been known as the mechanism by which some viruses replicate their circular genomes (*1–5*). In the laboratory, RCA has been exploited for a variety of applications (*6–9*). Recently, the use of random hexamers to non-specifically prime the highly processive in vitro replication of circular templates by Phi29 polymerase has been reported (*10*). Because random hexamers are employed as primers, the replication products themselves become templates for further replication, resulting in an exponential amplification of the source DNA. Sufficient quantities of DNA are produced so that sequencing of the unpurified product is possible.

The protocol described in this chapter is based on the use of a recently introduced kit from Amersham Pharmacia Biotech (TempliPhi; Piscataway, NJ). It

*From: Methods in Molecular Biology, vol. 255: Bacterial Artificial Chromosomes, Volume 1:
Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

describes a 384-well protocol, optimized for the amplification of high-copy-number plasmids (such as pUC-18) for capillary sequencing (*see* **Note 1**).

2. Materials

2.1. Culture Growth

1. 384-Well plates: 120 μ L, sterilized, clear polystyrene cell culture plates (Nalge Nunc, Naperville, IL).
2. Sterile Luria Bertani (LB) medium with 8% glycerol.
3. Plate seals (optional): Presto plate sealer system (Zymark, Hopkinton, MA).

2.2. Amplification

1. TempliPhi premix: TempliPhi DNA Sequencing Template Amplification Kit (product no. 25-6400-01) (Amersham Pharmacia Biotech).
2. Denaturation solution: supplied with the TempliPhi DNA Sequencing Template Amplification Kit (product no. 25-6400-01) (Amersham Pharmacia Biotech).
3. Purified water: RiOs reverse osmosis water purification system (Millipore, Bedford MA).
4. 2% Bleach solution: Mix 20 mL of concentrated bleach (PureBright, 5.6% sodium hypochloride; KIK Santa Fe Springs, CA) and 980 mL of purified water.
5. Hydra 384-well microdispenser (Robbins Scientific, Sunnyvale CA).
6. Water troughs (Robbins Scientific).
7. 384-Well Multidrop multidispenser (Titertek, Huntsville, AL).
8. ABgene ALPS-1000 heat-sealing system plate seals (ABgene, Rochester NY).
9. "A seals" plate seals (MJ Research, Waltham, MA).

2.3. Sequencing Reaction

1. ET Terminator Sequencing Mix (volumes for one reaction):
 - a. 2 μ L of DYEnamic™ (ET terminator sequencing kit; Amersham Pharmacia Biotech).
 - b. 1 μ L of primer at 2 pmol/ μ L (Integrated DNA Technologies, Coralville, IA).
 - c. 1.5 μ L of purified water: RiOs reverse osmosis water purification system (Millipore).
2. Big Dye Terminator Sequencing Mix (volumes for one reaction):
 - a. 1 μ L of Big Dye Terminator Sequencing Kit premix.
 - b. 0.5 μ L of Big Dye Terminator Sequencing Kit 10X buffer.
 - c. 1 μ L of primer at 2 pmol/ μ L (Integrated DNA Technologies).
 - d. 1 μ L of purified water: RiOs reverse osmosis water purification system (Millipore).
3. Hydra 384-well microdispenser (Robbins Scientific).
4. 384-Well PCR plates (Axygen, Union City, CA).
5. ABgene ALPS-1000 heat-sealing system plate seals (ABgene).
6. Biomek "Seal & Sample" aluminum foil lids, for plate seals (Beckman, Fullerton CA).

2.4. Sequencing Cleanup

1. 384-Well PCR plates (Axygen).
2. Washed bead reagent (store at room temperature and make fresh weekly):
 - a. Add 15 mL of thoroughly mixed Seradyn MG-CM carboxymethylated magnetic microbeads (50 mg/mL) (Seradyn, Indianapolis, IL) to a 50-mL plastic centrifuge tube (Corning, Acton, MA).
 - b. Fill tube to 50 mL with purified water.
 - c. Place on a magnet (MagneSil magnetic separation Unit; Promega, Madison, WI) long enough to allow the beads to migrate to the wall of the tube under the influence of the magnet (approx 15 min).
 - d. With the tube on the magnet, decant water.
 - e. Repeat **steps b–d** two times.
 - f. Add 15 mL of purified water, remove from the magnet, and invert until mixed.
3. BET solution (recipe for twenty 384-well plates; prepare fresh daily):
 - a. Add 64.0 mL of 200-proof ethanol (AAPER Alcohol & Chemical, Shelbyville, KY), 7.0 mL of purified water, and 6.4 mL of tetraethylene glycol (Aldrich, Milwaukee, WI) to a 100-mL flask. Cover with Parafilm® and mix by inverting.
 - b. Add 2.0 mL of washed bead reagent. Cover with Parafilm and mix by inverting.
4. Wash solution (prepare fresh daily; cover when not in use): 70% Ethanol made from 200-proof ethanol (AAPER Alcohol & Chemical).
5. Hydra 384-well microdispenser (Robbins Scientific).
6. 384-Well plate magnet (Prolinx, Bothell, WA).
7. 96-Well round-bottomed polystyrene plates from MegaBACE.
8. 384-Well Multidrop multidispenser (Titertek).
9. Clear seals.

3. Methods

3.1. Culture Growth

1. Pick colonies into a 384-well cell culture plate with 70 μ L of medium (LB + 8% glycerol + 50 μ g/mL of ampicillin)/well (*see Note 2*).
2. Grow overnight at 37°C, stationary, with a lid on the plate (*see Note 3*).

3.2. Amplification

1. Aliquot 2 μ L of overnight growth into each corresponding well of a 384-well PCR plate using a 384-well Robbin's Hydra. Between uses, decontaminate with a 2% bleach wash (three cycles of 20 μ L), followed by three rinsing steps (three cycles of 20 μ L each step) with purified water using three different troughs of water (*see Note 4*).
2. Add 8 μ L of denaturation solution to each well using a 384-well Multidrop. Spin down briefly if necessary (600 rpm, 1 min) (*see Note 5*).
3. Seal the plates using A seals (*see Note 6*).
4. Lyse the cells and denature the DNA by heating in a thermal cycler to 95°C for 5 min followed by 4°C.

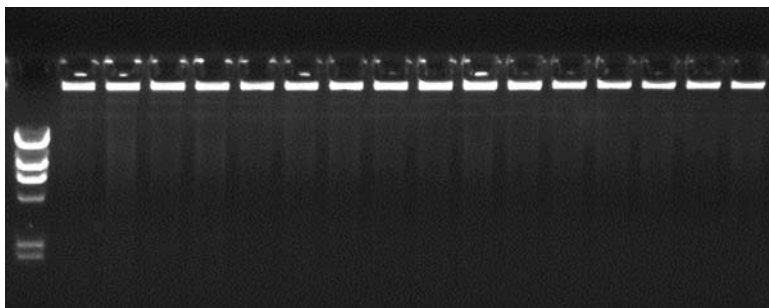


Fig. 1. Agarose gel image of RCA-amplified product produced as described in text. The left lane is the λ -HindIII digest control. Note that the RCA-amplified DNA barely enters the gel, remaining trapped near the well.

5. Thaw RCA mix on ice. Add 10 μ L of RCA mix to each well using a 384-well Multidrop (see **Note 7**).
6. Seal the plates. We use an ABgene heat sealer system, although a variety of other options exist. Spin down briefly if necessary (600 rpm, 1 min).
7. Heat to 30°C for 18 h (overnight) in an oven or a thermal cycler. Although the reaction can be complete in as few as 6 h, we get more consistent results letting the reaction proceed for 18 h (see **Note 8**).
8. Denature enzyme at 95°C for 10 min in a thermal cycler, and then hold at 4°C until the subsequent step (see **Note 9**).
9. Optional quality control check: To test for amplification, run 5 μ L of product on a 1% agarose gel (see **Fig. 1**) (see **Note 10**).

3.3. Sequencing Reaction

3.3.1. Energy Transfer Terminator Sequencing

1. Transfer 0.5 μ L of amplified product to corresponding wells in a 384-well PCR plate using a 384-well Hydra. Add 4 μ L of sequencing mix with a 384-well Multidrop. Spin down briefly if necessary (600 rpm, 10 s) (see **Note 11**).
2. Again, we use an ABgene heat sealer system, although there are other options (see **Note 12**). Run the plate on a thermal cycler with the following cycle conditions.
 - a. 95°C for 30 s.
 - b. 95°C for 25 s.
 - c. 50°C for 10 s.
 - d. 60°C for 2 min.
 - e. Repeat **steps b–d** 30 times.
 - f. 4°C indefinitely.

3.3.2. Big Dye Terminator Sequencing

1. Transfer 1.5 μ L of amplified product to corresponding wells in a 384-well PCR plate using a 384-well Hydra. Add 3.5 μ L of sequencing mix/well with a 384-well Multidrop. Spin down briefly if necessary (600 rpm, 10 s) (see **Note 11**).

2. Again, we use an ABgene heat sealer system, but a variety of other options are available, such as A seals (*see Note 12*). Run the plate on a thermal cycler with the following cycle conditions:
 - a. 95°C for 3 min.
 - b. 95°C for 35 s.
 - c. 50°C for 35 s.
 - d. 60°C for 4 min.
 - e. Repeat **steps b–d** 30 times.
 - f. 4°C indefinitely.

3.4. Sequencing Cleanup (*see Note 13*)

1. Add 10 μ L of BET solution to each well using a 384-well Multidrop. Incubate at room temperature for 15 min (*see Note 14*).
2. Place on a 384-well magnet until the beads are pulled to the bottom of the wells (approx 1 min).
3. Remove liquid using the Hydra. Bring the needles down as far as possible without touching the beads and slowly aspirate (*see Note 15*).
4. Add 15 μ L of 70% EtOH to each well using a 384-well Multidrop.
5. Remove liquid using the Hydra. Bring the needles down as far as possible without touching the beads and slowly aspirate. Allow to sit at room temperature until the beads are dry (approx 15 min) (*see Note 16*).
6. Add 15 μ L of purified water. Seal with clear seals. Vortex until the beads are resuspended. Let sit for 10 min at room temperature. Centrifuge briefly (600 rpm, 1 min) if necessary to remove air bubbles. Remove the seal.
7. Place on a 384-well magnet until the beads are pulled to the bottom of the wells (approx 1 min).
8. Transfer 10 μ L of liquid off of the beads and into a new plate. To transfer into a 384-well plate, use a 384-well Hydra. To transfer into four 96-well plates, use a 96-well Hydra equipped with a plate positioner or indexer (*see Note 17*).

3.5. Electrophoresis

1. For the MegaBACE 1000 capillary sequencer, use the following run conditions:
 - a. Run temperature: 45°C.
 - b. Injection voltage: 1.6 kV.
 - c. Injection time: 36 s.
 - d. Run voltage: 7 kV.
 - e. Run time: 160 min.
2. For the ABI Prism 3700 capillary sequencer, use the following run conditions:
 - a. Run temperature: 50°C.
 - b. Cuvet temperature: 45°C.
 - c. Injection voltage: 1000 V.
 - d. Injection time: 60 s.
 - e. Run voltage: 6500 V.
 - f. Sheath flow volume: 12,000.

- g. Sheath flow period: 942.
- h. Run time: 7000 s.

4. Notes

1. This protocol is written specifically for pUC-18 libraries. It works equally effectively with other high-copy-number vectors, but the antibiotic requirements may differ. Although there is some amplification of genomic DNA, the relative amount of genomic amplification is minimal with high-copy-number plasmids. More significant levels of genomic DNA are observed when amplifying from lower-copy-number vectors such as cosmids and BACs. In these cases, modification to the protocol is required.
2. LB medium is recommended. RCA is not as effective from cells grown in some other media, such as Terrific Broth.
3. Cells generally grow to a density of approx 1×10^8 CFU/mL under these conditions. They can be stored frozen at -80°C and thawed at 4°C prior to use if desired.
4. DuraFlex syringes are recommended, because they minimize or eliminate damage to syringes owing to plate-syringe misalignment.
5. There is sufficient mixing of cells with denaturation buffer during the buffer addition by Multidrop. If other means of adding buffer are used, it may be necessary to include a mixing step.
6. Use of a roller is recommended to ensure a good seal.
7. RCA premix should be stored in a -80°C freezer. Multiple freeze-thaws should be avoided. If small volumes are typically used, consider aliquoting and storing at -80°C . When thawing the premix, be sure to thaw on ice.
8. Subnanogram quantities of purified template will typically amplify in 2–4 h. However, amplification from cells is slower. Amplification is often complete at 6–8 h. However, we get more consistent results by letting the amplification reaction proceed 18 h. This helps ensure that almost all of the dNTPs are consumed during the reaction.
9. The amplification product can be used for sequencing without prior purification. The denaturation step is required to inactivate enzymatic activities that could interfere with the subsequent sequencing reaction. After this step, the RCA product can be stored at 4°C for several weeks.
10. Most of the amplified DNA will barely enter the gel, although it is possible to visualize a fraction of the DNA migrating in the high-molecular-weight range of the gel. Negative controls may show amplification of DNA, presumably the result of amplification of small amounts of contaminating DNA in the premix. It is possible to distinguish between amplification of template and this nonspecific amplification by digestion of the amplification product with a restriction enzyme known to cut within the plasmid vector (such as *EcoRI*) prior to running the agarose gel. In this case, discrete bands of predicted size will only be observed for template-amplified DNA.
11. The plate may need to be tapped to lightly touch the syringe needles in order to successfully transfer all the liquid. It is also possible to transfer template into

plates that already have the sequencing mix added (the ET mix is stable for >1 h at room temperature). Transfers of this nature (“wet transfers”) are easier to accomplish than “dry transfers” and more amenable to automation. However, it can be more difficult to detect poor transfer. Avoid multiple freeze thaws of sequencing mix. Prior to use of the plate, check that it is not significantly warped. For sequencing from pUC-18-based templates, we use the following primer sequences: (forward) gtttccagtcacgacgttgta and (reverse) aggaacagctatgacat.

12. It is important to have a good seal to avoid evaporation.
13. Other sequencing cleanup protocols, such as ethanol or isopropanol precipitation, can be used. However, we have consistently achieved longer read lengths with the BET protocol.
14. When adding BET solution with a Multidrop, the addition process should mix the components sufficiently. The brown beads should be uniformly distributed throughout the solution. The effective ethanol concentration range of BET solution is $\pm 5\%$. This makes it well suited to robotic platforms where the plates may sit uncovered for up to 1 h depending on the laboratory environment. When troubleshooting electrophoresis data, note that elevated ethanol or TEG concentrations can produce dye blobs. Low ethanol or TEG concentrations can cause blank lanes owing to failure of the labeled fragments to bind the beads.
15. Bead loss in 384-well plates can be problematic. Aspiration should be done at a low flow rate. Residual volumes of BET solution or ethanol left in the wells after aspiration are another potential source of problems. Complete removal is essential to a stable process.
16. Alternatively, it is possible to resuspend beads using the Hydra.
17. Make sure that the beads are not transferred along with the DNA.

Acknowledgments

We would like to thank the following individuals for their contributions to the initial development and large-scale testing of the protocols: Catherine J. Adam, Andre Arellano, Juanan Masako Boen, Christopher G. Daum, Chris Detter, Jennifer E. Grant, Nancy Hammon, Drew C. Ingram, Daisy P. Prado, Paul Richardson, Troy Smith, Kristina Tacey, and Marianne Vickers, all from the Joint Genome Institute & Lawrence Livermore National Laboratory; and John Nelson from Amersham Pharmacia. This work was performed under the auspices of the US Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48 and Lawrence Berkeley National Laboratory under contract no. DE-AC03-76SF00098.

References

1. Gilbert, W. and Dressler, D. (1968) DNA replication: the rolling circle model. *Cold Spring Harb. Symp. Quant. Biol.* **33**, 473–484.
2. Dressler, D. (1970) The rolling circle for phiX DNA replication. II. Synthesis of single-stranded circles. *Proc. Natl. Acad. Sci. USA* **67**, 1934–1942.

3. Schroder, C. H., Erben, E., and Kaerner, H. C. (1973) A rolling circle model of the in vivo replication of bacteriophagephiX174 replicative form DNA: different fate of two types of progeny replicative form. *J. Mol. Biol.* **79**, 599–613.
4. Doermann, A. H. (1973) T4 and the rolling circle model of replication. *Annu. Rev. Genet.* **7**, 325–341.
5. Kornberg, A. and Baker, T. A. (1992) *DNA Replication*. W. H. Freeman and Company, San Francisco.
6. Zhou, Y., Calciano, M., Hamann, S., Leamon, J. H., Strugnell, T., Christian, M. W., and Lizardi, P. M. (2001) In situ detection of messenger RNA using digoxigenin-labeled oligonucleotides and rolling circle amplification. *Exp. Mol. Pathol.* **70**, 281–288.
7. Zhong, X. B., Lizardi, P. M., Huang, X. H., Bray-Ward, P. L., and Ward, D. C. (2001) Visualization of oligonucleotide probes and point mutations in interphase nuclei and DNA fibers using rolling circle DNA amplification. *Proc. Natl. Acad. Sci. USA* **98**, 3940–3945.
8. Schweitzer, B., Wiltshire, S., Lambert, S., O'Malley, S., Kukanskis, K., Zhu, Z., Kingsmore, S. F., Lizardi, P. M., and Ward, D. C. (2000) Inaugural article: immunoassays with rolling circle DNA amplification: a versatile platform for ultra-sensitive antigen detection. *Proc. Natl. Acad. Sci. USA* **97**, 10,113–10,119.
9. Schweitzer, B. and Kingsmore, S. (2001) Combining nucleic acid amplification and detection. *Curr. Opin. Biotechnol.* **12**, 21–27.
10. Dean, F. B., Nelson, J. R., Giesler, T. L., and Lasken, R. S. (2001) Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11**, 1095–1099.

Transposon-Mediated Sequencing

Rachel Reeg and Anup Madan

1. Introduction

Transposon-mediated sequencing is an effective method for obtaining full-length high-quality DNA sequence. This method can be applied to the finishing stages of bacterial artificial chromosome (BAC) sequencing, allowing the user to expand a finishing repertoire of custom oligonucleotide sequencing, alternative chemistry sequencing, polymerase chain reaction (PCR), as well as other standard tools for acquiring high-quality data in low-coverage or low-quality regions. In addition to BAC subclones, transposon-mediated sequencing can be applied to cDNA or PCR products subcloned into vectors.

In genomes, transposons are sections of the chromosomes containing a gene for transposase and a segment of DNA that will be excised and transplanted to a new place in the genome in the absence of an RNA intermediate. Transposase enzymes interact with the transposon; a linear donor DNA molecule; and the adoptive, target DNA, in a variety of three-dimensional conformations for insertion of the transposon into the target DNA (1). Scientists have optimized these naturally occurring transposons from bacteria for inducing mutagenesis, producing gene knockouts, and randomly inserting primer binding sites and selectable markers into plasmids (2). Transposons are generally manipulated to have a selectable marker such as a gene for kanamycin or tetracycline resistance and flanking regions that include restriction enzymes, and sequencing primer sites for bidirectional sequencing (3). Transposase recognition sites are present at each end of the transposon, which are inverted repeats of each other. Transposon ends are gripped by the transposase while the target DNA that will adopt this fragment is cut, ligated to the transposon, and repaired to double strand-

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

edness. Insertion is either random, meaning that there is no sequence preference for insertion location, or it is not random, and there may be “hot spots” that preferentially facilitate transposon insertion to a particular region.

2. Materials

Listed are the materials needed for implementing the protocols discussed in this chapter. Because transposon-mediated sequencing involves many molecular biology techniques, one may find it useful to refer to other chapters in this book as well as a standard protocol book such as *Molecular Cloning: A Laboratory Manual* by Sambrook et al. (4).

1. Transposon kit including reagents for the transposon reaction and a manual (Epicentre Technologies or New England Biolabs are options).
2. Cloned DNA template(s) to have transposons inserted. This can be DNA cloned into any number of plasmid vectors and purified with an alkali lysis method, as described in **Subheading 3.3**.
3. Transformation reagents: competent cells, agar, medium, and antibiotics. Choose a transformation kit appropriate for the vector being used. Two highly efficient kits include XL-10 Gold competent cells (Stratagene), innately chloramphenicol resistant, and DH10B cells (Invitrogen). The DH10B cells can be used if the chloramphenicol resistance of XL-10 Gold cells interferes with a plasmid containing chloramphenicol resistance.
4. DNA miniprep kit or other protocol, solutions, and equipment. If not using a purchased kit, the following solutions can be used for the alkali lysis protocol:
 - a. Solution I: 50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA, pH 8.0.
 - b. Solution II: 0.2 N NaOH, 1% sodium dodecyl sulfate (SDS). Prepare fresh each time: make a stock of 10 N NaOH and 10% SDS; for 100 mL of solution use 88 mL of water, 10 mL of 10% SDS, and 2 mL of 10 N NaOH.
 - c. Solution III: 60 mL of 5 M potassium acetate, 11.5 mL of glacial acetic acid, 28.5 mL of H₂O (multiply volumes by 10 for 1 L). Autoclave.
 - d. Solution IV: 30% polyethylene glycol (PEG) (mol wt = 8000), 2.5 M NaCl. For 800 mL, dissolve 117 g of NaCl in 300 mL of ddH₂O; add 240 g of PEG (mol wt = 8000) while gently heating and stirring; pour into a 1-L graduated cylinder; bring to 800 mL with ddH₂O; and when the solution becomes clear, filter it.
5. Restriction enzymes, 1X TAE buffer, agarose, gel tray, and power supply. The enzyme used will depend on the vector and transposon being used, as discussed in **Subheading 3.4**.
6. Sequencing reagents including primers specific to the transposon (these may be included with the kit), Big Dye Terminator version 3, and isopropanol. See the transposon kit manual for the primer sequences, which will correlate to the 5'- and 3'-end sequences of the transposon. Depending on the sequencing equipment used, dye, water, or formamide will be needed for elution of the precipitated reaction.

7. Thermocycler, centrifuge, vortex, 95°C oven, and sequencing machine (ABI-377 or 3700, Amersham Biosciences).

3. Methods

Before proceeding with the following protocols, it is best to read and understand the notes section of this chapter (*see Subheading 4.*) in order to design the experiment that you will perform (*see Notes 1–10*). The Notes describe a process that involves mapping out the vector and transposon, as well as identifying restriction enzyme cutting sites (*see Notes 4–10*).

3.1. Transposon Reaction

The transposon reaction, in which the linear transposon is inserted into the target DNA, is a straightforward procedure. Review the literature accompanying the transposon kit you decide to use. Select a transposon that has a selectable marker different from the one already in the vector of the clones you are using. In addition, it is important to use equimolar amounts of the DNA templates if you are pooling them, in addition to using an equimolar amount of transposon to template DNA. If there is much more transposon DNA than template DNA, double insertions could occur where more than one transposon is inserted into a single template molecule, which will interfere with the sequencing reaction because of multiple priming sites. Although not necessarily recommended by the transposon kit manual, using less than an equimolar amount of transposon should be considered, since this is more cost-effective and will provide plenty of clones that do have transposons inserted.

1. Combine 0.2 µg of target DNA (one or more templates) (*see Note 11*), 1 µL of 10X buffer (provided in the kit), 0.2 µg transposon (provided in the kit), sterile water to 9 µL, and 1 µL of transposase (provided in the kit) for a final volume of 10 µL for the reaction.
2. Incubate at 37°C for 2 h.
3. If provided, use the stop solution to inactivate the transposon insertion step of the reaction.
4. Store the mix at –20°C to prevent DNA degradation.

3.2. Transformation

The transposon reaction is next transformed into an appropriate strain of competent cells (*see Notes 12 and 13*). Consult the transposon kit manual and competent cell manual for specific instructions on transformations. A general chemical transformation procedure is included in the following protocol. Electroporation, not discussed here, is another option for transformation.

1. Pour agar plates with the appropriate antibiotics for selection against clones without these antibiotic resistance markers. Autoclave 2X-YT or Luria Bertani broth with 1 to 2% agar. After cooling to 60°C, add the antibiotics to the proper concentration ($\mu\text{g/mL}$) for selection, and pour into Petri dishes. Let the plates dry and cool at room temperature overnight; store at 4°C.
2. Transform competent cells with the transposon reaction mix. It may take only a very small amount of the transposon reaction mix to efficiently transform cells because it is more highly concentrated with purified DNA than a typical ligation reaction. Incubate the competent cells and DNA on ice for 30 min. The competent cells may require incubation with β -mercaptoethanol prior to incubating them with the DNA.
3. Heat shock the cells at 42°C for 30–60 s to make the cell membrane permeable to small DNA molecules.
4. Incubate on ice for 1 to 2 min.
5. Add 1 mL of SOC or other medium without antibiotic. Incubate with shaking at 37°C for 30–60 min to promote expression of the antibiotic resistance genes (*see Note 14*).
6. Plate 50–350 μL of the transformation on the agar plates prepared in **step 1**. Incubate overnight at 37°C (*see Note 15*).

3.3. Preparation of DNA

If the transformation is successful and you have an estimate of the number of clones you will need to sequence for full coverage of the clone, then continue with this step of DNA purification. The same DNA preparation method as was used for BAC library subclones is also appropriate for preparation of transposon-inserted clones.

1. Inoculate 1 to 2 mL of medium (plus one of the antibiotics) with a single colony from the plates in **Subheading 3.2.**. Both the vector and transposon marker antibiotic need not be present in the medium because the clones with both of these resistances were already selected for on the plates (*see Note 13*).
2. Grow and shake overnight at 37°C.
3. Follow a procedure for preparing clones: an alkali lysis or other miniprep protocol (**4**). A protocol for preparation of alkali lysis plasmid is as follows:
 - a. Centrifuge the overnight culture at 500g for 5 min.
 - b. Dump the medium without dumping the pellet.
 - c. Add 4 μL of RNase I per reaction to 100 μL of solution I. Then add the solution to the pellet, cover, and resuspend the pellet by vortexing.
 - d. Add 100 μL of solution II. Tap the tube gently to lyse the bacterial cultures. Keep at room temperature for 5 min.
 - e. Add 100 μL of solution III. Cover the sample and mix by vortexing. Incubate on ice for 10 min.
 - f. Centrifuge the sample at 1300g for 30 min.
 - g. Transfer 250 μL of the supernatant to a new centrifuge tube.

- h. Add 125 μL of solution IV. Mix well by inverting the tube 20–30 times.
- i. Centrifuge at 980g for 10 min. Dump the supernatant, and blot upside down on paper towels to get rid of the PEG in solution IV.
- j. Add 100 μL of 70% ethanol. Centrifuge at 980g for 10 min. Dump the supernatant, and dry the plates at room temperature.
- k. Resuspend the DNA pellet with 40 μL of ddH_2O . Vortex briefly to mix and store at 4°C.

3.4. Screening Clones

As discussed, it may be cost-efficient to screen clones for a transposon insertion into the target DNA, and avoid sequencing clones that have a transposon insertion into the vector from which you will only get the vector sequence. Select restriction enzymes that will digest the clones near the 5' and 3' ends of the insert, cutting only once in the vector's multicloning sites so that the vector is in only one band. Use restriction enzymes from the same manufacturer that will efficiently work in the same buffer solution and at the same temperature. For example, *EcoRI* and *BamHI* may be compatible in buffer X, but using *EcoRI* from manufacturer B in the same reaction with buffer X from manufacturer A may not be effective, even if buffer X is a "universal" buffer. Digest 200–300 ng of DNA, or about 1 μL for many preparation methods. Use the following protocol for a 10- μL reaction.

1. Make a mixture of the following ingredients. For each item, multiply by the total number of reactions for the final amount or volume to put in the master mix:
 - a. 2–5 U of restriction enzyme A.
 - b. 2–5 U of restriction enzyme B.
 - c. 1 μL of 10X buffer.
 - d. Sterile water to 9 μL (if 0.5 μL of enzyme A and 0.5 μL of enzyme B were used, then this would be 7 μL of water per reaction).
2. Add 1 μL of DNA to individual tubes.
3. Add 9 μL of the master mix from **step 1** to the DNA.
4. Digest the vector with the same enzymes to use as a control.
5. Incubate at the appropriate temperature for the time specified in the enzyme packaging. Generally, 1 h at 37°C is sufficient since 1 U of enzyme is defined to cut 1 μg of DNA in 1 h at the appropriate temperature.
6. While the reactions are incubating, make a 1–1.5% agarose gel for electrophoresis of the digested samples. For a 1% gel, add 1 g of agarose for each 100 mL of buffer. Microwave on high for 2–5 min, until the agarose is dissolved. Carefully add ethidium bromide to a final concentration of 0.5 $\mu\text{g}/\text{mL}$ for visualization of the bands under ultraviolet (UV) light.
7. After incubating the samples long enough for full enzyme digestion, add a gel-loading buffer. Depending on the number of samples, use either a multipipet compatible with the distances between the gel slots, or a single pipet, to dispense the

samples into the gel that is solidified and immersed in running buffer. Include a standard 100-bp or 1-kb ladder in each row of samples for size estimation. For this specific procedure, also include a control band—the vector digested with the same restriction enzymes.

8. Electrophorese at a constant voltage for an appropriate amount of time for the bands to separate sufficiently and to prevent the gel from melting.
9. View and photograph the gel with UV light. Use proper eye protection.
10. Analyze the bands. If the transposon did not insert into the vector, a band will be present at the same length as the vector control band. If the transposon did insert into the vector, then this band will no longer exist (*see* **Notes 16–18**).

3.5. Sequencing

Sequence the plasmid DNA + transposon using the transposon kit sequencing primers or nonrepetitive custom oligonucleotides corresponding to the 5'- and 3'-end sequences of the transposon. Recall that the very ends of the transposon are inverted repeats of each other and are mutated during insertion, avoid selecting sequences for oligonucleotides from these regions. The most useful oligos are likely those included with the kit. Two sequencing reactions, in opposite orientations, can be produced from each template. Initially, you cannot identify the orientation the transposon inserts into the clone. However, if you wish to sequence in only one direction to begin with and go back and sequence in the opposite direction only when needed, it is important to specifically name clones for retrieval and resequencing. Since the transposon is linear and is randomly inserted, there is no preference for it to insert in a particular orientation relative to the initial clone. Investigate the manufacturer's literature accompanying the transposon kit for the mechanism of transposon insertion. Some transposases will have an excision and end-filling mechanism in which the same several bases are present at either end of the transposon, and when sequence data are assembled for reads sequenced from either end of the transposon, these several bases will overlap the two reads extending in opposite directions.

3.5.1. Big Dye Terminator Sequencing Protocol

1. Add the following reagents to a microtube for thermocycling (*see* **Notes 19 and 20**): 2 μL of purified DNA from above protocol (approx 300 ng total), 4 μL of BigDye Terminator version 2 reaction mix, 3 μL of ddH_2O , and 1 μL of 3.2 pmol/ μL primer.
2. Cycle with the following program
 - a. 96°C for 1 min.
 - b. 96°C for 10 s.
 - c. 50°C for 5 s.
 - d. 60°C for 2 min.
 - e. Repeat **steps b–d** 49 times.
 - f. 4°C hold.

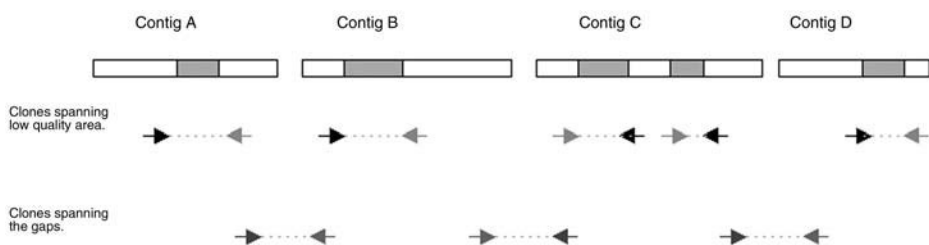


Fig. 1. Diagram of contigs A–D generated by assembly of sequencing reads of a BAC shotgun library made in plasmids. Low-quality regions are shaded, high quality appear white. End sequences of the same clone in adjacent contigs confirms the contigs' orientations in relation to each other and also confirm that the gaps can be sequenced from internal regions of the clones spanning these gaps.

3. Remove the sample from the centrifuge and briefly centrifuge to ensure that the contents are at the bottom of the tube.
4. Add 40 μ L of 75% isopropanol. Cover and vortex for 30–45 s. Keep at room temperature for 15 min.
5. Centrifuge at 1186g, 4°C for 30 min. Dump the supernatant.
6. Dry in a 95°C oven. This will take 2–10 min depending on how efficiently the supernatant was dumped.

4. Notes

1. It is important to prepare for transposon-mediated sequencing by reviewing the overall strategy you will use and by making calculations regarding the DNA to be sequenced. If applying this method to BAC finishing, the first step is to design a scaffold of the BAC and identify the clones that overlap any gaps between contigs (5,6). Scaffolding requires that data be produced from a clone supporting bidirectional sequencing, such as a shotgun plasmid library. Since both ends of the plasmids are sequenced, the data can be ordered by identifying end reads of plasmids in different contiguous sequences (contigs). If the 5'-end read of plasmid A is in contig 1 and the 3'-end read is in contig 5, then the contigs should be arranged so that the missing data from this plasmid, the gap between contigs, is spanned by this clone. It is good to confirm this orientation with end reads from other plasmids in the assembly. The third line in **Fig. 1** demonstrates this principle; the dashed lines represent data not yet sequenced but retrievable from the spanning clones. In addition to filling gaps, you may select clones overlapping low-quality regions of a contig, especially regions of extended low quality, e.g., 700–2000 bp or more, that would require several custom primer sequencing reactions for high-quality data (**Fig. 1**). Maintain a list of the clones selected for the transposon reactions. Using this method for BAC finishing requires access to the templates initially sequenced from the BAC library.

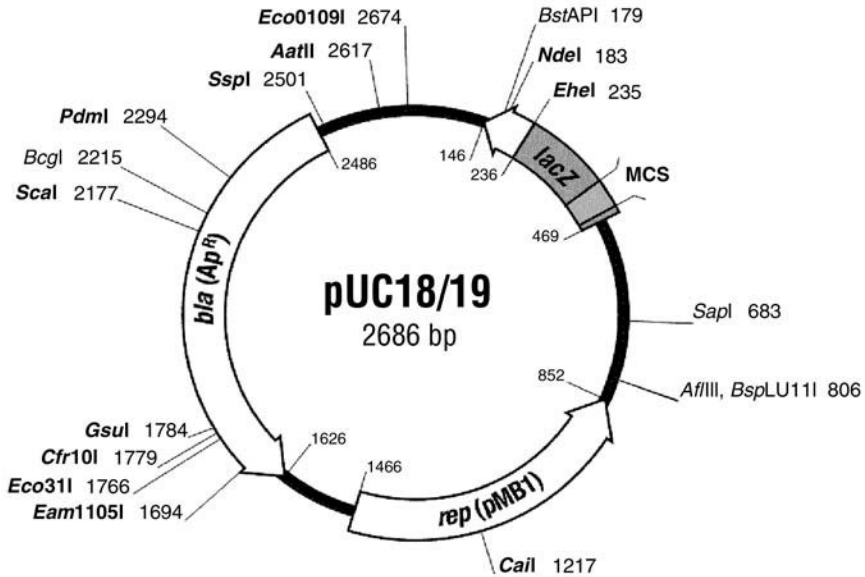


Fig. 2. Diagram of pUC18/19.

- Identify the vectors, cloning sites, and selectable markers of the clones. Use restriction enzymes that cut only once in the multicloning site, 5' and 3' to the inserted DNA, to determine the insert size. For example, if the BAC library was made in pUC18 (Fig. 2), cloned into the *SmaI* site, it would be suitable to do a double digestion with *Bam*HI and *Eco*RI, producing a vector band and one or several insert bands, depending on the frequency of internal *Eco*RI and *Bam*HI sites of the subclone. Calculate the total size of the insert bands for each clone and record this information in a table as demonstrated in Table 1 so that they can be grouped by size for transposon reactions.
- Clones from the same vector can be pooled according to comparable insert size, or each clone can be in an individual transposon reaction so that all derivative clones will be from this particular template. Using Table 1, you could place the clones into two groups: Group 1: clones A, F, H, I, and J; Group 2: clones B, C, D, E, and G. If the clones are pooled, it will not be until assembling the data that the source of the derived clones can easily be identified. When the data are assembled, reads from a particular clone will fall into the region of the clone's end reads used in the scaffolding exercise. Likewise, if, for example, 10 nonoverlapping cDNA clones of similar size for a single transposon reaction are grouped, on assembly of this transposon library sequence data the cDNA end reads and corresponding transposon sequence data will assemble into 10 individual contigs. Of course, this requires enough data for coverage of the insert sequence.
- Diagram the vector. For example, pUC18 is 2686 bp, the Amp^R gene is 2486–1626 bp (860 bp of the vector), the origin of replication is bases 1466–852

Table 1
Name and Size of Clones
for Transposon-Mediated
Sequencing to Finish BACA1

Clone	Insert size (kb)
A	1.2
B	2.6
C	2.8
D	2.9
E	2.7
F	1.3
G	2.5
H	1.5
I	1.2
J	1.4

(614 bp of the vector), and the lacZ operon and multicloning site are contained in bases 469–146. Note that plasmids with transposon insertion into the origin of replication or Amp^R region of the vector will not successfully transform *Escherichia coli* plated on ampicillin + transposon-selectable marker (i.e., ampicillin + kanamycin) agar plates. As a result, there is a theoretical and practical reduction of the vector-to-insert ratio, improving the likelihood of transposon insertion into the target DNA where priming sites are desired, rather than in the vector.

- Calculate the vector-to-insert ratio, or what the preferred location of transposon insertion will be. This is done assuming that there are no “hot spots” for insertion and insertion is a random process. For example, if clone G from **Table 1** is in the 2686 bp pUC18 vector, the working ratio can be determined as follows. Subtract from the total size of the vector the size of the origin of replication (replicon), and the ampicillin marker size

$$\text{pUC} - \text{replicon} - \text{marker} = 2686 \text{ bp} - 614 \text{ bp} - 860 \text{ bp} = 1212 \text{ bp}$$

for the remaining bases that can have a transposon insertion yet maintain the plasmid’s ability to transform *E. coli* for plating on doubly selective medium. Now the vector-to-insert ratio has been reduced from 2686:2500 (or 1.07:1) to 1212:2500 or (0.48:1), a ratio no longer favoring transposon insertion into the vector.

- Calculate the percentage of transformants with transposons in the vector rather than the desired insert DNA. Again, clone G has a total size of insert + vector = 2500 bp + 2686 bp = 5186 bp. Assuming random transposon insertion, the following percentages apply: 48% (2500/5186) of the insertion events will occur in the insert, 23% (1212/5186) will occur in regions of the vector allowing the clone to grow on the selective plates, and 28% will occur in the replicon or marker gene

that will not successfully transform a clone for selection on the plates. Since 23% of the transposon insertions will occur in the vector and 48% will occur in the insert, 32% ($23/[23 + 48]$) of the transformed clones will have transposons in the vector rather than the insert sequence.

7. Determine whether you will be pooling the clones or doing one transposon reaction per clone. Transposon insertion is analogous to making a shotgun library of a BAC—even more so when the clones are pooled—that is a mix of several clones of the same size and concentration used as the templates for one transposon reaction. When the initial BAC library is made, fragments of the large DNA molecule are ligated in a vector from which end sequences of the fragments can be obtained (7). Random insertion of transposons into the subclones provides new sequencing priming sites for sequencing coverage from the end and central areas of the clones. Pooling clones is a practical option when the sequence data from the resulting library are used in the same assembly. Until the clone is sequenced, it will not be apparent to the user what the clone source was if there were several templates in the transposon reaction. Pooled clones should be of similar size and concentration, or the concentrations may be adjusted according to their size, in effect their molar concentration in the pool. Overrepresentation of a clone will occur if it is of a higher concentration than clones of the same size or if it is a much larger clone than others in the template mix. If you decide to use this method, group the clones according to size, and mix them by adding 1 or 2 μL of the templates of the same concentration to a centrifuge tube. If it cannot be assumed that the templates are of the same concentration, then calculate the molar ratios with the aid of the following:
 - a. The spectrophotometric equivalent of double-stranded DNA (dsDNA) is as follows: 1 A_{260} unit is equivalent to 50 $\mu\text{g}/\text{mL}$; the molarity is 0.15 mM .
 - b. The average molecular weight of a base pair is 660.
 - c. For dsDNA, conversions between micrograms and picomoles are calculated using the following formulae:

$$\mu\text{g} \times (10^6 \text{ pg}/1 \mu\text{g}) \times (\text{pmol}/660 \text{ pg}) = \text{pmol}$$

$$\text{pmol} \times (\text{number of base pairs}) \times (660 \text{ pg}/1 \text{ pmol}) \times (1 \mu\text{g}/10^6 \text{ pg}) = \mu\text{g}$$

Use an appropriate volume of each template in the mix for unbiased insertion of the transposons into the templates.

8. Determine whether it is beneficial to screen the clones that you will be sequencing. Some clones will have a transposon insertion into the vector from which only vector sequence will be obtained from a sequencing reaction. To avoid these clones, screening the clones can be done with either of two simple methods. The first method is to do a double digestion of the clone with restriction enzymes that cut only once 5' and once 3' to the insert in the multicloning site of the vector. This separates the vector from its insert, producing a vector band and one or several bands from the insert, depending on how many restriction enzyme sites are in the insert. Separate the bands on a 1–1.5% agarose gel; take a picture for record keep-

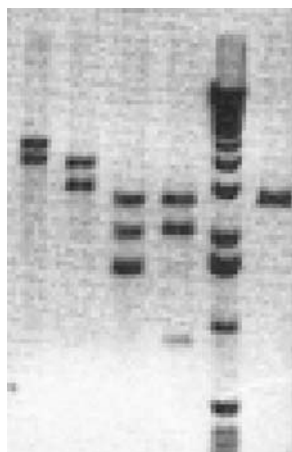


Fig. 3. Double digest of transposon-inserted clones pictured under UV light.

ing. From the gel picture it will be apparent if the vector band is present in its original size (pUC18 is 2.7 kb); or if a transposon successfully inserted for a band the size of the vector + transposon; or if the transposon had one of the restriction enzyme sites, the digestion of the clone broke the vector into smaller than normal size bands. It is important to note that inconclusive results will occur if the insert is the same size as the vector, or if the insert size + transposon size = vector size, and the restriction enzymes do not cut the transposon. These bands will appear as false positives, the same size as the vector band in the gel digest picture.

There are six lanes in the gel picture shown in **Fig. 3**. The first four lanes are BAC subclones in pUC18 that have undergone a transposon reaction and then were digested with *Bam*HI and *Eco*RI for excision of the vector band. Lane 5 is a 1-kb ladder. Lane 6 is pUC18 digested with *Bam*HI and *Eco*RI. Using the band in lane 6 as a control, it is easy to see that lanes 3 and 4 have a band the same size as the vector band; transposon insertion occurred in the subcloned BAC DNA in the multicloning site. Lanes 1 and 2 do not have a vector band at approx 2.7 kb, which implies that the transposon inserted into the vector, increasing the size of this band. Sequencing the clones digested for lanes 3 and 4 will provide sequence data for the DNA insert, rather than vector.

The second option for screening clones is PCR. Using a primer from the 5' and 3' ends of the vector going *into* the vector from the multicloning site, a PCR reaction can be performed to replicate the vector band. If the vector has not had a transposon insertion the band size will not be altered; if it has had a transposon inserted, the PCR band will be the size of the vector + transposon. As in the restriction enzyme digestion strategy, separate the bands on an agarose gel and take a picture for your records.

9. Calculate the coverage and total number of templates that will be needed for sequencing. For full-coverage sequencing of a region, it is appropriate to sequence to 5X coverage or more. This means that on average each base is sequenced five times. To calculate coverage, determine the total bases to be sequenced, i.e., the insert, not vector, DNA base pairs. Divide this by the total number of bases sequenced in the sequencing reactions, i.e., the number of sequencing reactions multiplied by read length (generally 500–700 bp of good-quality data). Therefore, coverage is a term with no units because it is just a ratio of bases to be sequenced to bases that are sequenced. Likewise, the number of templates needed to sequence can be determined by multiplying the bases to be sequenced (total size of inserts for subclone templates) by the amount of coverage desired (5–10X generally) and dividing this value by the number of bases per sequencing reaction. Then, this value is multiplied by the number of templates per sequencing reaction. It is more cost-effective if you plan on doing two sequencing reactions per template, both a forward and reverse read, rather than prepping twice the templates for one reaction per template (this term will be $\frac{1}{2}$ in the following formula; that is, one template serves for two sequencing reactions). Here are the two formulae:

$$\text{coverage} = \frac{\text{bases to be sequenced}}{(\text{no. of sequencing reactions} \times \text{average length of sequencing reaction})}$$

$$\text{templates} = [(\text{bases} \times \text{coverage}) \div (\text{bases/sequencing reaction})] \times (\text{no. of templates per sequencing reaction})$$

If you have determined the percentage of insertions into the vector, then you can include this source of “error” into the calculation of the number of templates to be prepared. If 32 of every 100 clones will be insertions into the vector, then you can compensate for this by preparing extra templates. Divide the number of clones needed by the percentage of clones with a transposon insertion into the nonvector DNA. If you need 125 templates for the right coverage, divide 125 by 68% ($125 \div 0.68$) for a total number of 184 clones to be prepared. (To check the validity of this formula, multiply 184 by the failure rate of 32%: $184 \times 0.32 = 59$ negative clones; $184 - 59 = 125$ “good” clones.) You may sequence all 184 clones, or screen them prior to sequencing in order to identify the approx 125 that have nonvector transposon insertion. Note that the percentage of insertion into the vector will decrease with an increase in the insert DNA-to-vector DNA ratio and also upon the increase in size of the elements in which transposon insertion into the vector makes the clone unable to successfully transform *E. coli* on selective medium, i.e., if there are few places for the transposon to insert into the vector without interrupting a drug resistance gene or origin of replication.

10. Revise the vector-screening portion of your assembly program to include the transposon sequence in order to prevent the genomic BAC sequence from being contaminated with transposon sequence. This sequence can be found in the transposon kit literature. Remember, the forward and reverse pair reads from the same clone may overlap by several base pairs; this is advantageous in some cases and

may allow you to join two contigs with this small overlap and continue finishing the sequence to high quality using custom oligonucleotides or sequencing of more transposon insertion templates. If you have not sequenced enough clones to give full coverage of a region, then return to the DNA preparation step and continue on in this method until you feel that enough data have been produced.

11. Is the target DNA nondegraded? This can be identified as a smear on the agarose gel. Retransform and prep if necessary.
12. Are the competent cells an appropriate strain for selecting for the markers in the vector and transposon? Check the manual for the competent cells and make sure there is no overlap between the cell antibiotic resistance and the vector and transposon. Also ensure that the drugs are included in the agar plates at an appropriate concentration for selection.
13. Are the selectable markers in the vector and transposon different? Similar to the competent cells, you need different drug resistance genes in the plasmid vector and transposon. Refer to the transposon manual.
14. Is the appropriate concentration of antibiotic being used in the medium? If you seem to be getting a high number of false positive clones when screening—i.e., there is no transposon insertion—increase the antibiotic concentration.
15. Are incubator temperatures accurate and stable? Use a reliable, calibrated thermometer in the incubator to confirm a digital reading.
16. Do the restriction enzymes cut in the places you originally predicted? Are they compatible in the same buffer? Check the vector map for restriction enzyme sites and the restriction enzyme product literature for buffer requirements.
17. Are you running the gel long enough to sufficiently separate the bands and predict sizes? The ladder should be well extended so bands between 1 and 5 kb will separate if they only have a 100-bp difference (i.e., 3.3 and 3.4 kb).
18. Did you digest the vector with the same enzymes as the clones to use as a control in screening? This will help in that you will not have to guess the size of the excised vector band, but will have a vector standard that can be run near the standard ladders for quick identification of the vector band.
19. Are you using the correct oligonucleotides for PCR screening? Double-check the sequence on the primer tube and compare to the transposon product literature.
20. Are you using the appropriate oligonucleotides for sequencing? Double-check the sequence on the primer tube and compare to the transposon product literature.

References

1. Hartwell, L., Lewis, R., et al. (2000) *Genetics: From Genes to Genomes*. McGraw-Hill, Boston.
2. Goryshin, I. Y. and Reznikoff, W. S. (1998) Tn5 *in vitro* transposition. *J. Biol. Chem.* **273**, 7367–7374.
3. Jendrisak, J., Meis, R., et al. (1998) High efficiency *in vitro* transposition for DNA sequencing projects, in *Proceedings of the 10th International Genome Sequencing and Analysis Conference*. Miami Beach, FL.

4. 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.
5. Roach, J. C., Boysen, C., et al. (1995) Pairwise end sequencing: a unified approach to genomic mapping and sequencing. *Genomics* **26**, 345–353.
6. Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868–877.
7. Rowen, L., Lasky, S., and Hood, L. (1999) Deciphering genomes through automated large-scale sequencing. *Methods Microbiol.* **28**, 155–191.

Pyrosequencing

A Tool for DNA Sequencing Analysis

Elahe Elahi and Mostafa Ronaghi

1. Introduction

Pyrosequencing, a bioluminometric DNA sequencing technique based on sequencing by synthesis, is emerging as a widely applicable tool for detailed characterization of nucleic acids (1–3). This technique relies on the real-time detection of inorganic pyrophosphate (PPi) released on successful incorporation of nucleotides during DNA synthesis. PPi is immediately converted to adenosine triphosphate (ATP) by ATP sulfurylase, and the level of generated ATP is sensed by luciferase-producing photons. Unused ATP and deoxynucleotide are degraded by the nucleotide-degrading enzyme apyrase (**Fig. 1**). The presence or absence of PPi, and therefore the incorporation or nonincorporation of each nucleotide added, is ultimately assessed on the basis of whether or not photons are detected. There is a minimal time lapse between these events, and the conditions of the reaction are such that iterative addition of nucleotides and PPi detection are possible. Prior to the start of the Pyrosequencing reactions, an amplicon is generated in a polymerase chain reaction (PCR) in which one of the primers was biotinylated at its 5' terminus. The biotinylated double-stranded DNA PCR products are then linked to a solid surface coated with streptavidin and denatured. The two strands are separated, and the strand bound to the solid surface is usually used as template. After hybridization of a sequencing primer to this strand, DNA synthesis under Pyrosequencing conditions can commence. In Pyrosequencing, 1 pmol of DNA template molecules can generate the same number of ATP molecules per nucleotide incorporated, which, in turn, can gen-

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

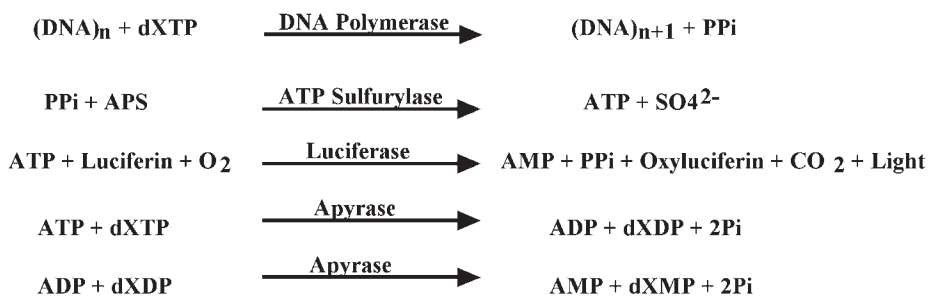


Fig. 1. Schematic representation of progress of enzymatic reactions in Pyrosequencing. DNA template with hybridized primer and four enzymes involved in Pyrosequencing are added to a well of a microtiter plate. The four different nucleotides are added stepwise, and incorporation is followed using the enzyme ATP sulfurylase and luciferase. The unincorporated nucleotides of each addition are continuously degraded by apyrase allowing addition of subsequent nucleotide.

erate more than 6×10^9 photons at a wavelength of 560 nm. This amount of light is easily detected by a photodiode, photomultiplier tube, or charge-coupled device (CCD) camera.

The methodological performance of Pyrosequencing has been established (3). Perhaps its most significant application thus far has been in the area of single nucleotide polymorphism (SNP) genotyping (4,5). Signature sequencing of stretches of <40 nucleotides has been used for cDNA clone identification and gene expression profiling (6) and microbial typing (7). Although the efficacy of Pyrosequencing is presently best for short sequences, longer stretches have been resequenced in disease genes (8) and microbial genes (9). Additionally, DNA regions with difficult secondary structure have been efficiently sequenced by this method (10). Finally, it has been shown that the inclusion of single-stranded DNA (ssDNA)-binding protein enhances the performance of Pyrosequencing especially for obtaining longer reads (11). We are now using this technique for clone checking and sequencing of bar code cassettes in the yeast strains from the yeast deletion project. In this chapter, we describe the steps involved in sample preparation and Pyrosequencing.

2. Materials

2.1. Preparation of Template Hybridized to Sequencing Primer

1. Materials needed for a conventional PCR reaction, with the exception that one of the primers is 5' biotinylated.
2. Streptavidin-coated magnetic beads (Dynal A.S., Oslo, Norway).

3. Magnetic rack for sedimentation of magnetic beads (Dynal A.S.; or Pyrosequencing AB, Uppsala, Sweden).
4. 2X Binding/washing buffer: 2 M NaCl, 0.1% Tween-20, 1 mM EDTA, 10 mM Tris-acetate, pH 8.0.
5. 10X Annealing buffer: 0.1 M Tris-acetate, pH 7.5, 20 mM magnesium acetate.
6. 1X Annealing buffer.
7. Sequencing primer.
8. ssDNA-binding protein for long-read sequencing (Amersham Pharmacia Biotech, Uppsala, Sweden).

2.2. Pyrosequencing Reaction

1. TAE buffer: 100 mM Tris-acetate, pH 7.75; 0.5 mM EDTA.
2. Enzyme mix (for 50- μ L Pyrosequencing reaction): 5 U of exonuclease-deficient (exo-) Klenow DNA polymerase (Amersham Pharmacia Biotech), 0.04 U of apyrase (Sigma, St. Louis, MO), 100 ng of purified luciferase (BioTherma, Dalarna, Sweden) or 1.5 μ g of recombinant luciferase (Promega, Madison, WI), and 0.015 U of purified recombinant sulfurylase (*see* **Notes 1** and **2**).
3. Substrate mix (final concentrations in Pyrosequencing reaction): 5 mM magnesium acetate, 0.1% bovine serum albumin, 1 mM dithiothreitol (DDT), 5 μ M adenosine 5'-phosphosulfate (APS) (Sigma), 0.4 mg/mL of polyvinylpyrrolidone ($M_r = 360,000$), and 100 μ g/mL of D-luciferin (Promega) (*see* **Notes 3** and **4**).
4. Solutions (2 mM) of each of the four nucleotides dATP- α S, dCTP, dGTP, and dTTP (*see* **Notes 5** and **6**).
5. Microtiter plates: Pyrosequencing AB provides special plates to be used with its instruments.
6. Inkjet cartridge for delivery of enzymes, substrates, and nucleotides. Pyrosequencing AB provides this.
7. Pyrosequencing machine: Two machines, PSQ 96 and PHS 96, are commercially available (Pyrosequencing AB).

3. Methods

3.1. Preparation of Template Hybridized to Sequencing Primer

1. Produce a biotinylated amplicon as in a conventional PCR reaction, except be sure that the primer that defines the 5' terminus of the strand acting as template during Pyrosequencing is biotinylated (*see* **Notes 7–10**).
2. For each sample to be sequenced, separate 200 μ g of streptavidin-coated paramagnetic beads from the manufacturer's suspension medium by sedimentation on a magnetic rack, wash once in 2X binding/washing buffer, and finally resuspend in 45 μ L of the same buffer (*see* **Notes 11** and **12**). Alternatively, streptavidin-coated sepharose beads can be used for template preparation.
3. Mix the resuspended beads with the same volume (45 μ L) of biotinylated PCR product, and for the most efficient immobilization, agitate for 15 min at 43°C (*see* **Notes 13–17**).

4. Sediment the beads and remove the supernatant. Incubate the immobilized template in 50 μL of 0.1 *M* NaOH for 3–5 min at 43°C to achieve denaturation (*see Note 18*).
5. Sediment the beads and remove the NaOH solution containing the nonbiotinylated DNA strand. Wash the beads once in 1X annealing buffer (*see Note 19*).
6. Resuspend the beads in 10 μL of annealing buffer containing 10 pmol of sequencing primer (*see Notes 20–22*).
7. Achieve hybridization of primer to template by incubating at 95°C for 30 s, 60°C for 5 min, and finally cooling to room temperature (*see Notes 23–25*).
8. After reaching room temperature, add 0.5–1 μg of ssDNA-binding protein to achieve better reads, particularly on templates with difficult secondary structure or when long reads are needed.

3.2. Pyrosequencing Reaction

1. Dilute DNA template with hybridized primer with 40 μL of TAE buffer and transfer to a well in a Pyrosequencing microtiter plate.
2. Transfer enzyme mix, substrate mix, and each of the four nucleotides to appropriate slots of a Pyrosequencing inkjet cartridge, and place the inkjet cartridge in a Pyrosequencing machine. All components should be allowed to reach room temperature (*see Notes 26–28*).
3. Program information indicating which wells of the microplate contain samples to be sequenced and the dispensation order of nucleotides for each well into the Pyrosequencing machine (*see Note 29*). After receiving the signal for the start of reactions, the Pyrosequencing machine will add to each well enzyme mix, then substrate mix, and then nucleotides according to the specified order.

3.3. Pyrosequencing Machine

The automated Pyrosequencing machine uses a disposable inkjet cartridge to deliver enzymes, substrates, and each of the four nucleotides into the wells of a microtiter plate. The plate is under continuous agitation to achieve optimal mixing of reagents. A lens array in the detection system of the instrument focuses luminescence from each well onto the chip of a CCD camera (**Fig. 2**). Successive nucleotides are dispensed in each well at 1-min intervals. A cooled CCD camera images the plate every second in order to detect progression of the Pyrosequencing reactions. Data acquisition modules and an interface for PC connection are used in the instrument. Two automated versions of the Pyrosequencing machine, PSQ 96 and TP 384, for simultaneous analysis of 96 and 384 samples, respectively, are currently available.

3.4. Data Analysis

The Pyrosequencing machine generates raw data in real time in the form of bioluminescence generated from the reactions (*see Note 30*). The data (pyrogram) are also retrievable at the end of all synthesis reactions in printed format

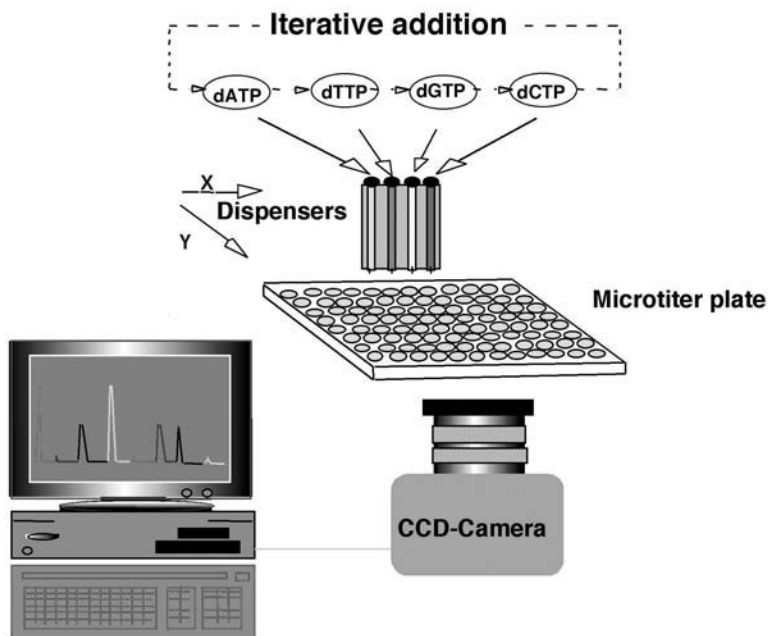


Fig. 2. Schematic drawing of automated system for Pyrosequencing. Four dispensers move on an X-Y robotics arm over the microtiter plate and add the four different nucleotides according to a prespecified order. The microtiter plate agitates continuously to properly mix the added nucleotide. Generated light is directed to the CCD camera using a lens array, which is located exactly below the microtiter plate.

(**Fig. 3**). The data are directly interpretable in terms of the identity of the nucleotide incorporated at each synthesis event. Successive incorporations of the same nucleotide are detectable because of generation of increased luminescence proportional to the number of nucleotides (*see Note 31*). Various types of software including SNP, Tag, allele quantification, haplotyping, and sequence determination are available or are under development at Pyrosequencing AB.

4. Notes

1. A prepared enzyme mix for direct placement into an inkjet cartridge (*see Sub-heading 3*.) is provided by Pyrosequencing AB.
2. We use exonuclease-deficient DNA polymerase such as mutated *exo-* Klenow DNA polymerase to obtain synchronized extension. High fidelity in the synthesis is nevertheless obtained owing to the addition of each of the natural nucleotides at a concentration above its K_M and the use of apyrase in the system. The apyrase degrades unincorporated nucleotides to a concentration well below the K_M at which incorporation of nucleotide by polymerase is very accurate.

Fig. 3. Pyrogram of raw data obtained from Pyrosequencing of a cloned DNA. The order of nucleotide addition is indicated below the pyrogram. The y-axis determines the luminescence intensity. The sequence of the clone is 5'-GATTTGGATTCCTTGGACCAGAGAATGTGAGCTGATT-CACGATTTGGA.

3. A substrate mix for direct placement into an inkjet cartridge (*see Subheading 3.*) is provided by Pyrosequencing AB. Alternatively, stock solutions for each of the enzymes, DTT, APS, and a mix containing other substrate components can be kept at -20°C and an enzyme-substrate mix in TAE prepared for each reaction (*see Subheading 3.*). Luciferin is light sensitive and therefore should not be exposed to light. The polymerase stock solution should be thawed only once, then kept at 4°C for future use.
4. Appropriate enzyme and substrate concentrations were defined empirically and with regard to the enzymes' K_{cat} and K_M values for respective substrates. Most important, the K_M of the DNA polymerase for the nucleotides is much lower than that of apyrase.
5. dATP- αS is used instead of dATP (*I*) because it serves as a good substrate for DNA polymerase and apyrase, but not for luciferase.
6. The nucleotides are provided ready to use by Pyrosequencing AB. Alternatively, 20 μL of 0.1 M solutions (Amersham Pharmacia Biotech) should be treated with 100 mU of inorganic pyrophosphatase isolated from baker's yeast (Sigma) to remove contaminating PPi, then diluted to 1 mL with 10 mM Tris-actate, pH 7.5. The nucleotide solutions can be kept at 4°C for up to 1 yr or at -20°C . Repeated thawing should be avoided.
7. It should be ascertained that unbound biotin was efficiently removed by the manufacturer of the biotin-labeled primer. The manufacturer will normally provide this information.
8. High concentrations of primers in the PCR should be avoided because unused biotinylated primers will later efficiently compete with the larger biotinylated PCR product for binding to streptavidin-coated magnetic beads. We normally use 5–10 pmol of each primer in a 50- μL PCR volume.
9. The amount of DNA template is extremely important for successful Pyrosequencing, particularly in order to obtain longer reads. For sequencing using the PSQ 96 instrument, 1 pmol of template is needed. The yield of PCR product is normally estimated by comparison of intensity of ethidium bromide staining of PCR fragment with staining of fragment of comparable size and known concentration after electrophoresis. We normally use 10 ng of genomic DNA for direct amplification by PCR for Pyrosequencing.
10. We routinely achieve successful sequencing on templates ranging in size from 50 to 2000 bp.
11. The beads should be treated gently; vortexing and centrifugation should be avoided. Furthermore, the beads should never be allowed to dry up.
12. Alternatively, streptavidin-coated Sepharose beads can be used for template preparation.
13. The recommended PCR volume is normally used in our laboratory. However, this volume can be decreased or increased depending on yield of PCR product. Use of more than 150 μL of PCR product may result in successful competition of unused biotinylated primers for binding sites on the beads. In any case, the final volume

should be adjusted with H₂O or 2X binding/washing buffer to achieve a final concentration of 1X binding/washing buffer.

14. When working with few samples, microfuge tubes can be used. With a large number of samples, routinely used microtiter plates can be used.
15. The agitation should be just enough to keep the beads constantly dispersed.
16. Fifteen minutes at 43°C achieves more than 90% binding. For more complete binding, incubation can be increased for up to 1 h. For immobilization of long PCR products (longer than 1 kb), incubation of more than 1 h is recommended.
17. The Eppendorf thermomixer (Eppendorf) that has holders for 96-well microtiter plates or microfuge tubes is convenient for the binding reaction.
18. Exposure to NaOH increases the tendency of the beads to clump. Excessive exposure to NaOH should be avoided. It is particularly important that the beads not be allowed to dry.
19. The nonbiotinylated strand, after neutralization with HCl, can also be used as template for Pyrosequencing.
20. One of the PCR primers or an internal oligonucleotide can serve as sequencing primer. Alternatively, a self-looping sequencing primer attached to the template in the PCR can be used as the sequencing primer, allowing omission of the hybridization step (12).
21. The primer mix can be prepared during one of the previous incubations to minimize time lapse. A 10X annealing buffer solution can be used to achieve a concentration of 1X in the primer mix.
22. Care must be taken to gather all the beads in the small volume of primer mix.
23. A temperature of 70°C instead of 60°C is used for templates with a high G/C content.
24. The incubations can be done in a conventional heating block or in a thermocycler.
25. Although template/primer can be kept for at least a couple of days before Pyrosequencing, it appears that optimal results are achieved with freshly prepared samples.
26. Alternatively, the enzymes and substrates are included in the 40 µL of TAE with which the template/primer hybrid was diluted. In this case, only nucleotides are added to the inkjet cartridge.
27. The Pyrosequencing reactions are carried out at room temperature because of heat sensitivity of the enzymes, particularly the luciferase.
28. If enzymes and substrates were added manually to the samples, the machine would go through the enzyme and substrate dispensation steps without actually making these dispensations.
29. An SNP analysis program is available that specifies optimal dispensation order for each SNP analysis. For sequencing purposes, a dispensation order consisting of up to 96 nucleotides is presently programmable. For sequencing of unknown sequences, cyclic dispensation of the four nucleotides is appropriate. For resequencing purposes, a dispensation order dictated by the known sequence may be more appropriate. Different dispensation orders may be programmed for different wells of the microtiter plate.

30. If inadequate signal is obtained despite having used a sufficient amount of PCR product, we suggest checking the concentration of the sequencing primer.
31. The proportionality between the amount of light generated and the number of nucleotides incorporated significantly diminishes when the template sequence demands the successive incorporation of the same nucleotide more than four or five times. Furthermore, in such cases it is possible that complete synthesis will not be achieved on a single dispensation of that nucleotide. This in turn can cause frameshifts and subsequent problems in reading the sequence. To overcome this problem, on resequencing, multiple (two or three) dispensations of the same nucleotide can be programmed at that site to ensure complete incorporation of the nucleotide.

References

1. Ronaghi, M., Karamohamaed, S., Pettersson, B., Uhlen, M., and Nyren, P. (1996) Real time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**, 84–89.
2. Ronaghi, M., Uhlen, M., and Nyren, P. (1998) A sequencing method based on real time pyrophosphate. *Science* **281**, 363–365.
3. Ronaghi, M. (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res.* **11**, 3–11.
4. Ronaghi, M. (2001) “Pyrosequencing for SNP genotyping,” in *Single Nucleotide Polymorphisms: Methods and Protocols* (Kwok, P.-Y., ed.), Humana, Totowa, NJ, pp. 189–195.
5. Nordstrom, T., Ronaghi, M., Forsberg, L., de Faire, U., Morgenstern, R., and Nyren P. (2000) Direct analysis of single-nucleotide polymorphism on double-stranded DNA by Pyrosequencing. *Biotechnol. Appl. Biochem.* **31**, 107–112.
6. Nordstrom, T., Gharizadeh, B., Pourmand, N., Nyren, P., and Ronaghi, M. (2000) Method enabling fast partial sequencing of cDNA clones. *Anal. Biochem.* **292**, 266–271.
7. Gharizadeh, B., Kalantari, M., Garcia, C. A., Johansson, B., and Nyren P. (2001) Typing of human papilloma virus by Pyrosequencing. *Lab. Invest.* **81**, 673–679.
8. Garcia, C. A., Ahmadian, A., Gharizadeh, B., Lundeberg, J., Ronaghi, M., and Nyren P. (2000) Mutation detection by Pyrosequencing: sequencing of exons 5-8 of the p53 tumor suppressor gene. *Gene* **253**, 249–257.
9. Monstein, H., Nikpour-Badr, S., and Jonasson, J. (2001) Rapid molecular identification and subtyping of *Helicobacter pylori* by Pyrosequencing of the 16S rDNA variable V1 and V3 regions. *FEMS Microbiol. Lett.* **199**, 103–107.
10. Ronaghi, M., Nygren, M., Lundeberg, J., and Nyren, P. (1999) Analyses of secondary structures in DNA by Pyrosequencing. *Anal. Biochem.* **267**, 65–71.
11. Ronaghi, M. (2000) Improved performance of Pyrosequencing using single-stranded DNA-binding protein. *Anal. Biochem.* **286**, 282–288.
12. Ronaghi, M., Pettersson, B., Uhlen, M., and Nyren, P. (1998) PCR-introduced loop structure as primer in DNA sequencing. *BioTechniques* **25**, 876–884.

Use of Fimers to Eliminate Polymerase Chain Reaction and Primer-Dimer Artifacts and to Increase Yield in BAC-Sequencing Reactions

**Andrei Malykh, Nikolai Polushin, Alexei Slesarev,
and Sergei Kozyavkin**

1. Introduction

Specificity of primer-based amplification reactions depends on the specificity of primer hybridization. Ideally, under the elevated temperatures used in a typical amplification, the primers should hybridize only to the target sequence. However, there is a relatively narrow range of conditions (temperature, concentrations of ions and denaturing agents) for specific annealing of an oligonucleotide to its complementary target, and often it does not coincide with other requirements of the cycle-sequencing reaction. The discrimination between specific and nonspecific hybridization is most challenging when DNA template is bacterial artificial chromosome (BAC), P1-derived artificial chromosome, yeast artificial chromosome, or other long DNA with a large number of potential nonspecific priming sites. A similar situation occurs when DNA contains perfect or nonperfect repeats. As a result, amplification of nonspecific primer extension products can compete with amplification of the desired target sequences and can significantly decrease the efficiency of the amplification of the desired sequence (1).

A particular type of artifact occurs in the linear cycle-sequencing reaction when the newly synthesized DNA strand contains both a primer sequence and the complement of a primer sequence in the appropriate orientation. Such DNA would then serve as a template for exponential amplification (polymerase chain reaction [PCR]) (2). Even if such molecules are produced only as rare side

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

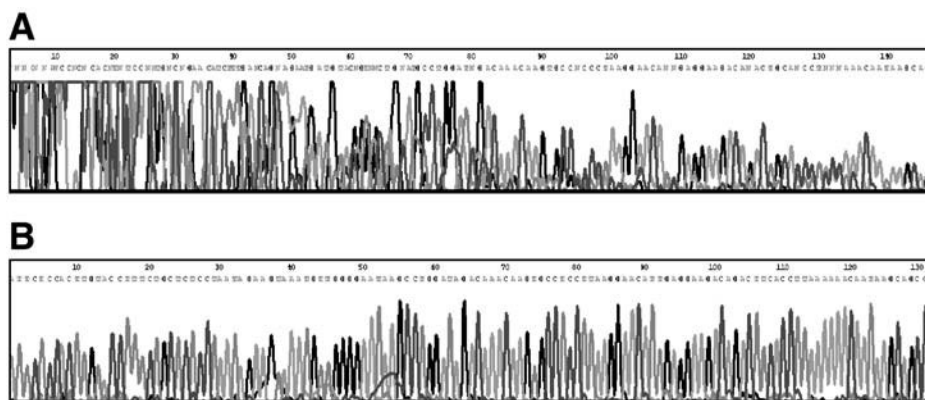


Fig. 1. Elimination of nonspecific PCR background during cycle sequencing from human BAC. Sequencing with primer 5'-GTGAGGGAGAATGCAGTATG-3' (A) and MOX-Fimer 5'-GTGAGGGAGAATGCAGU^{MOX-OH}ATG-3' (B).

products, their exponential growth may bring them to high concentrations relative to the linearly growing desired product (**Fig. 1**). The problem potentially worsens as the number of cycles exceeds 40.

Another situation that occurs often during directed sequencing of BAC DNA is when one cannot select an ideal primer in the particular sequence window. In this case, a primer may have two segments of sequences that are complementary to each other to form an internal hairpin structure preventing the primer's hybridization to its target; or, an oligonucleotide sequence selected allows it to be extended over another copy to form primer-dimer. The resulting concatenation forms an undesired template, which, because of its short length, is amplified efficiently.

A fundamental solution to the described problems is the use of oligonucleotides with one or more 2'-modified bases synthesized from methoxyoxalamido (MOX) and succinimido (SUC) precursors that cannot be replicated by DNA polymerase (*1,3*). Such modified oligonucleotides, we name Fimers, allow the desired linear amplification during cycle sequencing but prevent the undesired exponential growth of side products including primer-dimer formation.

Schematically the concept of Fimers is shown in **Fig. 2**. First, the precursor oligonucleotide containing 2'-MOX- or 2'-SUC-substituted nucleosides is synthesized. Second, the oligonucleotide is postsynthetically reacted with an appropriate nucleophile (primary aliphatic amine or hydroxide anion) and deprotected to form a modified oligonucleotide. The whole procedure is completely compatible with commonly used phosphoramidite chemistry.

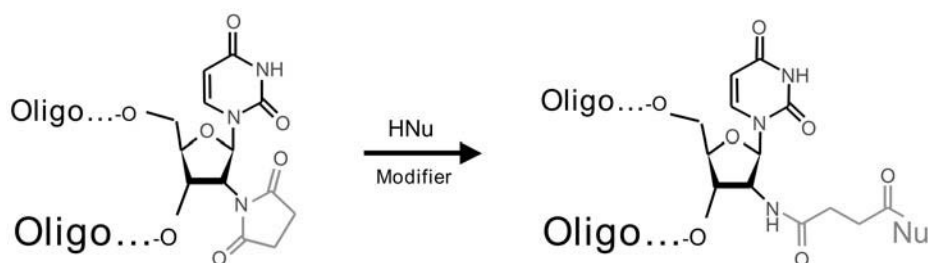


Fig. 2. Fimer synthesis from MOX precursor oligonucleotide. On postsynthetic treatment of the precursor with a desirable modifier, the reactive moieties are effectively derivatized and the final 2'-modified oligonucleotide (or modified primer library) is formed. This strategy enables one to synthesize a wide variety of modified primers from a single parent oligonucleotide.

We use Fimers to inhibit different nonspecific events that interfere with the sequencing of BAC DNA, including primer-dimer accumulation, exponential amplification of side products (PCR), and oligonucleotide annealing to secondary priming sites. Placing at least one modified nucleoside within a potential complementary site might inhibit formation of the dimer from the priming oligonucleotide (*1*).

Placing the modification close to the 3'-end prevents both nonspecific primer-dimer extension and nonspecific PCR during cycle sequencing (**Fig. 3**). The optimal effect was found when modified nucleotide is in the -5 position. In this case, linear amplification is efficient but artifacts owing to the exponential amplification disappear. We also found that primer extension reaction is completely inhibited when DNA polymerase approaches the modified nucleoside in the template strand.

Figure 4 gives an example of adjusting melting temperature (T_m) of Fimers using different modifiers. Selecting appropriate modifiers can easily modulate annealing properties of Fimers (*4*). This approach allows achieving higher specificity of Fimer hybridization with the template. In the experiment shown in **Fig. 4**, five nucleotides were modified. The change in T_m in this case depends significantly on the type of modifier.

For the past 5 yr, we have used more than 30,000 Fimers in combination with ThermoFidelase (*5–7*) in different DNA-sequencing applications involving directed sequencing off of BAC, phage, or genomic DNA templates. The most common tasks were finishing BAC draft shotgun sequences, BAC end sequencing, determining exon/intron boundaries in cloned genes in BACs, BAC sequencing in the presence of excessive amounts of *Escherichia coli* genomic

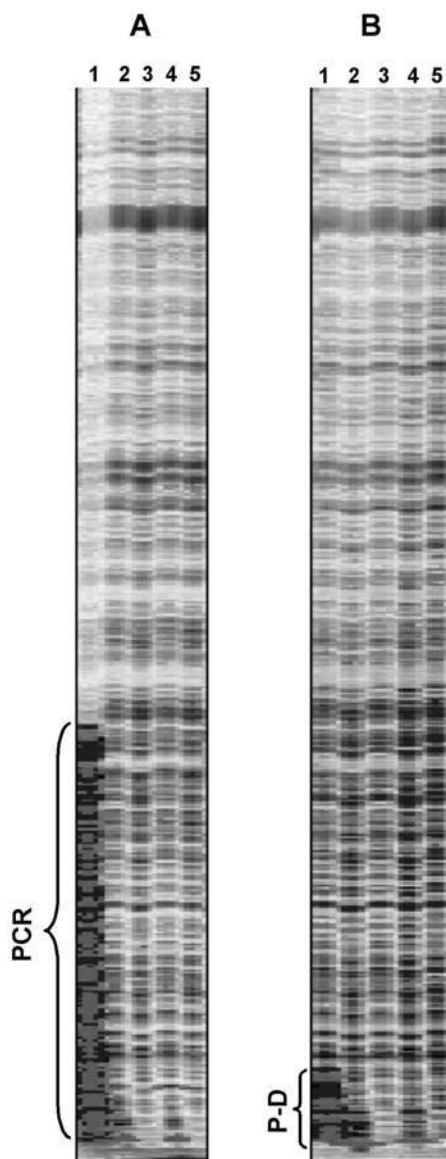


Fig. 3. Uses of Fimers to increase yield and prevent PCR and primer-dimer formation in BAC sequencing reactions. (A) Thirty nanograms and (B) 100 ng of human BAC DNA were end sequenced using the 35T7 primer 5'-CGGCCAGTGAATTGTAATACGACTCACTATAGGG-3 (lanes 1 in [A] and [B]). Lanes 2-5 in (A) and (B) were obtained by sequencing with Fimers having 2'-SUC-2'-deoxyuridine in -5 position from 3'-end of the oligonucleotide with -OH modifier (lanes 2), in -5 position with -NHCH₂CH₂OH modifier (lanes 3), in -5 and -11 positions with -OH modifier (lanes 4), and -5 and -11 positions with -NHCH₂CH₂OH modifier (lanes 5). Oligonucleotides were prepared on Bioset oligosynthesizer using a set of regular four nucleotide amidites and 2'-SUC-2'-dU specialty amidite from Fidelity Systems (www.fidelitysystems.com/SU.html).

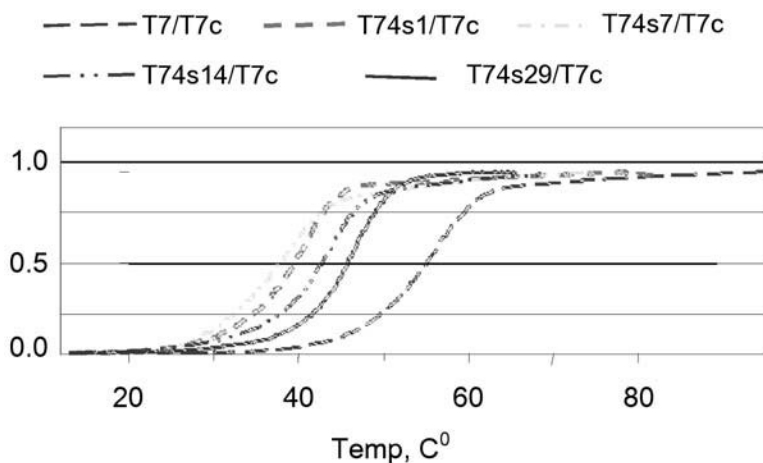


Fig. 4. Adjusting T_m of Fimers using different modifiers. The following oligonucleotides were synthesized: T7 (5'-GTAATACGACTCACTATAGGG-3'), T7c (5'-CCC TATAGTGAGTCGTATTAC-3'), T74s1, T74s7, T74s14, and T74s29 (5'-GU^sAAU^s ACGACU^sCACU^sAU^sAGGG-3'), in which s1, s7, s14, and s29 are different modifiers (7). A pair of complementary oligonucleotides, each at a concentration of 0.1 A_{260} optical units, was combined in 10 mM Tris-HCl buffer (pH 8.0 at 25°C), 2 mM $MgCl_2$ in a total volume of 400 mL; heated to 95°C; cooled to room temperature; and used in the melting experiment.

DNA, comparative sequencing of primate and rodent BACs, and no-error human BAC sequencing of individual samples.

2. Materials

2.1. Preparation of BAC DNA

1. Shaker-incubator.
2. Equipment for agarose gel electrophoresis.
3. Plasmid purification (Midi) kit (Qiagen).
4. Tabletop centrifuge.
5. Ultraviolet spectrophotometer.

2.2. BAC Sequencing

1. Automated DNA Sequencer (Applied Biosystems, Amersham Biosciences, LI-COR, or Beckman Coulter).
2. Thermocycler.
3. Heat sealer.
4. Centrifuge with rotor for microtiter plates.
5. Vacuum centrifuge with rotor for microtiter plates.

6. BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems).
7. ThermoFidelase 2 (Fidelity Systems).
8. Fimers (Fidelity Systems).
9. 7-Deaza-dGTP (Roche).
10. 96- or 384-well polyethylene or polypropylene plates (MJ Research).
11. Multichannel pipets.
12. Sealing foil (MJ Research).
13. MicroSpin G-50 columns (Pharmacia Biotech).
14. Sephadex G-50 Superfine (Sigma, St. Louis, MO).
15. Millipore MultiScreen 96-well filtration plates (Millipore, Bedford, MA).
16. Loading buffer.

3. Methods

3.1. Standard Cycle-Sequencing Protocol for BACs

3.1.1. Single-Sequencing Reaction Setup

For sequencing of a single BAC template with <96 Fimers, use the following for a single-reaction setup (*see* **Notes 1** and **2**):

1. BAC DNA: 100–500 ng.
2. ThermoFidelase 2: 0.1 μL .
3. 1 mM 7-deaza-dGTP: 0.1 μL .
4. Fimer: 1 μL .
5. Dye Terminator Mix: 2 μL .
6. dH_2O , to a total volume of 5 μL .

3.1.2. Multiple Reaction Setup

To compensate for pipetting errors we recommend mixing somewhat more reagents. An example of a 96-reaction setup is as follows:

1. BAC DNA (250 ng/reaction): 26.5 μg .
2. ThermoFidelase 2: 10.6 μL .
3. 1 mM 7-deaza-dGTP: 10.6 μL .
4. Dye Terminator Mix: 212 μL .
5. dH_2O , to a total volume of 424 μL .

3.1.3. Procedure

1. Label the tube that is going to be used for the reaction setup.
2. Add the desired amount of template to this tube based on the calculated DNA concentration.
3. Add ThermoFidelase 2 and mix by pipeting.
4. Add Dye Terminator Ready Reaction Mix.
5. Add 1 mM 7-deaza-dGTP. Use fresh solution every time.
6. Add calculated amount of water.

7. Mix well by pipetting. The total volume of premix should be equal to $4\ \mu\text{L} \times \text{number of reactions}$.
8. Dispense $4\ \mu\text{L}/\text{tube}/\text{well}$ (use 8- or 12-channel pipet).
9. Add Fimers ($1\ \mu\text{L}$ each) using a multichannel pipet. Mix by pipetting.
10. Close the tubes and seal the plate.
11. Spin the tubes and plate briefly.
12. Place the tubes and plate in a thermocycler. Use the following cycling conditions:
 - a. Heat the plate once at 95°C for 2 min.
 - b. 95°C for 5 s (denaturation).
 - c. 55°C for 20–30 s (annealing).
 - d. 60°C for 4 min (extension).
 - e. 100 cycles total, then 4°C .
13. Add $15\ \mu\text{L}$ of 20 mM EDTA to each well.

3.2. Sequencing From a Small Amount of BAC DNA

3.2.1. Reaction Setup

This protocol has been tested for sequencing from 10 to 100 ng of BAC DNA. It can be used for BAC end sequencing, since the yield of DNA in mini-preps is usually $<1\ \mu\text{g}$ and the concentration of samples varies in the range of 50–200 ng/ μL . ThermoFidelase 2BE is recommended when sequencing from $<100\ \text{ng}$ of BAC DNA. For sequencing of a single BAC template with <96 Fimers, use the following to calculate of the amount of each component (*see Note 2*):

1. BAC DNA: 10–100 ng.
2. ThermoFidelase 2BE: $0.5\ \mu\text{L}$.
3. 1 mM 7-deaza-dGTP: $0.1\ \mu\text{L}$.
4. Fimer: $1\ \mu\text{L}$.
5. Dye Terminator Mix: $2\ \mu\text{L}$.
6. dH_2O , to a total volume of $5\ \mu\text{L}$.

3.2.2. Procedure

1. Label the tube that is going to be used for the reaction setup.
2. Based on the calculated DNA concentration, add the desired amount of template to this tube.
3. Add ThermoFidelase 2BE. Mix by pipetting.
4. Add 1 mM 7-deaza-dGTP. Use fresh solution every time.
5. Add Dye Terminator Ready Reaction Mix.
6. Add the calculated amount of water.
7. Mix well by pipetting. The total volume of the premix should be equal to $4\ \mu\text{L} \times \text{number of reactions}$. For three reactions it will be $12\ \mu\text{L}$.
8. Dispense $4\ \mu\text{L}/\text{tube}/\text{well}$.

9. Add 1 μL of corresponding Fimer using a multichannel pipet. Mix by pipetting.
10. Close the tubes and seal the plate.
11. Spin the tubes and plate briefly.
12. Place the tubes and plate in a thermocycler.
13. Use the following cycling conditions:
 - a. Heat plate once at 95°C for 2 min.
 - b. 95°C for 5 s (denaturation).
 - c. 55°C for 20–30 s (annealing).
 - d. 60°C for 1 min (extension).
 - e. 400 cycles total, then 4°C .
14. Add 15 μL of 20 mM EDTA to each well.

3.3. Sequencing Through CT-Rich Regions in BAC

ThermoFidélase 2C is recommended for sequencing through CT repeats.

1. Use the following to calculate the amount of each component:
 - a. BAC DNA: 250–700 ng.
 - b. ThermoFidélase 2C: 0.5 μL .
 - c. Fimer: 1 μL .
 - d. Dye Terminator Mix: 2 μL .
 - e. dH_2O , to a total volume of 5 μL .
2. Use the following cycling conditions:
 - a. Heat plate at 95°C for 2 min.
 - b. 95°C for 5 s (denaturation).
 - c. 55°C for 20–30 s (annealing).
 - d. 60°C for 4 min (extension).
 - e. 100 cycles total, then 4°C .
3. Add 15 μL of 20 mM EDTA to each well.

3.4. Sequencing From BAC Contaminated With Chromosomal DNA

ThermoFidélase can enzymatically unlink (denature) closed, circular, supercoiled BAC DNA at temperatures at which linear chromosomal DNA remains double stranded. This procedure can be used for sequencing BAC DNA prepared by methods, that do not guarantee high purity of BAC DNA.

1. Use the following to calculate the amount of each component:
 - a. Total DNA: 500–1000 ng.
 - b. ThermoFidélase I: 1 μL .
 - c. Fimer: 1 μL .
 - d. Dye Terminator Mix: 8 μL .
 - e. dH_2O , to a total volume of 20 μL .
2. Use the following cycling conditions:
 - a. Heat plate at 90°C for 5 min.

- b. 90°C for 5 s (denaturation).
 - c. 55°C for 20–30 s (annealing).
 - d. 60°C for 4 min (extension).
 - e. 100 cycles total, then 4°C.
3. Add 15 μ L of 20 mM EDTA to each well.

4. Notes

1. The protocol given is optimized for reaction in a 5- μ L total volume. The sequencing reaction can be done in larger volumes as well. In this case, the amount of ThermoFidelase and Fimer is the same as for a 5- μ L reaction.
2. Addition of 7-deaza-dGTP helps overcome stops when sequencing GC-rich templates. Since the volume of ThermoFidelase 2 per reaction is very small, at least three sequencing reactions should be prepared at once. In this case, it is possible to handle 0.3 μ L of enzyme using a P2-type pipet. These reactions can be done in individual 0.2-mL PCR tubes, 8-tube strips, or 48- or 96-well plates. If a heat sealer is used, plates can be reused several times.

Acknowledgments

This work was supported in part by grants from the Department of Energy and the National Institutes of Health.

References

1. Polushin, N., Malykh, A., Malykh, O., et al. (2001) 2'-Modified oligonucleotides from methoxyoxalamido and succinimido precursors: synthesis, properties and applications. *Nucleosides Nucleotides Nucleic Acids* **20**, 75–78.
2. Stump, M. D., Cherry, J. L., and Weiss, R. B. (1999) The use of modified primers to eliminate cycle sequencing artifacts. *Nucleic Acids Res.* **27**, 4642–4648.
3. Polushin, N. N. (2000) The precursor strategy: terminus methoxyoxalamido modifiers for single and multiple functionalization of oligodeoxyribonucleotides. *Nucleic Acids Res.* **28**, 3125–3133.
4. Kozyavkin, S., Malykh, A., Polouchine, N., and Slesarev, A. (2003) US patent no. 6,548,251 B1.
5. Slesarev, A. I., Belova, G. I., Lake, J. A., and Kozyavkin, S. A. (2001) Topoisomerase V from *Methanopyrus kandleri*. *Methods Enzymol.* **334**, 179–192.
6. Kumaraswamy E., Malykh A., Korotkov K. V., et al. (2000) Structure-expression relationships of the 15-kDa selenoprotein gene: possible role of the protein in cancer etiology. *J. Biol. Chem.* **275**, 35,540–35,547.
7. Slesarev, A. I., Mezhevaya, K. V., Makarova, K. S., et al. (2002) The complete genome of the hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl. Acad. Sci. USA* **99**, 4644–4649.

Retrofitting BACs and PACs With LoxP Transposons to Generate Nested Deletions

Pradeep K. Chatterjee

1. Introduction

Physical mapping of markers in large-insert bacterial artificial chromosome (BAC) and P1-derived artificial chromosome (PAC) clones requires efficient methods for obtaining position-specific sequence information along the insert DNA. Because substantial quantities of high-quality clone DNA can be easily prepared from BACs and PACs, they are also potentially valuable reagents for direct use in biologic experiments that seek to functionally evaluate the sequences contained in them. Thus, both physical and functional mapping objectives would be facilitated by an efficient strategy for making nested deletions from one end of the insert DNA in a BAC or PAC clone.

Efficient procedures for sequentially deleting pieces from one or both ends of cloned genomic DNA have been developed using a variety of transposon systems (1–6). Using a loxP site-containing Tn10 minitransposon in BACs and PACs offers a unique advantage: the insert DNA in the large clone is directly amenable to transposition, and no subcloning of smaller fragments is required (4–6). This in turn obviates the need to map the ends of deletions back into the parent clone. Deletion ends generated with loxP transpositions into BACs or PACs can be mapped readily by analyzing the size of resulting clones on field inversion gel electrophoresis (FIGE) (7,8). The nested-deletion protocol described here also incorporates a mammalian cell-selectable antibiotic resistance gene into the resulting BAC or PAC deletion, and should facilitate subsequent steps in which the deletion plasmids are reintroduced into cells or animals for functional analyses.

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

The procedure involves introducing a small (7-kb) loxP transposon plasmid into the BAC or PAC clone by calcium chloride transformation. Transformed colonies are selected for resistance to antibiotic markers carried by both the BAC or PAC and the transposon plasmid (5). Transposition of loxP sites is induced with isopropyl- β -D-thiogalactopyranoside (IPTG) in cultures of transposon plasmid-transformed BAC or PAC clones. Clone DNA with transposon insertions is then transduced with P1 phage. Thus, although transposition is equally probable in one of two orientations and only one of these leads to a deletion, the limited packaging capacity of the P1 phage head ensures that deletions are recovered only when the starting BAC or PAC clone is >110 kb (5). BACs and PACs in most libraries are larger than this (7). The P1 phage additionally provides Cre protein *in trans* during transduction, and thus enables the transposed loxP site to recombine with the endogenous loxP sequence in BACs and PACs without necessitating the transfer of genomic clones to a Cre-expressing host (5,8).

2. Materials

All chemicals, unless specified, should be at least molecular biology grade. Solutions should be made with deionized water, autoclaved, and/or sterile filtered through a 0.2- μ Nylon Membrane Disposable Sterile Syringe Filter (Corning, Corning, NY).

1. Antibiotic stock solutions: make all solutions in ultrapure distilled water from Gibco-BRL (Life Technologies, Gaithersburg, MD). Following are the concentrations of the antibiotic stocks (reagents are biotech research grade from Fisher): 10 mg/mL of ampicillin (Amp) (penicillin-G potassium), 10 mg/mL of kanamycin (Kan) (kanamycin monosulfate), and 2.5 mg/mL of chloramphenicol (Cm). Store frozen at -20°C .
2. 100 mM IPTG stock solution in water; store in 1-mL aliquots and freeze at -40°C .
3. Bacto agar, Bacto tryptone, and yeast extract (Difco, Detroit, MI; Beckton Dickinson, Sparks, MD).
4. Glass capillary tubes (microhematocrit capillary tubes), for streaking bacteria and recovering phage plaques from agar plates (VWR; Drummond Scientific), and disposable plastic bacterial cell spreaders (sterile) (VWR). The use of capillaries and disposable spreaders obviates the need for a Bunsen flame and ethanol bath at the workbench.
5. Disposable plastic Petri dishes (VWR).
6. Polypropylene 14- and 5-mL snap-cap round-bottomed disposable tubes (sterile) (Falcon; VWR).
7. Phage and bacterial strains: All phage and bacterial strains used here have been described previously (5,9,10) and are available on request. NS3516 is a *lac* Iq host strain that does not express Cre protein. YMC (*rec+*, *supE*) is the bacterial host used to grow P1 phage stocks. The P1^{vir} used here is a virulent form of P1 phage. BACs and PACs are commonly in the DH10B host and work well.

8. Vectors and transposon plasmids: The commonly used vectors for BACs and PACs, pBeloBAC11 and pCYPAC2, respectively, work well in the procedure. Experiments with BACs constructed in pIndigoBAC536 vector were also successful. The transposon plasmids pTnBAC/loxP and pTnPGKpuro/loxP are used to introduce loxP sequences into BACs and PACs, respectively. They have been described before (5,7) and are available on request.
9. Luria Bertani (LB) growth medium, agar plates containing LB, and so forth: Procedures for these are described in **ref. 11**.

3. Methods

3.1. Nested-Deletion Procedure

3.1.1. Transforming BAC/PAC Clones With Transposon Plasmid

1. Streak out a small aliquot (2 μ L) from a glycerol stock of a BAC clone (or a PAC clone) on an LB agar plate containing 25 μ g/mL of Cm (or Kan for a PAC clone) and incubate overnight at 37°C (*see Note 1*).
2. Pick a single BAC colony and grow it as a suspension culture in LB containing Cm (or Kan for a PAC clone) at 25 μ g/mL overnight at 37°C.
3. Take 20 μ L of the overnight culture and inoculate 5 mL of LB containing 25 μ g/mL of Cm (or Kan for a PAC clone) in a 50-mL sterile polypropylene orange-cap tube (Corning) with a couple of small holes punched in the cap with a 22-gage needle in order to facilitate aeration (*see Note 2*).
4. Incubate the tubes with vigorous shaking at 37°C for 2 h (early log phase).
5. Remove 1.6 mL of culture into an Eppendorf tube (filled to the top), and spin the cells for 1 min gently at approx 6000 rpm in a microfuge (inside a chromatography refrigerator at 4°C, or a cold room). Pelleting cells at higher speeds prevents easy resuspension and results in reduced viability.
6. Discard the supernatant (remove as much of the supernatant as possible by shaking), and place the tube on an ice-water slurry.
7. Add 1 mL of prechilled (4°C) 0.1 M calcium chloride solution in water (sterile filtered), and resuspend the pellet by gentle vortexing (setting of 5). Try to prevent frothing (vortex the holding tube vertically), and keep the tube cold at all times. Pipetting with a chilled tip also works well.
8. Pellet the cells at 6000 rpm for 0.5 min at 4°C.
9. Remove all the supernatant as in **step 6**, and resuspend the cells in 0.5 mL of prechilled (4°C) 0.1 M calcium chloride solution in water.
10. Repeat **step 8**.
11. Resuspend the pellet in 150 μ L of chilled 0.1 M calcium chloride solution.
12. Incubate on ice for 10 min. The cells are now competent.
13. Dilute a crude preparation of plasmid DNA in distilled water (pTnBAC/loxP for BACs and pTnPGKpuro/loxP for PACs) such that approx 10 ng of DNA is in a 4- μ L volume in a prechilled Eppendorf tube. Add 50 μ L of competent cells to the DNA (*see Note 3*).
14. Mix well by gentle tapping (keeping chilled), and then leave on ice for 15 min.

15. Heat shock the suspension by placing each tube in a water bath at 37°C for exactly 5 min.
16. Immediately transfer the tubes to an ice bath, and leave for 2 min.
17. Place the tubes at room temperature for 2 min.
18. Add 500 μ L of room temperature S.O.C. (available from Gibco-BRL, cat. no. 15544-018), and gently rock the tubes in an air shaker at 37°C for 1 h.
19. Prewarm the LB agar plates containing Kan +Cm (25 μ g/mL of each antibiotic) to room temperature so as to dry (*see Note 4*).
20. Add approx 150 μ L of transformed cell suspension (S.O.C. culture) to the LB agar plates containing the two antibiotics, and spread well. Allow the plates to dry before inverting and placing in an incubator at 37°C overnight (16–20 h). As many as 30 BAC/PAC clones can be transformed at once with this procedure.

3.1.2. Transposition of *loxP* Sites into BAC/PAC Plasmid and P1 Transduction of Retrofitted Clone

1. Pick a swab of 100 colonies of a size that is well represented (i.e., neither the largest nor the smallest ones) from a plate, and grow in 2 mL of LB + Kan + Cm (25 μ g/mL of each antibiotic) overnight at 37°C with vigorous shaking. *See ref. 11* for a discussion on using a swab of colonies instead of a single clone. Alternatively, the pooled colonies can be resuspended in a small volume of warm medium and an aliquot from the suspension directly diluted and grown for IPTG induction (*see Note 5*).
2. Take 40 μ L of the overnight culture and add to 5 mL of LB + 12.5 μ g/mL of Cm + Kan (half concentrations of each antibiotic) contained in a 50-mL polypropylene orange-cap tube with several holes punched in the cap for aeration.
3. After 2 h of vigorous shaking at 37 to 38°C, add 95 μ L of 100 mM IPTG (dissolved in water, kept as frozen aliquots at –40°C), and continue shaking for a further 3 h (*see Note 6*).
4. Spin the cells down at 1600g for 5 min at room temperature in a swinging-bucket Sorval tabletop centrifuge. Let it come down without the brakes on.
5. Pour off the supernatant and resuspend the cell pellet in 500 μ L of fresh prewarmed LB + 5 mM calcium chloride (no antibiotics here). Calcium chloride stock is 0.5 M in sterile water.
6. Infect each culture with 250 μ L of P1 vir phage stock that has been dechloroformed (*see Note 7*).
7. Incubate without shaking for 30 min at 37°C, to adsorb the phage.
8. Add 4 mL of prewarmed LB containing 5 mM calcium chloride to each tube and shake vigorously for 2 h (*see Note 8*).
9. Add 200 μ L of chloroform (8–10 drops with a glass Pasteur pipet) to each culture, gently vortex the tube, shake for an additional 2 min at 37°C, and then store overnight at 4°C after sealing the holes in the cap with paraffin tape (*see Note 9*).
10. Spin the lysates in the polypropylene tubes at 1600g for 15 min at room temperature in a swinging-bucket rotor in the Sorval tabletop centrifuge. Carefully recover the clear supernatant with a 1-mL plastic tip attached to a pipet into a

new 50-mL orange-cap polypropylene tube, and leave the drops of chloroform and cell debris behind.

11. Dechloroform the clear lysate by shaking the tubes very gently (50 rpm) for 1 h at 37°C. The caps need to have several holes.
12. Recover the BAC/PAC deletion plasmids that are now packaged within the P1 heads by infecting freshly grown NS3516 Cre⁻ cells with the dechloroformed lysate from **step 11**.
13. Inoculate 10 mL (this volume of cells is good for each BAC/PAC clone) of LB with NS3516 Cre⁻ cells, and grow to mid log in a 50-mL orange-cap polypropylene tube with vigorous agitation.
14. Concentrate the NS3516 Cre⁻ cells by spinning at 1600g in a swinging-bucket rotor in a Sorvall tabletop centrifuge for 8 min and resuspending the cell pellet in 500 µL of LB + 5 mM calcium chloride.
15. Add concentrated cell culture to the 5 mL of dechloroformed lysate, and leave for 1 h at 37°C without shaking (phage adsorption).
16. Add 2 mL of fresh prewarmed LB + 5 mM calcium chloride and shake at 37°C of 30 min.
17. Add another 2 mL of LB + 5 mM calcium chloride and continue shaking for another hour.
18. Spin the cells down at 1600g for 7 min in the tabletop Sorval centrifuge at room temperature with the brakes left off.
19. Resuspend the pellet in 800 µL of prewarmed LB.
20. Plate the entire concentrated culture in three to four plates (~200 µL/plate) of LB + Cm + Kan (25 µg/mL of each antibiotic). Spread well and leave at 37°C. Good colonies should appear after about 36 h.
21. There should be about 300–500 colonies per plate. Pick these and directly streak them in parallel in LB agar plates containing either Cm or Amp (for BAC deletions) or (Kan or Amp for PAC deletions). This is the Ampicillin Sensitivity Screen described in **ref. 7**. Or grow them first in LB + Cm (or Kan for PAC deletions) and then do the Ampicillin Sensitivity Screen.

3.1.3. Ampicillin Sensitivity Screen

1. Pick BAC deletion colonies and array on 96-well plates.
2. Inoculate in parallel onto 96-well plates 150 µL/well of liquid LB medium or solid LB agar containing either Cm (25 µg/mL) or Amp (60 µg/mL). This can be done using a 96-pin replicator or sequential using a 12-channel pipettor.
3. After overnight growth, pick the colonies that fail to grow in Amp from the Cm plate. For PACs, deletion colonies are recovered from Kan plates.
4. Grow colonies and isolate BAC/PAC nested-deletion DNA by one of several methods described in this volume.

3.2. Preparation of P1^{vir} Lysate

Phage P1^{vir} stocks are made by the plate lysate method.

3.2.1. Preparation of LCa plates

1. Autoclave 1 L of LB medium containing 15 g of Bacto agar.
2. After cooling to 60°C, add calcium chloride (from 250 mM stock, sterile filtered) to a 2 mM final concentration (see **Note 10**).

3.2.2. Preparation of Magnesium Chloride Stock of YMC Strain

1. Streak out YMC strain onto an LB plate. Grow overnight at 37°C.
2. Pick a single colony and inoculate a 10-mL culture of YMC in LB with shaking at 37°C overnight.
3. Spin the cells for 5 min at 1600g at room temperature in a Sorvall tabletop centrifuge.
4. Resuspend the cells in 10 mL of 10 mM magnesium chloride solution (sterile filtered).
5. Spin the cells again for 5 min at 1600g.
6. Resuspend the cells in 2 mL of 10 mM magnesium chloride solution.
7. Store at 4°C. Keep on ice or cold at all times (see **Note 11**).

3.2.3. Preparation of Top Agar

1. Add 0.7 g of Bacto agar to 100 mL of distilled water and autoclave for 15 min.
2. Add MgSO_4 to a 10 mM final concentration after cooling to 55°C.

3.2.4. Creation of a Single Phage Plaque

1. Prewarm on LCa plate at 43°C for at least 30 min.
2. Prepare bacterial overlay in a 5-mL polypropylene round-bottomed tube (Falcon) by quickly mixing the following in a heat block set at 37°C: 2 mL of LB made up to 10 mM in MgSO_4 , 25 μL of YMC cell stock in magnesium chloride, 2 mL of molten top agar made up to 10 mM in MgSO_4 (see **Note 12**).
3. Cover the top of the tube with a piece of Parafilm®, invert the tube three times, pour the contents onto the LCa plate, and let solidify for 10 min at room temperature.
4. Streak P1^{vir} lysate (10^{-6} dilution of 6-mo-old stock) with a glass capillary across the surface once. (Remove chloroform from the lysate before streaking on the plate by shaking at 37°C for 30 min).
5. Incubate the plate uninverted overnight at 37°C (see **Note 13**). Plaques of P1 vir phage are clear and approx 0.7 mm in diameter (much smaller than those of λ phage).

3.2.5. Creation of a High-Titer Phage Lysate

1. Prewarm LCa plates at 43°C for at least 30 min.
2. Mix in a 5-mL polypropylene tube (Falcon) 50 μL of YMC stock made up to 5 mM in CaCl_2 , two similar sized plaques (see **Note 14**).
3. Incubate at 37°C for 5 min in a heat block.
4. Add to the tube of phage + cells 2 mL of LB containing 10 mM MgCl_2 + 5 mM CaCl_2 .

5. Quickly add 2 mL of molten top agar containing 10 mM MgSO_4 (kept molten at 50°C in a heat block).
6. Cover the mouth of the tube with a piece of Parafilm, quickly invert three times, and pour onto a warm LCa plate (see **Note 15**).
7. Immediately place in a 37°C incubator uninverted, and incubate for 5 to 6 h until a clear lysate covers the entire surface of the plate.

3.2.6. Harvesting of Phage Lysate

1. Add 2 mL of LB + 10 mM MgCl_2 to each plate containing the P1 vir lysate, resuspend the sloppy top agar layer with a plastic cell spreader (VWR), and pour into a polypropylene centrifuge tube (**caution:** polycarbonate tubes are corroded by chloroform). Do this once more to recover all the lysate.
2. Add 100 μL of chloroform and place in 37°C shaker for 5 min.
3. Spin for 5 min at 4000g in an SS-34 or SA-600 rotor at 4°C (no brake) to pellet the debris, agarose, and chloroform.
4. Carefully pipet the lysate (supernatant) into a 5-mL polypropylene snap-cap tube (keep cold and out of light).
5. Add 2 mL of additional LB + 10 mM MgCl_2 to the pellet in the centrifuge tube, mix well by shaking, and spin again.
6. Pipet into the previous tube of lysate.
7. Add one drop of chloroform, and store at 4°C protected from light (wrap aluminum foil around the tube).

3.2.7. Titering of Phage Lysate

1. Prewarm LCa plates at 43°C for at least 30 min.
2. Prepare a 10^{-5} and 10^{-6} dilution of phage lysate in LB + 10 mM MgCl_2 .
3. Mix the following in a 5-mL tube in a heat block at 37°C:
 - a. 10- μL dilution of phage to 50 μL of YMC magnesium chloride stock.
 - b. 5 mM CaCl_2 : Incubate at 37°C for 5 min and then add 2 mL of LB.
 - c. 10 mM MgCl_2 (prewarmed to 37°C) and then 2 mL of top agar.
 - d. 10 mM MgSO_4 (molten at 50°C).
4. Cover the top of the tube with Parafilm and invert three times to mix.
5. Pour onto an LCa plate. Let solidify for 5 min at room temperature. Incubate uninverted overnight at 37°C.
6. Count the plaques. They should be low (10^{10}) to high (10^9) or better. Homogeneous plaquing should be observed, i.e., the same general size of plaques.
7. Check the lysate for sterility by plating 50 μL on an LB plate; if colonies appear, repeat chloroform treatment and recheck.

3.3. Conclusion

The Tn10 minitransposon used in the nested-deletion procedure can insert into genomic DNA in either orientation, and thus potentially generate both deletions and inversions with equal efficiency. As discussed in several previous

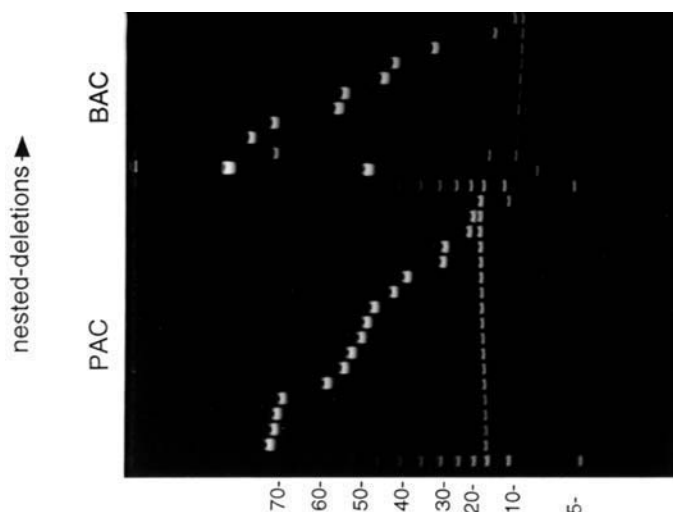


Fig. 1. FAGE analysis of DNA from deletion clones isolated from a 120-kb PAC (lanes 2–18), and a 140-kb BAC (lanes 21–30). Lane 20 shows the starting BAC clone, while lanes 1 and 19 show molecular weight standards comprising a 5-kb ladder. Clone DNA was linearized with *NotI* before electrophoresis. Lane 21 possibly contains DNA from two deletion clones.

articles (5,7,8), the procedure generates deletions exclusively only if the starting BAC or PAC clone is larger than the packaging capacity of a P1 phage head (110 kb). Examples of deletions generated in a BAC and a PAC clone >110 kb are shown in **Fig. 1**. Most BAC/PAC genomic libraries available today comprise clones that average 130–150 kb. Thus, this method is ideally suited for generating closely spaced deletions in clones from such libraries. BAC/PAC clones that are <110 kb also produce deletions with high efficiency. However, half of all clones in such pools are inversions, and these can be readily identified by their size (approx 3 kb larger than the starting clone owing to inserted transposon) on FAGE after being linearized with *NotI* enzyme.

The procedure presented herein can be scaled up to generate nested deletions in multiple BACs and PACs: as many as 30 clones can be processed simultaneously. The real hindrance in scaling it further appears to be clone-to-clone variation in growth characteristics of BACs and PACs. Some clones appear to grow very slowly, and require individual attention. Thus, induction with IPTG and the P1 transduction steps get out of phase with the rest.

A several thousand-member deletion library is readily generated from a BAC or PAC clone using the procedure in this chapter. Typically, size analysis of the DNA from 50 deletion clones by FAGE after *NotI* digestion indicates

that approx 72% are of unique length. An additional 3–8% are found to be unique after end sequencing, suggesting insertions at different sites on genomic DNA as their origin (8). Even more closely spaced deletions are obtainable if one is willing to screen a larger number of clones.

4. Notes

1. It is not necessary to thaw the BAC/PAC glycerol stock stored at -70°C , since the viability falters with repeated freeze-thaw. Scraping off a few specks with a glass capillary from the surface of the frozen stock is sufficient to produce a few hundred colonies. Make sure the agar plates are free of liquid condensate and relatively dry before streaking.
2. Making competent cells from BAC/PAC clones requires that they be grown with good aeration. Using 5-mL cultures in 50-mL tubes in an air shaker bath provides a larger air-liquid interface for the purpose.
3. Minipreps of transposon plasmid DNA are prepared by alkali/sodium dodecyl sulfate lysis and phenol/chloroform extraction and typically produce 2 to 3 μg of plasmid DNA/mL of culture (unpublished observations). The final pellet of plasmid DNA containing excess cellular RNA from a 1.6-mL culture is dissolved in 50 μL of 10 mM Tris-Cl (pH 8.0) plus 0.1 mM EDTA buffer containing 5 $\mu\text{g}/\text{mL}$ of RNase A. This crude preparation of supercoiled transposon plasmid DNA is diluted 10-fold with distilled water, and 4 μL of this diluted stock is used for transforming competent cells made from a BAC or PAC clone.
4. The loxP transposon plasmids introduced into a BAC or PAC clone carry a Kan or Cm resistance gene, respectively, and therefore the selection for transformed colonies from both BACs and PACs can be done on LB plates containing Kan + Cm.
5. Several hundred transposon plasmid-transformed colonies per plate will appear after 16–20 h of incubation. Colonies will be small at this stage because of the double antibiotic selection of somewhat slower-growing BAC/PAC clones, but this does not matter. Pick a swab of colonies (~30–100 colonies of a size that is well represented, i.e., neither the largest nor the smallest ones; *see ref. 11*) from each plate and grow them in 2 mL of LB + Kan + Cm (25 $\mu\text{g}/\text{mL}$ of each antibiotic) overnight at 37°C with vigorous shaking. (Start these cultures at the end of the working day, so they do not remain too long in stationary phase.) Alternatively, one could skip making the overnight cultures, and go directly to doing the inductions with IPTG. In that case, resuspend the mixture (swab) of colonies in 1 mL of warm LB + Kan + Cm (12.5 $\mu\text{g}/\text{mL}$ of each antibiotic) and grow for 1 h. One may need to use 40–100 μL of the mixed culture (depending on cell density) to inoculate 5 mL.
6. Induction of the *tac* promoter with IPTG to express transposase needs to be done during early log phase of cell growth. Measuring absorbance at 600 nm becomes tedious when large numbers of clones are being processed. Watch for a faint silky white appearance when swirling the tube of cells as an indicator of early log phase. The period of cell growth (2 h) to achieve this may vary depending on the BAC or PAC clone. Some large BACs and PACs are notoriously slow growing. Therefore, nested deletions in these should be done separately.

7. P1 vir phage stocks are stored at 4°C in the presence of a drop of chloroform to prevent anything from growing in them. Prior to infecting cells, the phage needs to be thoroughly dechloroformed to ensure good viability of the cells they infect. Approximately 250 μ L of phage suspension is placed in a 50-mL polypropylene tube with several holes in its orange cap and shaken gently for ~45 min at 37°C.
8. The medium for phage adsorption is made 5 mM in calcium chloride by adding an appropriate amount of a filter-sterilized stock solution of 0.5 M calcium chloride.
9. Cultures of BACs and PACs infected with P1 vir phage do not undergo cell lysis well because the host is deficient in certain recombination functions (which are otherwise beneficial for faithful propagation of genomic inserts) and is unable to support efficient growth of the phage (9). Therefore, the chloroform treatment at the end of infection is required to attain lysis. Even then, the suspension does not clear as in lysis with phage λ . Some cell membrane debris is visible when held against the light after chloroform treatment only if the cell density is low. The debris is somewhat stringy in appearance.
10. LCa plates = LB + 2 mM CaCl_2 . (Make sure that the plates are wet and moist. Freshness of plates may not be as important as moisture content.) Phage titers tend to be low when relatively dry plates are used.
11. The magnesium chloride stock of YMC cells is good for up to 2 wk if kept cold.
12. Top agar is conveniently made in a 100-mL Gibco serum bottle and can be used multiple times. Microwave the bottle of solidified top agar each time and shake its contents well to homogenize. Then place the bottle in a heat block set at 50°C for the duration of the experiment.
13. The plate containing YMC cells in top agar overlay should be incubated uninverted; otherwise, the entire overlay will slide out. This precaution is also necessary when growing the P1 phage stocks on plates.
14. Phage plaques are picked using VWR microhematocrit glass capillary tubes. After piercing the top and bottom agar layers of a plaque, the plaque is extruded by pulling out the capillary at an angle with the index finger sealing the top of the capillary. The agar plug containing the plaque is then ejected out of the capillary using a tip attached to a pipetman/Eppendorf pipettor. The 100 μ L of air pumped into the capillary by the pipet is sufficient to squirt the phage plaque into the receiving tube containing YMC cells.
15. The LCa plates need to be prewarmed to 43°C. To prevent them from cooling, place the plates on a foam surface (the lid of a dry-ice shipping container is ideal), not on the cool, stone benchtop, until the top agar layer is poured and the plate put back to incubate.

Acknowledgments

I wish to thank Drs. Ken Harewood, Richard Bukoski, Nancy Shepherd, and Goldie Byrd for encouragement and support. I thank Dr. Rachael Brake and Willie Wilson for testing out the protocol as described and providing valuable feedback. This work was supported in part by a pilot project MBRS grant from

the National Institute of General Medical Sciences (NIGMS) (grant no. GM08049) and supplement funding from the National Cancer Institute (NCI) (grant no. CA16086-24).

References

1. Ahmed, A. (1987) Use of transposon-promoted deletions in DNA sequence analysis. *Methods Enzymol.* **155**, 177–204.
2. Wang, G., Blakesley, R. W., Berg, D. E., and Berg, C. L. (1993) pDUAL: a transposon-based cosmid cloning vector for generating nested deletions and DNA sequencing templates *in vivo*. *Proc. Natl. Acad. Sci. USA* **90**, 7874–7878.
3. Chatterjee, P. K. and Sternberg, N. L. (1996) Retrofitting high molecular weight DNA cloned in P1: introduction of reporter genes, markers selectable in mammalian cells and generation of nested deletions. *Genet. Anal. Biomol. Eng.* **13**, 33–42.
4. Chatterjee, P. K. and Coren, J. S. (1997) Isolating large nested deletions in bacterial and P1 artificial chromosomes by *in vivo* P1 packaging of products of Cre-catalysed recombination between the endogenous and a transposed *loxP* site. *Nucleic Acids Res.* **25**, 2205–2212.
5. Coren, J. S. and Sternberg, N. (2001) Construction of a PAC vector system for the propagation of genomic DNA in bacterial and mammalian cells and subsequent generation of nested deletions in individual library members. *Gene* **264**, 11–18.
6. Chatterjee, P. K., Yarnall, D. P., Haneline, S. A., Godlevski, M. M., Thornber, S. J., Robinson, P. S., Davies, H. E., White, N. J., Riley, J. H., and Shepherd, N. S. (1999) Direct sequencing of bacterial and P1 artificial chromosome nested-deletions for identifying position-specific single nucleotide polymorphisms. *Proc. Natl. Acad. Sci. USA* **96**, 13,276–13,281.
7. Gilmore, R. C., Baker, J., Jr., Dempsey, S., Marchan, R., Corprew, R. N. L., Jr., Byrd, G., Maeda, N., Smithies, O., Bukoski, R. D., Harewood, K. R., and Chatterjee, P. K. (2001) Using PAC nested-deletions to order contigs and microsatellite markers at the high repetitive sequence containing Npr3 gene locus. *Gene* **275**, 65–72.
8. Sternberg, N., Smoller, D., and Braden, T. (1994) Three new developments in P1 cloning: increased cloning efficiency, improved clone recovery and a new P1 mouse library. *Genet. Anal. Tech. Applicat.* **11**, 171–180.
9. Sternberg, N. L. and Shepherd, N. S. (1996) Current protocols, in *Human Genetics* (Dracopoli, N. C., Haines, J. L., Korf, B. R., Moir, D. T., Morton, C. C., Seidman, C. E., Seidman, J. G., and Smith, D. R., eds.), Wiley, New York, pp. 5.3.1–5.3.26.
10. 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.
11. Chatterjee, P. K. and Briley, L. P. (2000) Analysis of a clonal selection event during transposon-mediated nested-deletion formation in rare BAC and PAC clones. *Analyt. Biochem.* **285**, 121–126.

Preparing Nested-Deletion Template DNA for Field Inversion Gel Electrophoresis Analyses and Position-Specific End Sequencing With Transposon Primers

Pradeep K. Chatterjee and Joseph C. Baker, Jr.

1. Introduction

Several protocols for direct sequencing of bacterial artificial chromosome (BAC) and P1 ends have been described recently (*1,2*). Primers located within a BAC or P1 vector have been used to sequence either end of a genomic insert to facilitate selection of minimally overlapping clones for analyzing large chromosomal regions (*1*). The entire insert in a BAC or P1-derived artificial chromosome (PAC) clone, however, is usually sequenced using shotgun procedures coupled with gap filling (*3*). Substantial redundant sequencing of 2-kb sub-clones appears unavoidable in such approaches. In certain circumstances, however, the sequence of only a small portion of a large BAC or PAC clone is required. Situations such as these arise either when known genetic markers need to be mapped or when new ones need to be identified in a small region of a large clone. Nested deletions become particularly useful in these applications.

Chromosomal regions high in repetitive sequence content also require iterative cycles of directed sequencing to overcome ambiguity in aligning sequence reads. Sequence contigs assembled from these regions are short, and it is therefore desirable to develop alternative approaches that can independently order such contigs and estimate the size of gaps between them. A dense, evenly spaced nested-deletion series generated in BACs/PACs spanning a region of interest appears ideally suited for this purpose. In addition, sequences from

*From: Methods in Molecular Biology, vol. 255: Bacterial Artificial Chromosomes, Volume 1:
Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

ends of deletion clones that have been sized by field inversion gel electrophoresis (FIGE) offer unique coordinates for ordering contigs and estimating gap lengths in areas of the genome high in repetitive DNA (4).

Usually two criteria need to be met for such applications: a high-quality sequence from a BAC/PAC deletion end, and a good-size estimate of the deletion clone so as to position the end-sequence read relative to another along the original BAC or PAC clone. Both results rely heavily on the quality of genomic clone DNA that is used. In this chapter, we describe three different but complementary procedures to prepare DNA from BAC/PAC nested deletions that are suitable for FIGE after *NotI* digestion. A subset of these procedures yields DNA that is also suitable for obtaining high-quality end sequences from deletion clones using a set of primers from the transposon end. The procedures range in throughput from low to high, with the quality of sequence reads correlating more or less inversely to throughput.

2. Materials

All chemicals, unless specified, should be at least molecular biology grade. Solutions should be made with deionized water and autoclaved or sterile filtered through a 0.2- μ Nylon Membrane Disposable Sterile Syringe Filter (Corning, Corning, NY).

1. All antibiotic stock solutions are made in ultrapure distilled water from Gibco-BRL (Life Technologies, Gaithersburg, MD). Antibiotic stock solutions are of the following concentrations (reagents are biotech research grade from Fisher): 10 mg/mL of ampicillin (penicillin-G potassium), 10 mg/mL of kanamycin (kanamycin monosulfate), and 2.5 mg/mL of chloramphenicol. Store reagent stock solutions in water frozen at -20°C .
2. BigDye Terminator mix (PE Biosystems).
3. Bacto agar, Bacto Tryptone, and yeast extract (Difco, Beckton Dickinson, Sparks, MD).
4. Cotton gauze for crude filtration of flocculant precipitate (Kendall Healthcare, Mansfield, MA).
5. FIGE Mapper Electrophoresis System (Bio-Rad, Hercules, CA).
6. Ultrapure agarose and 10X TBE buffer (Gibco-BRL, Life Technologies).
7. Ethidium bromide (EtBr), 6X loading dye (Sigma, St. Louis, MO).
8. Phenol (Gibco-BRL) was saturated with 50 mM Tris-Cl (pH 8.0) buffer containing 1 mM EDTA. Mix the buffer-saturated phenol with an equal volume of chloroform to give a 50:50 phenol/chloroform mixture. Store in a dark-colored bottle at 4°C .
9. Restriction enzymes, including *NotI* (New England Biolabs).

3. Methods

3.1. Miniprep Plasmid DNA From BAC/PAC Deletion Clones

The standard alkali/sodium dodecyl sulfate (SDS) lysis miniprep method (5,6) used for isolating DNA of small plasmids (between 5 and 10 kbp) with high copy number has been modified and adapted for single-copy large plasmids such as BACs and PACs (7,8). The procedure is simple, rapid, and of moderate throughput. DNA from about 120 deletion clones can be isolated and analyzed by FIGE in 2 to 3 d. It does not require any reagent kits or specialized incubator-shakers for growing cultures. BAC/PAC deletion clone DNA obtained by this procedure gives complete digests with *NotI* and several other restriction enzymes, and it is most suitable for subsequent FIGE analyses. The major drawback of the procedure appears to be its inability to produce good sequence reads reliably from BAC/PAC deletion ends.

1. Inoculate 2-mL cultures of Luria Bertani (LB) containing kanamycin and chloramphenicol (12.5 µg/mL each) with a single colony of a BAC/PAC deletion clone in 15-mL polypropylene snap-cap Falcon tubes. Leave the caps loose for good aeration. Grow overnight at 37°C with vigorous shaking (250 rpm) (*see Notes 1–3*).
2. Decant approx 1.7 mL of saturated culture into Eppendorf tubes. Leave the remaining 0.2 mL of culture in snap-cap tubes (sealed shut) at 4°C. Make sure that each is numbered, along with the Eppendorf tubes (*see Note 4*).
3. Spin the Eppendorf tubes in a microfuge at room temperature for 1 min at 10,000 rpm.
4. Decant the supernatant immediately, leaving behind the last 50 µL of medium.
5. Add 150 µL of a solution of 25 mM Tris-HCl, 50 mM glucose, and 10 mM EDTA (solution I; *see ref. 6* for preparation procedure); vortex gently (setting of 5 to 6) to resuspend the cell pellet thoroughly; and leave at room temperature. This should be done as quickly after pelleting as possible.
6. Set up an ice-water tray. Make up a fresh solution of 0.2 N NaOH + 1% SDS (solution II) (*see Note 5*).
7. Quickly add 310 µL of alkali/SDS solution to a set of 10 tubes at room temperature. Close the caps, and gently mix the contents of the tubes. Once suspension becomes clear (cell lysis is complete in about 30 s), immediately place the tubes in an ice-water tray (*see Note 6*).
8. After 5 min on the ice-water tray, open the caps of the tubes and add 250 µL of an ice-cold solution of KOAc/HOAc (solution III, *see Note 7*, and also *ref. 6* for preparation procedure) to neutralize. Close the caps securely, check that all tubes have equal volumes, and mix the contents of the tubes thoroughly before putting them back into the ice-water tray. You may take a break here and store the tubes in a cold room.

9. Spin the tubes (after wiping away moisture on the outside of the tubes) in a microfuge for 4 min at 10,000 rpm.
10. Carefully decant the supernatants into fresh numbered tubes. The supernatants contain the BAC/PAC plasmid DNA along with cellular RNA and small quantities of protein.
11. Add 200–300 μL of a 50:50 mixture of phenol/chloroform (the bottom layer from the bottle containing buffer-saturated phenol/chloroform) to each tube. Use a fume hood for the procedure. Cap tightly. Shake by gently inverting the rack of tubes back and forth (place an empty rack on top) for 1 to 2 min to mix the two immiscible layers.
12. Spin the tubes in a microfuge at room temperature for 2 min at 10,000 rpm.
13. Carefully recover the top layer (avoid any phenol/chloroform) and put in new numbered tubes.
14. Add isopropanol to fill the tubes. If recovery is good at each step, there should be about 750–800 μL of aqueous layer in each tube after the phenol/chloroform extraction, and thus an equal volume (800–900 μL) of isopropanol should be ideal to precipitate all nucleic acids.
15. Cap the tubes tightly and mix gently by inverting the tubes back and forth several times. Let stand at room temperature for 20 min. You may take a break here for the day (*see Note 8*).
16. Spin the tubes for 5 min at 14,000 rpm in a microfuge at room temperature. Discard the supernatant.
17. Add 1 mL of a solution of 70% ethanol in water to each tube. Wash the pellet by inverting the securely capped tubes several times. The pellet will dislodge.
18. Spin the tubes at 14,000 rpm for 2 min. Carefully discard the supernatant—the pellet is loose!
19. Remove as much of the wash solution as possible, and then tap the pellet to the bottom of the tube. This allows the dried pellet to be dissolved in a small volume of buffer.
20. Leave the pellet to air-dry overnight with the caps open. Cover with tissue paper.
21. Resuspend each pellet in 40 μL of a solution of 10 mM Tris-Cl (pH 8.0) + 0.1 mM EDTA + 5 $\mu\text{g}/\text{mL}$ of RNase A. Gently tap to mix after 15 min of incubation at room temperature. Spin lightly to get the liquid at the bottom. The BAC/PAC DNA should be resuspended and ready to use in 30 min. Aliquot 4 to 5 μL out of each for *NotI* digest and FAGE analysis. Save the remainder at -20°C (*see Note 9*).

3.2. Qiagen Tip Purification of BAC/PAC Deletions

This method produces the highest-quality DNA from BACs and PACs, and superior results are obtained for both FAGE analyses and fluorescent sequencing. The major drawback is its low throughput.

1. Streak out BAC/PAC deletion clone from a glycerol stock onto LB plates containing 25 $\mu\text{g}/\text{mL}$ each of kanamycin and chloramphenicol.

2. Pick a single colony and inoculate 85 mL of LB (enriched with yeast extract, 5 g/L) containing 12.5 µg/mL each of kanamycin and chloramphenicol. Or inoculate with a 1-mL overnight culture. Use 250-mL Kimax (Kimble) culture flasks with indentations on the bottom for superior agitation.
3. Shake vigorously in an air-shaker bath overnight.
4. Fill two Oak Ridge tubes (each can hold 40 mL) up to the mark on the neck. Cap the tubes tightly. Save the remaining 3–5 mL of culture in the flask in the cold room to make glycerol stock, on the same day.
5. Spin the tubes in an SA-600 rotor in a Sorvall centrifuge at 4°C and 3600g with the brakes on for 10 min.
6. Pour off the clear supernatant (bleach the pooled supernatants later to kill bacteria), and add 6 mL of Qiagen buffer P1 (make sure RNase A is included) to each tube.
7. Cap the tubes tightly and vortex (at a setting of 6) until all of the cell pellet is thoroughly resuspended. Leave the tubes at room temperature.
8. Add 6.5 mL of P2 buffer to a separate 15-mL white plastic tube (keep P2 buffer on a foam lid to avoid precipitation of SDS/NaOH over cool bench surface!).
9. Swirl the tube of cell suspension and quickly add the alkali/SDS (P2 buffer) all at once. Roll the tube after capping tightly. Be gentle, so as not to shear chromosomal DNA! Let sit at room temperature for 3 to 4 min (*see Note 10*).
10. Once the contents of the tube become clear (cell lysis is complete), place in an ice bath.
11. Add 6 mL of cold buffer P3 (KOAc + HOAc) and mix gently by inverting several times. Leave on ice for 20 min, mixing well by inverting every 5 min. Combine the contents of two tubes into one for each clone.
12. Spin in an SA-600 rotor at 4°C for 30–45 min at 9300g *without* brakes.
13. Get cotton gauze ready. Cut each piece into two.
14. The precipitate is light and some of it tends to float. Therefore, pour off and collect the supernatant through the cotton gauze into a clean Oak Ridge tube. If flocculant precipitate spills over, make a note of it and avoid loading the precipitate later into the Qiagen columns (tips). This precipitate does not dissolve in the QBT buffer used to dissolve the DNA pellet in a subsequent step (*see Note 11*).
15. Add $\frac{1}{20}$ vol of sterile filtered 3 M sodium acetate (approx 2 mL), mix by swirling, and divide into two Oak Ridge tubes. Add an equal volume of room temperature isopropanol. Cap tightly and mix well at room temperature by inverting.
16. Spin the tubes in an SA-600 rotor at 8000 rpm for 15 min at 4°C without brakes. Pour off the supernatant, and invert the tubes on a paper towel to drain the liquid completely. It is *not* necessary to dry the pellets.
17. Dissolve the wet pellet in each tube immediately in 5 mL of QBT (column equilibrating) buffer.
18. Set up Qiagen-tip 500 columns in 250-mL Erlenmeyer flasks with the blue plastic collar provided by the vendor. Mark each column clearly with the clone number. Once the DNA pellets are dissolved, equilibrate each column with 6 mL of buffer QBT.

19. Once the buffer has completely drained out of the tip, add 2×5 mL of the DNA solution from the two tubes into each column.
20. Wait for the liquid to drain through completely. Wash the column with 3×14 mL of wash buffer QC. Each time let the liquid drain through completely.
21. Make up Special Elution buffer. Take 3×5 mL of elution buffer QF for each column (i.e., $12 \times 15 = 180$ mL for eluting 12 tips), and add 1 mL of 5 M NaCl for each 35 mL of elution buffer (i.e., add $5\frac{1}{2}$ mL of 5 M NaCl to the 180 mL of QF buffer). Mix well, and microwave for 1 min on the highest setting. Check the temperature by inserting a thermometer into the solution, and microwave (after removing the thermometer) a few seconds more to get up to 70°C. Wrap the bottle in paper towels and aluminum foil, and put it in a heat block set at 85°C.
22. After the third 14-mL wash in QC buffer, remove the Qiagen tips from the flasks and mount on clean Oak Ridge tubes held upright on a foam rack. Apply three 5-mL aliquots of the hot elution buffer directly to the columns (*see Note 12*). Collect the flowthrough in the tubes, and precipitate the DNA with the addition of 12 mL of isopropanol at room temperature to each tube. Mix well by inverting the capped tubes several times.
23. Spin the tubes after 20 min of incubation at room temperature at 8000 rpm for 30 min at 4°C in an SA-600 rotor. Note the position on the tubes where the DNA pellet is expected.
24. Decant the supernatant and invert the tubes on a paper towel. After half an hour, wipe off the lips with a Kimwipe, and add 750 μ L of a 0.3 M sodium acetate solution in 10 mM Tris-Cl (pH 8.0) and 0.05 mM EDTA. (The DNA pellet need not be completely dry.)
25. Dissolve the DNA by swirling, spin the tubes at 5000 rpm for 2 min, and recover the solution into numbered Eppendorf tubes. Wash the Oak Ridge tubes with another 100 μ L of the same buffer, and add that to the Eppendorf tubes. The tubes should now have about 850 μ L of aqueous phase containing the DNA.
26. Extract with 300 μ L of a 25 mM Tris-Cl + 0.1 mM EDTA buffer-saturated 50:50 phenol/chloroform mixture. Collect the aqueous phase in a new tube.
27. Extract with 300 μ L of pure chloroform.
28. The volume of aqueous phase should be between 800 and 850 μ L. Fill the tube with isopropanol at room temperature. Cap the tube and mix. If the preparation is good, a good fibrous precipitate of pure BAC/PAC deletion clone DNA should be visible.
29. Spin the tube at room temperature in an Eppendorf centrifuge at 20,000g for 5 min. Decant the supernatant. Invert the tube on a paper towel.
30. Add 1.2 mL of 70% ethanol at room temperature. Cap and invert the tube several times to wash the pellet of excess salt, spin for 2 min, and decant the supernatant. (Careful: The pellet is loose.)
31. Let the pellet dry thoroughly (a couple of hours at least). Dissolve the pellet in 100–200 μ L of $\frac{1}{5}$ X TE buffer for BigDye Terminator fluorescent sequencing (*see Note 13*). The DNA pellet should be incubated at room temperature for at least

half an hour in buffer to dissolve. Gently tap and mix the contents of the tube several times during this period. The DNA concentration will be approx 0.5–1 $\mu\text{g}/\mu\text{L}$ and should be accurately determined by absorbance at 260 nm.

3.3. Qiagen R.E.A.L. Prep Kit Isolation of DNA

This procedure allows the highest throughput of samples and produces DNA that is suitable for analysis by both FIGE and fluorescent sequencing. The quality of sequence reads is not as high as with the Qiagen tip method: the signal intensities are approximately one-third of those obtained with the Qiagen tip procedure (9). The read lengths are similar by both methods.

Although *NotI* digests are complete for BAC/PAC deletion clone DNA obtained by this procedure, the pattern of DNA bands in FIGE gels is less sharp when compared with those obtained from the miniprep procedure. The Qiagen R.E.A.L. Prep Kit method of clone DNA preparation is useful when reliable but lesser-quality sequence reads and less-sharp images of bands in FIGE gels are tolerable during rapid analysis of large numbers of deletion clones (200–1000 clones).

BAC/PAC deletion clones are grown in 96-well blocks. Colonies are grown in 1.2 mL/well of LB (enriched with 5 g/L of yeast extract). Vigorous shaking is achieved in a special platform incubator-shaker. The host strain NS3516 containing the deletion clones grows consistently better than the DH10B host of BACs and PACs under these conditions (unpublished observations). The DNA-isolating procedure is exactly that described by Qiagen. An additional wash with 70% ethanol after the last step produces better sequence reads. The final pellet is dissolved in 35 μL of $\frac{1}{5}$ TE (2 mM Tris-HCl [pH 7.4] + 0.2 mM EDTA). FIGE analysis after *NotI* digestion is done with 5 μL of the sample. Fluorescent sequencing is performed with 12 μL of the sample in a 30- μL sequencing reaction.

3.4. *NotI* Digestions

Prepare a “stock reaction mix” that is 2X in restriction enzyme reaction buffer including bovine serum albumin (BSA) (reaction buffer) and contains 3 to 4 U of *NotI* enzyme per 5 μL of volume.

1. Thaw 10X reaction buffer and 100X BSA at room temperature. Multiply the number of enzyme digests (samples) by 5 to obtain the total volume of reaction mix required.
2. This volume of buffer needs to be 2X in *NotI* digestion buffer. Therefore, add the required volume of water to an Eppendorf tube and put on ice.
3. Add the required volume of 10X NE-buffer 3 and BSA after thawing (and mixing thoroughly).
4. After the 2X reaction buffer has been chilled in an ice-water bath (keep tubes capped tightly), add the calculated total volume of concentrated *NotI* enzyme-

(usually 10 U/ μ L), so that every 5 μ L of reaction mixture stock has 3 to 4 U of the enzyme. Keep on ice.

5. Mix thoroughly by tapping, but avoid creating bubbles. The restriction enzyme is usually stored in a high concentration of glycerol, and thus needs to be uniformly distributed in the reaction mix. Chill on ice in between.
6. Aliquot 5 μ L of enzyme mix to approx 10 tubes at a time kept at room temperature, and add deletion clone DNA directly to the enzyme droplet. The reaction mix tube with the enzyme should be on ice as much of the time as possible (*see Note 14*).
7. Mix the contents of each tube gently by tapping, spin in a microfuge for 30 s at 10,000 rpm, and incubate in a warm room at 38°C. Incubate for 2 to 3 h.
8. Add 3 μ L of 6X loading dye, and gently mix the contents of each tube by tapping.
9. Analyze the digested samples on FIGE gels. These gels are 1% agarose in 0.5X TBE.

3.5. FIGE Analysis

Each FIGE gel requires 150 mL of molten 1% agarose that is 0.5X in TBE.

1. Weigh out 1.5 g of ordinary agarose in a 200-mL orange-cap Pyrex bottle. Add 135 mL of distilled/deionized water. Melt in a microwave. *Be careful not to spill over.*
2. Once agarose is thoroughly melted, add 15 mL of 5X TBE slowly around the inside wall of the lip of the bottle. This is to prewarm the buffer and prevent agarose from solidifying on contact with it. Swirl well to mix. You may cool it in a water bath to approx 60°C.
3. Carefully tape a FIGE gel tray and set it on the laboratory bench. Set a 30-well comb in place.
4. Pour the agarose continuously and *not* in one spot (to prevent uneven heating of the tray), and avoid air bubbles (if they do form, take a pipet tip and bring it to the side or end). Let the agarose solidify at room temperature for 1 h.
5. Make up 2 L of 0.5X TBE buffer (for each gel) to run the gels.
6. Gently remove the comb. Then remove the tape. Set up the gel in the FIGE electrophoretic box. Fill with buffer slightly below the required level (marked). Start the circulation pump set to level 6. Let the circulation continue for a couple of minutes to reach equilibrium.
7. Stop the circulation pump. Load the samples without introducing bubbles. Start the program (Bio-Rad preprogram no. 6) for electrophoresis, and then restart the circulation pump (*see Notes 15 and 16*).
8. At the end of the run, take out approx 200 mL of the used electrophoresis buffer in a glass tray, add 20 μ L of concentrated EtBr stock solution (10 mg/mL), mix well, and then add the gel into the tray. Stain for 1 h by rocking the tray gently at room temperature and then add 500 mL of water to the tray. Take a picture after another 0.5 h at room temperature. You can leave the gel in the tray with buffer/water/stain overnight in a cold room.

3.6. Fluorescent Sequencing of BAC/PAC Deletion Ends Using Transposon End-Based Primers

One end of the loxP transposon used to create the nested deletions is left behind in the deletion clones. Primers from the transposon end suitable for sequencing the ends of insert DNA in these deletions have been described earlier (9).

BigDye Terminator-based fluorescent sequencing reactions are set up in 96-well MicroAmp reaction plates (PE Applied Biosystems, Foster City, CA). Each sequencing reaction is 30 μ L in total volume and contains 16 μ L of BigDye Terminator mix, 2–4 μ g of BAC/PAC deletion clone DNA template, and 16 pmol of primer (*see* **Notes 17** and **18**).

Cycling conditions are as follows: initial denaturation at 95°C for 5 min, followed by 50 cycles of 95°C for 30 s, 50°C for 30 s, and 60°C for 4 min. After standard cleanup to remove unincorporated nucleotides, the entire sample is loaded onto gel.

Note Added in Proof

A Magnesium based clean up procedure for sequencing reactions containing BAC and PAC templates and suitable for capillary DNA sequencers has recently been described in Gernon, A., Woldu, E., Godlevski, M., Wilson, W., Gilmore, R. C., Grant, D. J., Chatterjee, P. K., and Kephart, D. (2003) Automated Purification of dye terminator sequencing reactions: an approach to high throughput capillary electrophoresis sequencing of large templates. *J. Assoc. Lab. Automation*, in press.

4. Notes

1. Grow BAC/PAC deletion clone cultures with good aeration. More than 2 mL of culture per tube is therefore not desirable.
2. Note that although PAC clones contain an additional lytic replicon that is inducible with isopropyl- β -D-thiogalactopyranoside (IPTG) to produce greater quantities of PAC DNA per cell, this approach is seldom used when DNA from large numbers of clones is being prepared. In addition, the significantly greater cell density attainable in overnight cultures without induction somewhat compensates for the need to induce with IPTG.
3. Use of richer medium is not recommended here, because the capacity of the Eppendorf tubes (1.6 mL) is overwhelmed in subsequent steps when processing a substantially higher number of cells.
4. Aliquots of BAC/PAC cultures saved at 4°C are viable for about 1 wk. This time interval is sufficient to prepare and determine the DNA size from several hundred BAC/PAC deletion clones by FIGE so as to decide which ones to save.

5. Addition of 10% SDS to 10 *M* sodium hydroxide will lead to precipitation. Therefore, add SDS to a larger volume of water before adding alkali.
6. Vigorous shaking must be avoided after cell lysis with alkali-SDS so that shearing of chromosomal DNA does not occur. Contaminating chromosomal DNA pieces are a hindrance to obtaining good sequence reads.
7. Solution III is made by mixing 60 mL of 5 *M* potassium acetate, 11.5 mL of glacial acetic acid, and 28.5 mL of water.
8. Isopropanol precipitation of nucleic acids (BAC/PAC + cellular RNA) needs to be done at room temperature and not in the cold so as to avoid precipitating excess salt and proteins.
9. It is important to save BAC/PAC DNA obtained from the miniprep procedure at -20°C . The preparation is relatively crude, and prolonged storage at 4°C leads to degradation of the large plasmid DNA as judged by FIGE analysis. DNA isolated by the Qiagen tip method is much purer and can be stored at 4°C for longer periods of time (several months).
10. Alkali/SDS lysis of cells at high density is most efficient when all cells are exposed to reagent simultaneously. Thus, the measured aliquot of alkali/SDS solution is taken in a separate tube and added all at once to the cell suspension kept swirling gently. Once added, the tube containing cells is rocked very gently while rolling it slowly for thorough mixing of contents without shearing of chromosomal DNA. Pipetting the alkali/SDS solution gradually lyses cells on the periphery, and DNA from the lysed cells makes the remaining cells in the suspension impervious to fresh reagent.
11. A convenient way to filter the supernatant is to tightly wrap the cotton gauze over the mouth of the Oak Ridge tube containing solution and precipitate, and slowly pour out the liquid at a low angle. Care should be taken not to block the entire mouth of the tube with liquid while pouring. Otherwise, air will bubble into the tube to occupy displaced liquid, resulting in stirring up of the light precipitate.
12. The flow of warm elution buffer through the Qiagen tip stops occasionally owing to moisture condensing underneath the tip on the outside. The liquid forms an airtight seal with the mouth of the Oak Ridge tube. Wipe the bottom of the tip if necessary.
13. The final BAC/PAC DNA pellet is dissolved in a very low concentration of Tris-Cl (pH 7.4) plus EDTA buffer (2 *mM* Tris-Cl and 0.2 *mM* EDTA). Higher concentrations of Tris-Cl are thought to be detrimental for the BigDye Terminator sequencing reaction.
14. Five microliters of clone DNA from either the miniprep or the R.E.A.L. kit procedures is optimal for digestion with *NotI* and FIGE analyses. DNA from the Qiagen tip procedure is very pure and concentrated, and only 1 to 2 μL of a 10-fold dilution of that DNA should be used.
15. Program 6 set by Bio-Rad gives best resolution of DNA bands in the size range of 5–100 kb in a 1% agarose gel. DNA fragments larger than 100 kb tend to cluster and do not resolve under these conditions.

16. Molecular weight standards can be purchased from Bio-Rad. The 5- to 200-kb ladder is most useful in analyzing BAC/PAC nested deletions.
17. DNA templates prepared by the Qiagen tip method produce the best signal intensities and increase linearly with concentration of template DNA until about 4 to 5 $\mu\text{g}/30\ \mu\text{L}$ reaction. Thus, 4 $\mu\text{g}/30\text{-}\mu\text{L}$ reaction is used for 100-kb BAC/PAC deletion templates, and 2 $\mu\text{g}/30\text{-}\mu\text{L}$ reaction is used for those below 30 kb. Signal intensities of 1200–1400 on the Applied Biosystems scale are routinely obtained following these procedures with the Qiagen tip purified DNA templates (9).
18. When analyzing large numbers of samples, there arise situations when the origin/identity of a deletion clone is brought into question. The easiest way to track down the paternity of a deletion clone is to sequence the other end of the insert. The insert DNA end opposite the loxP site remains identical in all deletion clones and the parent. Thus, primers SP6II and T7L (described in ref. 9) can be used to sequence such an end in BACs and PACs, respectively.

Acknowledgments

We wish to thank Michele Godlevski, David Yarnall, Steve Haneline, and Simon Thornber for helpful discussions. We also thank Drs. Ken Harewood, Richard Bukoski, Nancy Shepherd, and Goldie Byrd for support and encouragement. This work was supported in part by a pilot project MBRS grant from National Institute of General Medical Sciences (NIGMS) (grant no. GM08049) and supplement funding from National Cancer Institute (NCI) (grant no. CA16086-24).

References

1. Kelley, J. M., Field, C. E., Craven, M. B., Bocskai, D., Kim, U. J., Rounsley, S. D., and Adams, M. D. (1999) High throughput direct end sequencing of BAC clones. *Nucleic Acids Res.* **27**, 1539–1546.
2. Boysen, C., Simon, M. I., and Hood, L. (1997) Fluorescence-based sequencing directly from bacterial and P1-derived artificial chromosomes. *Biotechniques* **23**, 978–982.
3. Dunham, I., Shimizu, N., Roe, B. A., et al. (1999) The DNA sequence of human chromosome 22. *Nature* **402**, 489–495.
4. Gilmore, R. C., Baker, J., Jr., Dempsey, S., et al. (2001) Using PAC nested-deletions to order contigs and microsatellite markers at the high repetitive sequence containing Npr3 gene locus. *Gene* **275**, 65–72.
5. Birnboim, H. C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513–1523.
6. (1989) Small-Scale Preparations of Plasmid DNA. *Molecular Cloning: A Laboratory Manual*, 2nd ed. (Sambrook, J., Fritsch, E. F., and Maniatis, T., eds.), Cold Spring Harbor, Woodbury, New York, pp. 1.25–1.28.

7. Chatterjee, P. K. and Sternberg, N. L. (1996) Retrofitting high molecular weight DNA cloned in P1: introduction of reporter genes, markers selectable in mammalian cells and generation of nested deletions. *Genet. Anal. Biomol. Eng.* **13**, 33–42.
8. Chatterjee, P. K. and Coren, J. S. (1997) Isolating large nested deletions in bacterial and P1 artificial chromosomes by *in vivo* P1 packaging of products of Cre-catalysed recombination between the endogenous and a transposed *loxP* site. *Nucleic Acids Res.* **25**, 2205–2212.
9. Chatterjee, P. K., Yarnall, D. P., Haneline, S. A., et al. (1999) Direct sequencing of bacterial and P1 artificial chromosome nested-deletions for identifying position-specific single nucleotide polymorphisms. *Proc. Natl. Acad. Sci. USA* **96**, 13,276–13,281.

BAC Finishing Strategies

Christine Bird and Darren Grafham

1. Introduction

The goal of the International Human Genome Project is to generate a reference sequence of the human genome with an accuracy of <1 error in 10,000 bases. The approach chosen by the International Human Genome Sequencing Consortium is a hierarchical mapping and sequencing strategy similar to that described for other large genomes, including *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Arabidopsis thaliana* (1–3). A physical map of the human genome has been assembled on the basis of “fingerprints” generated by restriction digestion of individual bacterial clones and mapping of individual clones to chromosomes using landmarks from existing genetic maps. At Washington University, St. Louis, more than 300,000 clones from the whole human genome bacterial artificial chromosome (BAC) libraries RPCI-11 and RPCI-13 from P. de Jong (4) were fingerprinted, and the resulting maps of these overlapping clones have been integrated with clone maps of individual chromosomes generated by other centers (5). Sets of minimal overlapping clones have been selected for sequencing from the resulting maps.

The sequencing process takes place in two stages. An initial random “shot-gun” phase entails generation of sequence reads of between 400 and 500 bases in length from 1- to 2-kb fragments for each bacterial clone subcloned into a double-stranded vector such as pUC18 or single-stranded M13 vector. Typically, random sequence is generated to deliver an average sequence depth across each bacterial clone of between 6- and 10-fold. At this stage the sequence reads will normally assemble into fewer than 20 pieces, or contigs, and $>95\%$ of each clone will be covered by sequence reads. Each clone is then subjected to a directed “finishing” phase, which involves an interactive process

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

of selection of additional sequencing reads to fill gaps and resolve ambiguities until the clone is covered with contiguous sequence of a defined accuracy agreed to by the International Human Genome Sequencing Consortium. Up-to-date information on the criteria used to define the finished sequence product can be found at <http://genome.wustl.edu/gsc/glbstand.php>.

To accelerate the production of a draft sequence of the entire human genome, the initial shotgun phase was halted when the average sequence depth of each clone reached fourfold coverage in Phred Q20 bases, i.e., bases that have a probability of at least 99% of being correct (6–8). By June 2001, the shotgun depth was increased to 6- to 10-fold coverage in preparation for the final finishing, which was completed in April 2003.

The efficiency of finishing is dependent on a number of variables. The quality of the shotgun sequence data has a profound effect on the quality and type of finishing required to finish the sequence of a BAC. Typically, the BAC libraries generated for the Human Genome Project contain insert sizes of between 150,000 and 220,000 bases. At the Sanger Centre each clone is sonicated to create fragments of 1400–2000 bases that are subcloned into the pUC18 sequencing vector, and sequence reads are generated from each end of the inserts using standard “universal” priming sites at positions –20 on the positive strand and –48 on the negative strand. Knowledge of the pairs of sequences (read pairs) generated from each insert can be used to provide ordering information in clone assemblies, and the DNA is available for further sequencing reads if paired reads span an assembly gap. Once a clone is covered with a depth of sequence that is on average between 6- and 10-fold across the clone insert, and the sequence reads are at least 500 bases in length with a Phred quality of at least 20, further shotgun data contribute little to the contiguation of the sequence. If sequence reads are shorter than 500 bases, some additional contiguation may be gained by increasing the depth of the shotgun sequence.

To finish a BAC clone, the first priority is to contiguate the insert sequence. When overlapping clones are being finished, the overlap region is only finished in one of the clones, and the sequence is annotated appropriately to reflect this when submitted to the public databases. The sequence is contiguous when the full length of the insert being finished has been assembled into one piece. At this stage, not all of the sequence reads will have been incorporated, but all of the major contigs will generally have been joined together correctly in the database. Major contigs have a length of >1000 bases and are composed of multiple overlapping sequence reads.

Every base in the sequence should be covered by sequence reads from two or more separate DNA templates. These two templates can be either sequence reads obtained in both directions, or one read being determined using terminator sequencing chemistry. For a sequence to be finished, it will be contiguous

with no unresolved bases, and it will be completely double covered. Furthermore, the sequence assembly will have been confirmed by comparison with a restriction digest of the clone.

Because all sequences are different, the process of finishing must be tailored to each individual BAC. The strategy employed for finishing is subjective and varies among sequencing centers. In this chapter, we do not, therefore, aim to cover all approaches and variables, but to set out an approach to finishing that is used at the Sanger Centre. Finishing is an iterative process, and once a new set of data has been generated, their contribution to resolving the sequence has to be assessed before further reactions are attempted. The skill of finishing lies in learning which approach to attempt first and then adjusting the strategy based on the reasons for success or failure of this directed sequencing.

In this chapter, we outline the equipment and protocols employed for finishing at the Sanger Centre. We also describe the various stages of the finishing process and highlight the iterative and interactive process among finisher, database, and wet laboratory work that turns a shotgunned clone into an accurate finished sequence.

2. Materials

2.1. Equipment

1. Sun Ultra 10 or equivalent with a 440-MHz processor and 1 gigabyte of memory, for assembling a single BAC database using the programs Phrap and GAP4.
2. ABI Prism™ 377 DNA Sequencer or PE3700 Capillary DNA Sequencer, for fluorescent sequencing (PE Applied Biosystems, Foster City, CA).
3. PTC-225 96-well DNA Engine Tetrad Thermal Cyclers, for sequencing and creating polymerase chain reaction (PCR) products (MJ Research, Watertown, MA).
4. Skirted Thermo-Fast® 96-well plates suitable for use with the PE3700 (ABgene).
5. Polypropylene plate lid, for storing 96-well plates (ABgene).
6. Hybaid Omniseal TD Mats, to seal wells during cycling (Hybaid).
7. Centrifuge capable of spinning 96-well microtiter dishes.

2.2. Reagents

1. 10X Reaction buffer: 40 mL of 1 M Tris-HCl, pH 9.0; 0.7 mL of 1 M MgCl₂; 59.3 mL of H₂O.
2. BigDye™ Terminator Cycle Sequencing Ready Reaction (8000 µL) (PE Applied Biosystems). Store at -20°C; light sensitive.
3. dGTP BigDye™ Terminator (800 µL) (PE Applied Biosystems). Store at -20°C; light sensitive.
4. SequenceR_x Enhancers A, E, F (3 × 100 µL) (Gibco-BRL, Life Technologies), Frederick, MD). Store at -20°C.
5. BigDye™ Primer (PE Applied Biosystems). Store at -20°C; light sensitive.

6. Custom primers (12 pmol/μL) (Sigma-Genosys). Store at -20°C once resuspended in water.
7. Universal pUC18 Sequencing Primer (-20) (17mer) 5'd(GTAAAACGACGGC CACGT)3' (Sigma-Genosys). Store at -20°C.
8. Universal pUC18 Reverse Sequencing Primer (-28) (24mer) 5'd(AGCGGATAA CAATTTCACACAGGA)3' (Sigma-Genosys). Store at -20°C.
9. DNA source subcloned into the pUC18 sequencing vector (pUC18 Ready-to-use vector cut by *Sma*I and dephosphorylated) (cat. no. CVPSM028; Q-Biogene, UK). Store at -20°C.
10. BAC DNA. Store at -20°C.
11. Ready-to-go™ PCR beads, 96 reactions (Amersham Pharmacia Biotech, Piscataway, NJ). Store dry at room temperature.
12. 60% Propan-2-ol (isopropyl alcohol).
13. 80% Propan-2-ol (isopropyl alcohol).
14. 100% Ethanol. Store at -20°C in a sparkproof freezer ready for use.
15. 3 M Sodium acetate, pH 4.8.
16. Dimethylsulfoxide (DMSO) (Sigma-Genosys). Store at -20°C.
17. ddH₂O, autoclaved.

2.3. Software

1. Phred base-caller (www.phrap.org).
2. Database assembly engine Phrap (www.phrap.org) or CAP2 (www.mrcimb.cam.ac.uk/pubseq/).
3. Finishing software capable of handling the data set size of a BAC such as GAP4 (www.mrc-lmb.cam.ac.uk/pubseq/) or Consed (www.phrap.org/consed/consed.html).

3. Methods

3.1. Database Assembly

The raw data from a sequencing machine (9) are processed using the Phred software (6,7). Phred is a base-calling program used as an alternative to the standard ABI base-calling software (10). Phred generates a base-quality index for each base it calls, indicating the probability of the call being correct. The assembly program Phrap can make use of these quality measures, both in assembly and when assessing the quality of the contigs' consensus sequence.

The processed reads are collated in a directory in which the assembly engine Phrap is run to produce a database that is accessible using an editor program such as GAP4 (11,12) or Consed (13).

3.2. Interacting With the Database

GAP4 and all the necessary hardware are covered in detail in ref. 12. The principles of finishing described in the rest of this chapter are not limited to one

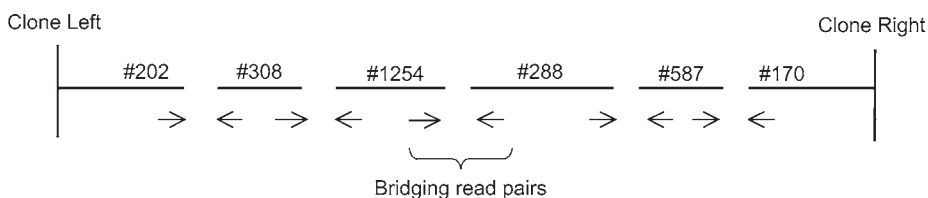


Fig. 1. Stick diagram representing reassembled BAC insert with ordered major contigs (contigs numbered as in GAP4, e.g., #170) and known bridging read pairs.

particular assembly package but are specific to the Sanger Centre. The finishers at the Sanger Centre use GAP4 from the Staden Package as an interactive viewer of database assemblies, but many of the partners in the Human Genome Project use Consed. These programs have evolved in parallel, and the most important feature that they share is the calculation of a base-quality index for each base, also known as a confidence value. The use of confidence values has been invaluable in speeding up the finishing process. The finisher is no longer required to correct every single base of an individual sequence read that disagrees with the consensus. GAP4 uses a probabilistic consensus algorithm, which utilizes the Phred base-called quality information from all the different traces covering any one base.

3.2.1. Major Contigs

The task of contigating a BAC sequence begins with ordering and orienting the major contigs to represent the full insert in the correct order from one end of the clone to the other. It is then much easier to close gaps in the sequence. The use of a stick diagram helps to visualize this with the inclusion of contig numbers or template names (*see Fig. 1*). GAP4 displays a graphic representation of the assembled BAC as a whole, called the *contig selector*.

The contigs containing the clone ends (cloning site) should now be found. They will have been screened out and tagged by Phrap and may be found by searching for the cloning vector tag (*see Note 1*) or by checking for this information in the output file created by Phrap. These two ends represent the start and end of the insert to which the rest of the major contigs will be anchored. There will be a left and a right end to the insert in relation to the chromosome, and, if possible, the insert should be oriented correctly using mapping and sequence data. If this is not possible, the ends may be chosen arbitrarily. Working from the nonvector end of these contigs, it is now possible to try to join another contig to this one.

Before attempting to join the next contig, it is important to ensure that the sequence at the ends of the contigs is of high quality without any obvious mis-assemblies. If the sequence is of low quality or there is a false end owing to a

misassembly during Phrap, this will mislead any sequence alignment tool. To check the quality of the ends, a list should be compiled of the major contigs, then each one opened using the editor display. Any bases below finished quality will need to be improved by adding more high-quality sequence, which will both solve these discrepancies and, it is hoped, bridge the sequence gap. A tag (see **Note 1**) should be added to the region of sequence using the general comment tags to highlight any weak areas. GAP4 uses cutoff data to allow the finisher to determine and select where the good-quality data end. The sequence in the database can then be manipulated to include or exclude the cutoff data.

3.2.2. Joining Contigs

Just like a jigsaw, there are lots of clues as to which major contigs should join together. Consider that the picture as the sequence and the nature of a region, such as a tandem repeat or a GC-rich region, may indicate that sequences are from the same locus or piece of the picture. The interlocking pieces may be compared to spanning read pairs of sequence from the same insert, indicating that these pieces go together. In finishing, however, a section of sequence may still be missing between the pieces.

Phrap will have made all the joins possible between the contigs it created within its working parameters. Using an assembly of force level 1 as default, with 0 being the most stringent, prevents most misassemblies unless a low-copy repeat is present. Some joins will not have been made because they are too short or between low-quality pieces of sequence and therefore considered not statistically significant. Using GAP4 and Consed, it is possible to search for any good sequence alignments (i.e., potential joins) and hence allow manual intervention to decide whether they are true joins. This process is subjective, and therefore it takes time to learn what makes a correct join, despite apparent sequence discrepancies. In GAP4, the alignment tool Find Internal Joins uses a mismatch percentage to indicate which joins have the least disagreement and then aligns the sequence to give the best possible match for the finisher to consider.

The main reason for using human intervention is that a confidence value of a base sometimes does not tell the full story. Studying the trace file for a read often reveals the cause of a discrepancy. This allows the finisher to decide on the quality of a region and to use editing and a minimum amount of resequencing to efficiently resolve a discrepancy.

A join can be made with a minimum of actual bases overlapping, but there should be about 100 bases that show a clear if not exact match between the two sequences. To assess whether a join is true, the following criteria are applied:

1. There should be a good sequence match between the contigs with few low-quality bases or disagreements. Any significant disagreement must be rational-

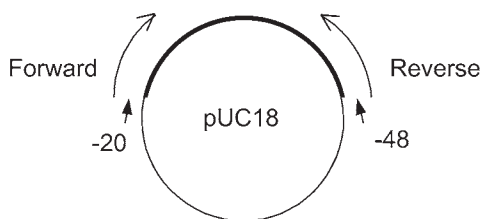


Fig. 2. Sequencing forward and reverse from a pUC18 plasmid using universal prim-

ized as either real or owing to a sequencing artifact; otherwise, the join may not be real, and, therefore, the trace files should be opened and checked. An experienced finisher can join contigs correctly using very weak data by identifying matching sequence motifs.

2. If there are any spanning read pairs that confirm these contigs join together, this is good proof of a join (*see Subheading 3.2.3.*).

All possible joins should be made to minimize the amount of resequencing required. It is relatively easy to strengthen a weak join by raising the quality of the sequence by directed sequencing. Once all the obvious sequence joins have been made, the number of major contigs should have decreased. The next section describes how to orient the remaining major contigs.

3.2.3. Orienting the Majors Using Read Pair Information

During sequencing, only a fraction of the whole subclone insert can be sequenced. This is limited by the read length possible owing to the chemistry used, the cycling conditions, and the time run on the sequencing machine. A plasmid vector such as pUC18 has a double-stranded insert of known size. During shotgun sequencing, the insert will be sequenced on both strands using the universal forward and reverse priming sites in the vector (*see Fig. 2*). The shotgun sequences are assembled as follows:

1. The resulting sequences are called read pairs and are stored in the database. These pairs will be assembled in the database so that they are facing toward each other. The sequence of the full insert will be covered not only by sequence from one insert, but by many others, which are staggered and overlapping with each other. This gives the sequence of the full insert despite the limited read length (*see Fig. 3*).
2. Two contigs may be oriented to face each other by arranging one half of a read pair to face the other half across a gap. If the sequence does not match, allowing a join, the gap size can be calculated from the number of bases missing from the subclone insert between these contigs. If there are two or more read pairs spanning a gap, it is very likely that these contigs will eventually join (*see Fig. 4*).

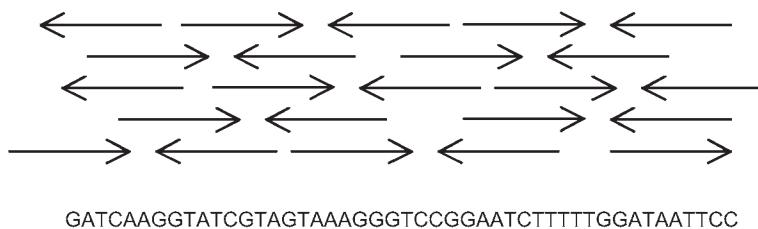


Fig. 3. Normal coverage of read pairs across a sequence.

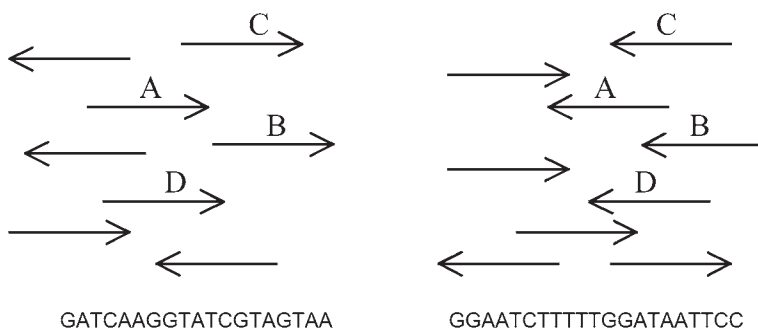


Fig. 4. Sequence gap with bridging read pairs.

3. The use of a double-stranded plasmid vector in a sequencing strategy is invaluable for the production of read pairs. The other most common vector is M13, but this creates single-stranded templates. Read pair information can be created using M13, but this first requires the template to be double stranded. To double-strand the M13 template, a PCR is required using the universal M13 forward and reverse primers. The other strand can then be sequenced using the M13 reverse primer. However, it is time-consuming and expensive to double-strand all M13 templates of a shotgunned BAC. Therefore, this approach should be used selectively, creating reverse reads only for templates that point into gaps and could generate spanning pairs to assist in orienting the contigs.
4. Read pairs are searched for one by one. Begin with one of the anchored contigs at the noncloning vector end (*see Subheading 3.2.1.*). Using the known insert size, create a list of all the reads extending toward the gap. To be sure that the templates could span the gap, only list those that begin priming up to 1500 bases away from the end of the contig (this is based on an insert size of 2000 bases). The other half of the respective pair can then be searched for in the database. It is hoped that there will be at least two that have sequenced successfully and that also sequence toward the gap.
5. Now orient the contigs so that the two halves of the read pair face toward each other. If the sequence matches make a join. If a join is not possible some sequence

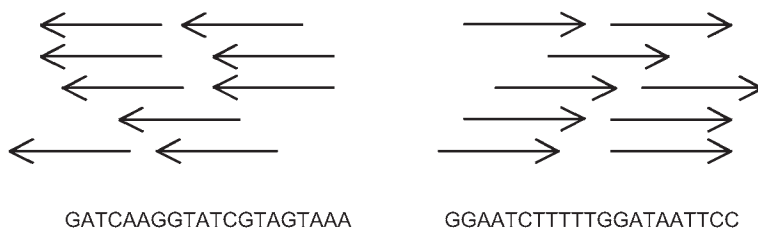


Fig. 5. Blunt-ended sequence gap with no bridging read pairs.

will still be missing. Add this contig information to the stick diagram (see **Fig. 1**) as the next anchored contig. Add a comment tag to the sequence at both ends with the information of the read pairs that they should join so that they can be checked and the join be made later once more sequence has been generated. Although a join has not been made, a variable has been removed. More sequence can now be generated specifically for this gap, which will now be of known size, and the type of sequence causing the discrepancy may be extrapolated from the sequence of the two ends.

6. The process can be continued by looking for read pairs with the other end of this newly anchored contig until all the major contigs are oriented to each other from one end of the clone to the other. Some contigs cannot be oriented because they have no read pairs because either the other half has consistently failed to sequence in a region or the reads may be blunt ended (all the reads end in sequencing vector) (see **Fig. 5**). Strategies for closing gaps without spanning read pairs are presented in **Subheading 3.3.5**.

3.3. Identifying Discrepancies and Selecting Solutions

For a sequence to be finished, every consensus base should be above the stated quality value and also be double covered. This means to cover every base with two different templates representing both strands or with one strand represented by a read of terminator chemistry, preferably a BigDye™ (they are not prone to compressions or “dropout” [14]). The aim of double coverage is to have a true representation of sequence from both strands so that one strand can resolve a sequence discrepancy on the other such as a compression. Using two templates that agree also prevents discrepancies from being introduced from a single subclone.

The principles for selecting reactions are similar for all internal discrepancies, with conditions varying for specific cases. All discrepancies should be identified and tagged with a general comment tag to allow an overview of the

regions requiring work. A record of the regions with the type of discrepancy and the chemistry used to solve them will be helpful if the discrepancies are not solved the first time and a new strategy needs to be developed.

3.3.1. Searching for Discrepancies

3.3.1.1. BASE QUALITY DISCREPANCIES

The identification of single base discrepancies is easy using both GAP4 and Consed because they provide search tools that allow you to define your quality threshold and to search at that value throughout a contig. Single base discrepancies may be resolved by editing if the double coverage criteria are fulfilled and two strong bases agree from the trace files. Using GAP4 to edit, we increase the confidence of the correct base so that the false base will no longer affect the consensus. It is important that these criteria are adhered to; otherwise errors may be introduced into the sequence. Edits are also done in lowercase so they are obvious and can be rechecked later. Most single base discrepancies occur in clusters and cannot be resolved by editing. It is a waste of time to edit one base if more sequence is required to resolve the base next to it.

3.3.1.2. DOUBLE COVERAGE

Regions that are not double covered can be found by scrolling through the editor and for each contig checking the quality of the traces in all regions with two or fewer reads. In GAP4 the strand display can be turned on so that all regions where only one strand is present will be obvious. Any regions with only one read will require a second read from the other strand or a terminator on either strand. Regions with multiple unidirectional reads that are of thermostable primer chemistry require the same treatment.

3.3.2. Choosing a Sequence Reaction

Any read that is selected must begin priming close enough to the discrepancy to allow the new read to cover it. To get good-quality sequence across the discrepancy, allow no more than 300 bases from the priming site to the discrepancy. Resequencing with the universal primers is only useful if they begin priming this close and can be used to resolve a simple sequence discrepancy. The best chance of solving a discrepancy is to select a custom primer 100 bases away (in either direction) from the discrepancy you wish to solve. This ensures that the strongest signal will be obtained from the new read over the problem. We use the GAP4 primer selection program OSP because it provides good primer options, but using the melting temperature (T_m) and the criteria described next you can do this without a selection program.

3.3.2.1. CRITERIA FOR A GOOD CUSTOM PRIMER

1. Use a length of 18–25 bases.
2. Use a T_m of about 50°. $T_m = 2(A + T) + 4(G + C)$.
3. Aim for a 50% GC content.
4. Select no more than three of any one base in a row.
5. Avoid repetitive sequences inside the primer.
6. Try to avoid making custom primers in Alu, LINE, or any low-copy repeat especially if making a BAC PCR.
7. Try to use only unique sequences, and search for the sequence in the major contig if necessary.

A primer selected to prime on the negative strand will have the reverse complement of the sequence on which it was chosen. Where available, up to four different subclones should be sequenced using BigDye Terminator and the chosen primer (either universal or custom). This is necessary to cover a discrepancy, because some reactions may fail. To be sure that the subclone covers the discrepancy, select reads that have universally primed no more than 1000 bases away from the discrepancy. This allows the other half of the insert to cover the region comfortably (*see* **Notes 2 and 3**).

3.3.2.2. SIMPLE RULES FOR SELECTING CHEMISTRY

1. Select subclones where the universally primed read starts no more than 300 bases away from the discrepancy. Otherwise, a custom primer will need to be generated.
2. Select up to four different subclones whose insert covers the discrepancy.
3. Select which strand to sequence on when designing the custom primer (*see* **Note 4**).
4. Select chemistry tailored to the type of discrepancy.

The protocols for these reactions are described in **Subheading 3.4**.

3.3.3. Common Types of Sequence Discrepancies and Tailored Chemistry Solutions

1. *Thin and weak regions of sequence.* Some regions will have thin template coverage or be covered only by reads that have a weak signal at the end of the read. Designing a custom primer and sequencing using BigDye Terminator should resolve this.
2. *Undefined single base sequence discrepancies.* Trace discrepancies can occur when the signal is weak or masked by another sequence product. For example, dye blobs can be produced when large amounts of unincorporated dyes are not removed and cause a large peak at 100 bases into the reads seen over the top of 10–20 true bases. On a slab gel image, this appears as a blue smear across the gel soon after the priming site. Repeating the reaction with particular attention to the wash step in the precipitation should resolve this. If the problem is not resolved,

try adding another wash step or try an alternative precipitation. If the discrepancy is not owing to the conditions under which the previous reads were run, a custom primer could be used instead; priming in a different position should resolve the discrepancy.

3. *Compressions.* GC compressions in traces can be seen as several signals either on top of each other or squashed together. These are caused by anomalies in the migration behavior of certain DNA fragments during electrophoresis because of intramolecular base pairing between short stretches of sequence with dyad symmetry, especially those containing a high proportion of guanine and cytosine residues (**15**). Thermosequenase primer chemistry and dGTP BigDye Terminator are prone to such compressions. One good standard BigDye Terminator read will resolve a compression. BigDye Terminator chemistry utilizes dITP nucleotides rather than dGTP. This analog forms I-C base pairs containing only two hydrogen bonds instead of the three normally formed by G-C base pairs, reducing the intramolecular bonding (**15,16**). Alternatively, sequencing on the other strand may solve the discrepancy.
4. *Repetitive regions.* Small units of tandem repeats are difficult to sequence through. Using a custom primer will help by decreasing the distance to the discrepancy. Use of either a BigDye Terminator or dGTP BigDye Terminator with the addition of SequenceR_x Enhancer A will assist in resolving the sequence (see **Notes 5** and **6**). Thermosequenase BigDye primer may also help solve the discrepancy (see **Note 3**).
5. *GC rich.* Regions rich in GC bases are particularly resistant to sequencing owing to high concentrations of these two bases and the compressions they can cause (**16**). In addition, it is difficult to design a unique custom primer close to the discrepancy with a T_m suitable for sequencing. If the T_m exceeds 50°C, the annealing temperature should be raised to 60°C. When the T_m becomes high, a point is reached where the primer loses specificity and therefore produces spurious results. The regions are often weak over several kilobases owing to these difficulties in sequencing and a cloning bias in pUC18. M13 subclones maintain GC-rich sequence more effectively than pUC18 subclones. Use either a BigDye Terminator or dGTP BigDye Terminator custom primer reaction with the addition of SequenceR_x Enhancer A or F that is designed to assist the resolution of GC-rich sequence. Thermosequenase BigDye primer may solve the discrepancy (see **Note 3**).

3.3.4. Structures

Secondary structures (hairpins) can be formed within a subclone on one or both strands. This is caused by an inverted repeat that could be between two very similar repeats such as a poly A and poly T tail or Alus or LINEs. The most difficult repeats to resolve are in CpG islands (**16,17**). Some structures are very obvious and can be picked up by looking at the traces most often ending in a sequence gap. Strong traces will stop prematurely either abruptly or with a rapid decline in trace height across 50–100 bases.

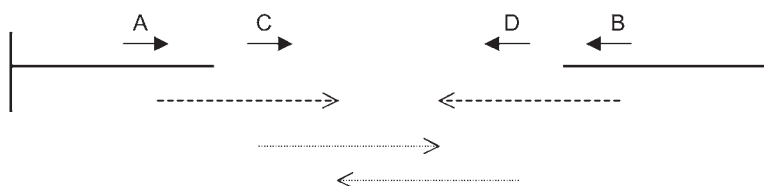


Fig. 6. Gap between two contigs being closed. Reads from the custom primers A and B extend into the gap. Custom primers C and D are generated on the new sequence, and the resulting reads then cross and close the gap.

To attempt to sequence an obvious secondary structure with BigDye Terminator is a waste of time and money because these will usually stop in exactly the same place. Using the alternative terminator kit dGTP BigDye should solve most structures and the addition of SequenceR_x Enhancer A may also improve the quality of the read. The dGTP BigDye kit uses a dGTP nucleotide rather than dITP. This nucleotide forms three hydrogen bonds in G-C base pairs rather than only the two bonds formed in I-C base pairs, which increases binding, allowing the sequence enzyme to continue around the hairpin (17). They can, however, introduce compressions (*see Subheading 3.3.3.*). Any compression should be resolved by sequencing on the other strand. (For other options, *see Notes 7–9.*)

3.3.5. Gaps and Large Regions of Internal Discrepancies

This section uses the principles of custom primers and read pairs described in the preceding sections. The closing of sequence gaps between major contigs and regions of internal discrepancies >300 bases is resolved using the same technique. The technique involves designing two custom primers to allow sequencing into the discrepancy from the left on the positive strand and from the right on the negative strand. The priming site should be 100 bases away from where the good sequence is required to start.

If the internal region is larger than 600 bases more custom primers can be generated. These new primers should be positioned to allow for a 100-base overlap between the end of the new read and the priming site of the next. This process can be continued until the new reads from the left and right cross in the center (*see Fig. 6.*).

In major contigs that have not been oriented by read pairs, it is possible to resequence the other half of any read pointing toward the gap with a universal priming site no greater than 1500 bases from the end. If these new reads are successful, this will orient the contig even if the gap is not closed the first time.

Custom primers for BigDye Terminator sequencing should still be chosen if there are no orientation data because the gap could still close, joining it onto an unknown contig.

If there are no reads pointing toward the gap and all sequence ends in sequencing vector, then the contig is blunt ended. If this occurs on both sides of the gap, then a BAC PCR from the original BAC DNA should be attempted to obtain a template for sequencing that spans the gap. BAC PCRs are used to bridge gaps or to serve as a second template when only one subclone covers an area. To create a BAC PCR, two unique custom primers are required. The sequence of these primers should be checked using a sequence search in the database to minimize mispriming. The PCR protocol for BAC DNA in **Subheading 3.4.5.** is reliable for a 2000-bp product. Therefore, the primers selected should be no farther apart than this. PCR can be used to bridge a gap of unknown size. In this case, the PCR product should be run on an agarose gel with a marker to give an accurate size of the gap. The resulting PCR can be sequenced with the custom primers used to create it or using nested primers (custom primers generated inside the region of the PCR) (*see Notes 10–13*).

3.4. Protocols

For the protocols in **Subheadings 3.4.1., 3.4.3., and 3.4.5.** the reactions should be prepared with the sequence plate on ice to prevent evaporation and to avoid premature and potentially unspecific base incorporation.

3.4.1. BigDye Terminator Sequencing

A total reaction volume of 10 μL is required for this protocol. The volume of water can be adjusted if other component volumes need to be varied.

1. Place 1 μL (12 pmol/ μL in ddH₂O) of each primer required into separate wells of a 96-well sequencing plate. *See Note 14* for information on primers.
2. Add 2 μL (60 ng/ μL) of each DNA sample required.
3. Add 1 μL of 10X reaction buffer to all reaction wells.
4. Add 2 μL of BigDye Terminator reaction mix to all reaction wells.
5. Add 4 μL of ddH₂O to all reaction wells.
6. Place a Hybaid lid on top of the plate and press down to seal all the wells.
7. Pulse centrifuge the samples at 300 rpm until the samples are collected at the bottom of the plate.
8. Place the reaction plate in a thermocycler programmed for the following conditions: 95°C for 15 s, 50°C for 5 s, and 60°C for 2 min for a total of 35 cycles.

For alternatives using a dGTP BigDye Terminator kit *see Note 15*. For use of SequenceR_X Enhancers *see Notes 5 and 6*.

3.4.2. Isopropanol Precipitation for Terminator Sequencing

1. Add 60 μL of 80% isopropanol (room temperature) to each reaction well of a 96-well plate.
2. Leave to stand at room temperature for 10 min.
3. Centrifuge at 4000 rpm at room temperature for 30–45 min.
4. Discard the supernatant onto a tissue.
5. Add 100 μL of 60% isopropanol (room temperature) to each reaction well.
6. Centrifuge at 2000g for 10 min.
7. Discard the supernatant onto a tissue.
8. Keeping the plate tipped up, place it on a tissue face down in the centrifuge and spin again for 1 min at 135g. The plate will now be dry. Store at -20°C until loading.

See **Note 16** for an alternative sodium acetate ethanol precipitation.

3.4.3. BigDye Primer Sequencing

1. Place 2 μL of each M13 or pUC template DNA to be sequenced in four wells of a 96-well plate (see **Note 17**).
2. Add 4 μL of a single base BigDye primer A, T, G, and C mix to each well (use one row of 12 wells for each base).
3. Place a Hybaid lid on top of the plate and press down to seal all the wells.
4. Pulse centrifuge the samples at 300 rpm until the samples are collected at the bottom of the plate.
5. Place the reaction plate in a thermocycler programmed for the following conditions: 95°C for 30 s, 55° for 30 s, and 70°C for 1 min for a total of 15 cycles, and then 95°C for 30 s, and 70°C for 1 min, for a total of 15 cycles.

3.4.4. Precipitation for BigDye Primer Sequencing

1. Pool the four reactions for each template into one well of a 96-well plate.
2. Add 50 μL of 100% ethanol (chilled) to each reaction well.
3. Let stand at room temperature for 10 min.
4. Centrifuge for 20 min at 2000g.
5. Discard the supernatant onto a tissue (alternatively pulse centrifuge upside down on a tissue at 200 rpm).
6. Leave the plate to dry at room temperature for 10–15 min.
7. Store the plate in dry conditions at room temperature until loading.

Thermosequenase BigDye primer chemistry requires a specific run module on the sequencing machine, and there is a different run module for BigDye terminators.

3.4.5. PCR of Cloned DNA Using PCR Beads

1. Aliquot 0.5 μL of BAC DNA (60 ng/ μL) into a 96-well sequencing plate.
2. Add 1 μL (12 pmol) of each of the two primers (suspended in ddH_2O) to each well.

3. Add one PCR ball to each reaction by tapping the ball gently into the well (store the beads dry at room temperature) (*see Note 18*).
4. Add 25 μL of ddH_2O to each well and allow to dissolve before mixing with a pipet.
5. Place a Hybaid lid on top of the plate and press down to seal all the wells.
6. Pulse centrifuge the samples at 300 rpm until the samples are collected at the bottom of the plate.
7. Place the reaction plate in a thermocycler programmed for the following conditions: hot start at 95°C for 5 min, then cycle at 95°C for 30 s, 50°C for 30 s, and 72° for 1 min for a total of 35 cycles.

Sequencing directly from this PCR product without a cleanup step is possible using 0.5 μL of the product from the method in **Subheading 3.4.1**.

3.5. Loading, Processing of Samples, and Assembly into Database

The compatibility of chemistry limits the type of sequencer that can be used. Most finishing samples use the BigDye chemistry and are loaded on an ABI Prism 377 slab gel or PE3700 capillary machine. Good results are obtained from both machines. Finishing samples often have a lower pass rate or read length than standard shotgun reads because they cover the most difficult regions of sequence.

Once run, the samples should be uploaded from the sequencer and processed through the same route as the original shotgun reads. The Staden package uses Pregap to process reads before their assembly into GAP4.

The method of assembly of new reads is dependent on the program used. The preferred method for GAP4 users is to add them directly to the GAP4 database using the “normal shotgun assembly” option. This is the method of choice for <300 new reads. The main alternative is to use Phrap, which will add the new reads but loses the tags and edits. An alternative to this is to use a modified version, “incremental Phrap,” which maintains tags and edits but also adds new reads.

The normal shotgun assembly option attempts to join all the new reads to the current contigs depending on the percentage of mismatched bases between the new reads and those already in the database. Adjusting the percentage mismatch allowed will vary the stringency of the assembly. It is important to make full use of the new reads. While checking that the discrepancies have been resolved, ensure that all the successful new reads have been assembled too. To do this, check all new reads created for the custom primer chosen for a discrepancy. They may be present either as a single contig on their own or attached to other reads. A manual join can be attempted to join these new reads, often solving the discrepancy. An alternative to checking each unsolved region is to use the join sequence alignment tools again (such as Find Internal Joins) and then join all good reads to the assembly in both solved and unsolved regions.

3.6. Checking and Selecting Further Directed Sequencing

Once the new reads have been added to the database, it is necessary to return to the discrepancy regions; this can be done by searching for the tags that were added to these regions. The quality should be reassessed, and if the discrepancies have been resolved or can now be edited, then the tag should be removed. In regions where a new join has been made across a gap, it should be checked that this is correct and the quality rechecked to ensure that no more sequence needs to be added to resolve this region. For gaps that have not been closed, consider whether the previous sequencing attempted was successful and extended into the gap or not. This may require another round of new custom primers to extend further into the gap. Any gaps that have failed to sequence or abruptly stopped sequencing should be retried using dGTP BigDye Terminator because there may be a structure that is not visible that prevents the gap from being bridged. The addition of SequenceR_x Enhancer A may improve the quality of sequence in a repetitive region.

Usually several rounds of reactions are required to resolve all regions of discrepancies. They should be tackled using the guidelines described in **Subheading 3.3**. Repeating the same reaction is unwise and instead a change of approach is required. This can be done in simple ways such as sequencing on the opposite strand, using a different group of templates when available, as well as changing the chemistry approach using a dGTP BigDye Terminator instead of adding SequenceR_x Enhancers. The flow diagram in **Fig. 7** outlines a strategy to tackle stubborn regions in a systematic manner.

3.7. Verifying of Assembly

Assembly of the BAC must be verified to ensure that no misassembly has occurred. This is achieved by comparing two or more restriction digests of the BAC with a virtual digest of the assembly. The most common restriction enzymes used are *HindIII*, *EcoRI*, and *BamHI*. These enzymes cut frequently enough so that most bands can be accurately sized using a marker. GAP4 provides a tool, Restriction enzyme map, that allows the user to carry out a virtual digest on the assembly. This produces a list of the fragment sizes expected, and the finisher then compares these with the bands sized from the real digest. The real and virtual digest results should match; otherwise, this indicates a misassembly.

3.8. Tackling Misassemblies

Phrap creates a database assembly by using a “greedy” algorithm, which essentially joins the most strongly overlapping reads first. Misassemblies are caused by matches between highly conserved repeats being confused with true sequence matches. Any repeat that is larger than the insert size is capable of

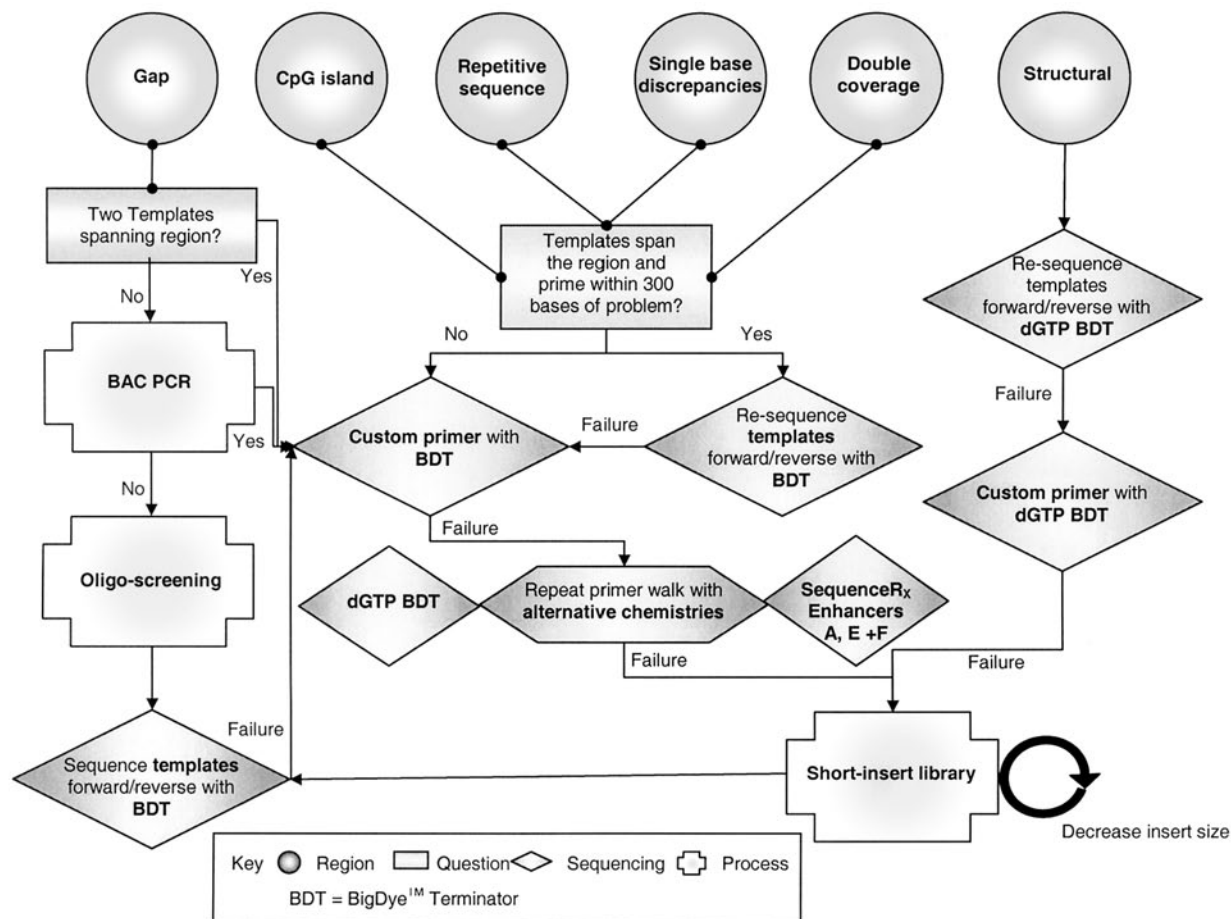


Fig. 7. Flow diagram of directed sequencing options.

causing a misassembly because one read pair cannot span the repeat. Misassemblies prevent the sequence from being a true representation of the insert and therefore must be resolved. Since there is currently no computerized method for doing this, a skilled finisher is required.

Misassemblies are detected by the appearance of high-quality single base discrepancies between two reads that otherwise appear to match. In the case of a human BAC, this could be a known repeat such as Alu or LINE or an unclassified low-copy repeat. Because known repeats can be tagged during Phrap, the region of similarity will be relatively easy to define and the two copies separated using read pairs anchored outside of the repeat region.

When the region of the repeat is not so obvious, a dot matrix analysis is invaluable. GAP4 plots matches between the two sequences along the *x*- and *y*-axis with a function called Find Repeats. From the region of the misassembly, two lists should be created of the reads thought to belong to the different copies of the repeat. To do this, begin with any reads containing high-quality base discrepancies. Add to this list any of their read pairs in this region. If a group of the read pairs is found to be in another region, this suggests where they may belong once the misassembly has been broken. By aligning the sequence of the misassembly with this second region, the continuation of the repeat region may become obvious.

Different tags can be used to highlight the length of the copies of the repeats; this will make them easier to compare. A diagram can also help to visualize the assembly as a whole and the consequences of breaking and rejoining a contig. The GAP4 Contig Selector can be used to do this with the tags displayed. Once the copies have been identified, they must be separated. The bulk of the reads can be separated using the lists. Care must be taken to separate the two copies completely. This may involve moving single reads from one copy to the other once the bulk has been separated. If no read pairs are available, the copies can still be separated, but this will be much more difficult, and in cases of completely identical copies, it may be impossible to reassemble the reads in the right copy. This is not a problem in this case since each subclone can exist in either copy.

3.9. Finishing Multiple Genomes

The finishing of whole genomes involves coordination of the finishing of multiple overlapping BACs. The use of a tracking database allows a BAC to be traced from selection to subcloning; sequencing; finishing; and, finally, submission and annotation. The finishing of overlapping BACs has to be coordinated to prevent a duplication of effort and the submission of redundant data. The data from overlapping BACs are used to assist the finishing of overlaps.

In addition, several tools have been written at the Sanger Centre to complement the use of GAP4. Readpairfind, written by David Harper of the Pathogen

Group, has two functions: ReadpairJoins displays suggested joins between contigs based on read pair information, and ReadpairGraph displays read pairs internally across a contig, which assists in resolving misassemblies. The tool gapNav, written by Mark Griffiths of the Production Software Group, creates a file to navigate through a GAP4 database, highlighting the weakest regions of the assembly based on read coverage and chemistry. Confirm, written by John Attwood, also of the Production Software Group, produces a visual display comparing the real restriction digest of a BAC with a virtual digest. This allows easy comparison of the results, highlighting any major discrepancies between them. A diagram representing the assembly with the restriction enzyme cut sites and fragments produced highlights any specific band showing disagreement and therefore a possible misassembly. For further details of all these programs visit www.sanger.ac.uk/Software/.

The following could improve and speed up the finishing process:

1. The use of capillary sequence machines, which have improved accuracy over slab gel machines by reducing the number of possible retracking errors.
2. The use of robotics to utilize DNA templates stored in 384-well plates rearranged for directed sequencing.
3. An automated finishing program for selecting directed sequencing reactions, such as Autofinish by Dave Gordon (18).
4. A checking procedure to ensure that quality of the finished BACs is maintained.

In the near future there will be: improvements to database assembly algorithms, thereby reducing misassemblies, as well as improvements to sequencing chemistries designed specifically for the most difficult sequences. As shown in this chapter, the finishing process is both complex and iterative.

4. Notes

1. Sequence editors such as GAP4 and Consed allow different visual tags to be added to lengths of sequence to highlight features that can then be searched for and edited by the user.
2. Using terminators or primers: For general discrepancy solving, we use BigDye Terminator reads in either a positive or negative direction. In normal regions of sequence their reads will be as long as the BigDye primer reads and are not prone to compressions.
3. Thermosequenase BigDye primers may sequence better than terminators in regions that are AT or GC rich, but they cannot be combined with the close priming site gained using a custom primer without first creating a special PCR product with the universal priming sites added.
4. Always consider which strand is not represented in a discrepancy region. Sequencing on the unrepresented strand may be more effective.

5. SequenceR_x Enhancers are available as a kit, with the three most frequently used being A, E, and F. Their use results in an improvement of quality. A is the most universally used to tackle GC-rich, direct repeats; short CT-rich repeats; and inverted repeats. E is used for direct repeats and short CT-rich repeats. F is suited to high GC-rich content regions around 80%. Two microliters of the chosen enhancer is added last after the H₂O (which is reduced by 2 μ L to compensate). Cycling conditions and precipitation remain the same.
6. dGTP BigDye Terminator with SequenceR_x Enhancer A will work more efficiently than BigDye terminators if tandem repeats are causing hairpins.
7. An alternative strategy to break a secondary structure is to add 1 μ L of 100% DMSO to the sequencing reaction. This is to achieve a 10% concentration of DMSO in the reaction.
8. If chemistry fails to resolve a structure, the best option is to create a short-insert library (**19**) of a piece of sequence, which spans the discrepancy.
9. For a repetitive CpG island, create a short insert library of the insert in M13 rather than pUC18. The combination of fragment size libraries at 100–300 and 300–500 bases allows a framework to rebuild the sequence and any structures to be broken.
10. A PCR product created with unique primers outside a repeat region can then be sequenced using nested primers in the repeat.
11. PCR may fail owing to a repeat or hairpin structure in the DNA. The technique oligoscreening (**20**) may be used to pick out a new template. This is more efficient than adding more shotgun reads in the hope that these will occur near the discrepancy region.
12. One and a half microliters of 100% DMSO can be added to help generate a PCR across a region containing a structure and also used in the sequencing reaction.
13. There are lots of PCR products on the market tailored to specific types of sequence.
14. The concentration of custom and universal primers used in these protocols is 12 pmol/ μ L. If using primers supplied dry, resuspend in sufficient ddH₂O to achieve this concentration. If the primers are supplied still in solvent, they must first be dried down; otherwise, they will inhibit the sequencing reaction. They can be dispensed as required in 1- μ L vol to the empty sequencing plate and dried down for 5 min in an oven or on a heated block and then the other components added, increasing the ddH₂O to compensate. Alternatively, a small stock sufficient for 10 reactions can be dried down and resuspended in 10 μ L of ddH₂O.
15. Replace the BigDye Terminator mix with 2 μ L of dGTP BigDye Terminator reaction mix. The following alternative cycling conditions are optional: 95°C for 30 s, 55°C for 5 s, and 72°C for 2 min for a total of 15 cycles.
16. Add 1 μ L of 3 M sodium acetate (pH 4.8) and 25 μ L of 100% ethanol. Then centrifuge at 4°C for 30 min at 2000g. Discard the supernatant and add 100 μ L of 70% ethanol. Then centrifuge at 4°C for 10 min at 2000g. Discard the supernatant and repeat this 70% ethanol wash. Air-dry the plate or heat it in an oven until dry.

17. Four reactions will be carried out for each template, one for each base in a separate well. These should be laid out on a 96-well plate using one row of 12 wells for each base and a column of 4 wells for each template being sequenced.
18. Add 25 μ L of ddH₂O directly to the PCR bead in the tube it is supplied in and allow to dissolve. Then add this solution to the DNA and custom primer already prepared in the plate.

Acknowledgments

We are extremely grateful to the following people who advised us in the writing of this chapter: Jane Rogers, Stephan Beck, Karen Barlow, and Alan Robinson. This work was funded by the Wellcome Trust.

References

1. (1997) The yeast genome directory. *Nature* **387**, 5–105.
2. The *C. elegans* Sequencing Consortium. (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018.
3. The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
4. Osoegawa, K., Mammoser, A. G., Wu, C., Frengen, E., Zeng, C., Catanese, J. J., and de Jong, P. J. (2001) A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.* **11**(3), 483–496.
5. The International Human Genome Mapping Consortium. (2001) A physical map of the human genome. *Nature* **409**, 934–941.
6. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**(3), 175–185.
7. Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**(3), 186–194.
8. International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
9. Mullikin, J. C. and McMurray, A. A. (1999) Sequencing the genome, fast. *Science* **283**(5409), 1867–1868.
10. Richterich, P. (1998) Estimation of errors in “raw” DNA sequences: a validation study. *Genome Res.* **8**(3), 251–259.
11. Staden, R. (1994) Computer analysis of sequence data, in *Methods in Molecular Biology*, vol. 25: *DNA Sequencing* (Griffin, A. M. and Griffin, H. G., eds.), Humana Press, Totowa, NJ, pp. 9–26.
12. Staden, R., Beal, K. F., and Bonfield, J. K. (2000) The Staden Package 1998, in *Methods in Molecular Biology*, vol. 132: *Bioinformatics Methods and Protocols* (Misener, S. and Krawetz, A., eds.), Humana Press, Totowa, NJ, pp. 115–130.
13. Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.* **8**(3), 175–185.

14. Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., Khan, S. H., Menchen, S. M., Heiner, C. R., and Chen, S. M. (1997) New dye-labelled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.* **25**(22), 4500–4504.
15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) DNA Sequencing, in *Molecular Cloning: A Laboratory Manual*, 2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Woodbury, NY. pp. 13.74–13.75.
16. Motz, M., Paabo, S., and Kilger, C. (2000) Improved cycle sequencing of GC-rich templates by a combination of nucleotide analogs. *Biotechniques* **29**(2), 268–270.
17. Hirao, I., Nishimura, Y., Tagawa, Y., Wataabe, K., and Miura, K. (1992) Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCAAAGC) and their effect on DNA sequencing. *Nucleic Acids Res.* **20**, 3891–3896.
18. Gordon, D., Desmarais, C., and Green, P. (2001) Automated finishing with Auto-finish. *Genome Res.* **11**(4), 614–625.
19. McMurray, A. A., Sulston, J. E., and Quail, M. A. (1998) Short-insert libraries as a method of problem solving in genome sequencing. *Genome Res.* **8**(5), 562–566.
20. Flint, J., Sims, M., Clark, K., Staden, R., and Thomas, K. (1998) An oligo-screening strategy to fill gaps found during shotgun sequencing projects. *DNA Sequence* **8**(4), 241–245.

Using the TIGR Assembler in Shotgun Sequencing Projects

Mihai Pop and Dan Kosack

1. Introduction

The TIGR Assembler (TA) (*1*) is the sequence assembly program used in sequencing projects at The Institute for Genomic Research (TIGR). Development of the TA was based on the experience obtained in more than 20 sequencing projects completed at TIGR (*see* www.tigr.org). This extensive experience led to a sequence assembler that produces few misassemblies (*2,3*) and has been used successfully in whole-genome shotgun sequencing of prokaryotic and eukaryotic organisms, bacterial artificial chromosome-based sequencing of eukaryotic organisms, and expressed sequence tag assembly.

The input to the assembler consists of a set of sequences and associated quality values. The quality of a base is defined in **refs. 4 and 5** as $-10\log_{10} P_e$, in which P_e is the probability that the base call is incorrect. These quality values can be computed by the Phred program base-caller (*4,5*). If unknown, the quality values may be omitted from the input.

As output, the assembler produces a set of contigs, each represented as a consensus sequence and a multiple alignment of input sequences. Ideally, a single contig will be produced, representing the molecule being sequenced.

A distinguishing feature of TA is the ability to build on previous assemblies in a process called contig jumpstarting. This procedure can be used to combine the outputs of previous assembly jobs into one single assembly. For more details refer to **Subheading 3**.

Like most other widely used sequence assemblers, the TA uses a greedy strategy for assembling individual fragments (*6,7*). The assembly algorithm

starts by computing all pairwise alignments between the input sequences. These rough alignments are computed by looking for exact 32-base words shared by each pair of sequences, and assigning a score to each such alignment. The alignment score takes into account not only the number of 32mers shared by the two sequences, but also the uniqueness of these 32mers. Intuitively, words that occur too many times in the input are indicative of repeat areas, and are therefore given a lower score. This ensures that unique regions will be assembled before potential repeats. The pairwise alignments (also termed matches) are considered in order, the highest scoring first. Each match is checked for feasibility using an implementation of the Smith-Waterman algorithm for sequence alignment (8). The assembler screens sequence alignments based on length of overlap, maximum length of the overhang, and the Smith-Waterman score of the alignment (*see Fig. 1* for details).

If an alignment satisfies all the constraints, the two sequences are merged into a single contig. The contigs corresponding to the matched sequences are merged into a single contig using a technique similar to that of Gribskov et al. (9). The procedure is repeated until no more merges are possible. At this point, the output consists of a set of contigs that cannot be merged any further.

Since the TA attempts to reduce the number of misassembled contigs, it may produce more contigs than other assembly programs such as Phrap (10) or CAP (11,12). We opted for this trade-off between assembly errors and number of contigs because it is generally more difficult to detect misassembled contigs than to determine those contigs that the assembler failed to merge. Because of the potentially large number of resulting contigs, it is important to build a scaffold using the forward-reverse relationships between the input sequences. Two sequences are said to be forward-reverse mates when they are sequenced from the two ends of the same clone.

We present herein the details of installing and using the TA.

2. Materials

This section contains a list of the prerequisites for running the TA. Besides the system requirements and a list of required programs, we describe the main file formats used as input or output by the TIGR assembler.

2.1. Installing the TA

1. UNIX- or Linux-based computer system: Solaris, Digital UNIX, Compaq Tru64 UNIX, and RedHat Linux were all tested. The system requires the following:
 - a. Available memory (RAM) greater than approx 72 times the total size of input sequence data.
 - b. Available disk space greater than approx six times the total size of input sequence data.

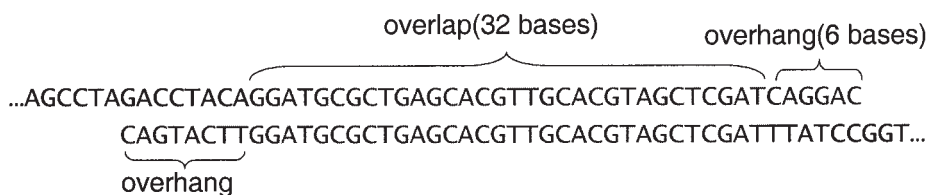


Fig. 1. Anatomy of an overlap.

- c. C compiler: Compaq/Digital C, SunPro C, Sun C, and GNU C (gcc) were all tested.
- d. Web browser: Microsoft Internet Explorer or Netscape Communicator should be sufficient.
- e. PDF viewer: Adobe Acrobat Reader is the industry standard and freely available.
- f. TA version 2.0 installation package: Instructions for retrieving this package are presented in **Subheading 3**.

2.2. Assembly From Scratch

This section presents the details of the files used when assembling a molecule from scratch, i.e., without incorporating any previously assembled contigs.

1. Sequence (.seq) file. A .seq file is formatted as a multi-fasta (**13**) file containing all the sequences that need to be assembled. Sequences appear in this file in a 60-character-per-line format. Each sequence is prefixed by a line containing its name and additional information in the following format:

```

name min_clone_len max_clone_len med_clone_len clear_
left clear_right

```

- a. name represents the name of the sequence. The name can have any format; however, if you want the assembler to take into account forward-reverse constraints, sequence names must have characters *F* or *R* starting after the seventh character in the name. The first seven characters of the name represent the name of the clone, while *F* and *R* represent the forward and reverse ends of the clone, respectively.
- b. The following three numbers (min_clone_len, max_clone_len, med_clone_len) are an estimate of the size of the clone—the minimum, maximum, and median clone sizes, respectively—and are usually standard across a shotgun library.
- c. clear_left and clear_right specify the “good” section of the sequence; that is, the bases outside this range (called clear range) are assumed to be of poor quality and will be ignored by the assembler. A program called Lucy (also available from TIGR) can compute the clear range for the sequences by removing the vector sequence and poor-quality sequence ends.

- d. If you do not wish to specify constraints on clone sizes you can simply set them to some arbitrarily high numbers, such as 1 1000000 500000. In addition, if you do not need to trim the sequences, you can provide just the name of the sequence, with no additional information (e.g., >GEFUB01TF). As an example, here are a few lines of an .seq file:

```
>GEFUB01TF 700 3200 1700 27 611
GAGACGCTCACTCTAGAGCATCCCCGTTCTAACGCTTTGATACTTAAAGCAATACGATGT
TCCTCTGGATTAACCTTCTAAAACCTTCACTTGTACTTGATCCCCCTTCATGAAGAACTCA
>GEFUB01TR 700 3200 1700 51 573
ATGGGNATGNGANAATATATGCCTCGCATCGAACCGTTTCGCGATAAGCTAGCAGGGGCTG
TAGGCGAT . . .
```

2. Quality (.qual) file. File containing the phred (4,5) quality values for the sequences contained in the .seq file. The format of the .qual file is similar to that of a .seq file, except that each line contains 17 quality values. The quality records are prefixed by the sequence name and must be in the same order as the corresponding sequences from the .seq file. It is possible to omit the quality values for some sequences, in which case no quality record needs to be written into the file. At the same time, the number of quality values in the .qual file must be exactly the same as the number of nucleotides in the corresponding sequence; otherwise, the assembler produces an error (see **Note 5** for error messages and warnings produced by the assembler). Here is an example of a .qual file:

```
>GEFUB01TF
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 15 00
00 00 00 00 19 21 24 24 25 25 25 25 30 31 31 31 36
33 40 23 23 00 00 00 21 18 00 00 00 20 20 30 28 22
>GEFUB01TR
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 15 19 . . .
```

3. Output: .asm file. The .asm file contains information about all the elements of the assembly. It consists of a “contig” record for each contig in the assembly, each such record being followed by a set of “sequence” records for each sequence belonging to the contig. The following are the most relevant parameters of a contig record:
- a. Sequence: the consensus sequence of the multiple alignment defining the contig after all the gap symbols (“-”) have been removed.
 - b. Lsequence: the consensus sequence without removing any gap symbols.

- c. **Quality**: the quality class for each base in the consensus; any base with a quality value <8 indicates that the assembler has high confidence in that base. A base with a quality value of 15 indicates a conflict between two aligned sequences and is usually an indication of a misassembly. We recommend that you carefully inspect any bases with a quality value of 9 or higher. Note that the quality classes are represented in hexadecimal notation, in which 0A through 0F stand for decimal numbers 10 through 15.
- d. **asembl_id**: the number of the contig in the .asm file.
- e. **Redundancy**: the average coverage, i.e., average number of fragments overlapping at any point in the assembly.
- f. **perc_N**: percent of bases having an IUB ambiguity code (anything other than A, C, T, or G).
- g. **seq#**: the number of sequences in the contig.

Here is an example of a typical contig record from an .asm file:

- a. **Sequence**: CAAGAAAAAATGAGTTTGACACAGCCGATCTGTTT
- b. **Lsequence**: CAAGAAAAAATGAGT-TTGACACAGCCGATCTGTTT
- c. **Quality**: 0x0B0B0B0B09090909090909090909090909090909
- d. **asembl_id**: 1
- e. **seq_id**
- f. **com_name**
- g. **Type**
- h. **methodasmg**
- i. **ed_status**
- j. **Redundancy**: 14.10
- k. **perc_N**: 0.00
- l. **seq#**: 36
- m. **full_cds**
- n. **cds_start**
- o. **cds_end**
- p. **ed_pn**: GRA
- q. **ed_date**: 06/08/00 13:10:18
- r. **Comment**
- s. **Frameshift**

The parameters defined by the sequence record are as follows (*see Fig. 2*):

- a. **seq_name**: name of sequence (from .seq file).
- b. **asm_lend**, **asm_rend**: coordinates of sequence in ungapped consensus (sequence entry from contig record).
- c. **seq_lend**, **seq_rend**: coordinates within sequence that align to **asm_lend** and **asm_rend**; these coordinates are relative to the sequence listed in the .seq input file. Thus, **asm_lend** and **asm_rend** must be within the clear range given in the input.
- d. **Offset**: offset of sequence from beginning of contig consensus (lsequence entry from contig record).
- e. **Lsequence**: gapped representation of the aligned portion of the sequence.

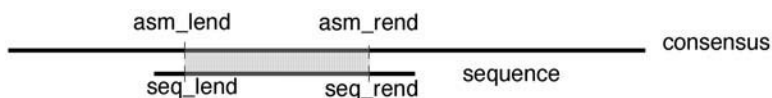


Fig. 2. Coordinates of a sequence within a contig.

Here is an example of a typical sequence entry from an `.asm` file:

- a. `seq_name`: GEFUB94TF
- b. `asm_lend`: 1
- c. `asm_rend`: 605
- d. `seq_lend`: 44
- e. `seq_rend`: 648
- f. `Best`: 0
- g. `Comment`
- h. `db`
- i. `Offset`: 0
- j. `Lsequence`: CAAGAAAAAATGAGTTT-ACACAGCCGATCTG

2.3. Jumpstart Assembly

1. Sequence file (see previous section). This file must contain *all* sequences appearing in your assembly, including those that occur in the `.contig` file.
2. Quality file (see previous section).
3. Contig (`.contig`) file. This is the file containing the contigs you want held together during the current assembly. You need to run the assembler with the `-a` option in order to create a set of `.align` files for each of the contigs (for more details refer to **Subheading 3.2.**). You can create the `.contig` file from the `.align` files by concatenating the `.align` files corresponding to the contigs you want to preserve. You can use the UNIX program “cat” to concatenate a set of `.align` files into a single `.contig` file:

```
cat myseqs_1.align myseqs_2.align > myseqs.contig
```

You can specify as many `.align` files as you want, with the caveat that no two `.align` files may contain the same sequences.

Similar to the `.asm` file, the `.contig` file contains two types of records: contig records and sequence records. There is one contig record for each contig in the file and one sequence record for each sequence contained in each contig.

4. Contig record. A contig record contains the consensus sequence for the contig, along with aggregate information about the contig. The format is similar to that of a `.seq` file. Sequence information is included in a 60-characters-per-line format and is prefixed with a header row. In the case of the contig record, the header is prefixed by two hash signs and contains the following information:

```
##name num_seqs num_bases bases, checksum checksum
```

name is the name of the contig and it is followed by the number of sequences in the contig (`num_seqs`) and the number of bases in the consensus (`num_bases`). The `checksum` is a signature of the contig that can be used for consistency checks. Here is an example of a contig record from a `.contig` file:

```
##UB_1 36 1355 bases, CB99078D checksum.
CAAGAAAAAATGAGTTTGACACAGCCGATCTGTTTACGGCTATGTCAAAC TCATAAATTT
CAAGAAAGTAACGTGTTATTCCTCTTCTTTTCGCATCAGATAAAGCGTCACCTAAAATAT
CGCCCATGGTAAAGCCAGTATTTTCTTCAGGCAATTCATATTCCTGTTCTTCTTTTGGTT
```

Each contig record is followed by sequence records for each of the sequences contained in the contig.

5. Sequence record. Sequence records contain information about the position of each sequence within the contig. The header of a sequence record starts with a single hash sign and contains the following parameters:

```
#name(offset) [ ] num_bases bases, checksum
checksum{seq_lend seq_rend}<asm_lend asm_rend>
```

The name of the sequence is followed by its `offset` within the consensus. If **RC** is mentioned within the brackets, the sequence is a reverse complement. `Num_bases` represents the number of bases in the aligned portion of the current sequence. As in the case of the contig record, the `checksum` is a signature for the sequence. The remaining four parameters (`seq_lend`, `seq_rend`, `asm_lend`, `asm_rend`) are identical to the similar parameters from the `.asm` file (see **Subheading 2.2**). An example of a sequence record follows:

```
#GEFUB94TF(0) [RC] 609 bases, 5BCEC289 checksum. {648 44} <1 605>
CAAGAAAAAATGAGTTTGACACAGCCGATCTGTTTACGGTTATGTCAAATTTATAAATTT
CAAGAAAGT-AACGGGTTATTCCTCTTCTTTTCGCATGAGATAAAGCGTCACCTAAAATAT
```

In the `.contig` file there will be 36 such records following the contig record described in **item 4**.

3. Methods

This section details the procedures for running the TA. We mention some of the files used by the assembler. A detailed description of the file formats can be found in **Subheading 2**.

3.1. Installing the TA

1. Download software from <ftp://ftp.tigr.org/pub/software/Assembler>. The required file is `TIGR_Assembler.v2.tar.gz`

2. After downloading the software onto a UNIX or Linux machine, you must unpack the distribution from `TIGR_Assembler_v2.tar.gz`. To do this, type the following commands on the UNIX command line:

```
gzip -dN TIGR_Assembler_v2.tar.gz
tar xvf TIGR_Assembler_v2.tar
```

If the archive extraction was successful, a `TIGR_Assembler_v2` directory folder should exist with the following subfolders: `bin`, `src`, `obj`, and `data`. A `README` file should exist as well. You can check for these files with the UNIX command

```
ls -l TIGR_Assembler_v2
```

3. Move into the `TIGR_Assembler_v2` directory by typing

```
cd TIGR_Assembler_v2
```

4. Read the `README` file in that directory.
5. Move into the `src` directory.
6. Read the `README` file in that directory, and then build the assembler by typing
`make`

If the build process completed successfully, the `bin` directory should contain the file

```
TIGR_Assembler
```

7. Before you can use it, the TA executable must be in your path. If your shell is `csh` or `tcsh` type the following command:

```
setenv PATH `pwd`:${PATH}
```

Otherwise:

```
PATH=`pwd`:${PATH}; export PATH
```

8. Move into the `data` directory using the `cd` command:

```
cd .. /data
```

9. To test the TIGR Assembler, move into the `201.pre` directory and execute the following command:

```
run_TA -C 201.contigs -q 201.qual 201.seq
```

The program may take several minutes to execute. When the command prompt returns, the TA has finished executing. Look for the files `201.asm`, `201.fasta`, `201.align`, and `201.error`. The presence of all these files, without `201.scratch`, indicates a successful assembly. The presence of `201.scratch` indicates a failure in the assembly software.

10. Following the testing phase, copy the `TIGR_Assembler` and `run_TA` files in the `bin` directory into a globally accessible location. Discuss this topic with your system support staff to determine the best strategy for your environment.

3.2. Running the Assembler

The best way to run the assembler is by using the script `run_TA` distributed with the assembler. If you wish to manually run the TA, a typical command line is as follows:

```
TIGR_assembler -q test.qual -a test.align -f test.fasta  
-n test -g 8 -e 15 -l 40 -p 97.5 test.scratch <  
test.seq > test.asm
```

1. `-q test.qual`: denotes the file containing the quality values.
2. `-a test.align`: the name of the folder where the `.align` files will be created. If omitted, no `.align` files are created. The `.align` files can be used as input to Genetic Data Environment (GDE) (14).
3. `-f test.fasta`: the name of the multi-fasta file that contains the consensus sequence of all contigs. If omitted, the file is not created.
4. `-n test`: the prefix for the names given to contigs in both the `.fasta` file and the `.align` directory; in this case the contigs are called `test_1`, `test_2`, and so on and the files in the `.align` directory are called `test_1.align`, `test_2.align`, and so on.
5. `-g 8`: allows at most eight mismatches in any 32-base window; this parameter is meant to prevent false matches in long repeat sequences.
6. `-e 10`: the maximum length of mismatch at the end of the sequence (also called maximum overhang); see **Fig. 1**.
7. `-l 40`: the minimum length of overlap between two fragments.
8. `-p 97.5`: the minimum percent identity in the overlap region between two fragments.
9. `test.scratch`: the name of the scratch file—temporary file created by the assembler.
10. `test.seq`: the name of the input file (see **Subheading 2.** for more details).
11. `test.asm`: the name of the output file.

3.3. Assembly From Scratch

1. Before launching the assembler, you need to create a sequence and a quality file in the format specified in **Subheading 2.** Let us call these files `myseqs.seq` and `myseqs.qual`.
2. Using the `run_TA` script (see previous section) run the following command:

```
run_TA '-q myseqs.qual' myseqs.seq
```

Note the quotation marks in the command. They are necessary to pass the correct parameters to the TA. If you are not using the `run_TA` script, you do not need to include the quotes in the command.

3. The output from the assembler can be found in the following files:
 - a. `myseqs.fasta`: the multi-fasta file containing the consensus sequences of all contigs produced by the assembler.
 - b. `myseqs.asm`: the main output of the TA (see **Subheading 2.**).

3.4. Contig Jumpstart

Contig jumpstart allows you to hold together a set of previously assembled contigs. For example, you may have assembled a small set of sequences that span a repeat. If you try to assemble the whole genome without holding this repeat together, the sequences may be assembled on top of other copies of the repeat. Contig jumpstart allows you to avoid such a situation.

1. Before starting a contig jumpstart, you must obtain all the `.align` files corresponding to the contigs you want to hold together. Assume you have files `contig_1.align` and `contig_2.align`. You need to create a `.contig` file that will be used to jumpstart the assembly. Let us call this file `all.contig`; it is the concatenation of the corresponding `.align` files:

```
cat contig_1.align contig_2.align > all.contig
```

Note that you can use the same command for as many `.align` files as you wish.

2. At this point, `all.contig` contains all the contigs needed for the jumpstart. You next need to create a `.seq` and a `.qual` file containing all the sequences present in the `all.contig` file, plus any additional sequences you would like to add to the assembly. For example, `contig_1` and `contig_2` contain two repeats that you want to hold together, but you now want to assemble the whole genome. The `.seq` and `.qual` files must therefore contain all the sequences in the genome, besides those contained in the two held contigs.

Let us assume that we have all the sequences spanning the two repeats in the files `repeat1.seq`, `repeat1.qual`, `repeat2.seq`, `repeat2.qual` and all the other sequences in the genome appear in the files `others.seq`, `others.qual`. You need to first make sure that no two files contain any of the same sequences; that is, each sequence must appear in exactly one file. The following commands create the files necessary for running the assembler.

```
cat repeat1.seq repeat2.seq others.seq > all.seq
```

```
cat repeat1.qual repeat2.qual others.qual > all.qual
```

3. At this point you have all the necessary files: `all.seq`, `all.qual` and `all.contig`. The command for running the assembler is as follows:

```
run_TA '-C all.contig -q all.qual' all.seq
```

The `-C` option specifies the sequences that need to be kept together. The output is the same as in the case of an assembly from scratch.

3.5. Interfacing With Phrap and CAP3

The TA has the ability to create and read “old” format ACe files (15). This feature allows the user to read the output of the assembler into any program that can process ACe files. It is also possible to use an ACe file for contig jumpstart. Therefore, you can use the output of Phrap (10) or CAP (11,12) as input to the TA. The only caveat is that the TA can only handle the old ACe format; thus, you must be sure to specify that as an output option to Phrap or CAP (in the case of Phrap you can use the `-old_ace` option).

3.5.1. Creating ACe Output

The following command line produces a file `all.ace` in addition to the normal assembler output:

```
run_TA '-q all.qual -A all.ace -d' all.seq
```

When writing ACe output you can specify the PHD files (created by Phred) that are needed by Consed (16). This way the ACe file will correctly point to the appropriate PHD files.

```
run_TA '-q all.qual -A all.ace -d -D all.phd' all.seq
```

The relevant command line options are as follows:

1. `-A`: This is the “old” ACe format output. If the parameter was not passed, no ACe file will be created.
2. `-d`: Use the description line in the fasta file as the description line in the ACe file.
3. `-D`: This is the directory for `.phd` files.

3.5.2. Jumpstarting From an ACe File

Similar to the normal contig jumpstart, you can use the command

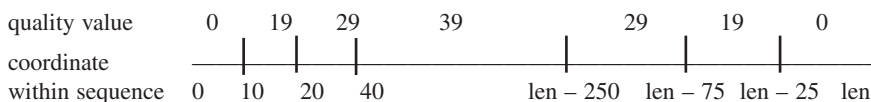
```
run_TA '-q all.qual -P all.ace' all.seq
```

to jumpstart on the output of Phrap or CAP. The `-P` command line option specifies the name of the ACe file.

4. Notes

1. The TA cannot process any sequence that is shorter than 32 bases because of the algorithm for computing rough alignments between input sequences.
2. If quality values are not provided, the TA assigns the following default quality values to the bases in the clear range. These values simulate the quality values you

would expect to come from the sequencer. It is preferable to use actual quality values if you have them.



3. If estimates of clone sizes are not specified in the `.seq` file, the assembler assigns them values 0, 10 000 000, and 1 600 for minimum, maximum, and median clone size, respectively.
4. The assembler produces a `.scratch` file while it runs. This file is erased on normal completion. If it remains after the assembler has exited, this is a clear indication that an error has occurred. You need to examine the `.error` file to find out more about the error condition.
5. The `.error` file contains a variety of warning and error messages. Error messages (starting with the string `ERROR`) indicate error conditions that cause the assembler to abort its execution. These error messages are fatal and generally cause the assembler to exit. Warning messages are for information only; however, they indicate inconsistencies in the input data and therefore should be checked because they may also indicate problems in the output. Here is a summary of the most common errors:

```
ERROR: Could not allocate memory for ...
```

The assembler needs more memory than available on your system. Try running the assembler on a computer with more memory.

```
ERROR: Could not make directory ...
```

```
ERROR: ... output file ... is not writeable
```

```
ERROR: ... input file ... is not readable
```

```
ERROR: could not create scratch file
```

These error messages refer to either incorrect permissions in the current directory or to a full disk.

```
ERROR: Sequence header line is not properly
formatted in ...
```

```
ERROR: Contig header line is not properly formatted!
```

The input file formats do not follow the format specified in **Subheading 2**.

```
ERROR: Contig ... was not supported by any
underlying sequences
```

The `.contig` file does not contain any sequences, other than the consensus. You should make sure you have the correct file.

```
ERROR: alignment range ... is not within clear
range ... for ...
```

The `seq_lend` and `seq_rend` parameters specified in the `.seq` file are inconsistent with the clear range specified in the `.seq` file. This problem is usually created when the `.seq` files are edited after the `.align/.contig` files are generated.

```
ERROR: Fewer quality values than nucleotides:
ignoring ...
```

```
ERROR: More quality values than nucleotides:
ignoring ...
```

For a sequence, the entry in the `.qual` file has a different number of quality values than the number of bases in the `.seq` file.

```
ERROR: seq_name ... in contig ... not found in
input!
```

A sequence specified in a `.contig` file was not found in the corresponding `.seq` file. You must always include all sequences from the `.contig` file into the `.seq` file.

```
ERROR: Contig header length ... does not agree with
actual contig length ...
```

```
ERROR: Contig header line num_seqs field ... is
greater/less than the actual number of sequences for
the contig ...
```

The data in the `.contig` file contain inconsistencies. Make sure that this file is obtained from an assembly run and is not corrupted.

```
ERROR: Line too long in ...
```

```
ERROR: Can't handle sequences longer than ...
```

```
ERROR: Can't handle more than ... sequences ...
```

You have reached the limits of the assembler. Sequences cannot be longer than 65,536 bases, and the maximum number of sequences is 524,288.

Here is a summary of the most common warnings:

```
WARNING: Input file sequence names and quality values
file sequence names are not in the same order ...
```

```
WARNING: Assuming no quality values for sequence ...
```

```
WARNING: Fewer sequences in quality values file than
in sequence file!
```

These warnings are common if you omit the quality values for a particular sequence.

```
WARNING: Sequence is too short ...
```

The assembler cannot handle any sequences shorter than 32 bases.

```
WARNING: Unexpected character ... in ...
```

The assembler found characters it does not understand in the input. Make sure your files are not corrupted. This error could happen if you edited one of your files with a text editor that does not save files in plain-text format (such as Word).

```
WARNING: sequence ... in contig ... was already  
present ...
```

A sequence in your .contig file occurs in more than one contig. You must remove it from all but one contig or else the output will be inconsistent.

```
WARNING: characters in sequence ... exceeds number  
in alignment range ...
```

```
WARNING: characters in sequence ... exceeds number  
in contigs ...
```

```
WARNING: sequence ... in contig ... has fewer  
characters than expected!
```

The .contig file contains inconsistent information. Make sure that record headers agree with the sequence data following them.

```
WARNING: sequence ... in contig ... appears to  
have been edited.
```

The sequence appearing in the .contig file disagrees with the sequence from the .seq file, thus indicating that the .seq file was edited.

```
WARNING: first ... positions of contig were not  
supported by any underlying sequences ...
```

```
WARNING: last ... positions of contig were not  
supported by any underlying sequences ...
```

```
WARNING: contig ... positions ... were not  
supported by any underlying sequences ...
```

Some of the sequences from the .contig file were removed and therefore the consensus is no longer supported by sequence data. This is not a fatal error and can sometimes be useful in breaking up contigs that have been misassembled. If you remove some sequences in the middle of the contig, the output of the assembler will contain two contigs, one for each contiguous piece of your contig.

6. The progress of the assembler can be gauged by examining the error file using the UNIX command: `tail -f asm.error` (in which asm is the name of the

.seq file passed as input). The assembler reports how many sequences were merged, and how many potential alignments are being resorted. The corresponding .error file entries have the following format:

```
merged(29)
resorting 668 matches
```

The assembler finishes its job when all sequences are merged; thus, the difference between the current number of merges and the total number of input sequences is an indication of the amount of work left. The total number of sequences being assembled is present on the first line of the .error file:

```
input stats: num_seqs 69, tot_length 36278,
max_length 763, min_length 105, ave_length 525
```

7. The percent similarity parameter specified on the command line gets internally converted into a Smith-Waterman alignment score, which is adjusted further by the algorithm. Therefore, in the output of the assembler, you may find pairs of sequences that have a lower percent similarity than that passed as the parameter to the assembler. The same observation holds for the length of the maximum overhang.
8. If seq_lend is greater than seq_rend in either the .asm file or the .align or .contig file, the sequence is reverse complemented in the alignment. At all times asm_lend is smaller than asm_rend.
9. When performing contig jumpstarts you must be careful to have consistent data between the clear range specified in the .seq file and the alignment range specified in the .contig file. Inconsistencies occur when you edit the sequence records after a .align file is created. In case of inconsistency, the assembler exits with an error. The only solution, in this case, is to reassemble each of the contigs using the new sequence records.
10. The order of contigs in an assembly may vary between assembly runs. You can use the checksum to find the correspondence of contigs between assemblies.
11. Although the .align files produced by the assembler are compatible with the input to the GDE program, the reverse is not true: the files created by GDE cannot be used to jumpstart the assembler.

References

1. Sutton, G. G., White, O., Adams, M. D., and Kerlavage, A. R. (1995) TIGR Assembler: a new tool for assembling large shotgun sequencing projects. *Genome Sci. Technol.* **1**, 9–19.
2. Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S. L., and Quackenbush, J. (2000) An optimized protocol for analysis of EST sequences. *Nucleic Acids Res.* **28(18)**, 3657–3665.
3. Pevzner, P., Tang, H., and Waterman, M. S. (2001) A new approach to fragment assembly in DNA sequencing, in *Proceedings of the Fifth Annual International Conference on Computational Biology (RECOMB)*. (Istrail, Lengauer, Pevzner,

- Sankoff, and Waterman, eds.), Association for Computing Machinery, Montreal, Canada. pp. 256–265.
4. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185.
 5. Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.
 6. Staden, R. (1979) A strategy of DNA sequencing employing computer programs. *Nucleic Acids Res.* **6**(7), 2601–2610.
 7. Tarhio, J. and Ukkonen, E. (1988) A greedy approximation algorithm for constructing shortest common superstrings. *Theoret. Comput. Sci.* **57**, 131–145.
 8. Smith, T. F. and Waterman, M. S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197.
 9. Gribskov, M., McLachlan, A. D., and Eisenberg, D. (1987) Profile analysis: detection of distantly related proteins. *Proc. Natl. Acad. Sci. USA* **84**, 4355–4358.
 10. Green, P. <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>.
 11. Huang, X. (1996) An improved sequence assembly program. *Genomics* **33**, 21–31.
 12. Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.* **9**(9), 867–877.
 13. Pearson, W. R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **183**, 63–98.
 14. Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W., and Gillevet, P. M. (1994) The genetic data environment and expandable GUI for multiple sequence analysis. *Comput. Appl. Biosci.* **10**(6), 671–675.
 15. Thierry-Mieg, J. and Durbin, R. (1992) ACEDB—a C. elegans database: syntactic definitions for the ACEDB data base manager, 1992 (www.acedb.org).
 16. Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202.

Finishing “Working Draft” BAC Projects by Directed Sequencing With ThermoFidelase and Fimers

Andrei Malykh, Olga Malykh, Nikolai Polushin,
Sergei Kozyavkin, and Alexei Slesarev

1. Introduction

A typical “working draft” bacterial artificial chromosome (BAC) project consists of approx 2000 shotgun reads obtained from a library of M13 or plasmid subclones that provide 2–5X coverage of BAC sequence. The reads are assembled into 5–50 long contigs (>2 kb), and hundreds of smaller contigs and singletons. The majority of the remaining gaps in the sequence are shorter than 1 kb. The utility of the working draft sequence is limited by the presence of the low-quality islands in the contigs, misassemblies, and by the presence of contaminating reads and contigs that originate from different BAC clones. One of the methods of finishing BAC projects consists of the production of additional shotgun libraries, obtaining 2000–4000 additional shotgun reads followed by the specialized finishing methods. The finishing methods require careful storage of the subclone libraries, optional shipping of them to the specialized facility, cherry picking of hundreds of subclones for end or directed sequencing, polymerase chain reaction (PCR) sequencing, and potentially even more specialized techniques.

In this chapter, we describe an alternative finishing method that does not require subclone libraries and can be accomplished with 100–300 directed sequencing reactions off a BAC DNA template. A key component of the directed sequencing procedure is the robust sequencing reaction customized for BAC DNA that gives high-quality reads. The reaction is based on highly sensitive BigDye terminator sequencing chemistry (*I*) and incorporates ThermoFidelase

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

(2) (*see Note 1*) and specially modified primers (Fimers) (3) (*see Note 2*) with greatly improved specificity and template-annealing characteristics (4).

The principal component of ThermoFidelase is a thermostable topoisomerase (5) that has a unique combination of activities that are not found in any other protein: it enzymatically unlinks DNA double helix, stimulates primer annealing and extension, and protects DNA from thermal decomposition (6). Incorporation of ThermoFidelase into sequencing protocols has resulted in successful sequencing of many DNA samples that for different reasons have not been sequenced previously. Examples include GC- and AT-rich plasmid samples; strong stop regions; long, simple mono-, di-, and trinucleotide repeats; multiple hairpin repeats; and direct sequencing of microbial genomic DNA (www.fidelitysystems.com)

2. Materials

2.1. Selection of Primers and Assembly of Shotgun and Directed Sequencing Data

1. Software: Phred/Phrap/Consed sequence assembly software is required for the procedures described in this chapter. These programs can be obtained from www.phrap.org. Windows and Macintosh versions of Phred and Phrap (but not Consed) are offered by CodonCode (www.codoncode.com). A primer-picking program, Primou, can be downloaded via ftp from the University of Oklahoma ACGT ftp site (<ftp://ftp.genome.ou.edu/pub>). The Image program is required for DNA restriction fragment analysis. It is freely distributed by the Sanger Center (www.sanger.ac.uk/Software/Image).
2. Hardware: The Phred/Phrap/Consed package is available for major Unix platforms: Sun SPARC Solaris (2.5.1 or better), Compaq Alpha Digital Unix (OSF1 V4.0 or better), HP HP-UX (11.0 or better), and SGI Irix (6.2 or better). The programs can also be installed on an Intel PC running a Linux operating system (RedHat 5.2 or better).

2.2. Preparation of BAC DNA

1. Shaker-incubator.
2. Equipment for agarose gel electrophoresis.
3. Plasmid purification (midi) kit (Qiagen).
4. Tabletop centrifuge.
5. Ultraviolet (UV) spectrophotometer.

2.3. BAC Sequencing

1. Automated DNA Sequencer (Applied BioSystems, Amersham Pharmacia Biotech, Li-Cor, or Beckman).
2. Thermocycler.
3. Heat sealer.

4. Centrifuge with rotor for microtiter plates.
5. Vacuum centrifuge with rotor for microtiter plates.
6. BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems).
7. ThermoFidelase 2 (Fidelity Systems).
8. Fimers (Fidelity Systems).
9. 7-Deaza-dGTP (Roche).
10. 96- or 384-Well polyethylene or polypropylene plates (MJ Research).
11. Multichannel pipets.
12. Sealing foil (MJ Research).
13. MicroSpin™ G-50 columns (Pharmacia Biotech).
14. Sephadex G-50 Superfine (Sigma, St. Louis, MO).
15. Millipore MultiScreen 96-well filtration plates (Millipore, Bedford, MA).
16. Loading buffer: 85% formamide, 5 mM EDTA (pH 8.0), 10 mg/mL of Blue dextran.

2.4. Restriction Digestion of BAC DNA

1. Restriction enzymes *Hind*III, *Eco*RV, and *Xho*I (New England BioLabs).
2. Controlled-temperature water bath.
3. LE agarose (Sigma).
4. 10X TAE buffer (Life Technologies, Gaithersburg, MD).
5. TE buffer: 10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA.
6. Agarose gel-loading dye (Quality Biologicals).
7. High-molecular-weight DNA markers (Invitrogen).
8. 1-kb Plus DNA ladder (Life Technologies).
9. SYBR Green (Molecular Probes).
10. Sub-Cell model 192 apparatus for agarose gel electrophoresis (Bio-Rad, Hercules, CA).
11. Peristaltic pump.
12. FluorImager (Hitachi, Applied BioSystems, or Bio-Rad).

3. Methods

We have implemented our finishing strategy with ThermoFidelase 2 and Fimers in a high-throughput environment on 10 human BAC “working draft” sequences (i.e., with 3–5X coverage) produced at the Baylor College of Medicine Human Genome Sequencing Center. Our approach consisted entirely of constructing primers at the draft contig ends and around low-quality regions, followed by sequencing directly off BAC DNA to extend into gap regions and cover low-quality areas. For contig assembling, we used Phred (7,8), Phrap (unpublished), and Consed (9,10) programs. The draft projects were in different stages of redundant shotgun sequencing, with Phred quality 20 (Q20) coverage ranging from 2.25 to 4.84. In analyzing the quality of final contigs, we relied on a cumulative error calculated by Consed and the number of low-quality regions in contigs, i.e., those having Phrap quality <30. Completed

sequences were verified by comparing the calculated *Hind*III map with experimental ones produced by the Washington University Genome Sequencing Center and found to have no detectable misassemblies.

3.1. Preparation of BAC Shotgun Sequencing Reads for Finishing

This job can be performed in several different ways, depending on the computer setup in your laboratory. For the purpose of this chapter, it is assumed the user has the Phred/Phrap/Consed (v.11) package installed on a computer running a Unix/Linux operating system in the default directory `/usr/local/genome`, and that a directory with the Primou program is in the system path. It is also assumed that a BAC project has $\geq 4X$ Phred Q20 coverage (see **Note 3**). Novice users should read the Consed manual for details on how to install and run Phred, Phrap, and Consed programs.

1. Create the following directory structure:
 - a. Top-level directory (named after the BAC).
 - b. Subdirectory `chromat_dir`—chromatograms are stored here.
 - c. Subdirectory `phd_dir`—this will be filled with `phd` files automatically.
 - d. Subdirectory `edit_dir`—various Phrap and Cross_match files will be put here.
 - e. Subdirectory `primou_dir`—primou and auxiliary files should be put there.
2. Change to the `edit_dir` directory, and type `phredPhrap -forcelevel 0`. This perl script will run Phred, Cross_match, and Phrap programs that assemble sequencing chromatograms into contigs.
3. Make soft links in the `primou_dir` directory to `<BAC_project_name>.fasta.screen.contigs` and `<BAC_project_name>.fasta.screen.contigs.qual` files located in the `edit_dir` by typing the following from the `primou_dir` directory:
 - a. `ln -s ../edit_dir/<BAC_project_name>.fasta.screen.contigs`
`<BAC_project_name>.fasta.screen.contigs`
 - b. `ln -s ../edit_dir/<BAC_project_name>.fasta.screen.contigs.qual`
`<BAC_project_name>.fasta.screen.contigs.qual`

Make sure that files `human.rep` and `oligo.cri` (both come with the Primou program) are in the `primou_dir` directory. The file `human.rep` contains L1, THE, and ALU repeats. Add any other repeats you think may be present in your BAC sequence.
4. Modify Primou's preference file `oligo.cri` as follows (see **Note 4**):

PRIMOU'S PREFERENCES FILE

25 22 28	#oligo length
60.0 55.0 85.0	#Tm
30.0 80.0	#CG content

50.0 50.0	#salt and DNA concentration
14	#acceptable primer self-complementarity
3	#acceptable 3' end primer self-complementarity
0	#required number of GC clamps in primer
0	#number of low-quality bases (Ns) allowed in primer
20	#minimum acceptable value of base quality
30	#number of bases in sliding window for clipping
3	#number of low-quality bases allowed in window for clipping
60	#number of bases in buffer zone from end of clip
250	#maximum size of region for primer selection
1	#number of primers to select (per direction)
15	#acceptable primer complementarity to preselected oligos
0	#bases to skip after each acceptable primer
500	#read length (used with -cover option)
50	#start of read (used with -cover option)
20 30	#low- and high-quality cutoff scores (used with -cover option)
5. Run Primou from the `primou_dir` directory by typing the following command:

```
primou <BAC_project_name>.fasta.screen.contigs -all
-global -garbage
```

The program will create the file `<BAC_project_name>.fasta.screen.contigs.primers` with primer sequences selected at the ends of contigs of more than 2 kb.

6. Contact Fidelity Systems by e-mail to place an order for Fimers. Include the `<BAC_project_name>.fasta.screen.contigs.primers` file with your order.
7. Run directed sequencing reactions with Fimers on BAC DNA as described in **Subheadings 3.5.–3.7.**
8. Copy new chromatograms into the `chromat_dir` directory and run `phredPhrap` script as in **step 2**.
9. Repeat **steps 2–9** until Primou selects no new primers.
10. While in the `edit_dir` directory, bring up Consed and identify contigs that do not contain directed sequencing reads (*see Note 5*). These contigs are likely assembled from sequencing reads that are not related to your BAC project (“floating” contigs).
11. Create in the `edit_dir` directory a file named `.consedrc` containing the following resources (*see Note 6*):


```
consed.autoFinishMinNumberOfErrorsFixedByAnExp: 0.5
consed.primersMaxMeltingTemp: 80
consed.primersPickTemplatesForPrimers: false
consed.primersSubcloneFullPathnameOfFileOfSequences
  ForScreening: /usr/local/genome/lib/screenLibs/vector.seq
consed.primersCloneFullPathnameOfFileOfSequencesFor
  Screening: /usr/local/genome/lib/screenLibs/vector.seq
```

```
consed.primersMinMeltingTemp: 55
consed.primersMinMeltingTempForPCR: 60
consed.primersNumberOfBasesToBackUpToStartLooking: 50
consed.primersOKToChoosePrimersInSingleSubcloneRegion:
    true
consed.autoFinishCallReversesToFlankGaps: false
consed.autoFinishAllowWholeCloneReads: true
consed.autoFinishAllowCustomPrimerSubcloneReads: false
consed.autoFinishAllowDeNovoUniversalPrimerSubclone
    Reads: false
consed.autoFinishAllowMinilibraries: false
consed.autoFinishAllowPCR: false
consed.autoFinishAllowResequencingAUniversalPrimerAuto
    finishRead: false
consed.autoFinishAlwaysCloseGapsUsingMinilibraries:
    false
consed.autoFinishMaximumFinishingReadLength: 450
consed.primersWindowSizeInLooking: 300
consed.primersAssumeTemplatesAreDoubleStrandedUnless
    Specified: true
consed.autoFinishAllowResequencingReadsToExtendContigs:
    false
consed.autoFinishCloseGaps: false
consed.autoFinishContinueEvenThoughReadInfoDoesNotMake
    Sense: true
consed.autoFinishCostOfCustomPrimerCloneReaction:
    25.000
consed.autoFinishCoverSingleSubcloneRegions: true
consed.autoFinishCoverLowConsensusQualityRegions: true
consed.autoFinishDoNotAllowWholeCloneCustomPrimerReads
    CloserThanThisManyBases: 150
consed.autoFinishExcludeContigIfOnlyThisManyReadsOr
    Less: 8
consed.autoFinishExcludeContigIfTooShort: true
consed.autoFinishExcludeContigIfThisManyBasesOrLess:
    2000
consed.primersMinNumberOfTemplatesForPrimers: 100
consed.autoFinishMinBaseOverlapBetweenAReadAndHigh
    QualitySegmentOfConsensus: 50
consed.autoFinishPrintForwardOrReverseStrandWhen
    PrintingSubcloneTemplatesForCustomPrimerReads: false
consed.autoFinishPrintMinilibrariesSummaryFile: false
consed.autoFinishNearGapsSuggestEachMissingReadOfRead
    Pairs: false
```

```

consed.primersMinimumLengthOfAPrimer: 20
consed.primersMaximumLengthOfAPrimer: 30
consed.primersMaxLengthOfMononucleotideRepeat: 6
consed.primersChooseTemplatesByPositionInsteadOf
  Quality: false
consed.primersWhenChoosingATemplateMinPotentialRead
  Length: 10000
consed.autoFinishEmulate9_66Behavior: false
consed.autoFinishDoNotAllowWholeCloneCustomPrimerReads
  CloseTogether: true
consed.autoFinishMinNumberOfSingleSubcloneBasesFixed
  ByAnExp: 10
consed.primersMaxMatchElsewhereScore: 20
consed.primersMaxSelfMatchScore: 14
consed.primersMinQuality: 20
consed.primersPrintInfoOnRejectedTemplates: false
consed.primersScreenForVector: true

```

12. Select primers to cover single-stranded and low-quality regions by running Consed with the following options:

```

consed -autofinish -ace <BAC_project_name>.fasta.
  screen.ace.<N>

```

in which <N> is the number of the latest ace file. This will create several files in the edit_dir directory, from which you need only the *.customPrimers file.

13. Edit the *.customPrimers file to remove all primers targeted “floating” contigs (see **step 10**) as well as any previously selected primers and save it as follows:

```

<BAC_project_name>.customPrimers.ace.<N>

```

14. Contact Fidelity Systems by e-mail to place an order for Fimers. You will need to include the <BAC_project_name>.primers.ace.<N> with your order.
15. Run directed sequencing reactions with Fimers on BAC DNA as described in **Subsections 3.5–3.7**.
16. Add new reads to the chromat_dir directory and reassemble the project with new reads by running phredPhrap script with default options. Depending on the complexity of your BAC sequence, your finishing standards, and the initial Q20 shotgun coverage, you may wish to repeat **steps 12–15** (see **Note 7**). At the end you should have one long contig flanked at both ends on BAC cloning vector sequences as well as several “floating” short contigs and unassembled reads.

3.2. Verification of Finished BAC Sequences

The correctness of an assembled BAC sequence can be verified by digestion with restriction enzymes and/or by BLAST analysis (using public or commercial databases).

1. Use the following restriction enzymes: *Hind*III, *Eco*RV, and *Xho*I. For each of these enzymes set up the following reaction: 1.0 μ L of 10X buffer, 1.0 μ L of enzyme (10 U/ μ L), BAC DNA (0.1 μ g of DNA) not to exceed 1 μ L, and 7 μ L of sterile water. When setting up the reactions keep the enzymes on ice or in a cooler.
2. Incubate the reactions at 37°C for 1h.
3. Prepare a large 1.0% agarose gel (20 wells) using 1X TAE. The gel is 250 mL and is made by adding 2.5 g of LE agarose to 250 mL of 1X TAE.
4. Prepare 2 L of 1X running buffer.
5. Prepare DNA markers for the gel as follows: 20 μ L of DNA standard, 17.0 μ L of TE (10:0.1), and 3 μ L loading dye. (DNA standard: Mix 1 μ L of 1-kb plus ladder (1 μ g/ μ L), 1 μ L of high-molecular-weight marker, 92 μ L of TE buffer, and 6 μ L of loading dye.)
6. To each 9- μ L digest add 2 μ L of loading dye. Load a lane of marker, then the three digests for each BAC DNA, followed by another lane of marker. Marker should be in every fifth lane. Repeat for as many BACs as you have. Run the gel at 90 V for 8 h using water circulation.
7. Stain the gel in 400 μ L of distilled water with 50 μ L (1 tube) of SYBR Green for 1 h.
8. Scan the gel on a FluorImager.
9. Process the scanned gel using an image program (see **Note 8**).
10. Compare obtained experimental restriction maps of your BAC with theoretical ones generated by your favorite DNA analysis software (e.g., VectorNTI from Informax).

3.3. Isolation of BAC DNA

Qiagen kits are the most effective tools for BAC DNA purification. A typical yield of DNA from 100 mL of overnight *Escherichia coli* culture using a Qiagen midiprep kit is about 30 μ g. This amount is sufficient for more than 100 sequencing reactions. BAC DNA is isolated according to the manufacturer's protocol. We recommend resuspending DNA in 100–200 μ L of H₂O or 10 mM Tris-HCl (pH 8.0) after the isolation of DNA according to Qiagen protocols.

3.4. Estimation of BAC DNA Quantity

DNA quantity can be estimated either by agarose gel electrophoresis or by measuring the absorbance of a DNA solution at 260/280 nm. A concentration of BAC DNA between 200 ng/ μ L and 1 μ g/ μ L is convenient for small-volume sequencing reactions.

1. Make a 1:100 dilution of DNA with water.
2. Measure the absorbance at 260 nm in a laboratory spectrophotometer.
3. Calculate the DNA concentration using the following formula (see **Note 9**):

$$\text{OD}_{260} \times 50 \times 100 = \text{ng}/\mu\text{L}$$

4. Load three wells with 0.25, 0.5, and 1 μg of DNA on a 1% agarose gel and run at 10–15 V/cm for at least 30 min. As a control, use DNA ladder with known concentration (e.g., DNA Marker II from Roche). By comparing the intensity of BAC DNA lanes and marker lanes you can estimate the quantity of BAC DNA in the sample. If this estimate is close to the optical density-based calculation proceed with sequencing.
5. If you see less DNA on the gel than expected, repeat UV measurement. You may want to precipitate DNA with ethanol and then repeat **steps 1–5**.
6. Test the DNA preparation by sequencing with one of the Fimers designed for end sequencing. Standard procedure includes sequencing three quantities of DNA (0.25, 0.5, and 0.75 μg) in order to estimate the optimal amount of template per reaction.

3.5. Cycle Sequencing With BigDye Terminators

For sequencing of a single BAC with a few Fimers (<96), use the following to calculate the amount of each component.

1. BAC DNA: 0.1–0.5 μg .
2. ThermoFidelase 2: 0.1 μL .
3. 1 mM 7-Deaza-dGTP: 0.1 μL .
4. Fimer: 1.0 μL .
5. Dye Terminator Mix: 2.0 μL .
6. dH_2O to a total vol of 5.0 μL .
7. Total volume: 5.0 μL .

Since the volume of ThermoFidelase 2 added to the reaction is very small, a single mix for at least three sequencing reactions should be prepared (*see Note 10*). These reactions can be done in individual 0.2-mL PCR tubes, 8-tube strips, or 48- or 96-well plates. If a heat sealer is used, plates can be reused several times. To compensate for pipetting errors, we recommend adding 10% more reagents. An example of a 96-reaction setup is as follows:

1. BAC DNA (250 ng/reaction): 26.5 μg .
2. ThermoFidelase 2: 0.6 μL .
3. 1 mM 7-Deaza-dGTP: 10.6 μL .
4. Dye Terminator Mix: 212.0 μL .
5. dH_2O to a total vol 424.0 μL .
6. Total volume: 424.0 μL .

Then follow these steps for the reaction:

1. Label the tube that is going to be used for the reaction setup.
2. Add the desired amount of template to this tube based on the calculated DNA concentration.

3. Add ThermoFidelase 2 and mix by pipetting.
4. Add Dye Terminator Ready Reaction Mix.
5. Add the calculated amount of water.
6. Mix well by pipetting. The total volume of the premix should be equal to $4\ \mu\text{L} \times$ number of reactions.
7. Dispense $4\ \mu\text{L}$ per tube/well (*see Note 11*).
8. Add Fimers ($1\ \mu\text{L}$ each) using a multichannel pipet. Mix by pipetting.
9. Close the tubes and seal the plate.
10. Spin the tubes and plate briefly.
11. Place the tubes and plate in a thermocycler.
12. Use the following cycling conditions:
 - a. Heat the plate at 95°C for 2 min.
 - b. 95°C for 5 s (denaturation).
 - c. 55°C for 20–30 s (annealing).
 - d. 60°C for 4 min (extension).
 - e. 100 cycles total (*see Note 12*), then 4°C .
 - f. Add $15\ \mu\text{L}$ of 20 mM EDTA to each well.

3.6. Reaction Cleanup Using Millipore MultiScreen 96-Well Filtration Plates

Prepare Millipore MultiScreen 96-well filtration plates with Sephadex G-50 Superfine according to the manufacturer's recommendations.

1. Place the MultiScreen plate on top of a clean 96-well plate.
2. Transfer the entire 20- μL reaction mix onto a Sephadex column.
3. Centrifuge at 910g for 5 min.
4. Check the sample volume eluted from the columns. It should be about 20 μL /well.
5. Dry the samples for 20–30 min in a vacuum centrifuge.
6. Resuspend each pellet in 3 μL of loading buffer (85% formamide; 5 mM EDTA, pH 8.0; 10 mg/mL of Blue dextran).
7. Load 1 μL of each sample per lane on sequencing gel.
8. Run the sequencing gel as recommended by Applied Biosystems.

3.7. Reaction Cleanup Using MicroSpin™ G-50 Columns

Prepare the necessary number of MicroSpin G-50 columns (Pharmacia Biotech) according to the manufacturer's instructions.

1. Place a MicroSpin column in a new 1.5-mL Eppendorf tube.
2. Load the entire 20- μL reaction mix onto the column.
3. Spin the column for 2 min at 735g.
4. Discard the column.
5. Dry the sample in a vacuum centrifuge for 20 min.
6. Resuspend the pellet in 3 μL of loading buffer.

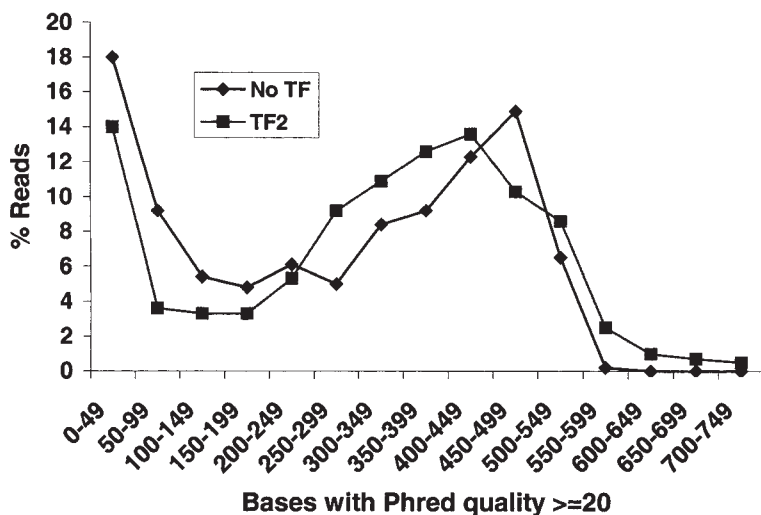


Fig. 1. Effect of TF1 on chromatogram quality and read length. Four hundred eighty-nine directed reads from BAC clones without ThermoFidelase were compared with directed reads from the same BAC DNA with ThermoFidelase 2.

4. Notes

1. A new version of ThermoFidelase (ThermoFidelase 2) was developed to be compatible with Fimers and a small-volume BAC sequencing reaction. It retains the properties of ThermoFidelase I, and, in addition, these proteins accelerate the kinetics of Fimer annealing to complementary targets. **Figure 1** shows that ThermoFidelase 2 decreases the number of low-quality chromatograms (reduces the failure rate of sequencing reactions) and increases the quality of directed BAC reads.
2. To increase the outcome of directed BAC sequencing reaction and to reduce the consumption of BAC DNA, we extended the number of thermal cycles to 100. We use Fimers to inhibit two nonspecific events that interfere with the sequencing reaction: primer-dimer extension and nonspecific PCRs (4). The chemical modifications allow one to choose a Fimer sequence to target the exact site dictated by the project for which most of the regular primers would not work (and will not be chosen by standard primer selection programs). Fimer chemistry is based on the synthesis of a precursor oligonucleotide containing highly reactive methoxyoxalamido (MOX) or succinimido (SUC) moieties attached to the 2' positions of ribose residues (3). On postsynthetic treatment of the precursor with a desirable modifier, the reactive moieties are effectively derivatized and the final 2'-modified oligonucleotide (Fimer) is formed (**Fig. 2**). This strategy enables one to synthesize a wide variety of modified Fimers from a single parent oligonucleotide. Both MOX- and SUC-reactive moieties are stable enough to allow efficient incorpora-

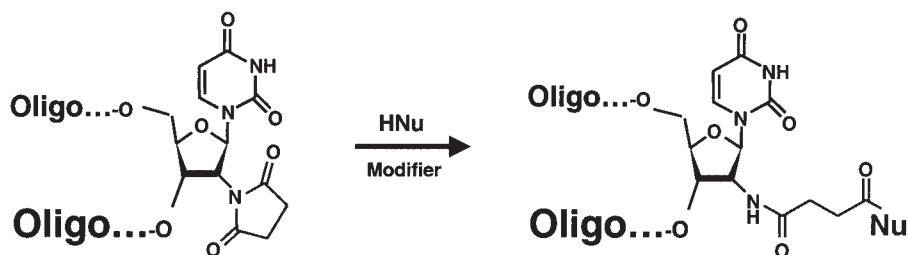


Fig. 2. Strategy for making Fimers with diverse properties in one oligonucleotide synthesis step. Precursor oligonucleotide containing highly reactive MOX or SUC moieties at the 2' positions of ribose residues is treated with a desirable modifier to form a 2'-modified oligonucleotide (or modified primer library).

tion of the corresponding monomer by the commonly used phosphoramidite solid-phase approach. On the other hand, these groups react rapidly and nearly quantitatively with strong nucleophiles, such as primary aliphatic amines, ammonia, and hydroxyl anion, to form stable 2' derivatives. **Figure 3** schematically illustrates the inhibition of nonspecific PCR by Fimers in a cycle sequencing reaction. We found that one modification located from three to seven residues from the 3' end of an oligonucleotide shows excellent priming ability without yielding exponential amplification byproducts. We have extensively tested Fimers and ThermoFidase 2 in BAC sequencing reactions, and the results indicate that the effects of chemical modifications and protein addition are synergistic and that the protein binds duplexes with Fimers.

3. This is not an absolute requirement. We were able to finish a BAC sequence with as low as 2.25 initial Q20 shotgun coverage.
4. These parameters are good starting points. It may be necessary to change some of them to accommodate the specifics of your sequencing project(s).
5. The necessary information is also contained in a text file <BAC_project_name>.phrap.out.
6. You can view current Consed resources by clicking the Info button in the Consed main window. Copy and paste Consed parameters into your .consedrc file and then edit them as required.
7. If the error rate is still high but autofinish no longer suggests new primers, you may wish to raise the Consed parameter consed.primersMaxMatchElsewhereScore up to 30. In practice, it is this parameter that eliminates most primers. However, if you raise it, you should be aware of the danger of mispriming.
8. Consult the Sanger Web site (www.sanger.ac.uk/Software/Image) on how to use the Image program.
9. This estimate of DNA concentration, however, is not sufficient in our opinion. The reason for the gel electrophoresis estimate is that it shows the contamination

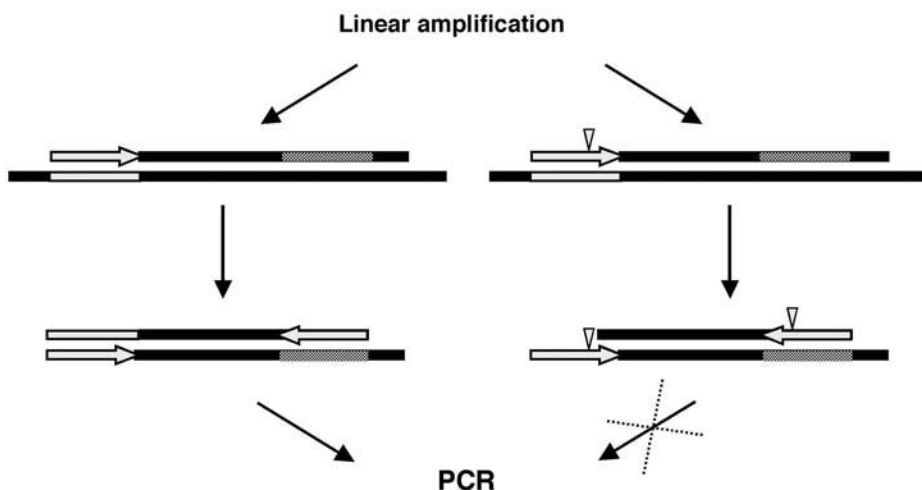


Fig. 3. Scheme illustrating suppression of PCR in sequencing reaction using Fimers. Solid dark lines are the template and newly synthesized Sanger fragments. Gray arrows and gray arrows with attached triangles are primers and fimers, respectively. Gray and checker-filled rectangles represent the primary and secondary priming sites, respectively, in the Sanger fragments.

of BAC DNA by chromosomal DNA and RNA. BAC preparations may contain *E. coli* chromosomal DNA. Although its presence is not critical, you may wish to decrease the denaturation temperature during cycle sequencing to 90°C if it is difficult to get rid of genomic DNA.

10. We suggest using a Rainin R2 pipet that can handle submicroliter volumes.
11. Dispensing the reaction mix into 96- and 384-well plates using an 8-channel pipet takes just a few minutes. First, dispense the reaction mix into an 8-tube strip. Then dispense the reaction mix into a 96- or 384-well plate from the strip. This procedure can be done in a small laboratory with no automation. Dispensing the reaction mix into a 384-well plate can be done with a 16-channel pipet from the tray.
12. More than 100 cycles is recommended if <50 ng of BAC DNA is used. In the 200 cycle sequencing protocol, extension time is 2 min, and for 400 cycles we recommend a 1-min extension. A 400 cycle sequencing reaction takes about 17 h.

Acknowledgments

We thank Steve Scherer, Donna Muzhny, and Richard Gibbs for providing working draft data on 10 BAC projects and helpful discussions. This work was supported in part by grants from the Department of Energy and the National Institutes of Health.

References

1. Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., et al. (1997) New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.* **25**, 4500–4504.
2. Slesarev, A. I., Belova, G. I., Lake, J. A., and Kozyavkin, S. A. (2001) Topoisomerase V from *Methanopyrus kandleri*. *Methods Enzymol.* **334**, 179–192.
3. Polushin, N. N. (2000) The precursor strategy: terminus methoxyoxalamido modifiers for single and multiple functionalization of oligodeoxyribonucleotides. *Nucleic Acids Res.* **28**, 3125–3133.
4. Polushin, N., Malykh, A., Malykh, O., et al. (2001) 2'-Modified oligonucleotides from methoxyoxalamido and succinimido precursors: synthesis, properties and applications. *Nucleosides Nucleotide Nucleic Acids* **20**, 75–78.
5. Belova, G. I., Prasad, R., Kozyavkin, S. A., Lake, J. A., Wilson, S. H., and Slesarev, A. J. (2001) A type IB topoisomerase with DNA repair activities. *Proc. Natl. Acad. Sci. USA* **98**, 6015–6020.
6. Kozyavkin, S. A., Pushkin, A. V., Eiserling, F. A., Stetter, K. O., Lake, J. A., and Slesarev, A. I. (1995) DNA enzymology above 100 degrees C. Topoisomerase V unlinks circular DNA at 80–122 degrees C. *J. Biol. Chem.* **270**, 13,593–13,595.
7. Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.
8. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185.
9. Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202.
10. Gordon, D., Desmarais, C., and Green, P. (2001) Automated finishing with autofinish. *Genome Res.* **11**, 614–625.

Optimized Multiplex Polymerase Chain Reaction

An Effective Method for Rapid Gap Closure

Diana Radune and Hervé Tettelin

1. Introduction

Implementation of the whole-genome shotgun sequencing approach to prokaryotes and eukaryotes necessitates methods for rapid gap closure. During a whole-genome shotgun sequencing project, DNA sequences are aligned against each other to create continuous assemblies or contigs (**1**). However, some areas of the genome are not represented owing to repeats, secondary structure, or toxicity in *Escherichia coli*. When no cloned template is available for sequencing, these gaps must be resolved by polymerase chain reaction (PCR), testing primers from each contig end against primers from all the other contig ends in all the possible combinations. When dealing with large genomes, generating templates for numerous gaps by only combinatorial PCR is cumbersome and inefficient. Optimized Multiplex PCR (**2**) provides an effective new method for the rapid closure of a large number of gaps. Its framework for minimizing the number of PCR pipetting steps is achieved by combining Multiplex PCR (**3**) with a mathematical method (**4**). Unlike combinatorial PCR, Optimized Multiplex PCR minimizes the number of reactions by combining primers into pools and testing these pools against each other. This method has been successfully implemented in the closure of numerous genomes at The Institute for Genomic Research. In this chapter, we provide a simple guide to the Optimized Multiplex PCR method by describing in detail the steps involved in primer design, primer pooling, Multiplex PCR, and multiplex sequencing. We also include a detailed interpretation of the Optimized Multiplex PCR results from the successful closure of a microbial genome.

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

2. Materials

Reagents for PCR and sequencing are available as kits (*see Note 1*). Use deionized autoclaved or Millipore (Bedford, MA) Milli-Q water to prepare all the solutions. Keep all reagents on ice at all times.

2.1. Optimized Multiplex PCR Reaction

1. Platinum *Taq* DNA Polymerase High Fidelity (5 U/ μ L (cat. no. 11304011; Invitrogen). This kit includes 10X High Fidelity PCR Buffer: 600 mM Tris-H₂SO₄ (pH 8.9), 180 mM ammonium sulfate, and 50 mM magnesium sulfate. Store this kit at -20°C.
2. 10 mM dNTP mix: 10 mM each of dATP, dCTP, dGTP, dTTP (cat. no. 18427013; Invitrogen). Store at -20°C.
3. 3 M Betaine anhydrous (C₅H₁₁NO₂) (EC no. 203-490-6; Sigma, St. Louis, MO), hygroscopic. Store at 2-8°C.
4. 5.8 μ M of each primer in the primer pool.
5. Genomic DNA (0.5 μ g).

2.2. Optimized Long-Range Multiplex PCR Reaction for Large Products

1. TaKaRa LA PCR Kit Ver.2.1 (cat. no. RR013; Intergen). This kit includes TaKaRa LA *Taq* (5 U/ μ L), dNTP mix (2.5 mM of each), 10X LA PCR Buffer II (25 mM Mg²⁺). Store this kit at -20°C.
2. 5.8 μ M of each primer in the primer pool.
3. Genomic DNA (-0.5 μ g).

2.3. Multiplex Sequencing

1. QIAquick PCR Purification Kit (250) (cat. no. 28106; Qiagen).
2. BigDye[®] Terminator V3.1 Cycle Sequencing Kit (cat. no. 4337455). This kit includes BigDye[®] Terminator V3.1 Cycle Sequencing Kit and BigDye[®] V1.1/3.1 Sequencing Buffer (5X). This reagent is light-sensitive. Store the kit at -25°C and the sequencing buffer at 4°C.
3. 0.6 μ M of each primer in the primer pool.
4. Purified PCR product (25 ng).

3. Methods

3.1. Optimized Multiplex PCR Primer Design

1. Design primers 200-500 bp from the ends of all contigs, pointing into the gap. Primers can be designed by hand following the melting temperature calculation rule, $2 \times A/T + 4 \times G/C$, or by any primer-designing software such as *primer3* (Copyright © 1996, 1997 Whitehead Institute for Biomedical Research. All rights reserved).

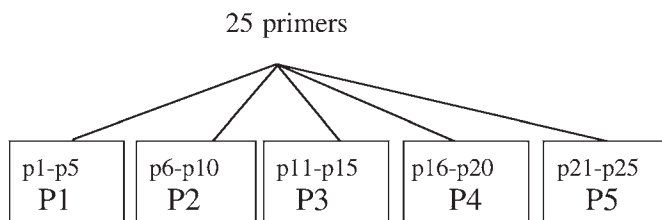


Fig. 1. Schema for primer pool preparation, in which p1–p25 are individual primers and P1–P5 are primer pools. The primers are arranged randomly between primer pools.

2. The GC content of the primers should be similar to the overall GC content of the genome. Design primers with two or three Gs or Cs at the 3' end. Avoid designing primers in areas with stretches of three or more purines or pyrimidines.
3. Primers 22–26 nucleotides in length with melting temperatures between 62 and 68°C are best for obtaining specific products in Multiplex PCR. Be sure to design all primers for Multiplex PCR with similar melting temperatures ($\pm 2^\circ\text{C}$) to allow annealing at the same temperature.
4. Avoid designing primers in the repetitive areas of the genome. To check for primer uniqueness, align them against the genome using any sequence alignment software.
5. Use the contig's number or ID in the naming of the primer designed from it. This will establish the location of the primer and simplify further primer tracking.

3.2. Creation of Multiplex Primer Pools

1. Use the following formula to calculate (K), the number of primers per multiplex PCR (2):

$$K = 2 \times \sqrt{n}$$

in which n is the total number of primers. For example, if the total number of primers is 25, there will be 10 primers per multiplex PCR, or 5 primers per pool for a total of 5 pools (*see Note 2*).

2. Arrange primers randomly into pools where each pool contains $K/2$ primers (*see Fig. 1*). To create a primer pool, add an equal amount of each primer so the concentration of each primer in the pool is $5.8 \mu\text{M}$. Keep very accurate notes of which primers are in which primer pool.
3. Set up Multiplex PCR reactions between primer pools. Each primer pool should be reacted with the other primer pools in all the possible combinations. For example, pool P1 should be reacted with pools P2–P5, pool P2 should be reacted with pools P3–P5, and so on (*see Fig. 2*).

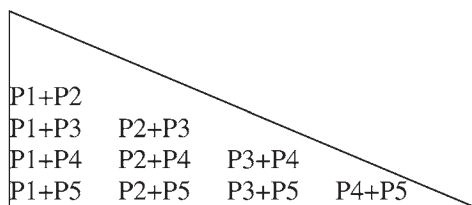


Fig. 2. Schema for setting up Multiplex PCR reactions. In the first reaction, primer pool P1 is reacted with primer pool P2; in the second reaction, pool P1 is reacted with pool P3; and so on.

3.3. Optimized Multiplex PCR Reaction

1. Prepare a reaction mix for one reaction with the following reagents: 1X High Fidelity PCR buffer, 2 mM MgSO₄, 0.2 mM of each dNTP mixture, 0.1–0.5 µg of genomic DNA, 0.5 M Betaine, 5 U of Platinum *Taq* DNA polymerase High Fidelity, and an equal amount of each primer pool so that the final concentration of each primer in the reaction tube is 0.35 µM. Add dH₂O to give a final volume of 50 µL (*see Note 3*). Mix the reagents gently; do not vortex.
2. Cap the tubes, quick spin in a centrifuge, and place in a thermocycler.
3. Set up cycling conditions as follows: an initial denaturation step of 95°C for 1 min, followed by a denaturation step of 94°C for 45 s, an annealing step of 60°C for 10 s, and an extension step of 68°C for 6 min (*see Note 4*). The latter three steps are repeated 30 times, followed by a final extension step of 72°C for 10 min. The reaction tubes should be kept at 4°C until they are removed from the thermocycler.

3.4. Optimized Long-Range Multiplex PCR Reaction for Large Products

1. If the expected products are larger than 5000 bases and the GC content of the genome is high, the following reaction conditions can be used: Make a reaction mix of 1X LA PCR Buffer II, dNTP mixture (400 µM of each), 2.5 U of TaKaRa LA *Taq*, 0.1–0.5 µg of genomic DNA, and an equal amount of each primer pool so that the final concentration of each primer in the reaction tube is 0.35 µM. Add dH₂O to give a final volume of 50 µL (*see Note 3*). Mix the reagents gently; do not vortex.
2. Cap the tubes, quick spin in a centrifuge, and set in a thermocycler.
3. Set up cycling conditions as follows: an initial denaturation step of 94°C for 1 min, followed by 14 cycles consisting of a denaturation step at 98°C for 20 s and annealing and extension steps at 68°C for 8 min (*see Note 4*). This is followed by another 16 cycles consisting of a denaturation step at 98°C for 20 s, an annealing step at 63°C for 15 s, and an extension step at 68°C for 8 min with 15s increments per cycle. The cycling is completed by a final extension step at 72°C for 10 min. Keep reaction tubes at 4°C until they are removed from the thermocycler.

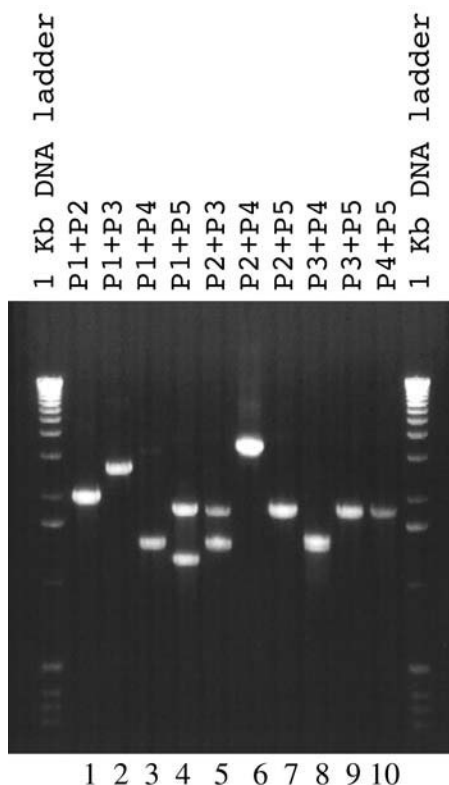


Fig. 3. Optimized Multiplex PCR results for 25 primers. Primers were arranged in five pools with five primers in each primer pool. Each lane represents a product made between two primer pools; for example, lane 1 is a product made by primer pools P1 and P2. A single product in lanes 1, 2, 3, and 6 represents a unique interpool product made by one primer from each primer pool. A top product in lane 4 and a single product in lanes 7, 9, and 10 are all of the same size, which suggests that this is an intrapool product, a product made by two primers from the same primer pool. In this case, the common primer pool present in all reactions is primer pool P5.

4. To view the results from the PCR, electrophorese the samples on a 1% agarose gel with ethidium bromide at 80 V for 1 h or long enough to obtain adequate separation of the PCR products. Use the 1-kb ladder as a marker to correctly size the PCR products.

3.5. Interpretation of Optimized Multiplex PCR Results

Several outcomes are possible from the Optimized Multiplex PCR (*see Fig. 3*):

1. No product is observed. This means that none of the primer pairs in the two pools involved in the reaction made a product. In this case, no further action is necessary.

2. A single PCR product is observed. This could be a unique product that is made by one primer from each pool, an interpool product, such as the products seen in lane 1 (P1 + P2), lane 2 (P1 + P3), lane 3 (P1 + P4), and lane 6 (P2 + P4) in **Fig. 3**. In this case, the PCR product could be multiplex sequenced (*see Subheading 3.6.*). When both primers involved in the reaction have been identified, run a confirmatory PCR with only these two primers using the same conditions from **Subheading 3.3.** or **3.4.** A single band could also be an indicator of an intrapool product, a product made by two primers from the same pool (*see Note 5*). In this case, a PCR product of the same size will be found in every reaction in which the particular pool is present. For example, in **Fig. 3** two intrapool primers from pool five (P5) create a product that is seen in every reaction where pool five (P5) is present; i.e., in lane 4 (P1 + P5, the top product), lane 7 (P2 + P5), lane 9 (P3 + P5), and lane 10 (P4 + P5). To determine which two primers in one pool made this product, run a secondary set of PCR reactions by eliminating one primer at a time from the pool (*see Fig. 4*). The disappearance of the band on the agarose gel will suggest that the missing primer is necessary for product formation. For example, in **Fig. 4** products are absent from lane 2 (primer p22 is missing) and lane 3 (primer p23 is missing); therefore, primers p22 and p23 were involved in synthesis of the P5 intrapool product. Run a confirmatory PCR with only these two primers using the same PCR conditions from **Subheading 3.3.** or **3.4.**
3. When two PCR products are observed, they could both be a result of intrapool primers, or one product made by intrapool primers and the other by interpool primers, such as in lane 4 (P1 + P5) in **Fig. 3**. Both products can also be made by interpool primers, such as lane 5 in (P2 + P3) in **Fig. 3.** (*see Note 6*). Set up the secondary PCR reactions (*see Fig. 5*) by eliminating one primer at a time to determine the primer pairs. For example, in **Fig. 6** the larger product is absent in lane 2 (primer p12 is missing) and lane 7 (primer p7 is missing); therefore, primers p12 and p7 are responsible for making the top band in lane 5 in **Fig. 3**. In **Fig. 6** the smaller product is absent in lane 1 (primer p11 is missing) and lane 9 (primer p9 is missing); therefore, primers p11 and p9 are involved in the synthesis of the bottom band in lane 5 in **Fig. 3** (*see Note 7*). Run a confirmatory PCR with the identified primer pairs using the PCR conditions from **Subheading 3.3.** or **3.4.**
4. If the first round of Optimized Multiplex PCR does not resolve all of the contig ends, a second round of Optimized Multiplex PCR may be required. Some minor adjustments in the PCR cycling conditions such as decreasing the annealing temperature or increasing the extension time could produce the missing PCR products.

3.6. Multiplex Sequencing

1. Clean the Multiplex PCR products using the QIAquick PCR Purification Kit following the manufacturer's instructions. Any other available PCR purification techniques can also be used.
2. Set up a sequencing reaction with each primer pool using the PCR product as a template.

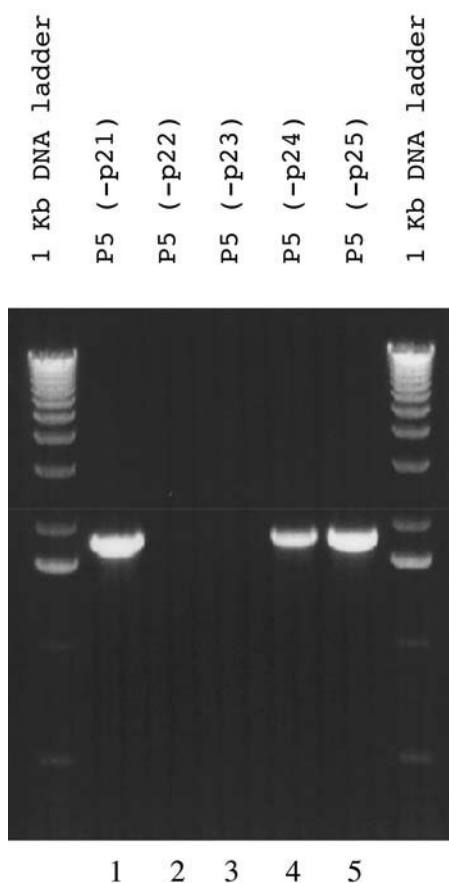


Fig. 4. Secondary Optimized Multiplex PCR reactions for intrapool P5, which comprises primers p21–p25. Lane 1 contains a product made by intrapool P5 with primer p21 absent. Lane 2 represents a result of the intrapool P5 with primer p22 absent, and so on. The disappearance of PCR products in lane 2 and 3 suggests that primers p22 and p23 are involved in synthesis of the P5 intrapool product.

3. For one reaction, add 2 μ L of BigDye Terminator mix, 2 μ L of 5X Sequencing Buffer, 25 ng of purified PCR product, and 0.6 μ M of each primer from the pool. Add dH₂O for a final volume of 10 μ L.
4. Set up cycling conditions as follows: an initial denaturation step of 96°C for 2 min, followed by a denaturation step of 96°C for 10 s, an annealing step of 55°C for 10 s, and an extension step of 60°C for 4 min. The latter three steps are repeated 40 times. Keep the tubes at 4°C until they are removed from the thermocycler.

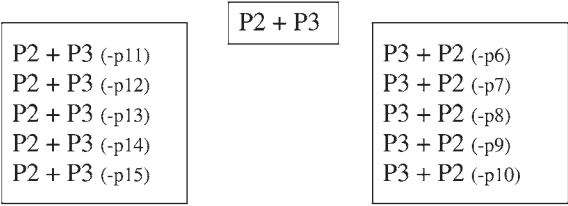


Fig. 5. Schema for setting up a secondary set of Multiplex PCR reactions for double products (products in lane 5 in **Fig. 3**) in which one primer gets eliminated at a time from the primer pool. For example, in the first reaction, primer pool P2 is reacted with primer pool P3 in which primer p11 is absent.

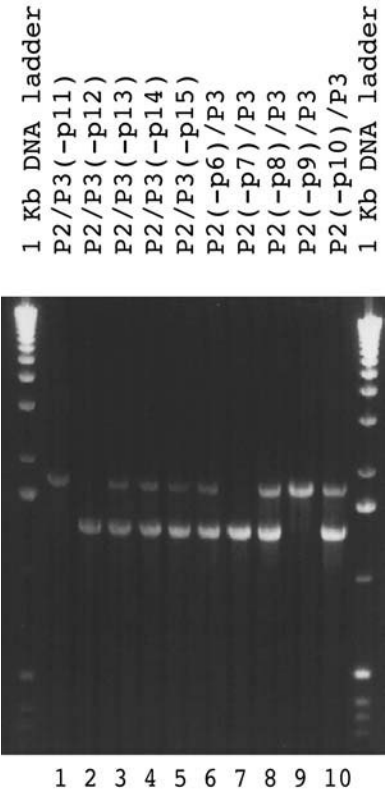


Fig. 6. Secondary Optimized Multiplex PCR reactions for interpretation of two products. The top product is absent from lane 2 and lane 7, therefore primers p12 and p7 are involved in the synthesis of this product. The bottom band is absent from lane 1 and lane 9, therefore p11 and p9 are involved in the synthesis of this product.

5. Precipitate the products with 100% isopropanol and 70% ethanol.
6. Load the precipitated samples on a sequencing machine (ABI Prism™ 377 DNA Sequencer, ABI Prism® 3700 DNA Analyzer, or ABI Prism® 3100 Genetic Analyzer, Applied Biosystems).
7. After the sequence is obtained, align it to the contigs from the genome using any sequence alignment software to identify the contig end it matches to and therefore the primer involved. When both primers are identified, run a confirmatory PCR with only these two primers using conditions from **Subheading 3.3.** or **3.4.**

4. Notes

1. The materials and protocols described in this chapter were used in our laboratory to conduct the experiments, but any other available PCR or sequencing kits can be used following the manufacturer's instructions.
2. The efficiency of the Optimized Multiplex PCR seems to reduce as the number of the primers in the primer pool exceeds 15. This could be owing to excessive primer-primer interaction during PCR.
3. When setting up a large number of reactions, the volume of all the reagents can be halved to give a final reaction volume of 25 μ L. This has been shown to work as well as 50- μ L PCR reactions.
4. Usually, when running Multiplex PCR, the size of the gaps or expected PCR products is not known. If after the first round of multiplex PCR reactions there are still some PCR products unaccounted for, the extension time in the cycling conditions can be increased up to the maximum manufacturer's recommended time for the DNA polymerase used in the reaction. This will allow the formation of larger PCR products.
5. An intrapool product, a product made by two primers from the same pool, can be identified prior to the Optimized Multiplex PCR reaction to simplify interpretation of the PCR results. Set up a PCR reaction in which each primer pool is reacted with itself. For example, primer pool P1 is reacted with primer pool P1, primer pool P2 is reacted with primer pool P2, and so on. If intrapool products are formed, run a secondary set of PCR reactions by eliminating one primer at a time from the pool to identify the primers involved (*see Fig. 4*). Also *see Subheading 3.5*. Identified intrapool primers can then be excluded from further Optimized Multiplex PCRs.
6. When the same PCR reaction produces two interpool products of different size, each product can be cut out from the agarose gel. Each PCR product can then be purified using available gel purification techniques and then multiplex sequenced (*see Subheading 3.6.*).
7. Try to resolve all of the single products before working on double products. This will allow for the elimination of these primers from further reactions and will reduce the number of secondary PCR reactions.

References

1. Fleischmann, R., Adams, M., White, O., et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
2. Tettelin, H., Radune, D., Kasif, S., Khouri, H., and Salzberg, S. (1999) Optimized Multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **69**, 500–507.
3. Burgart, L., Robinson, R., Heller, M., Wilke, W., Iakoubova, O., and Chevill, J. (1992) Multiplex polymerase chain reaction. *Modern Pathol.* **5**, 320–323.
4. Hall, M. (1996) *Combinatorial Theory*, 2nd ed. Wiley-Interscience, New York.

Assembly of DNA Sequencing Data

Jeremy Schmutz, Jane Grimwood, and Richard M. Myers

1. Introduction

At this writing, public databases contain the completed, contiguous sequences of a large number of bacterial genomes (e.g., www.tigr.org/tdb/mdb/mdbcomplete.html) (1), the yeast *Saccharomyces cerevisiae* genome (2); and the genomes of the nematode (3), *Drosophila* (4), and *Arabidopsis* (5). Public sequencing projects for many other genomes, including several large ones, are in progress. In the case of the publicly funded human genome sequencing effort, more than 45% of the >3-Gb genome is in finished form, including several completed chromosomes, with the remaining 55% expected to be finished by the spring of 2003. The term *finished* to describe sequence has special meaning and significance in the public genome-sequencing arena, where there is a general agreement that a large-insert clone, chromosome, or genome is considered finished when the accuracy of the sequence exceeds 99.99% (i.e., <1 error in 10,000 bp) and all gaps that can be filled by known techniques are filled, so that the sequence is either completely contiguous or has very few, well-annotated gaps.

With the exception of most of the bacterial genomes and the *Drosophila* genome, these public genome efforts rely heavily on sequencing large-insert clones, most typically bacterial artificial chromosome (BAC) clones, as a major part of the sequencing strategy. Although there are variations in the way that different sequencing groups produce finished sequences of BAC clones, all strategies have similar steps. Shotgun, or draft, sequence is produced from sheared subclone libraries; sequences are computationally assembled, analyzed, and submitted to databases as unfinished sequences; and various automated and manual experiments, guided by reassembly of the sequence after each

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

round of new data acquisition, are done until the clone is finished. In this chapter and in Chapters 24 and 25, we describe three of these critical steps in the sequencing process. First, in this chapter, we describe the process of sequence assembly, which is important at every stage of sequencing, not just during finishing. In Chapter 24, we describe the experimental methods used to perform the finishing process. Finally, in Chapter 25, we discuss methods that can be used to assess the quality of the sequences produced at the shotgun and finished stages.

Assembly is a complex computational process in which a computer attempts to reconstruct a BAC sequence from a large number of sequencing reads. When finishing an entire BAC or a section of a BAC, it is essential to have the assembly correct before continuing through the finishing process. Here, we describe the use of the Phred/Phrap/Consed (6–8) suite of computer programs from the University of Washington (hereafter referred to as the U. WASH suite). This software package is the one most widely used for assembly and assembly editing among laboratories doing genomic finishing from BAC clones and is relatively easy to learn and configure. In this chapter, we give some background about this package, explain the Phred program and how it can be used as a quality control tool, describe how to use Phrap to assemble sequence, and discuss how to identify and correct errors in the assembly process with Phrap and Consed. For convention, an asterisk is used to indicate a user-definable name, and statements appearing in single quotes are program commands or are program-specific files or options.

2. Materials

This chapter makes extensive use of the computational software package from the University of Washington, which includes Phred, Phrap, PhrapView, and Consed. This suite of programs is available free to academic users and for a licensing fee to commercial users. See www.phrap.org for availability and licensing information. Phred and Phrap currently compile on many computational platforms. Consed and PhrapView, however, are available only precompiled for the following platforms: Solaris, Red Hat Linux, SGI, DEC-Alpha, or HP-UX. Access to one of these platforms is necessary to be able to use the full functionality of the U. WASH software suite. It is also necessary to have a knowledgeable systems administrator install the software suite. The U. WASH suite can be installed on an Intel-compatible PC running Linux, but for a more robust system for assembly and analysis, a low-end Sun Microsystems workstation (Ultra5 or Ultra10) is preferable. The Suns are similar in cost, but once configured are extremely reliable and can be shared between several PCs each running an X-window emulator. The minimum RAM requirement for Phrap and Consed is about 256 megabytes, but consider purchasing 1 gigabyte of

RAM to support this package. Once installed, a complete assembly with base-calling can be run by using the 'phredPhrap' script provided with Consed.

3. Methods

3.1. Information on Phred, Phrap, and Consed

There are three main stages in the assembly and finishing process: (1) base-calling, (2) read alignment and assembly, and (3) assembly viewing and finishing read selection. The Phred program is used to call the bases in the raw trace files generated by automated sequencing machines (more information about Phred is provided in the next section). The Phrap program aligns the basecalls from Phred and constructs a consensus sequence of the BAC clone. The PhrapView program displays an overview of the Phrap assembly and shows paired-end relationships in the cases in which sequences of both ends of plasmid templates are used by the sequencing group. The Consed program graphically displays the aligned reads from the Phrap assembly and can also be used for picking primers and selecting finishing reads. For further information and in-depth details about these programs please see the "readme" files and ".doc" files that accompany the software distributions.

3.2. Read-Naming Conventions

Several read-naming conventions are required for using the U. WASH suite of programs. The minimum requirement for Phrap is that a name contain the template name followed by a period, followed by an 's' or 'x' for a forward read or an 'r' or 'y' for a reverse read ('s' and 'r' indicate dye-primer chemistry; 'x' and 'y' indicate dye terminator). These names tell Phrap not to use confirming reads from the same template but to allow two different sequencing chemistries used on the same strand to confirm one another (*see Note 1*).

We recommend selecting a naming convention that gives the maximum amount of information from the read name alone while still conforming to a Phrap-compatible naming scheme. Our read names at the Stanford Human Genome Center contain information about the BAC clone, the sheared library from which the subclone template was derived, the microtiter plate and well location, the primer used, and the number of times that sample was attempted. For example, for the name 'EYL02-B07.y1d-R', the read is from library EYL, microtiter plate 2, well B07, and is the first attempt with a reverse primer.

By including information about how a read was created in its name, analyzing assembly issues or identifying systematic sequencing errors is easier. The finishing process is also much smoother because it is possible to identify experiments that have been previously attempted to solve the particular problem. We build on this basic read name convention to construct read names for each of our finishing steps.

3.3. Assessment of Raw Data Quality With Phred

The first step in the assembly process is base-calling of the trace files using the Phred program. Phred can be run on a set of traces in a directory, and the information can then be used to create a fasta sequence file and a quality file for each trace by typing 'phred -q -s -id <directory name>'. Phred takes the raw sequence chromatograms and identifies the peaks of each base pair, calls the bases, and creates a 'phd' file, which is readable by Consed. Phred also assigns a number to each base pair—its phred score (6–7). The phred score is a statistical measurement of certainty that the “real” base in the sequenced DNA molecule is the same as the base called by Phred. The phred score is calibrated to a variety of sequencing platforms on the basis of a large amount of sequencing data collected on each platform (*see Note 2*). The score is used by Phrap in the assembly process, but it can also be used as a measurement of sequencing quality for each read. The score provides a quantifiable measure of sequence quality that is comparable between different sequencing reads. The phred score scale is logarithmic. A phred score of 10 indicates that 1 of 10 bases with that score was called incorrectly. A score of 20 indicates that 1 of 100 bases with that score is incorrect, whereas for a score of 30, 1 of 1000 is incorrect, for a score of 40, 1 of 10,000 is incorrect, and so on. In other words, if a trace has 1000 bases with a phred score of 30, about one base in those 1000 will be incorrect.

The Human Genome Project has adopted the measurement of phred20 bases as an indication of sequencing read qualities for each sequencing project. The phred20 measurement is calculated by adding the number of bases in a sequencing read with a phred score ≥ 20 .

In **Fig. 1**, the phred scores are plotted across a typical sequencing read. In the sections below the plot, the quality of sequences from three regions in the sequence read are shown in the trace file sections, and these qualities track with the phred scores in the plot.

The quality of an assembled sequence of a BAC clone is determined by the quality of the raw data that goes into the assembly. It is important to maintain a consistent quality for each sequencing read during the shotgun phase (*see Note 3*). By counting the number of phred20 bases, the raw data used for an assembly can be quantified.

We explore each BAC project's raw data by binning the number of phred20s per read in bins of 50 bp. **Figure 2** provides examples for BAC projects sequenced on an ABI 3700 capillary sequencer and an ABI 377 slab gel sequencer.

Sequence quality histograms such as those in **Fig. 2** provide important information about the raw data that are used for an assembly of a BAC-sequencing

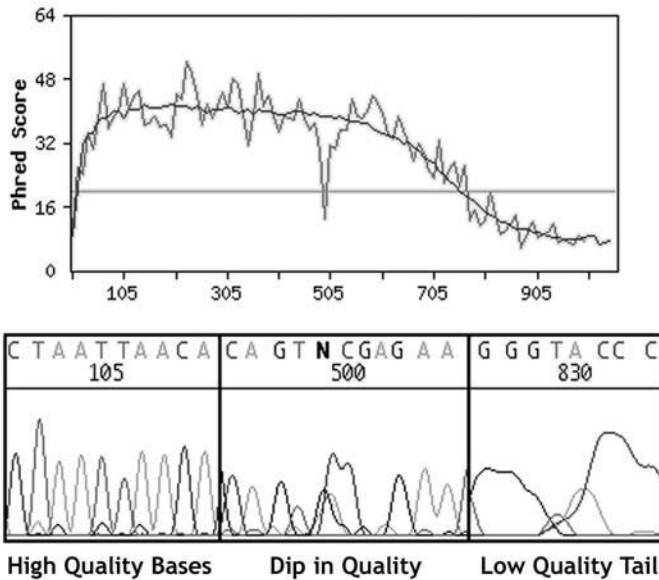


Fig. 1. Phred scores compared to various regions in underlying sequence trace.

project. In **Fig. 2**, the 3700 project has a higher failure rate, but almost all the reads that did not fail have 600 bp with scores of phred20 or greater. The 377 project, which was electrophoresed on short 30-cm glass plates at 4× run speed, has a wide distribution of phred20 bases, but a lower overall failure rate. In assembly, the 377 project will form fewer and larger contigs than the 3700 project but will have more low-quality regions within the contigs because of the long, low-quality tails from the slab detection.

By examining the distribution of phred20s in a large number of traces, it is possible to identify systematic problems in a sequencing process. For example, a dip in the phred20s at the beginning of a sequencing read is a good indicator of insufficient cleanup of the sequencing reaction or the presence of a salt front across a gel run. In our sequencing process, we have optimized the conditions of the sequencing reactions and run conditions to maximize the phred20 score for each sequencing read. The phred20 score also provides a good metric to compare sequencing from standard, control templates and new, test templates. In our center, we track the average phred20s for each gel and report the data on an HTML-based Web page so that we can quickly track down and eliminate problems in our sequencing pipeline. Phred20s provide a consistent quality control measurement for the sequencing and finishing process.

3700 Project				377 Project			
Phred20_Bin	Reads	%of Reads		Reads	%of Reads		
0- 49	360	19.7	XXXXXXXXXX	420	12.4	XXXXXXX	
50- 99	35	1.9	X	45	1.3	X	
100-149	18	1.0	X	35	1.0	X	
150-199	28	1.5	X	56	1.7	X	
200-249	36	2.0	X	92	2.7	XX	
250-299	36	2.0	X	112	3.3	XX	
300-349	40	2.2	XX	221	6.5	XXXX	
350-399	43	2.4	XX	399	11.8	XXXXXX	
400-449	61	3.3	XX	497	14.7	XXXXXXXX	
450-499	84	4.6	XXX	556	16.5	XXXXXXXXX	
500-549	150	8.2	XXXXX	455	13.5	XXXXXX	
550-599	188	10.3	XXXXXX	315	9.3	XXXXX	
600-649	302	16.5	XXXXXXXXX	150	4.4	XXX	
650-699	369	20.2	XXXXXXXXXXXX	22	0.7	X	
700-749	76	4.2	XXX	4	0.1	X	
750-799	2	0.1	X	0	0.0		

Fig. 2. Examples of BAC projects sequenced on an ABI 3700 capillary sequencer and an ABI 377 slab gel sequencer.

3.4. Assembling With Phrap

Before Phrap can be used to assemble sequence, *Escherichia coli* and vector sequences must first be screened and removed. This is done in the 'phredPhrap' script by using Cross_match, a tool included in the Phrap distribution. Cross_match compares each read to the sequence of the *E. coli* genome (see **Note 4**) and replaces any matching bases with an 'X'. Phrap then ignores these parts when assembling the consensus sequence. The sequence of the subclone vector must be added to the vector screen file so it can be removed from the beginning of each read. In addition, the BAC or other large-insert cloning vector sequence, as well as the sequences of any special cloning vectors that are used in the finishing process, should be added to the vector screen file as fasta-formatted sequences.

After the screening is complete, Phrap calculates an alignment score by using a modified Smith-Waterman algorithm (9) between each of the read segments. It then uses what is known as a greedy algorithm to put together reads into contigs.

Phrap starts with the highest score, builds small contigs, then compares between those small contigs, to build bigger contigs. After Phrap has assembled the largest contigs possible, it combs through each contig and creates a consensus sequence out of the highest-quality bases at each base position. Phrap also combines the phred scores for each read that crosses each position and calculates a new phrap score measured on the same logarithmic scale as the phred scores. It does this by combining the phred scores for each base on each strand of the sequence and making adjustments based on having both strands covered or reads from two different sequencing chemistries on the same strand. In addition, Phrap attempts to remove chimeric subclones that could cause problems in the assembly. **Figure 3** shows three passes of a greedy assembly algorithm, each pass joining fragments of a successively lower comparison score.

An assembly with real data is much more complicated. Phrap has to manage to assemble sequence data with noise, incomplete data, repeats, cloning problems, as well as other problems. Actual data from a sequencing project are rarely a perfect representation of the underlying BAC clone. A large part of the finishing process is designed to clean up the raw data, verify and correct discrepancies, and address problems with the generation of the subclone library.

The Phrap parameters used for an assembly should be selected on the basis of the desired stringency of the output consensus sequence. In its simplest form, Phrap can be run with 'phrap -new_ace <*>'. Phrap assumes that a quality file that corresponds to the read file named '<*>.qual' has been generated. The '-new_ace' flag tells Phrap to create an output file, the 'ace' file, that Consed

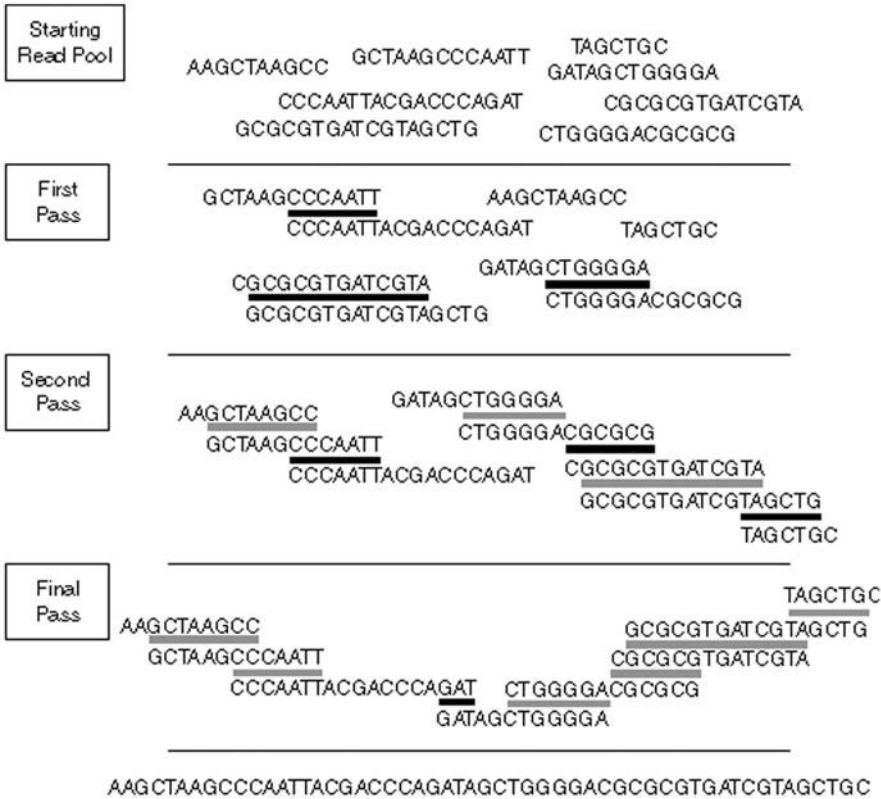


Fig. 3. Simplified example of Phrap assembly.

can read. All of the other parameters are the default parameters for the Phrap program.

The default parameters can assemble BAC sequences correctly, but we find that more stringent parameters can help work out misassembled regions and provide a good starting point from which to finish a BAC clone. Our parameters ‘phrap <*> -view -revise_greedy -new_ace -minmatch 30 -maxmatch 55 -minscore 55’ (see **Note 5**) break the assembly at the weak comparison points, allowing a finishing read to be selected to verify the joins before letting Phrap attach them together in a single contig. The ‘-view’ flag is necessary to use PhrapView with the assembly; the ‘minmatch’, ‘maxmatch’, and ‘minscore’ parameters control the stringency of the Phrap comparison. We recommend that the ‘-revise_greedy’ flag always be used; with this flag turned on, Phrap breaks the assembly at weak join points and looks for alternative, higher-scoring connections. For more Phrap options, see the file ‘phrap.doc’ that comes with the Phrap distribution.

After running a Phrap assembly with the 'phredPhrap' script, several files are produced as output in the directory. In addition to the '<*>.fasta.screen' file, which contains the initial reads screened for vector and *E. coli* sequences, and the '<*>.fasta.screen.qual' file, which is the corresponding quality scores for the reads, there are several other important files related to an assembly:

1. The '<*>.fasta.screen.ace' is the file to be viewed by Consed, while the '<*>.fasta.screen.view' is the PhrapView-formatted file.
2. The '<*>.fasta.screen.contigs' file contains the output consensus for each contig in a fasta format, and the '<*>.fasta.screen.contigs.qual' file contains a phrap quality score for each base represented in the 'contigs' file.
3. The '<*>.fasta.screen.singlets' is a fasta-formatted list of the sequencing reads that Phrap could not place within the assembly. Any read achieving a significant score with any other read in the assembly, but not ending up in a contig, appears in the assembly as a single-read contig. The remaining single reads are in the 'singlets' file.
4. Two files are related to the screening out of vector and *E. coli* sequences from the reads. They are both outputs from Cross_match and represent the screened sequence on a single line like this:

```
745 0.90 0.13 0.00 GOB05-H13.x1d-F 54 827 (0) C ecoli
(402512) 10 6388 105614
```

This line indicates that the read 'GOB05-H13.x1d-F' hit from bases 54 to 827 in the *E. coli* genome and was screened out as a result. The *E. coli*-screened read information is in the '<*>.screen-ecoli.out' file and the vector screen information appears in the '<*.screen.out' file (see **Note 6**).

After assembly, there will be some joins that Phrap was not able to join with the designated parameter set but they can be made by a human. If these joins are close to identical sequence, they will be tagged in Consed with a green 'match elsewhere' tag. The join can be searched and evaluated in the Consed join editor. Frequently, there are discrepancies in these areas of the assembly owing to dye fronts, compressions, or subclone mutations. These joins should be cleaned up by editing the traces, after which Phrap will join them appropriately in successive assemblies.

For clones in which a large amount of assembly editing (breaking and rejoining contigs) has been done, new reads can be added to the assembly without rephrapping and changing the edited contigs by using the script included with Consed called 'addReads2Consed.pl'. This script is run from the main Consed window and accepts a list of trace file names to be added to the assembly. The script creates a new 'ace' file and attempts to find a high match score and place each read into the current contigs. Except for special cases, such as BAC clones with near-identical repeats, we recommend running Phrap again and creating a new assembly whenever new traces are added.

When finishing regions of a genomic BAC clone that contain exons in a known cDNA, the utility ‘mktrace’ (included with Consed) can be used to create a fake trace file and ‘phd’ file. After removing the low-quality trace file created with ‘mktrace’, Phred will recall the sequence at high quality. In cases in which genomic resequencing and polymorphism identification are being performed, ‘mktrace’ can also be used to guide the assembly. The areas of the BAC clone that still need work will have a low-quality consensus in the combined assembly.

When sequencing from paired plasmid ends, the finishing process can be guided by ordering and orienting the contigs. To achieve this, open the PhrapView display and select ‘show paired ends’. If there are viable subclones linking the contigs, they will show up as black parabolas between contig ends. By starting with a contig that has a clone end, each contig can be linked together successively until the other clone end is reached, as follows:

END-1-4C-2-3-5-END

In this example, contig 4 is “complemented” (i.e., the sequence has to be flipped to put it in the proper orientation according to the subclone links) and the clone ends are on contigs 1 and 5. To close this clone, the gaps between 1-4, 4-2, 2-3, and 3-5 must be bridged. If there are no subclones linking contigs, these gaps are considered to be uncaptured gaps in the subclone library.

3.5. Identifying and Correcting Misassemblies

Sometimes, because of limited information or repeat content in a BAC clone, Phrap misassembles the consensus sequence. A misassembly is defined as any place where one part of a BAC clone is joined improperly during the assembly process to another part of the clone. This usually occurs with reads that cover a large or small repetitive sequence. False joins, another type of misassembly, are usually caused by chimeric library clones. These misassembled areas can lead to the mistaken interpretation that a clone is contiguous or that a gap is present when the sequence is actually contiguous. It is critical to identify and correct these misassembled regions early in the finishing process so that the problems do not have to be addressed later on in the process.

Many misassemblies are easy to diagnose by examining paired-end structures and contig matches in PhrapView. To do this, open the ‘view’ file with PhrapView. If paired plasmid end sequences are being used, select “Show Fwd-Rev Links.” PhrapView displays “good” links in black and “bad” links in red. A misassembly is characterized by a set of red links from a contig that link to one or more other contigs. For example, in **Fig. 4**, contig 19 has a false join and should be split, such that the first part of the contig is attached to the middle of contig 25 and the rest is attached to the end of contig 25.

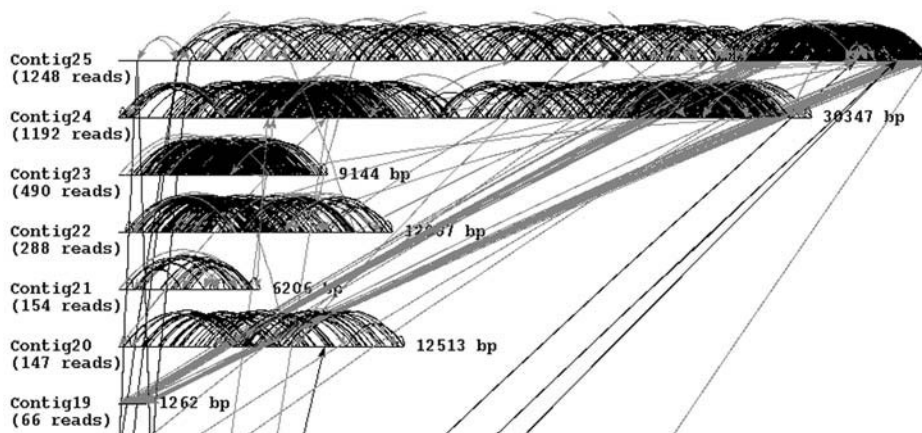


Fig. 4. False join with paired ends in PhrapView.

If sequences from paired plasmid ends are not being used, “Show Contig Matches” with PhrapView can be used, although in some cases, it is more difficult to diagnose an assembly problem based just on the similar sequence in the consensus. The same false join shown in **Fig. 4** is shown with contig matches in **Fig. 5**. Contig matches can be a good indicator of a misassembly or repeat assembly problem, but it is more challenging to identify misassemblies with contig matches alone.

After identifying a possible misassembly in PhrapView, open the same assembly with Consed and examine the area of the questionable join. If there is a misassembly of a perfect repeat longer than a sequencing read, there will be few, if any, discrepancies, but for a typical repeat misassembly or false join, some evidence of discrepancies should be visible in Consed. In these cases, navigate to high-quality discrepancies (HQDs) around the area of the misassembly. These will have sequences from multiple subclones that disagree at a particular base. In these cases, Phrap can be instructed not to overlap these bases in the subsequent assemblies through the Consed interface. Select the discrepant base, and by using the pull-down menu from the middle button, select “Tell phrap to not overlap . . .,” as shown in **Fig. 6**. This will make each base around the discrepancy a high-quality base, and Phrap will not assemble them the same way again. Rerun the assembly and reexamine the area in PhrapView to determine whether the misassembly has been corrected (*see Note 7*). If a single subclone is all that holds the assembly together in this region, attempt to change the read or reads to “n”s or to remove the reads altogether and reassemble. Because Phrap identifies likely chimeric candidates based on the assembly, it misses some of the chimeric clones that integrally

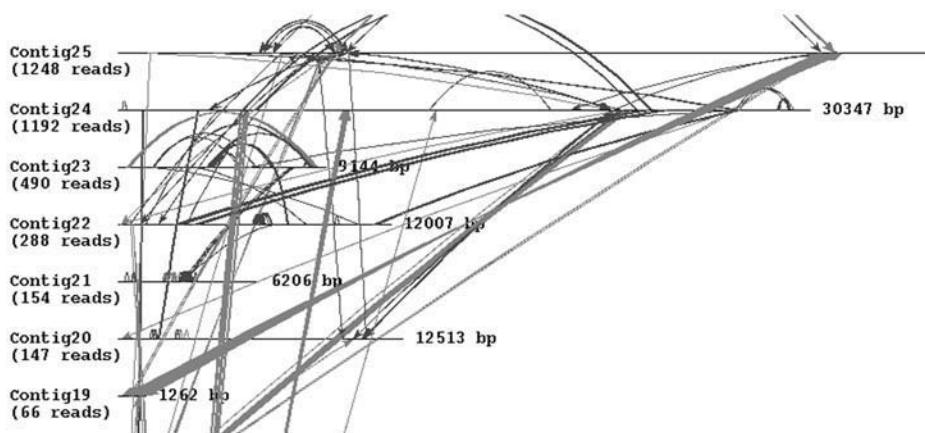


Fig. 5. False join with contig matches in PhrapView.

hold together parts of the assembly. Although such problems are difficult to identify, once found, a false join misassembly owing to a chimeric subclone can be quickly corrected.

Instead of instructing Phrap not to overlap the reads, most reads with a high-quality discrepancy can be broken apart by increasing the phrap parameter 'repeat_stringency' to '.99'. This action creates many contigs that must be hand-joined to obtain a contiguous assembly, but it allows the human finisher to make the decisions about which similar pieces to join instead of letting Phrap make the decisions. For complex repeats, contigs can be torn apart manually by using the paired subclone end sequences to identify incorrect joins. This is tedious work but sometimes yields full consensus coverage through an identical repeat that can then be further verified experimentally.

When this type of data analysis indicates a misassembly, we recommend that additional sequencing reactions be done to reinforce the electronic editing. Placing a universal resequencing read or performing primer walking with divergent subclones or the BAC clone often sorts out the repeat problem and supports the new joins with additional sequence data. The new information allows the repeat to be accurately assembled in subsequent iterations.

4. Notes

1. Currently, Consed receives the information about reads from the script 'determineReadTypes.pl', which must be configured correctly to extract the template and primer type from the read name.
2. If Phred gives the error message 'unknown chemistry', it is because a chemistry name that has not been predefined is being used. To correct this problem, add the

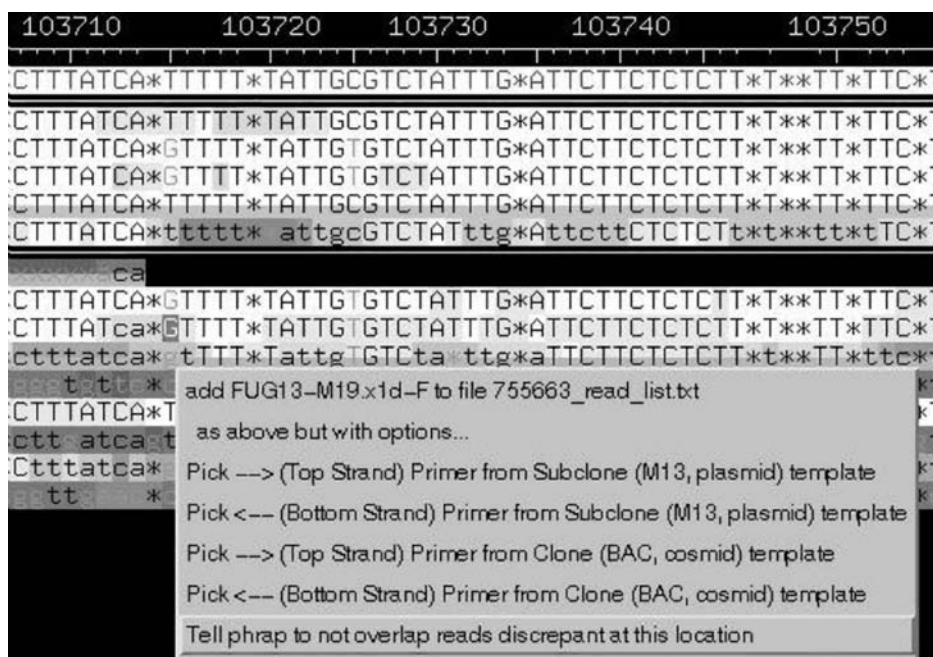


Fig. 6. Fixing a misassembly with Consed.

new chemistry name to the 'phredpar.dat' file, and then Phred will be able to identify the sequencing platform.

- For finishing reads, it is equally important to track and address quality issues. For example, the number of phred20s in a read can be used to evaluate the success of a custom primer. Generally, a misprimed reaction generates fewer than 50 phred20 bases in the read.
- By splitting the *E. coli* genome into 10 equal parts as fasta sequences in the screen file, the time to screen against *E. coli* by Cross_match will be greatly reduced.
- When running a phrap assembly, always direct the output to a file in the directory. Phrap will save valuable information in this file, usually referred to as the 'phrap.out' file.
- If you believe a read should be included in an assembly, but it does not appear when Consed is opened, first determine from the *E. coli* and vector screen files whether the read completely comprises vector or *E. coli* sequences. Then determine whether the phred20 score for that read is not high enough quality to be included. Look in the '<*>.fasta.screen.qual' file to check the quality and make sure that the read was base-called properly.
- A bottom-up approach can be used to identify possible areas of misassembly by navigating to HQDs in Consed. Set the threshold low, to about 30, and look for a concentration of discrepancies across several reads from several different subclones.

References

1. Tettelin, H., Nelson, K. E., Paulsen, I. T., et al. (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506.
2. Goffeau, A., Barrell, B. G., Bussey, H., et al. (1996) Life with 6000 genes. *Science* **274**, 563–567.
3. The *C. elegans* Sequencing Consortium. (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018.
4. Adams, M. D., Celniker, S. E., Holt, R. A., et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.
5. The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
6. Ewing, B., Hillier, L., Wendl, M., and Green, P. (1998) Base-calling of automated sequencer traces using Phred I. Accuracy assessment. *Genome Res.* **8**, 175–185.
7. Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using Phred II. Error probabilities. *Genome Res.* **8**, 186–194.
8. Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202.
9. Smith, T. F. and Waterman, M. S. (1981) Identification of common molecular sub-sequences. *J. Mol. Biol.* **147**, 195–197.

Sequence Finishing

Jeremy Schmutz, Jane Grimwood, and Richard M. Myers

1. Introduction

Sequence “finishing” is the process of turning a rough draft assembly composed of shotgun sequencing reads into a highly accurate finished DNA sequence with a defined maximum allowed error rate. The standard established by the international publicly funded sequencing community for considering a sequence finished is that it be completely contiguous, with no gaps in the sequence, and that it have a final estimated error rate of <1 error in 10,000 bases. Various groups have produced finished sequences of entire genomes—in some cases with whole-genome shotgun strategies and in others with large-insert clones covering the genome—as well as of several entire human chromosomes.

The amount of time and effort, as well as the sequencing strategy, required to finish a bacterial artificial chromosome (BAC) clone varies depending on whether finished sequence is desired for an entire BAC or only a portion of a BAC. The latter case may exist when there is already overlapping finished sequence for a part of a BAC, or the researcher is interested in obtaining finished sequence for only a particular gene in a BAC clone.

Because much of the effort to produce finished sequence has involved the use of BAC clones as the source of genomic DNA for sequencing, this chapter is devoted to methods used to perform the finishing process on BACs. However, these methods are applicable to finishing DNA sequences cloned into any large-insert cloning vector, such as cosmid, fosmid, P1-derived artificial chromosome, and YAC vectors. We describe the experimental design and laboratory steps in the finishing process, including the selection of directed sequence finishing reads, the selection of custom primers for sequencing reactions, the types of sequencing chemistry used in finishing, and the finishing platform (i.e., the

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

choice of sequencing instrumentation). These methods should be supplemented with the processes we describe in Chapters 23 and 25 for assembling sequences and for quality assessment of finished sequences, respectively.

2. Materials

All chemicals should be of molecular biology grade, and all solutions should be made with ddH₂O.

2.1. Resequencing With Universal Primers

1. Plasmid DNA.
2. Universal primer: All stock primers are at 100 μ M. Typical examples that are in common use at the Stanford Human Genome Center (SHGC) are SP6, T7, T3, and M13 forward and M13 reverse primers. Store primers frozen at -20°C . Dilute primers prior to use to a final concentration of 1 μ M with ddH₂O.
3. Sequencing mix: BigDye Terminator Ready Reaction Kit (#4390242; ABI) or dGTP BigDye Terminator Ready Reaction Kit (4307176; ABI). Use accompanying buffers.
4. 70% Isopropanol.
5. 70% Cold ethanol.
6. Blue Loading Dye: 25 mL (0.5 M) of EDTA, 25 g of Ficoll, 0.0625 g of bromophenol blue, and 175 mL of ddH₂O. Add Ficoll to the EDTA. Stir overnight to allow the Ficoll to dissolve. Add the bromophenol blue and filter with a 0.2- μ filter.

2.2. Custom Primer Walking on Plasmid Subclones

1. Plasmid DNA.
2. Custom oligonucleotide primer: 10 μ M stock. Dilute primers prior to use to a final concentration of 1 μ M with ddH₂O. As discussed, the SHGC uses subclone primers that are on average 18 nt in length.
3. Sequencing mix (*see Subheading 2.1., item 3*).
4. 70% Isopropanol.
5. 70% Cold ethanol.
6. Blue Loading Dye (*see Subheading 2.1., item 6*).

2.3. Sequencing Directly From BAC Clones

1. Luria Bertani (LB) broth: 10 g of Bacto Tryptone, 5 g of yeast extract, 5 g of NaCl. Dissolve ingredients in ddH₂O and bring up to 1 L. Autoclave for 45 min.
2. Terrific Broth (TB): 47.6 g of Bacto Terrific Broth, 8 mL of 50% glycerol. Dissolve ingredients in ddH₂O and bring up to 1 L. Autoclave for 45 min.
3. Chloramphenicol (12.5 mg/mL): 500 mg of chloramphenicol. Bring up to 40 mL with 100% ethanol. Store in the dark at -20°C . Use at 12.5 μ g/mL in agar and cultures.
4. LB/chloramphenicol agar plates.

5. Qiagen Maxiprep Kit (#12163): The maximum binding efficiency of the Qiagen-tip 500 is 500 μg of plasmid DNA.
6. Isopropanol.
7. 70% Ethanol.
8. TE buffer: 40 mL of 1 M Tris-HCl (pH 8.0), 8 mL of 0.5 M EDTA (pH 8.0). Dissolve ingredients in ddH₂O and bring to a final volume of 4 L. Autoclave for 45 min.
9. Orange Ficoll loading dye: 125 mL (0.1 M) of EDTA, 18.75 g of ficoll, 0.0315 g of Orange G powdered dye. Add the Ficoll to the EDTA and stir overnight until the Ficoll is dissolved. Add the Orange G dye and filter through a 0.2- μ filter.
10. Universal primers: All stock primers are at 100 μM . We use SP6, T7, and T3 routinely. Store frozen at -20°C . Dilute primers prior to use to a final concentration of 1 μM with ddH₂O.
11. Custom Oligo primers: Dilute primers prior to use with ddH₂O to a final concentration of 3.2 pmol/mL (32 μL of primer + 68 μL of ddH₂O). Our BAC primers are on average 22 nt in length.
12. Sequencing mix (*see Subheading 2.1., item 3*).
13. Princeton Separations Centri-Sep Columns (CS-901).
14. Blue Loading Dye (*see Subheading 2.1., item 6*).

2.4. Sequencing Polymerase Chain Reaction Products to Size and Close Gaps

1. BAC DNA (600 ng to 1.5 μg per reaction).
2. Custom oligonucleotide primer: 10 μM stock. Store frozen at -20°C .
3. dNTPs: stock 2.5 mM (#18427013; Gibco-BRL, Gaithersburg, MD).
4. Taq Polymerase: We use Platinum Taq (#10966-034; Gibco-BRL) for routine amplification or Platinum High Fidelity (#11304-029; Gibco-BRL) for long-range polymerase chain reaction (PCR) (12 kb and above). Use accompanying buffers.
5. Exonuclease I (#70073Z; USB), shrimp alkaline phosphatase (USB #70092Y), for PCR cleanup.
6. Na acetate, pH 5.2.
7. 100% Ethanol.
8. 70% Ethanol.
9. Sequencing mix (*see Subheading 2.1., item 3*).
10. Blue Loading Dye (*see Subheading 2.1., item 6*).

2.5. Alternative Strategies for Difficult Regions

1. SequenceRx Enhancer System (12237-012; LTI).
2. In vitro transposons (EZ1982K; EpiCenter Technologies).
3. Erase-a-Base System for nested deletions (E5850; Promega, Madison, WI).

3. Methods

This section discusses four issues that need to be considered before embarking on sequence finishing: the methods used to select finishing reads; the choice

of the type of chemistry used for finishing reactions; the choice of the type of sequencing instrumentation, or platform, to use in finishing; and the use of the computer program Consed to select custom primers for sequencing reactions. These discussions are followed by the laboratory protocols that are used to perform the finishing experiments themselves.

3.1. Selection of Finishing Reads

Drafted or shotgun-sequenced BAC clones that are not yet finished have several types of problems that must be addressed if they are to be turned into finished sequences. The following problems, which are identified by using the sequence assembly methods described in the previous chapter, need to be considered:

1. *Low-quality regions of the shotgun sequence.* These are bases with phred scores of <30 in the assembled sequence. In these areas of the BAC clone, the accurate determination of a base pair or multiple base pairs is not possible on the basis of the available shotgun data.
2. *Single subclone areas.* These are sections of the BAC clone that are covered by only one subclone. To ensure that the consensus sequence is not confounded by a mutation or deletion that occurred in a subclone during its growth in *Escherichia coli*, it is important to have confirming sequence from another subclone or directly from the BAC clone.
3. *High-quality discrepancies (HQDs).* These are regions, typically at a single base pair in a segment, that show two different consensus sequences, both of which have high-quality scores. HQDs suggest the presence of subclone mutations or a misassembly.
4. *Captured gaps.* These are gaps in the BAC clone shotgun or partially finished sequence that are spanned by subclones, but do not have DNA sequence data that extend to close the gap.
5. *Uncaptured gaps.* These are gaps in the clone that are not spanned by subclones.

To address these issues, it is usually necessary to add carefully selected sequencing reads to an assembly. Several criteria need to be considered when deciding which type of finishing read to perform:

1. *Universal resequencing reads.* A “resequencing read” is defined as a repeat of a sequencing read initially done in the shotgun phase that is performed using universal primers (recognizing the vector sequences near the cloning junction). Such reads are chosen when the initial read was of low quality and additional coverage of the base pairs covered by the read are of low quality, and they are used to increase quality and to extend sequence contigs. Resequencing reads are usually chosen if the base pairs in question are within a reasonable distance, usually 400 bp, of the universal priming site in the vector segment of a subclone.

2. *Subclone custom primer reads.* Subclone primer reads can be used to increase quality in areas where resequencing reads are not possible and to close “captured” gaps. These reads are also used to cover single subclone reads in cases in which another subclone is available for sequencing.
3. *BAC custom primer reads.* These reads, which are done by sequencing directly from the BAC clone as a template, are used in cases in which sequence gaps are not captured by subclones. BAC reads are also used to confirm consensus sequence in areas of subclone discrepancies or subclone deletions and to confirm areas of single subclone coverage in cases in which no other subclones are available.
4. *PCR sequencing.* In some instances, gaps cannot be closed by using routine sequencing reads, and PCR amplification from the BAC clone or other clones can be used to generate sequencing templates. By designing and then amplifying with unique primers on either side of the sequence gap, a PCR fragment is generated and its size is determined by agarose gel electrophoresis to estimate the gap size. The same PCR fragment is then used as a sequencing template. PCR products do not need to be cloned prior to sequencing, so regions that may be toxic to *E. coli* can be sequenced.
5. Alternative strategies can be attempted for sequencing through difficult regions of a BAC clone:
 - a. Sequencing enhancers, such as dimethylsulfoxide, betaine, and various commercially available kits (*see Subheading 2.5.*), often increase the quality of sequencing reads through strings of simple sequence repeats and dinucleotide runs.
 - b. Transposon-mediated sequencing helps to sequence through highly repetitive regions. Transposons hopped into the subclone DNA provide universal primer sites in regions where it is often impossible to choose custom primers. In addition, because the transposition events are so precise, it is possible to obtain a virtual read of more than 1000 bp at a hopped site because the two reads extending from each end of the transposon can be combined into a single contiguous stretch of sequence. This feature is often very helpful in obtaining accurate assemblies in repetitive regions. Commercial in vitro transposon kits are available (*see Subheading 2.5.*).
 - c. Small-insert “shatter” libraries (*1,2*), constructed by sonicating subcloned DNA templates into small fragments, size-selecting fragments of 100–200 bp, and ligating the fragments into a plasmid vector, can be used to verify the DNA sequence in small regions that are particularly difficult to sequence (e.g., in long stretches of CT repeats). Although it is often difficult or impossible to assemble the sequence obtained from these short reads, shatter library sequencing is still useful in such cases because it can determine whether any unique sequences are present within a repeat.
 - d. Nested deletions are a set of sequencing templates generated by processive, timed, unidirectional deletion of any inserted DNA. This technique can be used to place a universal primer site near a region of interest, which often results in higher-quality sequence than that obtained with a custom primer. In some cases, nested deletions can also be used for sequencing through long stretches

of DNA with only universal primers (*see Note 1*). There are commercially available kits for using this method (*see Subheading 2.5.*).

3.2. Choice of Finishing Chemistry

At the SHGC, we perform all our finishing reads with dye terminator chemistry. Our typical reactions use standard BigDye terminators, which provide high-quality sequence with minimal compressions. Our final finishing reactions, which are performed in problem areas where standard chemistry has failed to verify the sequence, are performed with dGTP BigDye terminators. In some cases, this chemistry provides sequence through hard stops or simple-sequence repeats. However, GC compressions and pileups are common when dGTP BigDye terminators are used, so it is sometimes impossible to call the correct consensus with reads from this chemistry alone.

3.3. Choice of Sequencing Platform for Finishing

The choice of sequencing instrumentation has an impact on the performance of the sequence finishing process. Although many large sequencing groups use only capillary sequencing machines, we have chosen to use ABI 377 slab gel sequencers for much of our finishing process. Although the use of slab gel machines is more labor-intensive and data tracking is more of a challenge than with capillary machines, gel sequencing leads to higher pass rates and longer read lengths. The longer read lengths are particularly useful, because good slab gel reads typically have at least as long a stretch of high-quality bases as capillary reads, but they also have an additional long stretch of lower-quality sequence after the good stretch, resulting in a total much longer segment than in a capillary read. These longer reads are valuable in finishing because they often are needed to provide contiguity of sequence during assembly that is not obtained with shorter reads. In our standard finishing pipeline, we use 12-h ABI 377 gel runs and typically achieve 700 phred20 bases, an additional 100–300 bases of lower-quality sequence, and 95% pass rates. Unlike reads in the shotgun-sequencing phase of a project, every finishing read is a custom read and a high failure rate significantly hinders the finishing process.

We load all of our gels using 96-well membrane combs from The Gel Company (San Francisco, CA) according to its recommended protocol. In our hands, gels loaded with membrane combs give straighter lanes and less lane-to-lane noise than standard combs.

3.4. Choice of Custom Primers With Consed

To perform finishing reactions with custom primers, the user must choose which type of template to use, which direction to sequence from on the template, and which primers to use for each sequencing reaction. As discussed,

templates can be plasmid subclones, BAC clone DNA, or PCR products, and the direction to sequence depends on where the desired sequence information resides relative to the primer. We use the program Consed to choose custom oligonucleotide primers. We select primers an average of 18 nt in length for sequencing from subclone templates, and an average of 22 nt in length both for sequencing from BAC clone templates directly and for PCR product generation and sequencing.

Consed uses the following criteria when selecting a custom primer:

1. The primer melting temperature is in an acceptable range.
2. Every base in the sequence to which the primer will hybridize has a quality score above a threshold value.
3. The primer has a low likelihood of annealing elsewhere on the template. When selecting a primer for a subclone template, Consed checks for possible annealing within 3000 bp from the site of the primer. When selecting a primer from a large-insert template, the program checks for possible annealing within the entire assembly of all contigs (*see Note 2*).
4. The primer does not have self-complementary sequences and, therefore, should not anneal with itself.

3.5. Laboratory Protocols

The protocols given next have been optimized for slab gel electrophoresis. They may be used for capillary electrophoresis, but the user should be aware that certain conditions may need to be adjusted for them to be optimal.

3.5.1. Resequencing With Universal Primers

1. Set up each sequencing reaction as follows: 1 μL of sequencing mix ($\frac{1}{2}$ reaction), 2 μL of primer, 2 μL of DNA, 2 μL of 5X sequencing buffer (provided by the manufacturer), and 3 μL of ddH₂O.
2. Run the sequencing reaction using the following conditions:
 - a. 96°C for 5 min.
 - b. 96°C for 10 s, 25 cycles.
 - c. 50°C for 5 s.
 - d. 60°C for 4 min.
 - e. Hold samples at 4°C.
3. Clean up each reaction by adding 40 μL of 70% isopropanol. Vortex and centrifuge at 3000g for 30 min. Wash the pellet with 250 μL of 70% ethanol. Centrifuge for 15 min at 3000g. Dry the pellet, add 3 μL of Blue Loading Dye, denature at 95°C for 3 min, and load the samples.

3.5.2. Custom Primer Walking on Plasmid Subclones

Sequencing reactions, cycle conditions, and cleanup are all as in **Subheading 3.5.1.**, except that a custom primer is used instead of a universal primer.

3.5.3. Sequencing Directly From BAC Clones

3.5.3.1. PREPARING BAC DNA TEMPLATE

1. Streak a BAC clone from the glycerol stock on an LB/chloramphenicol agar plate and grow for 14–16 h at 37°C.
2. The next day, pick a single colony, inoculate a 10-mL TB/chloramphenicol culture, and grow it for 14 h at 37°C with agitation at 300 rpm (*see Note 3*).
3. The following morning, use the culture to inoculate a 500-mL TB/chloramphenicol culture in a 1-L flask. Grow at 37°C with agitation at 250 rpm for about 4 h or until the OD₆₀₀ is 1.0 (*see Note 4*).
4. Prepare DNA using a Qiagen Plasmid Purification Maxi Kit and following the manufacturer's instructions (*see Note 5*).
5. Resuspend the pellet in 600 µL of TE buffer. Place tubes at 37°C for 20 min with the lids open to help resuspend the DNA and remove any residual ethanol.
6. To check the quality of the prep, run 2 µL of DNA + 5 µL of ddH₂O + 3 µL of Orange G Ficoll dye on a 0.3% LE agarose gel at 120 V for 45 min.
7. Dilute 2 µL of DNA in 198 µL of ddH₂O and measure the absorbance at OD₂₆₀. Calculate the DNA concentration using the following formula:

$$\text{OD at } (260 \times 50 \text{ (g/mL)} \times 100 \text{ (dilution factor)} \times 0.001 \text{ mL/}\mu\text{L} = \text{concentration (g/}\mu\text{L)}$$

3.5.3.2. PERFORMING BAC SEQUENCING REACTIONS

1. Set up each sequencing reaction as follows: 4 µL of dGTP (½ reaction), 4 µL of custom primer, 5 µL of DNA, 2 µL of 5X sequencing buffer, and 5 µL of ddH₂O.
2. Run the sequencing reaction using the following conditions:
 - a. 95°C for 3 min.
 - b. 95° C for 30 s, 35 cycles.
 - c. 50°C for 20 s.
 - d. 60°C for 4 min.
 - e. Hold samples at 4°C.
3. Clean up each reaction using a Centri-Sep Column. Perform the cleanup according to the manufacturer's instructions (*see Note 5*).
4. Dry the samples in a Speedvac (high setting) for 45 min.
5. Resuspend the samples in 2 µL of Blue Loading Dye. Denature for 5 min at 97°C. The samples are now ready to be run.

3.5.4. Sequencing PCR Products to Size and Close Gaps

1. Set up each PCR reaction as follows: 2 µL of BAC DNA (600 ng to 1.5 µg), 5 µL of 10X PCR buffer, 1.5 µL of MgSO₄, 4 µL of dNTPs, 1 µL of primer 1, 1 µL of primer 2, 0.5 µL of Platinum Taq, and 35 µL of ddH₂O.
2. Perform the sequencing reaction as follows:
 - a. 94° C for 2 min.
 - b. 94°C for 30 s, 30 cycles.
 - c. 55°C for 30 s.

- d. 72°C for 5 min.
- e. Hold samples at 4°C.

When using HiFi Taq, reduce the extension temperature to 68°C and increase the time to 10 min.

3. Before attempting to sequence, resolve the PCR fragments on an agarose gel. Experiment with PCR conditions if you do not obtain a distinct, sharp band.
4. Clean up the PCR reactions as follows:
 - a. 40-μL PCR reaction.
 - b. 1.6 μL of Exonuclease I.
 - c. 8 μL of shrimp alkaline phosphatase.

Mix well and incubate at 37°C for 1 h. Heat to 75°C for 15 min to inactivate the enzyme.

5. To precipitate add 50 μL of ddH₂O, 220 μL of 100% ethanol, and 20 μL of sodium acetate (pH 5.2). Centrifuge at 14,000g for 10 min. Wash the DNA pellet in 500 μL of 70% ethanol. Centrifuge again for 5 min. Air-dry the pellet and resuspend the DNA in 30 μL of ddH₂O.
6. Set up each sequencing reaction as follows: 2 μL of dGTP, 2 μL of primer, 2 μL of DNA, 2 μL of 2.5X sequencing buffer (5X), 2 μL of ddH₂O.
7. Run the sequencing reaction and perform cleanup using the protocol in **Subheading 3.5.3.2**. Denature the samples and resuspend in 1.7 μL of Blue Loading Dye. The samples are now ready to run.

3.5.5. Alternative Strategies for Difficult Regions

Protocols are provided for each of the kits listed in **Subheading 2.5**, and should be followed according to the manufacturer's instructions. A protocol for producing shatter libraries is in the Notes below.

4. Notes

1. The rate of deletion by Exonuclease III activity varies according to the composition of the DNA. For example, deletion through dinucleotide runs is much faster than through typical DNA regions, so it is often not possible to obtain a useful deletion with an end point within these regions.
2. If you have problems picking a BAC primer in a nonrepetitive region, check to see whether small contigs exist in the assembly that match the area you are trying to fix. These "contiglets" can be low-quality reads, chimeras, or deleted subclones that Phrap has chosen not to assemble. You will need to hand join these into the main contig to be able to pick a primer.
3. Grow the initial culture in a 50-mL conical Corning tube sealed with aeropore tape.
4. To make a glycerol stock, add 60 μL of 80% glycerol to a 240-μL aliquot of culture and store at -80°C.
5. Centri-Sep Columns increase the cost by \$2 to each reaction, and using the columns is labor-intensive. However, despite the cost, we have found that cleanup

of reactions with precipitation methods is not as effective, and because BAC reads are so important in the finishing process, we use columns in all cases.

References

1. Andersson, B., Wentland, M. A., Ricafrente, J. Y., Liu, W., and Gibbs R. A. (1996) A “double adaptor” method for improved shotgun library construction. *Anal. Biochem.* **236**, 107–113.
2. McMurray, A. A., Sulston, J. E., and Quail, M. A. (1998) Short insert libraries as a method of problem solving in genome sequencing. *Genome Res.* **8**, 562–566.

Quality Assessment of Finished BAC Sequences

Jeremy Schmutz, Jane Grimwood, and Richard M. Myers

1. Introduction

After the sequence of a bacterial artificial clone (BAC) clone is finished, it is important to assess the completeness and accuracy of the consensus sequence that was constructed from the underlying shotgun sequencing and finishing reads. Not all sequencing projects require that the entire BAC clone be finished to a specific high standard. For instance, in cases in which a BAC clone overlaps an already completely finished clone, the region of overlap need not be finished to the level that the nonoverlapping segment is finished. However, for this chapter, we refer to finished sequence to mean the complete BAC clone from end-to-end finished to high quality and meeting or exceeding the finished sequence standards set by the publicly funded Human Genome Project consortium. Although using some type of quality assessment technique is important at all stages of the sequencing/finishing process, and can be useful in determining a point at which to stop finishing a clone, our group at the Stanford Human Genome Center (SHGC), as well as most other sequencing groups, use quality assessment methods primarily for determining the quality of and performing quality control on finished sequence. Many of these techniques rely on the use of the Phred/Phrap/Consed system to help identify problem areas in the sequence of a BAC clone. In this chapter, we discuss the generally accepted standards of finished sequence and then describe the computational examination of the finished sequence, coupled with some laboratory techniques, for this quality assessment/quality control process.

From: *Methods in Molecular Biology*, vol. 255: *Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing*
Edited by: S. Zhao and M. Stodolsky © Humana Press Inc., Totowa, NJ

2. Materials

All chemicals should be of molecular biology grade, and all solutions should be made with double-distilled water.

1. BAC DNA.
2. Restriction enzymes *Eco*RI, *Hind*III, and *Xho*I and the accompanying buffers.
3. 1% Agarose gel.
4. Midrange and 1-kb DNA size markers.
5. Pulsed-field gel electrophoresis apparatus (PFGE) using the following conditions: 6-V/cm field strength, 120° internal angle, 1-s initial pulse time, 11-s final pulse time, linear ramping factor, 14-h run time.
6. 5% Ethidium bromide (EtBr) in 0.5% TBE.

3. Methods

3.1. Current Quality Standards for Finished Human Genomic Clones

The publicly funded groups that produced the draft sequence of the human genome and that are at a the midway point in finishing these sequences established a set of rules, sometimes referred to as the Bermuda Standards, for considering a sequenced large-insert clone finished (*see Note 1*). These standards state that finished sequence of a BAC clone must attain at least an accuracy on average of less than one error in 10,000 bp throughout the clone. This rate is measured so that the average base pair quality in the BAC clone has a Phrap score of 40. In addition, finished sequence must be contiguous from clone end to clone end with no gaps or ambiguous bases. The standards have also been expanded to include a requirement that areas of the sequence derived from a single subclone, regions where sequence is derived from only polymerase chain reaction (PCR) products, and regions with low-quality single-stranded sequences be annotated in the GenBank submission for the BAC clone.

3.2. Identifying Low-Quality Areas and High-Quality Discrepancies in BAC Sequences With Consed

Consed can be used to examine a limited number of potential problem areas in a sequence assembly, which is much more practical than checking each base pair in an entire BAC sequence from the sequence traces. By navigating to the low-quality areas of a sequence with the default parameters of Consed, each area in the sequence with a phrap score below 30 (corresponding to <1 error per 1000 bp) is visited. For each of these areas, several traces can be opened and used to examine the data that underlie the consensus sequence. On this basis, the user can make judgment calls as to whether the traces support the consensus sequence calls and correct the traces and consensus if necessary. If

there is any question about the validity of the consensus sequence, another sequencing reaction should be done to verify that the sequence is correct.

In addition to low-quality regions, sequencing projects sometimes have regions with high-quality discrepancies (HQDs), which are positions that have different base calls with high scores from two or more different reads. To identify and correct these problems, navigate to the HQDs in a sequencing project with Consed and examine each one. Open the trace contributing the discrepant base and the traces that match the consensus and attempt to determine what is causing the discrepant base.

Following is a list of possible reasons that HQDs appear in an assembly and how to classify them:

1. *Subclone point mutation.* Subclone mutations that occur during propagation in *Escherichia coli* result in a single high-quality base that disagrees with the consensus sequence derived from all other subclones in that region. This is not a problem with sequence quality but, rather, to a biologic problem that occurs relatively rarely. However, it is frequent enough that the sequencing community has agreed that every base in a consensus finished sequence must be derived from multiple subclone templates.
2. *Sequencing artifact.* These can cause Phred to miscall a base. In these cases, when the trace is examined, the correct base can be occluded by the miscalled base. This problem is sometimes caused by a problem with the gel or with the sequencing reaction cleanup procedure. In addition, an artifact can appear after the DNA polymerase traverses through a simple sequence repeat. These types of problems can be fixed simply by examining the trace data and typically do not require additional experimentation. An example of background noise interfering with base-calling is shown in **Fig. 1**.
3. *Misassembled repeat.* The same HQD in several different subclones can be an indication that a similar repeat has been misassembled. Check the paired end read structure over such an area using PhrapView, and check for repeated occurrences of HQDs in the same sequencing reads in other parts of the consensus. Refer to Chapter 23 for more strategies to correct misassemblies.
4. *Chimeric subclone.* These occur when ligation events in the sheared library construction process bring two unrelated DNA fragments together into the same plasmid subclone. A chimeric subclone is recognized in an assembly as an unaligned high-quality portion of a read at the end of the subclone that matches consensus somewhere else on the BAC clone. Although Phrap identifies these problems, some small chimeric subclones can slip through the Phrap screening process. Once a chimeric clone is identified, it can be removed and the BAC reassembled.
5. *Deleted subclone.* Plasmid subclones from BAC clones sometimes have small deletions, resulting from short multicopy repeats or other sequences that are detrimental when in high-copy plasmids in *E. coli*. In these cases, the unaligned portion of a sequence read matches the consensus sequence further downstream. In Consed, these are usually tagged green, a "Match Elsewhere High Quality" tag.

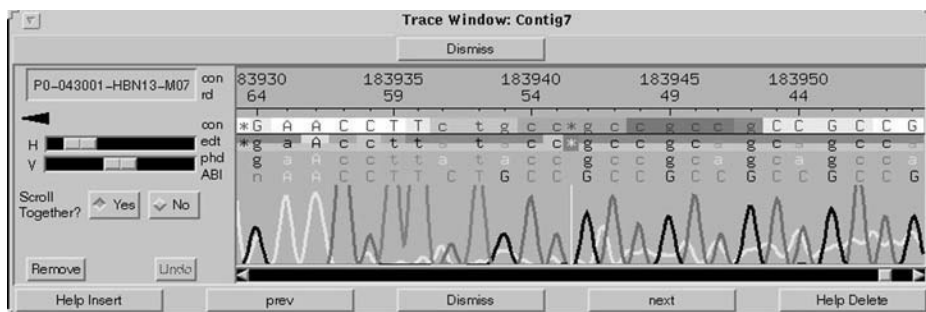


Fig. 1. Trace considered low quality owing to background noise. Several bases have been miscalled an “A” but can be called correctly from the trace data.

Occasionally, several deleted subclones are present for a region. In these cases, we sequence across the deleted region directly from the BAC clone.

6. *Unremoved sequencing vector*. High-quality unaligned bases that appear in a sequencing read at the junction between insert and vector are typically the unremoved subclone vector sequences. This can occur if Cross_match is unable to identify the complete vector sequences owing to many miscalled bases. In these cases, identify the vector junction and edit the bases out of the read.
7. *Misplaced sequencing read*. Occasionally, Phrap misplaces a sequencing read in an incorrect contig, and the read partially matches the consensus sequence. These misplaced reads are recognized by the presence of many HQDs throughout the entire length of the read. In these cases, remove the read from the incorrect contig and place it in its correct location.

After examining each HQD, the problem can be repaired computationally, or additional sequencing reads can be performed to support the hypothesis for the discrepancy.

3.3. Checking the Global Assembly, Paired Ends, and Restriction Digests

To verify the overall assembly of sequence in a BAC-sequencing project, open the assembly with PhrapView and examine the paired end links. Examine the distribution of paired ends across the clone. An example of good coverage of a BAC clone based on analysis of the paired ends is shown in **Fig. 2**. Light sequence coverage of paired ends or no coverage in a region is often evidence that the region is underrepresented or not represented in the subclones used for sequencing, possibly because of some biologic constraint effected by the sequence in the region. Pay special attention to assembly joins made in these areas and consider verifying the assembly by performing additional sequencing



Fig. 2. Good paired ends across a contig. PhrapView displays the reads from opposite ends of a subclone as a parabola with the ends terminating at the start of each read. The entire contig is covered by subclones that are within the expected size range.



Fig. 3. Paired end view of assembly with a low-coverage area. There are no appropriately sized subclones that cover the area indicated by the arrow.

reactions directly from the BAC clone or by determining the size of the region in question by PCR amplification from the BAC clone.

One of the quality requirements of genome sequencing is that the BAC clone is covered across most of the sequence by multiple subclones or has additional sequencing reads directly off the BAC clone. Generally, when an area of a large clone is covered by only one subclone, there is a possibility that the subclone has misrepresented the actual sequence of the BAC. It is unlikely that multiple subclones containing the same sequence have the same mistake or mutation, but this can happen in some cases. In **Fig. 3**, the arrow points to a potential problem area that needs further examination. After checking the paired ends, examine contig matches. In this case, look for sets of red lines terminating within the large contig, or from the small leftover contigs to the finished contig. Examine each of these small contigs in Consed and attempt to place them back into the large contig. A small contig that matches in two places can indicate the presence of a chimeric subclone or several subclones that have deleted a similar portion of sequence. This occurs with a reasonable frequency in areas that contain short nucleotide repeats.

Global verification of an assembly can be done in the laboratory by digesting the BAC clone with restriction enzymes and determining the sizes of the resulting fragments by agarose gel electrophoresis. These sizes then can be compared with the sizes of fragments expected from the sequence itself by computationally producing the set of restriction fragments based on the assembled finished sequence (*see Note 2*).

The following experimental protocol can be used to verify an electronic digest:

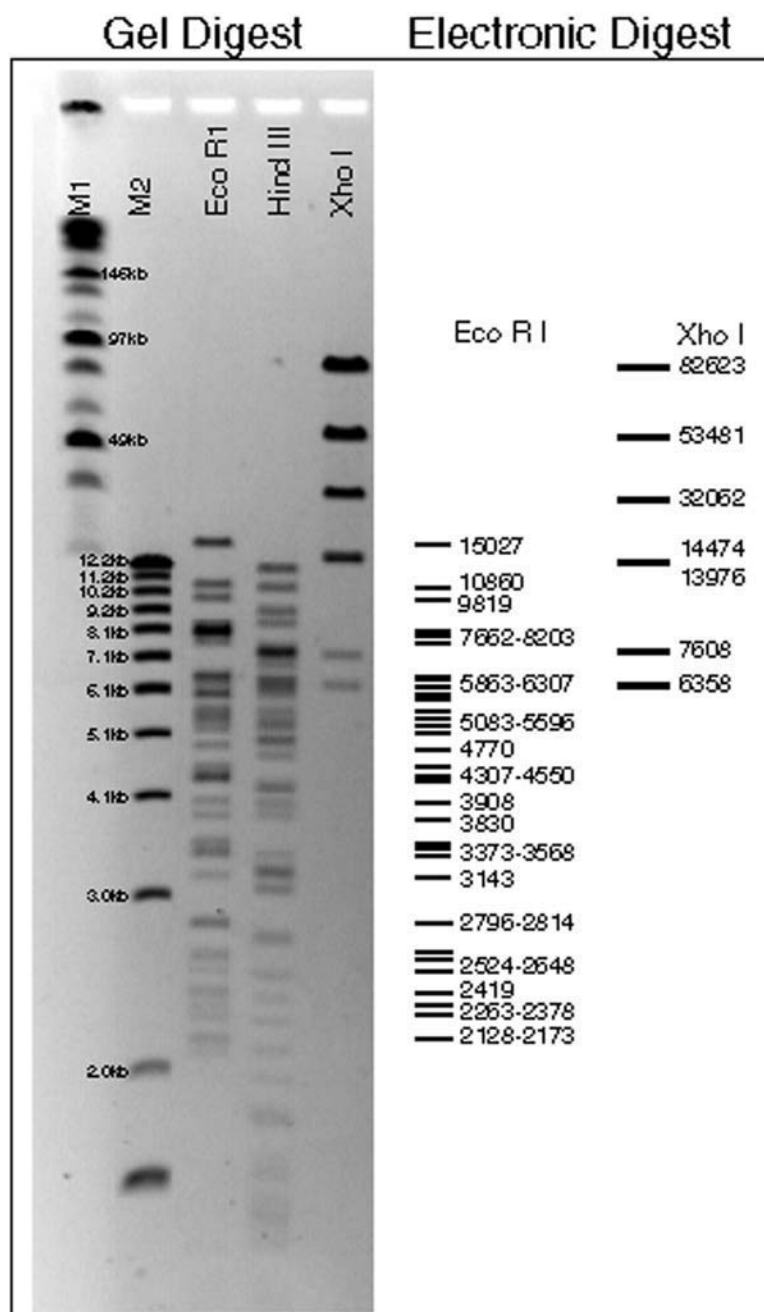


Fig. 4. Comparison of a gel and an electronic digest of the same BAC clone.

1. Perform a restriction digest of a BAC sample as follows: 1 μ L of BAC DNA, 5.5 μ L of ddH₂O, 2 μ L of 10X buffer, and 1 μ L of restriction enzyme (20 U). Perform separate reactions for each restriction enzyme. Mix well and incubate at 37°C for 3 h.
2. Run the samples for 14 h on a 1% agarose gel on a PFGE apparatus. Use DNA size markers in control lanes.
3. Stain the gel for 2.5 h in EtBr, and then destain it in water for 15 min. Take an image of the gel.

3.4. Completing Assessment and GenBank Submission

Following the processes of shotgun, assembly, finishing, and quality assessment, the BAC clone sequence is ready to be submitted to GenBank, the American repository for biologic sequence information at the National Center for Biotechnology Information. See the center's Web site (www.ncbi.nlm.nih.gov) for submission information.

4. Notes

1. The current set of rules for finished human genomic clones can be found on the Web site for the Washington University Genome Sequencing Center (<http://genome.wustl.edu>). Whenever possible, our group at the SHGC attempts to exceed these finishing standards when finishing BAC clones.
2. At the SHGC, we do this type of analysis by using three restriction enzymes—two that are common cutters and one that cuts rarely in human DNA—and then examine the restriction digests for any significant missing or additional bands (*see Fig. 4*).

