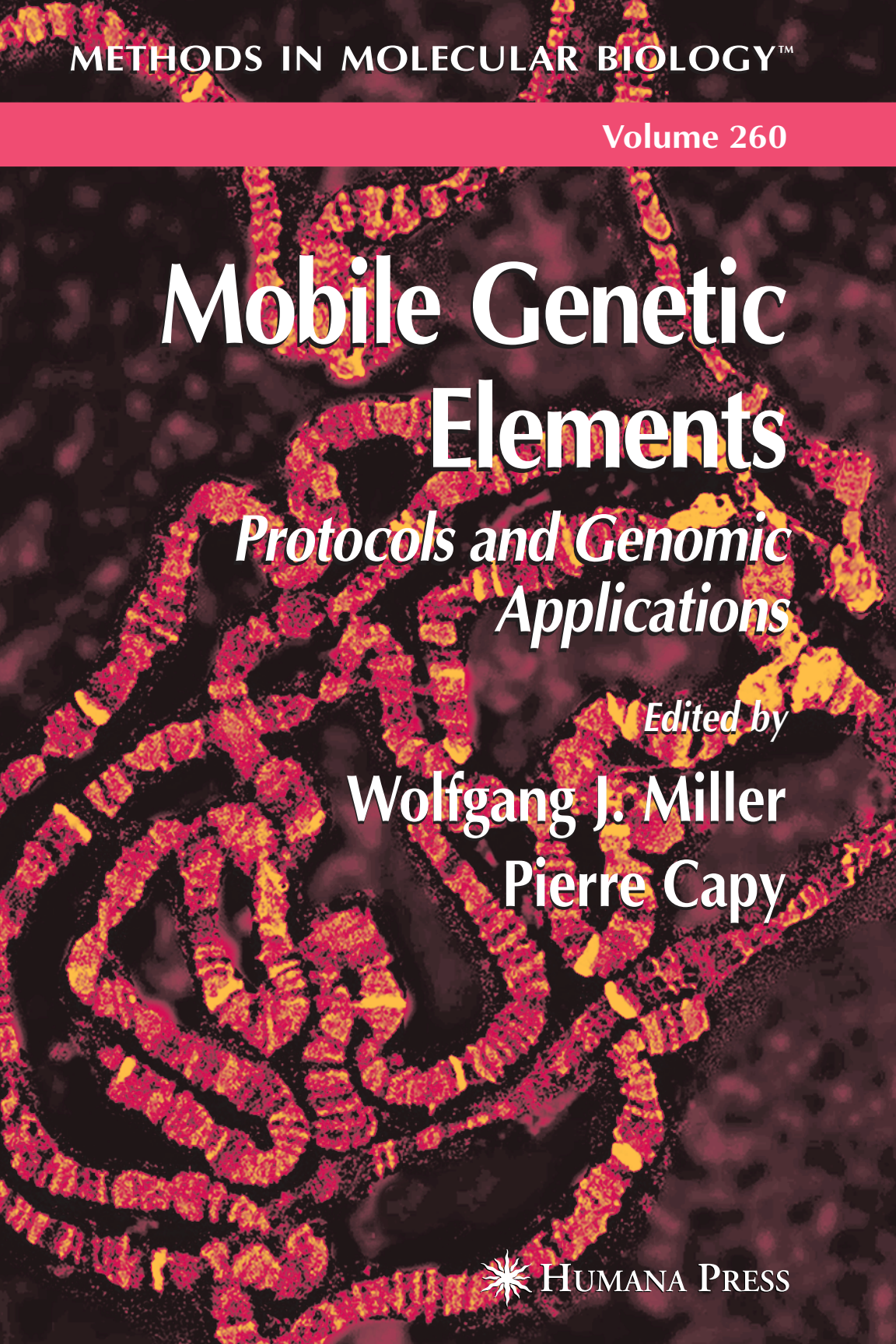
METHODS IN MOLECULAR BIOLOGY™

Volume 260



Mobile Genetic Elements

Protocols and Genomic Applications

Edited by

Wolfgang J. Miller

Pierre Capy

HUMANA PRESS

1

Mobile Genetic Elements as Natural Tools for Genome Evolution

Wolfgang J. Miller and Pierre Capy

Summary

Transposable elements (TEs) are ubiquitous components of all living organisms, and in the course of their coexistence with their respective host genomes, these parasitic DNAs have played important roles in the evolution of complex genetic networks. The interaction between mobile DNAs and their host genomes are quite diverse, ranging from modifications of gene structure and regulation to alterations in general genome architecture. Thus over evolutionary time these elements can be regarded as natural molecular tools in shaping the organization, structure, and function of eukaryotic genes and genomes. Based on their intrinsic properties and features, mobile DNAs are widely applied at present as a technical "toolbox," essential for studying a diverse spectrum of biological questions. In this chapter we aim to review both the evolutionary impact of TEs on genome evolution and their valuable and diverse methodological applications as the molecular tools presented in this book.

Key Words: Transposable elements; selfish DNAs; genome evolution; neogene formation; heterochromatin; stress induction; molecular tools.

1. Introduction

Many organisms contain far more repetitive DNA sequences than single-copy sequences. Repetitive sequences include mobile genetic DNAs that are universal components of all living genomes. Transposable elements (TEs) are gene-sized segments of DNA with the special ability to move between different chromosomal locations in their hosts' genome. Today the genomes of virtually all eukaryotic and prokaryotic species are known to contain significant numbers of TEs.

1.1. Occurrence and Classification

In some bacterial species, up to 10% of the genome is composed of insertion sequences (IS elements), while in eukaryotes these elements can make up more

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

than 50%. In genetic model systems like *Drosophila melanogaster*, *in silico* analyses have recently indicated that approx 22% of its genome is built up by TEs and their remnants ***(1)***. Even in humans, about half of the genome is derived from transposable elements—in particular, long interspersed elements (LINEs), short interspersed elements (SINEs), LTR retrotransposons, and DNA transposons ***(2)***.

When compared to the genomes of other eukaryotic organisms such as mouse, fly, worm, and mustard weed, the human genome has a much higher density of TEs in the euchromatin. This difference is based on the finding that the vast majority of TEs in humans seem to be more ancient and mainly transpositionally inactive, while in the model organisms mentioned above mobile DNAs are younger and thus still more active ***(2)***.

TEs are classified into two major groups based on their transposition mechanism ***(3)***. Class I elements, such as LTR-retrotranposons and LINEs, are characterized by DNA sequences with homology to reverse transcriptase, and they are often referred to as retroelements or retrovirus-like elements. Their mobility is achieved through an RNA intermediate that is reverse-transcribed prior to reinsertion, thus mediating a "copy-and-paste" mechanism. This group also includes the SINE elements that use the reverse transcriptase of LINEs.

Class II elements are characterized by terminal inverted repeats (TIRs), and they use DNA as a direct-transposition intermediate. They are therefore called DNA transposons and move by a conservative "cut-and-paste" mechanism catalyzed by a transposase. This enzyme is element-encoded in the autonomous DNA transposons and is provided *in trans* for internally deleted, nonautonomous elements.

1.2. Historical Overview

In the course of the twentieth century, our vision of the genome dramatically evolved from that of a stable and almost fixed structure to that of a highly flexible and dynamic information storage system. In the first half of the last century, the genome was basically considered as a stable chain of genes located in a head-to-tail organization along chromosomes, slowly evolving by the accumulation of random mutations at constant frequencies. Today such a conception is outdated, but it took more than 30 yr to change this dogma ***(4)***.

Based on her pioneering work on chromosome breakage in maize in the early 1940s, Barbara McClintock provided the first direct experimental evidence suggesting that genomes are not static but highly plastic entities ***(5)***. Elements involved in these phenomena were initially called "controlling elements." Based on her observations that some breakage events were always observed at the same chromosomal region, McClintock assumed that these events were due to a particular genetic element named *Ds* for "Dissociation."

In later work she deduced that the instability of *Ds* elements causing chromosomal breakage is dependent on the presence of another type of element designated as *Ac* for "Activator." Later on in the 1980s, molecular techniques revealed that the *Ac–Ds* system is composed of autonomous (*Ac*) and nonautonomous (*Ds*) copies, whereas only *Ac* encodes the functional transposase enzyme required for the mobility of both elements ***(6)***. Although McClintock's genetic work was the first clear indication of the existence of mobile DNA elements serving as a major genetic source for genome plasticity, it took more than 30 years before her concept of a dynamic genome became generally accepted ***(7–9)***.

Between the 1960s and 1970s the following observations paved the way for the discovery and the molecular characterization of mobile DNAs in prokaryotes (*reviewed in* **ref. *4***): The discovery of the bacteriophage *Mu* ***(10)***; and the elucidation of IS elements ***(11–13)*** as causative agents of mutations, along with their capacity for transmitting antibiotic resistance ***(14,15)***. As soon as appropriate molecular tools were developed for eukaryotic systems in the early 1980s, TEs were recognized immediately as universal components of all living organisms.

Two types of theories have been suggested to explain the ubiquitous presence of TEs as well as their high genomic proportions. Soon after the initial discoveries regarding TEs, researchers influenced by the "phenotypic paradigm" of the neo-Darwinian theory broadly speculated that mobile DNAs provide a direct selective advantage to their host organisms. Alternatively, in the light of the emergence of the neutral theory at the end of the 1970s and early 1980s, mobile DNAs were classified as "selfish DNAs" or "ultimate parasites" ***(16,17)***. The authors of both classic papers pointed out that the presence and spread of mobile DNAs could be explained solely by their ability to overreplicate the genes of the host genome without invoking a positive selection advantage at the level of the individual organism. As dogmatically stated by Dawkins ***(18)***, mobile DNAs are "…genes or genetic material which spread by forming additional copies of itself within the host genome and do not contribute to the phenotype. …"

During the last two decades, detailed molecular analyses of transposable elements, focusing on their dynamics and evolution within the host genomes, have modified our perception. Although it is generally accepted at present that mobile DNAs can be regarded as genomic parasites producing mainly neutral and deleterious effects, some of their induced mutations and genomic changes have made significant contributions to the evolution of their hosts ***(19–21)***. In this respect these elements can be regarded as a useful genetic load or even as useful parasites ***(22)***.

Today, it seems increasingly obvious that genomes can profit from the presence and action of mobile DNAs at various levels bringing about acceleration of genome evolution, as will be detailed in the sections that follow. Of course, mobile DNAs are not the only factor driving genome evolution, but it seems that they could be present at the origin of important events. Therefore, mobile DNAs can be regarded as evolution accelerators, particularly when genomes are facing population and/or environmental stresses ***(23)***.

In general, TEs are found in all kinds of genomic compartments, such as pericentromeric heterochromatin, telomeres, regulatory regions, exons, and introns. *A priori*, they can move everywhere in a genome, because their actual genomic target sites consist of a few base pairs only. However, they are not randomly distributed since they are frequently observed in heterochromatin and in regulatory regions. It remains difficult to demonstrate whether they preferentially target such regions by target sequence specificity or chromatin accessibility, or instead integrate randomly in the genome with natural selection then retaining and accumulating insertions at particular genomic compartments.

In the following sections, we discuss several aspects of the dynamics and evolution of TEs and their interactions with the host genome. Extensive reviews have been published recently covering in detail the broad spectrum of TE–host interactions and their evolutionary consequences ***(9,19–21)***. Thus we will first review briefly some of the most important impacts of TEs acting as natural tools on host genome evolution, so that we may then introduce their technical applications as molecular tools and molecular marker systems in modern biology.

2. The Role of TEs as Natural Tools for Shaping Genome Evolution

2.1. Heterochromatin: Only a Wasteland for Transposable Elements?

The evolutionary relationship between TEs and heterochromatin is still controversial. In general, TEs and their derivatives are found as highly enriched clusters in genomic regions close to the centromere and telomere, and along the chromosomal arms within the intercalary heterochromatin. Obviously TE insertions in heterochromatin are less deleterious than euchromatic insertions, and their concentration in these regions of low gene density might be mainly due to selection against ectopic recombination ***(24)***. Indeed, theoretical models have implied that TEs should accumulate in regions with low rates of recombination, such as in the heterochromatin ***(25,26)***. Recent experimental data obtained from *Drosophila*, however, have provided no sufficient support for the hypothesis that the primary reason for the accumulation of mobile DNAs in the heterochromatin is selection against TEs in the euchromatin ***(27,28)***. As suggested by Dimitri and Junakovic, "Their accumulation in heterochromatin does not seem to be related to intrinsic properties of transposon families … [but could be] determined by some sort of interaction between each transposon

family and the host genome" ***(29)***. The authors conclude that the heterochromatin might attract *de novo* insertions of mobile elements mediated by host factors that provide a safe haven to the elements themselves, and thus minimize their mutagenic effects in the euchromatin.

Moreover, there is also accumulating evidence for direct contribution of TEs in the evolution of heterochromatin. Tandem arrays of engineered *P* elements inserted in euchromatic positions are sufficient to cause *de novo* formation of heterochromatin-like structures ***(30,31)***, whereas 5S genes do not. Thus, formation of heterochromatin seems to have some sort of sequence requirement that is met by at least some sorts of TEs. Although the nature of these proposed special requirements is still unknown, it seems likely that only their structural repetitions are important, thus serving only some structural roles for modifying chromatin. Indeed, in *Zea mays* the *Huck* retrotransposon seems to provide a structural component the centromeric regions ***(32–34)***.

Consistent with this conclusion, for example, is the massive insertion of *TRIM* and *TRAM* retroelements that has been correlated with heterochromatinization of the neo-Y chromosome of *D. miranda* ***(35)***; another example is provided by the functional transition of a formerly active SGM transposon into the structural repetition unit of the main heterochromatic satellite of *D. guanche* ***(36)***.

2.2. TEs and Their Role in Restructuring Chromosomes

Barbara McClintock originally discovered mobile DNAs in *Zea mays* because of their potential to cause chromosomal mutations such as deletions, translocations, and inversions ***(5)***. In *Drosophila*, TEs can be found at the breakpoints flanking chromosomal inversions in both natural populations and laboratory strains ***(37,38)***. The *hobo* element was reported at the breakpoints of three endemic inversions from Hawaiian populations of *D. melanogaster* ***(39)***. In the laboratory strain of the Hikkone line transformed with an active copy of *hobo* (HFL1), inversions were detected after 50 generations, some of them similar to endemic ones found in natural populations ***(40)***. In addition, rare inversions flanked by *P* elements at the breakpoint were also observed in natural populations collected in the southeastern U.S. ***(41)***. Such phenomena are not restricted to *D. melanogaster*; similar events have been reported from other *Drosophila* species such as *D. buzatii* ***(42)***. Moreover, it has been shown that all classes of mobile DNAs are capable of causing chromosomal inversions ***(43,44)***.

2.3. Emergence of New Genes or New Functions

In general, class I elements are defined as using reverse transcriptase (RT) for their own propagation, but in some cases a specific RT enzyme can be recruited for other purposes, such as *trans*-mobilization of other TEs and

pseudogene formation. For example, SINE elements do not encode the proteins require for their retroposition, but use RT encoded from other elements, i.e., LINEs ***(45,46)***. Moreover, L1-encoded RT is able to give rise to retroprocessed pseudogenes in humans ***(47)***.

Most of these retrotransposed host gene sequences will evolve like classical pseudogenes, but in some cases such events can initiate the formation of neogenes, which provide a new function to the host. Indeed, retroposition has been viewed as sowing the "seeds" for the evolution of novel gene function ***(48)***. As one example, the presence of the *Jingwei* neogene is restricted to the closely related species *D. teissieri*, *D. yakuba*, and *D. santomea*, belonging to the *melanogaster* subgroup, and is absent in all other species of *Drosophila*. This neogene has been originated by the reverse transcription of a spliced *Adh* mRNA fused to the exons and introns of the *yellow emperor* gene ***(49,50)***. In primates, the chimerical *PMCHL* neogenes emerged from the initial reverse transcription of the *AROM* sequence ***(51,52)***. Additional cases supporting the important evolutionary role of retroposition in gene evolution have been recently reviewed in detail ***(53,54)***.

In contrast to the above-mentioned indirect effects on neogene formation induced by retroposition, even the coding section of mobile DNAs can co-opt new host functions, a mechanism designated as "molecular domestication" ***(20,55,56)***. For instance, the non-LTR retrotransposons *TART* and *Het-A* are exclusively found at the telomeric positions of *Drosophila* chromosomes ***(57–59)***. Because *Drosophila* lacks conventional telomeres and telomerase, these retroelements play an essential role in counteracting the erosion of chromosomal ends and thus providing a substitute for telomerase function to the host.

Molecular domestication of mobile DNAs is not restricted to class I elements. As deduced from the initial sequence analyses of the human genome, at least 45 human host genes with currently unknown function unequivocally stem from the coding region of formerly active class II elements ***(2)***. So-called transposase-derived neogenes were earlier isolated from various *Drosophila* species belonging to the *obscura* and *montium* species group ***(55,56,60)***. In this case, *P* element-related neogenes have evolved at least two times independently from coding derivatives of once-mobile *P* element transposons in separate lineages of *Drosophila*. Although the functional properties of the *P* element-derived neogenes are still unknown in their respective hosts, this system provides the first case for a multiple independent acquisition of the same type of TE-derived coding section during *Drosophila* evolution ***(56)***. Moreover, both independent cases of *P* element domestication were accompanied by further TE-induced events giving rise to (1) the formation of novel *cis*-regulatory section by multiple insertions of non-*P* element-related TEs in the

obscura group ***(36)*** and (2) the *de novo* synthesis of a new intron by the noncoding sections of the *P* element in the *montium* subgroup ***(60)***.

The most spectacular example of molecular domestication of TEs is the recent finding that a key function of the vertebrate immune system most likely evolved directly from a formerly active DNA transposon approximately 100 mya ***(61–63)***. The recombination of the V(D)J locus is catalyzed by two enzymes, *RAG1* and *RAG2*, with significant functional and structural similarities to Tc1 transposons. Furthermore, the binding sites for the major centromere-binding protein (CENP-B) of mammals, the "CENP-B box," have been shown to match the terminal inverted repeats of the *pogo* DNA transposon (*reviewed in* **ref. *64***), and the protein CENP-B itself is an ancient descendant of a *pogo*-like transposase with a well-conserved DNA-binding domain ***(65)***. These data strongly imply that derivatives of once-mobile DNAs can play important roles in the evolution of essential hosts' cellular functions, such as telomere elongation, immune response, and chromosome segregation.

2.4. Transposable Elements Are the Wild Cards of the Genome

Under stable or slightly variable genomic and ecological conditions, the transposition rate of TEs seems to be relatively low. In natural populations of *D. simulans* the transposition rate of *412* retrotransposons ranges between 10^{-3} to 2×10^{-3} independent of the copy number in their respective genomes ***(66)***. These values are significantly higher than earlier estimations (10^{-5} to 10^{-3}), which were mainly deduced from laboratory strains ***(66–68)***. Therefore, the transposition rates in laboratory strains seem to be one or two orders of magnitude lower than those derived from natural populations.

As suggested by McClintock as early as the late 1970s, genome restructuring mediated by TE activity can be seen as an essential component of the hosts' response to stress, facilitating the adaptation of populations and species facing changing environments ***(69)***. Following this assumption, three essential conditions must be fulfilled: (1) TEs have to be capable of responding to stress by enhancing their transcriptional and transpositional activity; (2) the enhanced TE mobility has to be sufficient for generating broad genetic variation within the host genome; and (3), this new genetic variability has to be transmissible from one generation to the next.

Several lines of arguments are in agreement with the first criterion. Transcription of the *Tnt1* retrotransposon of *Nicotiana tabacum*, for instance, seems to be inducible by several biotic and abiotic stress factors ***(70–72)***, followed by an actual enhanced mobility of the retrotransposon ***(73)***. Moreover IS elements in bacteria may also play an important role in adaptive mutagenesis ***(74,75)***. Significant differences of transposition rates are detectible between natural populations within a given species of *Drosophila*. Some of these differences

are structured according to the geographical origin of the populations. For instance, the activity of *mariner* and *412* elements exerts a latitudinal variation pattern along an African–European axis. Whereas *mariner* shows latitudinal variations of the somatic excision rate ***(76)***, *412* varies with respect to its copy number ***(77)***. Furthermore, developmental temperature ***(76,78–80)*** and exposure to insecticides seem to increase the somatic excision rate of *mariner* from a reporter gene (Meusnier, Guichou, and Capy, unpublished results).

Fewer experimental data are available to in order to support the second and third criteria. Mackay studied hybrid dysgenesis in *D. melanogaster*, finding that it was induced by bursts of *P* element transpositions ***(81)***. In the progeny of dysgenic crosses, the response to selection, i.e., to increase or decrease the abdominal bristle number, is higher than in progeny of nondysgenic parents, suggesting that the mutational activity of the *P* element is sufficient for causing genetic variability on which selection can operate. Based on this pioneering work, several groups have shown that a number of other traits can be affected by transpositional activity ***(82–86)***.

Although the concept of stress response seems conclusive, some problems still remain to be solved. First, not all types of TEs might be capable of activating transposition due to stress. This specificity probably results from particular small nucleotide motifs located within the regulatory section of the TE. Indeed, such binding site motifs, similar to the plant defense-response elements, were detected in the *Tnt1* element ***(71)***. Within the untranslated leader region of the *Drosophila copia* element, sequence motifs were found similar to the core sequence of the *SV40* enhancer ***(87)***. Therefore, the potential of a specific TE to respond to specific stress might be caused by the presence and accumulation of specific inducible enhancers in their regulatory regions. As stated by McDonald et al. ***(87)***: "inter-element selection may favor the evolution of more active enhancers within permissive genetic backgrounds. We propose that LTR retroelements and perhaps other retrotransposons constitute drive mechanisms for the evolution of eukaryotic enhancers which can be subsequently distributed throughout host genomes to play a role in regulatory evolution."

The fact remains that most of the reported cases of stress-induced TE mobilizations were assayed in somatic tissues only. However, a long-term adaptation of the host to environmental changes requires germline modifications ***(23)***. In *Drosophila*, it was assumed that a product derived from the activity of an element might be transferred to the next generation via the egg cytoplasm, causing maternal effects and in some case even grand-maternal effects ***(88–92)***.

3. The Taming of TEs and Their Technical Applications

At present a deep and detailed understanding of the complex biology of mobile DNAs and their short-term as well as long-term evolutionary fate and

consequences within genomes is essential for their successful technical application. Based on their exceptional biological features, TEs provide a valuable collection of molecular tools and experimental strategies appropriate for elucidating a diverse spectrum of biological questions.

The most prominent features of TEs are obviously their invasiveness, the structural and functional consequences caused by their genomic insertions, and their potential ability to cross species boundaries. Therefore, TE-based experimental strategies serve as standard key molecular tools in modern biology for investigating the structure, organization, and function of genes and genomes. However, prior to the successful application of a given TE as a mutagenic agent, a marker system, or a genetic vehicle for transgenesis, a detailed analysis of the structure, function, and dynamics of the mobile element itself is essential. In this respect several protocols for studying the biology of mobile elements by in vivo, in vitro, or *in silico* approaches are presented in detail in Chapters 2–7 of this book, ranging from high-resolution detection approaches such as *in situ* hybridization and Southern blot techniques to biochemical and computational *in silico* whole-genome analyses. In the rest of this book, a large spectrum of technical applications is provided, including protocols for insertional mutagenesis, gene tagging, gene silencing, molecular marker analyses, and genetic transformation systems in arthropods and vertebrates.

Transposable elements were initially discovered because of their ability to disrupt genes spontaneously, thus acting as natural mutagenes. In the early 1980s the transposon tagging technique was developed in *Drosophila* as a strategy to clone genes, representing the very first transposon-based, genome-wide approach to study gene function in eukaryotes. In later research the systematic extension of this *P* element-induced, gene disruption technique finally resulted in a compendium of thousands of *P* insertion lines, covering one-fourth of the vital genes of *D. melanogaster* ***(93)***. Similar genome-wide, TE-based gene disruption strategies were successfully designed and established for a number of other genetic model systems, ranging from *Saccharomyces cerevisiae* to mouse.

TE-based insertional mutagenesis systems can be applied both to localize and isolate a gene involved in a known function, and to infer the function of a gene known only from its sequence. Finally, the objective is to target a TE into a specific gene of interest for analyses of loss or even gain of function. For a long time the technical ability to target DNA sequences to a specific locus were restricted to genetic systems such *S. cerevisiae* and mouse, but not available for *Drosophila*. Currently, *Drosophila* biologists can choose between two different methods for gene targeting, both utilizing the natural tendency of the cell to repair DNA double-strand breaks left behind after the excision of a DNA transposon. The first method, named the "gene conversion technique," depends on the presence of a *P* element insertion close to the gene of interest

(94,95). More recently, a second method was developed, designated as the "homologous gene targeting technique" ***(96)***. This strategy is a combination of *P* element-mediated transformation, FLP-FRT recombination, and the *I-SceI* endonuclease system, the latter two derived from yeast. Protocols for applying both methods in *Drosophila* are provided by Gregory Gloor in Chapter 8.

Today, insertional mutagenesis techniques serve as the standard reverse genetics tool for characterizing the function of a given gene in a diverse set of organisms. However, insertions in specific genes belonging to large gene families often do not change a phenotype, simply due to redundancy. In Chapter 9, Vandenbussche and Gerats present a newly developed TE-based mutagenesis protocol for plants in order to overcome this problem by designing a gene-family-specific primer for rapid PCR screening.

In the course of their long-term coexistence with mobile elements, host genomes might have evolved mechanisms counteracting the mobility and mutability of TEs. A growing body of research suggests that epigenetic regulatory mechanisms such as methylation, heterochromatization, and cosuppression arose originally as defense mechanisms against mobile DNAs ***(97,98)***. These findings opened for discussion the question of whether TEs might be regarded as the driving force in the evolution of epigenetic regulatory mechanisms in eukaryotes (*see* **ref. *97***) and thereby might have contributed to two main macroevolutionary transitions in the history of life, namely chromatin formation for the prokaryotic/eukaryotic transition, and methylation for the invertebrate/vertebrate transition ***(99)***. Today the evolutionary relationship between TEs and epigenetic silencing mechanisms is generally appreciated by investigators.

Post-transcriptional gene silencing (PTGS) was first discovered as a subset of cosuppression in plant transgenesis experiments when the transgene was integrated as multiple copies or was identical to endogenous sequences ***(100)***. Contrary to expectations, the increased gene dosage did not result in enhanced expression, but in gene silencing. Subsequent work identified distinct nucleic sequence homology-based mechanisms that lead to transcriptional or post-transcriptional gene silencing designated as TGS and PTGS, respectively.

The technical application of RNA interference (RNAi) provides a tremendously powerful knockout tool for the selective ablation of gene expression for reverse genetics in various organisms. Originally discovered in *Caenorhabditis elegans* ***(101)***, RNAi is a post transcriptional gene silencing mechanism targeting double-stranded RNAs (dsRNAs) leading to the specific degradation of mRNAs with homology to the dsRNA source. Subsequent mutagenesis experiments have identified various genes that are involved in regulating RNAi, but some of these mutants also reactivate otherwise-silenced transposable elements ***(102,103)***. These data strongly suggest that at least some components of RNAi might serve a critical role in silencing genetic parasites.

The main objective of RNAi-based methodologies is to reveal the phenotype of a given gene by providing dsRNAs derived from the coding section of the gene of interest. Today, there are several methods available to deliver dsRNA into a broad range of organisms (*see* Chapter 10). Clearly, the most efficient method is to generate stable transgenic organisms by microinjecting a construct producing hairpin dsRNAs in vivo under the control of an inducible promoter system. Following this technique strategy, heritable gene-silencing mutants can be generated and maintained over generations.

The "copy-and-paste" mode of transposition is a characteristic feature of retrotransposons (class I). Thus retroelements once inserted at any specific locus in the genome are incapable of excising actively, leaving a fixed mark in the genome. Rare but incomplete excision events can be caused by ectopic recombination between LTRs of *Pseudoviridae* or *Metaviridae*, or between two neighboring copies of the same type of element when they are in the same orientation. In both cases, ectopic recombination gives rise to a deletion of the genomic region originally spacing the two repeated sequences, whereas the remaining sequence left behind is composed of a hybrid structure of the two initial copies. For Ty elements in the *Saccharomyces cerevisiae* genome, the complete list of full-length copies and solo-LTRs that have resulted from ectopic recombination between the terminal repeats is well documented ***(104)***. These studies conclude that insertions of retroelements are relatively stable over evolutionary time, thus providing an excellent set of highly polymorphic molecular marker systems. In Chapter 11, Schulman and colleagues provide a collection of retrotransposon-based PCR protocols for plants, but the rationale of these techniques is easily adaptable for animals and humans as well. The technical application of mobile DNAs for serving as polymorphic marker systems is not limited to LTR elements. In Chapter 12, Wessler and collaborators present a detailed protocol for the usage of another group of TEs named miniature inverted transposable elements (MITEs).

Okada and collaborators have developed a retrotransposon-based technique for the vertebrate system (*see* Chapter 13). This so-called retroposon-mapping technique is mainly based on the features of SINEs. These elements are widely distributed as well as highly abundant throughout vertebrates, making up, for instance, more than 12% of the human genome.

It seems obvious that each of the TE-based protocols provided here can be easily applied to a broad range of investigations, for the analyses of population structures and for phylogenetic analyses of species. In addition, the polymorphism of the TE insertion sites provides highly informative sets of chromosomal marker for QTL mapping strategies. Depending on the group of organisms under examination the most informative type of retroelement will be selected according to its abundance, mobility, and genomic distribution.

Based on these criteria, SINE elements are the marker system of choice for analyzing vertebrates, whereas for arthropods LINEs and LTR retrotransposons as well as MITEs are useful candidates. For instance, insertions of the LTR element *roo*/B104 were successfully applied in *Drosophila* for QTL mapping of chromosomal regions involved in fitness-related traits such as reproductive success, ovariole number, body size, and early fecundity ***(105)***.

In the course of extensive evolutionary surveys on the distribution of mobile elements within and between eukaryotic species, it has been clearly shown that TEs have the capacity to cross species boundaries followed by their successfully propagation in a new host environment (reviewed in **refs.** ***22,106***). This so-called "horizontal transfer hypothesis" is frequently proposed as soon as inconsistencies are observed between phylogenies of the host species and TEs. However, in some cases, alternative models such as variable evolution rates, stochastic loss, or comparisons between orthologous and paralogous sequences might serve as more appropriate explanations for these inconsistencies ***(107,108)***. Nevertheless, various unequivocal cases for lateral transfer events of TEs between distantly related hosts are well documented (*see* **refs.** ***109–113***). Based on their intrinsic abilities to integrate actively into genomes and to invade other species, mobile DNAs provide powerful molecular tools for cross-species transgenesis. In the past two decades various TE-based vector systems were designed and successfully applied in a broad range of organisms.

For almost 20 years the *P* element provided the standard genetic transformation system for *Drosophila*, but its mobility seems to be restricted to the family of *Drosophiliae* (*see* **refs.** ***114–116***). Thus, more universal vectors systems were developed according to two main strategies. First, natural TEs were isolated and characterized as appropriate for transgenesis on a much broader spectrum of species. Today, *Hermes*, *PiggyBac*, *minos*, and *mariner* elements are among those frequently used, at least in arthropods (***117***; *see* Chapter 14). Second, natural elements have been artificially modified in order to improve their transfer efficiency. Such an approach has been successfully developed for the *Sleeping Beauty* transposon in vertebrates (***118***; *see* Chapter 15) as well as for *mariner* elements ***(119,120)***. However, in spite of the fact that TE-based techniques serve as standard tools for transgenesis at present, several open problems remain to be solved. For instance, transgenes in plants and other organisms are often found to become epigenetically silenced by processes that are best interpreted as cellular defense reactions to parasitic sequences ***(121)***. In addition, the stability of a transgene once inserted into a specific genomic position in its new host has to be assured. Studies in insects, for example, have shown the ability of the *hAT* DNA transposons *hobo* and *Hermes* to interact and cause cross mobilization. Using plasmid-based and chromosome-based element mobility assays, it was found that the terminal sequences of *hobo* and

Hermes were almost equally good substrates for *hobo* transposase ***(122)***. This suggests that a detailed screening of the recipient host genome for functionally related TEs is required prior to the selection of the vector system in order to avoid cross mobilization. Finally, in human gene therapy, the problem of targeting a transgene into a specific insertion site in order to replace a defective homologous gene remains unsolved ***(123)***.

As briefly reviewed in the first part of this chapter, mobile DNAs serve a number of important functions as natural molecular tools for hosts' genome evolution. Based on their intrinsic properties, TEs immediately became an essential "tool box" for all scientists interested in a broad range of biological and medical questions. The detailed protocols for each technique are presented in the following chapters. All of them have been developed and further improved within the last two decades. It is expected that in the next few years novel TE-based techniques will be developed, expanding the repertoire of the "tool box" dramatically. Indeed, based on the rapidly accumulating data obtained from more and more whole-genome sequencing projects, TEs should no longer be considered as purely parasitic genetic elements or even "junk DNAs," but as essential components driving genome evolution. Therefore, with a further expansion of our understanding on TE biology in the very near future, new characteristics of TEs will be discovered that will be useful in innovative technical applications.

References

1. Kapitonov, V. V. and Jurka, J. (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. USA* **100,** 6569–6574.
2. International Human Genome Sequencing Consortium (2001) A physical map of the human genome. *Nature* **409,** 934–941.
3. Berg, D. E. and Howe, M. M., eds. *Mobile DNA*. American Society for Microbiology, Washington, DC, 1989.
4. Shapiro, J. A. The discovery and significance of mobile genetic elements. In *Mobile Genetic Elements* (Sherrat, D. J., ed.). IRL Press, Oxford, 1995, pp. 1–17.
5. McClintock, B. *The Discovery and Characterization of Transposable Elements: The Collected Papers of B. McClintock*. Garland, New York, 1987.
6. Fedoroff, N., Wessler, S., and Shure, M. (1983) Isolation of the transposable maize controlling element *Ac* and *Ds*. *Cell* **35,** 243–251.
7. McDonald, J. F., ed. *Transposable Elements and Evolution*. Contemporary issues in Genetics and Evolution. Kluwer Academic Publishers, Dordrecht, Netherlands, 1993.
8. Britten, R. J. (1996) Cases of ancient mobile element DNA insertions that now affect gene regulation. *Mol. Phylogenet. Evol*. **5,** 13–17.
9. Kidwell, M. G. and Lisch, D. R. (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution Int. J. Org. Evolution* **55,** 1–24.

10. Taylor, A. (1963) Bacteriophage-induced mutation in *E. coli. Proc. Natl. Acad. Sci. USA* **50,** 1043.
11. Adhya, S. L. and Shapiro, J. A. (1969) The galactose operon of *E. coli* K-12. I. Structural and pleiotropic mutations of the operon. *Genetics* **62,** 231–247.
12. Shapiro, J. A. (1969) Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli. J. Mol. Biol.* **40,** 93–105.
13. Shapiro, J. A. and Adhya, S. L. (1969) The galactose operon of *E. coli* K-12. II. A deletion analysis of operon structure and polarity. *Genetics* **62,** 249–264.
14. Heffron, F. Tn*3* and its relative. In *Mobile Genetic Elements* (Shapiro, J., ed.). Academic Press, New York, 1983, pp. 223–260.
15. Kleckner, N. Transposon Tn*10*. In *Mobile Genetic Elements* (Shapiro, J., ed.). Academic Press, New York, 1983, pp. 261–298.
16. Doolittle, W. F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284,** 601–603.
17. Orgel, L. E. and Crick, F. H. C. (1980) Selfish DNA: the ultimate parasite. *Nature* **284,** 604–607.
18. Dawkins, R., ed. *The Selfish Gene*. Oxford University Press, UK, 1976.
19. McDonald, J. F. (1995) Transposable elements: possible catalysts of organismic evolution. *Trends Ecol. Evol.* **10,** 123–126.
20. Miller, W. J., McDonald, J. F., Nouaud, D., and Anxolabehere, D. (1999) Molecular domestication—more than a sporadic episode in evolution. *Genetica* **107,** 197–207.
21. Kidwell, M. G. and Lisch, D. R. (2000) Transposable elements and host genome evolution. *Trends Ecol. Evol.* **15,** 95–99.
22. Capy, P., Bazin, C., Higuet, D., and Langin, T., eds. *Evolution and Impact of Transposable Elements*. Kluwer Academic Publishers, Dordrecht, Netherlands, 1997.
23. Capy, P., Gasperi, G., Biémont, C., and Bazin, C. (2000) Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85,** 101–106.
24. Charlesworth, B., Sniegowki, P., and Stephan, W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371,** 215–220.
25. Charlesworth, B., Langley, C. H., and Stephan, W. (1986) The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* **112,** 947–962.
26. Charlesworth, B. and Langley, C. H. (1986) The evolution of self-regulated transposition of transposable elements. *Genetics* **112,** 359–383.
27. Biemont, C., Vieira, C., Hoogland, C., Cizeron, G., Loevenbruck, C., Arnault, C., et al. (1997) Maintenance of transposable element copy number in natural populations of *Drosophila melanogaster* and *D. simulans. Genetica* **100,** 161–166.
28. Dimitri, P., Arca, B., Berghella, L., and Mei, E. (1997) High genetic instability of heterochromatin after transposition of the LINE-like I factor in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **94,** 8052–8057.
29. Dimitri, P. and Junakovic, N. (1999) Revising the selfish DNA hypothesis: New evidence on accumulation of transposable elements in heterochromatin. *Trends Genet.* **15,** 123–124.

30. Dorer, D. R. and Henikoff, S. (1994) Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila. Cell* **77,** 993–1,002.
31. Dorer, D. R. and Henikoff, S. (1997) Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics* **147,** 1181–1190.
32. Ananiev, E. V., Phillips, R. L., and Rines, H. W. (1998) Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. *Proc. Natl. Acad. Sci. USA* **95,** 13,073–13,078.
33. Ananiev, E. V., Phillips, R. L., and Rines, H. W. (1998) Complex structure of knob DNA on maize chromosome 9. Retrotransposon invasion into heterochromatin. *Genetics* **149,** 2025–2037.
34. Ananiev, E. V., Phillips, R. L., and Rines, H. W. (1998) A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? *Proc. Natl. Acad. Sci. USA* **95,** 10,785–10,790.
35. Steinemann, M. and Steinemann, S. (1997) The enigma of Y chromosome degeneration: *TRAM*, a novel retrotransposon is preferentially located on the Neo-Y chromosome of *Drosophila miranda. Genetics* **145,** 261–266.
36. Miller, W. J., Nagel, A., Bachmann, J., and Bachmann, L. (2000) Evolutionary dynamics of the *SGM* transposon family in the *Drosophila obscura* species group. *Mol. Biol. Evol.* **17,** 1597–1609.
37. Lim, J. (1988) Intrachromosomal rearrangements mediated by *hobo* transposons in *Drosophila melanogaster. Proc. Natl. Acad. Sci. USA* **85,** 9,153–9,157.
38. Lim, J. and Simmons, M. J. (1994) Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster. BioEssays* **16,** 269–275.
39. Lyttle, T. W. and Haymer, D. S. (1992) The role of transposable element *hobo* in the origin of the endemic inversions in wild populations of *Drosophila melanogaster. Genetica* **83,** 113–126.
40. Ladeveze, V., Aulard, S., Chaminade, N., Periquet, G., and Lemeunier, F. (1998) *Hobo* transposons causing chromosomal breakpoints. *Proc. R. Soc. Lond. B. Biol. Sci.* **265,** 1157–1159.
41. Kusakabe, S., Harada, K., and Mukai, T. (1990) The rare inversion with a *P* element at the breakpoint maintained in a natural population of *Drosophila melanogaster*. *Genetica* **82,** 111–115.
42. Caceres, M., Ranz, J. M., Barbadilla, A., Long, M., and Ruiz, A. (1999) Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285,** 415–418.
43. Eggleston, W. B., Rim, N. R., and Lim, J. K. (1996) Molecular characterization of *hobo*-mediated inversions in *Drosophila melanogaster*. *Genetics* **144,** 647–656.
44. Lim, J. K. and Simmons, M. J. (1994) Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. *BioEssays* **16,** 269–275.
45. Okada, N., Hamada, M., Ogiwara, I., and Ohshima, K. (1997) SINEs and LINEs share common 3' sequences: a review. *Gene* **205,** 229–243.
46. Kajikawa, M. and Okada, N. (2002) LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* **111,** 433–444.

47. Esnault, C., Maestre, J., and Heidmann, T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24,** 363–367.
48. Brosius, J. (1991) Retroposons—seeds of evolution. *Science* **251,** 753.
49. Long, M. and Langley, C. H. (1993) Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila. Science* **260,** 91–95.
50. Wang, W., Zhang, J., Alvarez, C., Llopart, A., and Long, M. (2000) The origin of the *Jingwei* gene and the complex modular structure of its parental gene, *yellow emperor*, in *Drosophila melanogaster. Mol. Biol. Evol.* **17,** 1294–1301.
51. Viale, A., Courseaux, A., Presse, F., Ortola, C., Breton, C., Jordan, D., et al. (2000) Structure and expression of the variant melanin-concentrating hormone genes: Only *PMCHL1* is transcribed in the developing human brain and encodes a putative protein. *Mol. Biol. Evol.* **17,** 1,626–1,640.
52. Viale, A., Ortola, C., Richard, F., Vernier, P., Presse, F., Schilling, S., et al. (1998) Emergence of a brain-expressed variant melanin-concentrating hormone gene during higher primate evolution: a gene in search of a function. *Mol. Biol. Evol.* **15,** 196–214.
53. Long, M. (2001) The evolution of novel genes. *Curr. Opin. Genet. Dev.* **11,** 673–680.
54. Long, M., ed. *Origin and Evolution of New Gene Functions. Genetica* **118** (Special Issue). Kluwer Academic Publishers, Dordrecht, Netherlands, 2003.
55. Miller, W. J., Hagemann, S., Reiter, E., and Pinsker, W. (1992) *P* element homologous sequences are tandemly repeated in the genome of *Drosophila guanche. Proc. Natl. Acad. Sci. USA* **89,** 4018–4022.
56. Nouaud, D. and Anxolabéhère, D. (1997) *P* element domestication: A stationary truncated *P* element may encode a 66-kDa repressor-like protein in the *Drosophila montium* species subgroup. *Mol. Biol. Evol.* **14,** 1132–1144.
57. Biessmann, H., Valgeirsdottir, V., Lofsky, A., Chin, C., Ginther, B., Levis, R. W., et al. (1992) *HeT-A*, a transposable element specifically involved in "Healing" broken chromosome ends in *Drosophila. Mol. Cell. Biol.* **12,** 3910–3918.
58. Levis, R. W., Ganesan, R., Houtchens, K., Tolar, L. A., and Sheen, F. M. (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* **75,** 1083–1093.
59. Pardue, M. L. *Drosophila* telomeres: another way to end it all. In *Telomeres* (Greider, C. and Blackburn, E. H., eds.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1995, pp. 339–370.
60. Nouaud, D., Boeda, B., Levy, L., and Anxolabéhère, D. (1999) A *P* element has induced intron formation in *Drosophila. Mol. Biol. Evol.* 1503–1510.
61. Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394,** 744–751.
62. Hiom, K., Melek, M., and Gellert, M. (1998) DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations [see comments]. *Cell* **94,** 463–470.
63. Plasterk, R. (1998) Ragtime jumping. *Nature* **394,** 718–719.

64. Kipling, D. and Warburton, P. E. (1997) Centromeres, CENP-B and *Tigger* too. *Trends Genet.* **13,** 141–145.
65. Smit, A. F. and Riggs, A. D. (1996) *Tiggers* and other DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. USA* **93,** 1443–1448.
66. Vieira, C. and Biemont, C. (1997) Transposition rate of the *412* retrotransposable element is independent of copy number in natural populations of *Drosophila simulans*. *Mol. Biol. Evol.* **14,** 185–188.
67. Harada, K., Yukuhiro, K., and Mukai, T. (1990) Transposition rates of movable genetic elements in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **87,** 3248–3252.
68. Suh, D. S., Choi, E. H., Yamazaki, T., and Harada, K. (1995) Studies on the transposition rates of mobile genetic elements in a natural population of *Drosophila melanogaster*. *Mol. Biol. Evol.* **12,** 748–758.
69. McClintock, B. (1984) The significance of responses of the genome to challenge. *Science* **226,** 792–801.
70. Grandbastien, M.-A., Lucas, H., Morel, J.-B., Mhiri, C., Vernhettes, S., and Casacuberta, J. M. (1997) The expression of the tobacco *Tnt1* retrotransposon is linked to the plant defense responses. *Genetica* **100,** 241–252.
71. Grandbastien, M.-A. (1998) Activation of plant retrotransposons under stress conditions. *Trends Plant. Sci.* **3,** 181–187.
72. Mhiri, C., Morel, J.-B., Vernhettes, S., Casacuberta, J. M., Lucas, H., and Grandbastien, M.-A. (1997) The promoter of the tobacco *Tnt1* retrotransposon is induced by wounding and by abiotic stress. *Plant Mol. Biol.* **33,** 257–266.
73. Melayah, D., Bonnivard, E., Chalhoub, B., Audeon, C., and Grandbastien, M.-A. (2001) The mobility of the tobacco *Tnt1* retrotransposon correlates with its transcriptional activation by fungal factors. *Plant J.* **28,** 159–168.
74. Hall, B. G. (1998) Adaptive mutagenesis: a process that generates almost exclusively beneficial mutations. *Genetica* **102-103,** 109–125.
75. Hall, B. G. (1999) Mobile elements as activators of cryptic genes in *E. coli*. *Genetica* **107,** 181–187.
76. Giraud, T. and Capy, P. (1996) Somatic activity of the *mariner* transposable element in natural populations of *Drosophila simulans*. *Proc. R. Soc. Lond. B. Biol. Sci.* **263,** 1481–1486.
77. Viera, C. and Biémont, C. (1996) Geographical variation in insertion site number of retrotransposon *412* in *Drosophila simulans*. *J. Mol. Evol.* **42,** 443–451.
78. Chakrani, F., Capy, P., and David, J. R. (1993) Developmental temperature and somatic excision rate of *mariner* transposable element in three natural populations of *Drosophila simulans*. *Genet. Sel. Evol.* **25,** 121–132.
79. Hartl, D. L. Transposable element *mariner* in *Drosophila* species. In *Mobile DNA* (Berg, D. E. and Howe, M. M., eds.). American Society for Microbiology, Washington D.C., 1989, pp. 531–536.
80. Lampe, D. J., Grant, T. E., and Robertson, H. M. (1998) Factors affecting transposition of the *Himar1 mariner* transposon *in vitro*. *Genetics* **149,** 179–187.

81. Mackay, T. F. C. (1984) Jumping genes meet abdominal bristles: hybrid dysgenesis-induced quantitative variation in *Drosophila melanogaster*. *Genet. Res.* **44,** 231–237.
82. Currie, D. B., Mackay, T. F., and Partridge, L. (1998) Pervasive effects of *P* element mutagenesis on body size in *Drosophila melanogaster*. *Genet. Res.* **72,** 19–24.
83. Lai, C. and Mackay, T. F. (1993) Mapping and characterization of *P*-element-induced mutations at quantitative trait loci in *Drosophila melanogaster*. *Genet. Res.* **61,** 177–193.
84. Long, A. D., Lyman, R. F., Langley, C. H., and Mackay, T. F. (1998) Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149,** 999–1017.
85. Long, A. D., Lyman, R. F., Morgan, A. H., Langley, C. H., and Mackay, T. F. (2000) Both naturally occurring insertions of transposable elements and intermediate frequency polymorphisms at the *achaete-scute* complex are associated with variation in bristle number in *Drosophila melanogaster*. *Genetics* **154,** 1255–1269.
86. Lyman, R. F., Lawrence, F., Nuzhudin, S. V., and Mackay, T. F. C. (1996) Effects of single *P*-element insertions on bristle number and viability in *Drosophila melanogaster*. *Genetics* **143,** 277–292.
87. McDonald, J. F., Matyunina, L. V., Wilson, S., Jordan, I. K., Bowen, N. J., and Miller, W. J. (1997) LTR retrotransposons and the evolution of eukaryotic enhancers. *Genetica* **100,** 3–13.
88. Bryan, G. J. and Hartl, D. L. (1988) Maternally inherited transposons excision in *Drosophila simulans*. *Science* **240,** 215–217.
89. Bucheton, A. (1979) Non-Mendelian female sterility in *Drosophila melanogaster*: influence of aging and thermic treatments. III. Cumulative effects induced by these factors. *Genetics* **93,** 131–142.
90. Coen, E. S., Robbins, T. P., and Almeida, J. Consequences and mechanisms of transposition in *Antirrhinum majus*. In *Mobile DNA* (Berg, D. E. and Howe, M. M., eds.), American Society for Microbiology, Washington, DC, 1989, pp. 413–436.
91. Ho, Y. T., Weber, S. M., and Lim, J. K. (1993) Interacting *hobo* transposons in an inbred strain and interaction regulation in hybrids of *Drosophila melanogaster*. *Genetics* **134,** 895–908.
92. Kidwell, M. G. (1981) Hybrid dysgenesis in *Drosophila melanogaster*: the genetics of cytotype determination in a neutral strain. *Genetics* **98,** 275–290.
93. Spradling, A., Stern, D., Beaton, A., Rhem, E., Laverty, T., Mozden, N., et al. (1999) The Berkeley *Drosophila* Genome Project gene disruption project: Single *P*-element insertions mutating 25% of vital *Drosophila* genes. *Genetics* **153,** 135–177.
94. Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R., and Engels, W. R. (1991) Targeted gene replacement in *Drosophila* via *P* element-induced gap repair. *Science* **253,** 1110–1117.

95. Nassif, N., Penney, J., Pal, S., Engels, W., and Gloor, G. (1994) Efficient copying of nonhomologous sequences from ectopic sites via *P*-element-induced gap repair. *Mol. Cell. Biol.* **14,** 13–25.
96. Rong, Y. and Golic, K. (2000) Gene targeting by homologous recombination in *Drosophila*. *Science* **288,** 2013–2018.
97. Henikoff, S. and Matzke, M. (1997) Exploring and explaining epigenetic effects. *Trends Genet.* **13,** 293–295.
98. Yoder, J., Walsh, C., and Bestor, T. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13,** 335–340.
99. McDonald, J. F. (1998) Transposable elements, gene silencing and macroevolution. *Trends Ecol. Evol.* **13,** 94–95.
100. Napoli, C., Lemieux, C., and Jorgensen, R. (1990) Introduction of a chimeric chalcone synthase gene into *petunia* results in reversible co-suppression of homologous genes in *trans*. *Plant Cell* **2,** 279–289.
101. Kelly, W. and Fire, A. (1998) Chromatin silencing and the maintenance of a functional germline in *Caenorhabditis elegans*. *Development* **125,** 2451–2456.
102. Ketting, R., Haverkamp, T., van Luenen, H., and Plasterk, R. (1999) *Mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99,** 133–141.
103. Tabara, H., Sarkissian, M., Kelly, W., Fleenor, J., Grishok, A., Timmons, L., et al. (1999) The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99,** 123–132.
104. Kim, J. M., Vanguri, S., Boeke, J. D., Gabriel, A., and Voytas, D. F. (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.* **8,** 464–478.
105. Wayne, M., Hackett, J., Dilda, C., Nuzhdin, S., Pasyukova, E., and Mackay, T. (2001) Quantitative trait locus mapping of fitness-related traits in *Drosophila melanogaster*. *Genet. Res.* **77,** 107–116.
106. Bushman, F., ed. *Lateral DNA Transfer: Mechanisms and Consequences*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2002.
107. Capy, P., Anxolabéhère, D., and Langin, T. (1994) The strange phylogenies of transposable elements: are horizontal transfers the only explanation? *Trends Genet.* **10,** 7–12.
108. Cummings, M. P. (1994) Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. *Trends Ecol. Evol.* **9,** 141–145.
109. Daniels, S. B., Peterson, K. R., Strausbaugh, L. D., Kidwell, M. G., and Chovnick, A. (1990) Evidence for horizontal transmission of the *P* element between *Drosophila* species. *Genetics* **124,** 339–355.
110. Kidwell, M. G. (1992) Horizontal transfer of *P* elements and other short inverted repeat transposons. *Genetica* **83,** 275–286.
111. Robertson, H. M. and Lampe, D. J. (1995) Recent horizontal transfer of a *mariner* transposable element among and between *Diptera* and *Neuroptera*. *Mol. Biol. Evol.* **12,** 850–862.

112. Hagemann, S., Haring, E., and Pinsker, W. (1996) Repeated horizontal transfer of *P* transposons between *Scaptomyza pallida* and *Drosophila bifasciata. Genetica* **98,** 43–51.
113. Jordan, I. K., Matyunina, L. V., and McDonald, J. F. (1999) Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc. Natl. Acad. Sci. USA* **96,** 12,621-12,625.
114. Khillan, J. S., Overbeek, P. A., and Westphal, H. (1985) *Drosophila P*-element integration in the mouse. *Dev. Biol.* **109,** 247–250.
115. Miller, L. H., Sakai, R. K., Romans, P., Gwadz, R. W., Kantoff, P., and Coon, H. G. (1987) Stable integration and expression of a bacterial gene in the mosquito *Anopheles gambiae*. *Science* **237,** 779–781.
116. O'Brochta, D. A. and Handler, A. M. (1988) Mobility of *P*-elements in Drosophilids and non-Drosophilids. *Proc. Natl. Acad. Sci. USA* **85,** 6052–6056.
117. Atkinson, P. W., Pinkerton, A. C., and O'Brochta, D. A. (2001) Genetic transformation systems in insects. *Annu. Rev. Entomol.* **46,** 317–346.
118. Ivics, Z., Hackett, P. B., Plasterk, R. H., and Izsvak, Z. (1997) Molecular reconstruction of *Sleeping Beauty*, a Tc*1*-like transposon from fish, and its transposition in human cells. *Cell* **91,** 501–510.
119. Auge-Gouillou, C., Hamelin, M. H., Demattei, M. V., Periquet, M., and Bigot, Y. (2001) The wild-type conformation of the *Mos-1* inverted terminal repeats is suboptimal for transposition in bacteria. *Mol. Genet. Genomics* **265,** 51–57.
120. Sherman, A., Dawson, A., Mather, C., Gilhooley, H., Li, Y., Mitchell, R. F. D., et al. (1998) Transposition of the *Drosophila* element *mariner* into the chicken germ line. *Nat. Biotechnol.* **16,** 1050–1053.
121. Matzke, M., Mette, M., Aufsatz, W., Jakowitsch, J., and Matzke, A. (1999) Host defenses to parasitic sequences and the evolution of epigenetic control mechanisms. *Genetica* **107,** 271–287.
122. Sundararajan, P., Atkinson, P. W., and O'Brochta, D. A. (1999) Transposable element interactions in insects: cross mobilization of *hobo* and *Hermes*. *Insect Molec. Biol.* **8,** 359–368.
123. Zhu, Y., Dai, J., Fuerst, P. G., and Voytas, D. F. (2003) Controlling integration specificity of a yeast retrotransposon. *Proc. Natl. Acad. Sci. USA* **100,** 5891–5895.

2

Detection of Transposable Elements in *Drosophila* Salivary Gland Polytene Chromosomes by *In Situ* Hybridization

Christian Biémont, Laurence Monti-Dedieu, and Françoise Lemeunier

Summary

In situ hybridization is particularly appropriate for mapping specific DNA sequences on polytene chromosomes of *Drosophila* and other dipterans. This technique is based on the recognition and binding of one labeled sequence (the probe) to homologous sequences on chromosomes fixed on a microscope slide. The probes are labeled with biotin or other nonradioactive products, and the probe signal can be detected as a thin line on the chromosomes, following the shape of the classical Giemsa-stained chromosome bands, thus allowing the detection of TE insertions within the range of 50 to 200 kb. In our laboratory we work on many individuals from natural populations, and as a result we process high numbers of slides hybridized with various DNA probes of transposable elements every day. Therefore, the *in situ* hybridization technique we use is a simplification of earlier published protocols. This chapter presents our simplified standard *in situ* hybridization protocol for labeling polytene chromosomes of *Drosophila* with biotin and a fluorescence stain (FISH).

Key Words: Transposable elements; *in situ* hybridization; FISH; *Drosophila*; polytene chromosomes.

1. Introduction

In situ hybridization is a powerful technique for localizing specific DNA sequences on chromosomes. It has been used in many experiments since the 1970s, and it is particularly appropriate for mapping specific DNA sequences on polytene chromosomes of *Drosophila* and other dipterans. This technique is based on the recognition and binding of one labeled sequence (the probe) to homologous sequences on the chromosomes fixed on a microscope slide. Although radioactive probes were initially used, it is now more common to use probes labeled with biotin or other nonradioactive products. These advanced

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

labeling methods allow more precise localization of the probe on polytene chromosomes because the probe signal can be detected as a thin line on the chromosomes, following the shape of the classical Giemsa-stained chromosome bands, thus allowing the detection of TE insertions within the range of 50 to 200 kb. We process high numbers of slides hybridized with various DNA probes of transposable elements every day in our laboratory. Therefore, the *in situ* hybridization technique we use is a simplification of earlier published protocols ***(1–7)***. The present chapter presents our simplified standard *in situ* hybridization protocol for labeling polytene chromosomes of *Drosophila* with biotin ***(8–10)*** and a fluorescence stain (FISH) ***(11)***. *In situ* techniques used for mitotic chromosomes are discussed by P. Dimitri in the following chapter.

2. Materials

1. Giemsa solution (prepare immediately before use). Add 3 mL Giemsa (Merck), 3 mL phosphate buffer (buffer tablets, pH 6.8, GURR Merck) to 94 mL of water.
2. 10X PBS: 1.3*M* NaCl, 0.07*M* Na_2HPO_4, 0.03*M* NaH_2PO_4.
3. 20X SSC: 3.0*M* NaCl, 0.3*M* trisodium citrate $2H_2O$, adjusted to pH 7.0 with NaOH.
4. 1X SSC: 0.15*M* NaCl, 0.015*M* sodium citrate.
5. Triton X100 in 1X PBS: add 1 mL of Triton X100 to 1 L of 1X PBS. Stir until completely dissolved.
6. 50% dextran sulfate (w/v): dissolve 1 g of dextran sulfate in 1.3 mL of distilled water for at least 6 hrs. Complete melting is essential for high-quality *in situ* hybridization. Do not hesitate to work with fresh product. Store at 4°C.
7. 10X BSA stock (bovine serum albumin): 10% BSA in 10X PBS. Store at 4°C. For preparation of the 1X BSA solution prewarm the 10X stock solution at 37°C before dilution.
8. Extravidin-horseradish peroxidase conjugate (we usually use Sigma, cat. no. E 2886). Mix 4 µL conjugate with 996 µL of 1X BSA solution.
9. DAB solution: Just before use dissolve 5 mg DAB (diaminobenzidine tetramine—Life Technology, ref. 15 972-011) in 10 mL of 1X PBS. **Caution:** This compound is a carcinogen. Use gloves and carry out all manipulations in the hood. Just before treating the slides with the DAB solution add 3.33 µL of a 30% H_2O_2 stock. DAB solutions are light sensitive, thus keep it in dark bottles, and also treat the slides in the dark.
10. Sodium Tris buffer (STB 5): For 1 mL add 200 µL 20X SSC, 50 mg BSA (stored at 4°C), 1 µL Triton X 100 (stored at room temperature), to 800 µL UHQ water.
11. Sodium Tris buffer (STB 1): For 1 mL add 200 µL 20X SSC, 10 mg BSA, and 1 µL Triton X 100, to 800 µL UHQ water.
12. Extravidin-FITC (50 µg/mL final).
13. Anti-DIG-rhodamine (4 µg/mL final).
14. Phosphate buffer albumine (PBA): add 398 mL 4X SSC, 1.6 mL 30% BSA (stored at 4°C), and 400 µL Triton X 100, store at room temperature).

15. Wash buffer (pH 7.2–7.3): add 40 mL 20X SSC, 200 μL Triton X 100 to 160 mL of UHQ water.

3. Methods

We presently use biotinylated probes for *in situ* hybridizations without any troublesome effects on the quality of the chromosomes. This outcome may be due to our simplified method. In contrast to other *in situ* protocols we omit additional steps like acetylation ***(8)*** or RNAse treatments. In addition, third-instar larvae are dissected directly in 45% acetic acid, and the salivary glands are placed in a clean drop of this acid before being squashed (*see* **Note 1**).

3.1. Slides and Coverslip Treatment

The slides and the coverslips are washed but not siliconized: First place the slides in chromic acid, then rinse them in 95% ethanol, and finally in distilled water. Slides can, however, be rinsed in ethanol only and wiped clean with thin paper, but this method depends on the slides and must be checked carefully. Clean the coverslips with lens paper or use them as they are. They are not siliconized because we found that siliconized coverslips might cause breakage of the spread chromosomes when the coverslip is removed after freezing in liquid nitrogen.

3.2. The Squash

1. Place one pair of salivary glands in a drop of 45% acetic acid, cover with a coverslip, and tap gently with an eraser without holding the coverslip. This dissociates the cells and makes the chromosomes flow in the liquid. Check the quality of the squash visually—the cells should be well spread out. If the squash does not appear good enough, the coverslip should be tapped again gently. Squashing is finished by slightly scratching the whole area of the coverslip in a zigzag motion with a blunt needle (or a pencil, as preferred). This improves the quality of the chromosome spread.
2. Place the slide and coverslip on blotting paper and crush firmly under the thumb, or in the jaws of a vise. We now use a vise because this avoids damage to fingers from acetic acid, and it is easy for young or less strong students to perform. This stage is essential because it completely flattens the chromosomes (*see* **Note 2**).

3.3. Squash Dehydration

1. Immerse the squashed slides in liquid nitrogen for at least ten minutes. Afterward, flip off the coverslips with a razor blade. Removal of the coverslip must be quick; otherwise, parts of the chromosomes will stick to the coverslip.
2. Immediately dehydrate the slides in ethanol at room temperature. Two baths of 70% ethanol followed by two baths of 95% ethanol can be used, but immersion in one bath of 95% ethanol for 10–15 min is usually sufficient.

3. Air-dry the slides and place them in boxes for later hybridization. At this step the slides should be reexamined under the microscope, and only good squashes should be selected for further analyses. They are stable for months at room temperature, but the best *in situ* hybridization results are generally obtained with 2- to 3-d-old slides. We have had successful results even with 2-yr-old preparations but only with long probes, such as those made from long retrotransposon sequences.

3.4. Preparation of DNA Probes

We currently use probes (1 μg of DNA) labeled by nick translation ***(12)*** because it is a simple technique that does not require DNA denaturation or extraction of the insert from plasmids. We use the Bionick™ Labeling System kit from Life Technology based on biotin-14-dATP, which requires the mixing of only two vials. We have also worked with biotin-11-dUTP and biotin-16-dUTP, obtaining good results (*see* **Note 3**). Biotinylated DNA can be kept at 4°C or at –20°C for months. Random prime labeling techniques are appropriate for short DNA probes.

1. For homologous high-stringent probes prepare the hybridization mix in 50% formamide as following (*see* **Note 4**):

biotinylated DNA	10 μL
sterile 20X SSC	10 μL
50% dextran sulfate	10 μL
formamide	30 μL

   For heterologous probes (to improve the detection of signals from homologous sequences that have diverged, or when working with species that have diverged from the probe species) prepare the hybridization mix in 35% formamide. The *in situ* hybridization is then said to be under heterologous conditions.

labeled DNA	10 μL
sterile 20X SSC	8 μL
50% dextran sulfate	8 μL
formamide	14 μL

2. Heat the vial containing the hybridization mix in boiling water for 8–10 min (do not forget to make a small hole in the cap), and cool it quickly in a mixture of ice and ethanol. The vial can be store at 4°C for months (*see* **Note 5**).

3.5. In Situ *Hybridization*

1. Warm the slides in a bath of 2X SSC at 70°C for 30 min, and dehydrate them in 95% ethanol. They can be kept in boxes at 4°C for months after this treatment.
2. Soak slides in 0.07*N* NaOH for 2–3 min to denature the chromosomes, wash them in 95% ethanol, and air-dry. These slides are ready for hybridization (*see* **Note 6**).

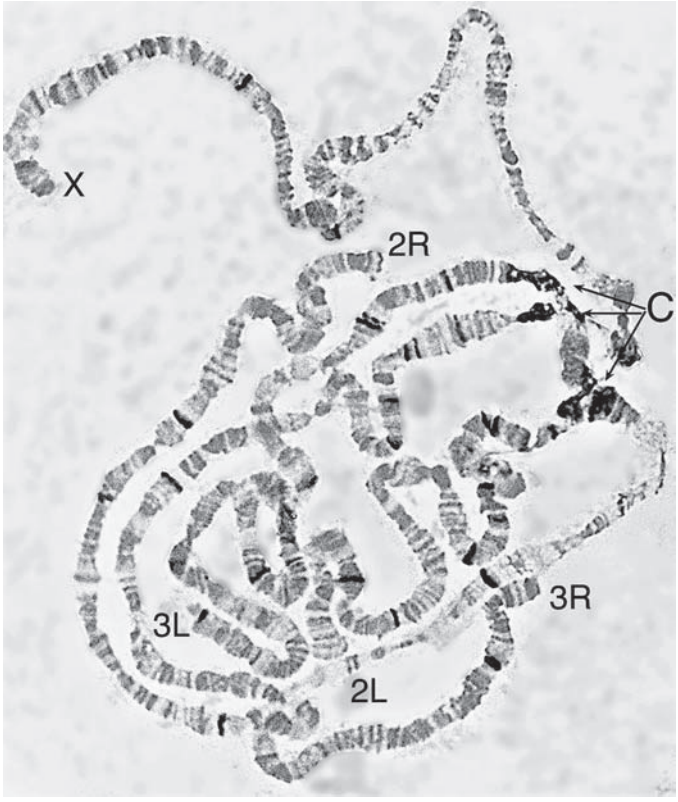


Fig. 1. *In situ* hybridization of *D. melanogaster* salivary gland chromosomes with a biotinylated DNA probe for a transposable element. The brown-labeled insertions are easily distinguished from the blue, Giemsa-stained bands of the chromosomes. The chromosome arms (*X, 2L, 2R, 3L, 3R*) are noted, as are the highly labeled chromocenters (*C*).

3. Add one drop per slide of biotinylated DNA probe (*see* **Subheading 3.4.**), cover with a cleaned coverslip, and place slides overnight in a humid chamber at 37°C.
4. The following morning, wash slides for 10 min in 2X SSC, followed by two quick washes for 3 s each in 0.1% Triton in 1X PBS, and then in 1X PBS (*see* **Note 7**). Add one drop per slide of the extravidin-horseradish peroxidase solution and cover with a cleaned coverslip. The reaction is allowed to proceed in a humid chamber at 37°C for 30 min.
5. Rinse slides in 0.1% Triton in 1X PBS for 3 s followed by a second wash for a few sec in 1X PBS. After these washes cover the slides with the DAB solution for 3–4 min and incubate them in the dark (*see* **Note 8**). Finally rinse the slides quickly in 1X PBS and stain them with freshly prepared Giemsa solution for 4–8 min.

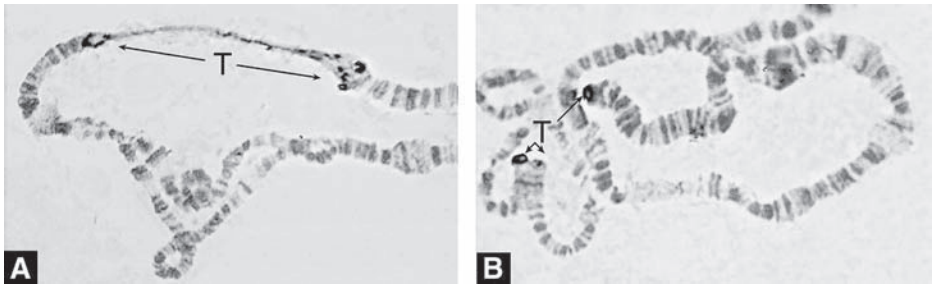


Fig. 2. *In situ* hybridization of salivary gland chromosomes of *D. melanogaster* with the telomeric *HeT-A* element. **(A)** *HeT-A* hybridizes across the length of the stretched sequences of thin DNA pulled out between the telomeres of the 2R and 2L chromosomes. **(B)** Hybridization follows the morphology of the extreme end of the chromosome, and the intensity of the labeling differs according to the chromatides, which are separated.

6. Mount the slides in EUKITT resin (Merck, ref. 82601) under a coverslip. The chromosome preparation must be completely dry before the resin is added. The slides can then be kept at room temperature for years without any significant alteration (*see* **Figs. 1** and **2**).

3.6. Fluorescence In Situ *Hybridization on Polytene Chromosomes With Two Probes*

The protocol presented below follows the procedure described by Muleris et al. ***(13)*** with some modifications. The hybridization mixture (70 µL) using two different probes simultaneously under homologous conditions contains:

deionized formamide	30 µL
sterile 20X SSC	10 µL
50% dextran sulfate	10 µL
probe 1: biotin-labeled DNA, about 500 ng	10 µL
probe 2: digoxigenin-labeled DNA, 500 ng	10 µL

The steps for preparation of the hybridization mixture and denaturation are identical to those described in **Subheadings 3.4.** and **3.5.**

1. Add 10 µL/slide of the hybridization mixture, cover with a cleaned coverslip, and place them overnight in a moist chamber at 37°C.
2. Remove the coverslips and incubate the slides in 2X SSC at 39°C for 2× 10 min.
3. Place 80 µL of STB5 solution on the marked squash, cover with a large coverslip, and incubate slides in a humid chamber for 30 min at 37°C.
4. Remove the coverslip, pour off the STB5 solution, and add 80 µL of STB1 detection solution. Add anti-DIG rhodamine and extravidin FITC just before use.

5. Incubate the slides for 30 min at 37°C in a humid chamber. Remove coverslips and wash the slides for 10 min at 37°C in the wash buffer.
6. Stain chromosomal DNA by adding 15 µL of DAPI (4'-6-diamidino-2-phenylindole)-Vectashield mounting solution (*see* **Note 9**).

4. Notes

1. Big salivary glands are obtained from well-fed larvae raised under uncrowded conditions. Adding a solution of fresh yeast on first-instar larvae improves the future quality of the polytene chromosomes.
2. We usually encircle good squashes by scratching the surface of the slide with a needle. It helps to limit the amount of liquid used.
3. There is no need to remove the TE probe from its plasmid for nick translation. A longer probe will always give rise to a stronger signal compared to a short one. Two labeled nucleotides can be used for nick translations of very short probes. If the DNA sequence of the probe is too rich in long stretches of the same nucleotide, try mixing the cold nucleotide with the labeled one at a 1 : 1 ratio. The purity of the probe DNA is essential.
4. We never use Denhardt's solution, but dextran is absolutely necessary, and the freshness of the dextran sulfate powder is important. Do not hesitate to use a fresh vial from time to time.
5. The nick translation kit, the extravidin, or the dextran sulfate should be checked if there is no hybridization signal or when the signal gets fainter and fainter with successive runs of hybridization. *In situ* hybridization must always be done with a previously tested DNA probe control.
6. It is often stated that *in situ* hybridization of *Drosophila* polytene chromosome squashes using biotin leads to deterioration of the chromosomes. Lakhotia et al. ***(14)*** even suggest treating the slides with gelatin to overcome this problem. We have never used subbed slides and our protocol does not cause chromosomes to deteriorate. The quality of the chromosomes after hybridizations is as it was when checked after squashing. Exact timing of the denaturation step is crucial. The treatment with Triton X is optional, although it helps to obtain clean preparations as the detergent removes unspecific hybridizations as well as dust.
7. The chromosomes and cytoplasm may come loose on the slide after hybridization. When the cytoplasm does not adhere well to the slides, try to be very gentle during the washes (SSC, PBS, Triton, etc.). Do not shake the slides, not even during the final Giemsa-staining step. Slow movements of the rack containing the slides are generally sufficient to insure efficient washing and homogeneous staining.
8. Because DAB is sensitive to light and humidity, the usually white powder can sometimes be yellow or even brown. Although this color change does not seem to be a vital problem, we prefer to use a new, fresh vial when the DAB color is too strong. The DAB solution can be aliquoted and kept frozen at –20°C, but we have had problems with this method because the solution was sometimes too colored. We now make up fresh DAB solution as required.

9. DAPI can be conserved as stock solution (100 μg/mL) at –20°C in the dark. Vectashield (Vector Laboratories, Burlingame, CA 94010) is a mounting medium added to DAPI to prevent fluorescence fading and also to favor a better conservation of the slides. Prepare DAPI–Vectashield (500 ng/mL) as follows: 2.5 μL DAPI (100 μg/mL) and 497.5 μL Vectashield. Store at –20°C. Use gloves for the manipulation of DAPI.

References

1. Pardue, M. L. and Gall, J. G. (1975) Nucleic acid hybridization to the DNA of cytological preparations. *Methods Cell Biol.* **10,** 1–16.
2. Bhadra, U. and Chatterjee, R. N. (1986) A method of polytene chromosome preparation of salivary gland of *Drosophila* for *in situ* transcription and *in situ* hybridization technique. *Droso. Infor. Ser.* **63,** 140.
3. Engels, W. R., Preston, C. R., Thompson, P., and Eggleston, W. B. (1986) *In situ* hybridization to *Drosophila* salivary chromosomes with biotinylated DNA probes and alkaline phosphatase. *Focus* **8,** 6–8.
4. Bruneval, P. (1988) Hybridation moléculaire *in situ*. *Rev. Fr. Histotechnol.* **1,** 17–22.
5. Blackman, R. K. (1996) Streamlined protocol for polytene chromosome *in situ* hybridization. *Biotechniques* **21,** 226–230.
6. Kim, W. and Kidwell, M. G. (1994) *In situ* hybridization to polytene chromosomes using digoxigenin-11-dUTP-labelled probes. *Droso. Infor. Ser.* **75,** 44–46.
7. Philips, A. M., Martin, J., and Bedo, D. G. *In situ* hybridization to polytene chromosomes of *Drosophila melanogaster* and other dipterans. In In Situ *Hybridization Protocols* (Choo, K. H. A., ed.), Humana, Totowa, NJ, 1994, pp. 193–209.
8. Langer-Safer, P. R., Levine, M., and Ward, D. C. (1982) Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proc. Natl. Acad. Sci. USA* **79,** 4381–4385.
9. Neumann, R., Rudloff, P., and Eggers, H. J. (1986) Biotinylated DNA probes: sensitivity and applications. *Naturwissenschaften* **73,** 553–555.
10. Pliley, M. D., Farmer, J. L., and Jeffery, D. E. (1986) *In situ* hybridization of biotinylated DNA probes to polytene salivary chromosomes of *Drosophila* species. *Droso. Info. Ser.* **63,** 147–149.
11. Wiegant, J., Ried, T., Nederlof, P. M., van der Ploeg, M., Tanke, H. J., and Raap, A.K. (1991*) In situ* hybridization with fluoresceinated DNA. *Nucleic Acids Res.* **19,** 3237–3241.
12. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* **113,** 237–251.
13. Muleris, M., Richard, F., Apiou, F., and Dutrillaux, B. Hybridation *in situ* en cytogénétique moléculaire—Principes et techniques. In *Techniques et Documentation*, Editions Médicales Internationales, Lavoisier, Paris, 1996, pp. 1–180.
14. Lakhotia, S. C., Sharma, A., Mutsuddi, M., and Tapadia, M. G. (1993) Gelatin as a blocking agent in Southern blot and chromosomal *in situ* hybridizations. *Trends Genet.* **9,** 261.

3

Fluorescent *In Situ* Hybridization With Transposable Element Probes to Mitotic Chromosomal Heterochromatin of *Drosophila*

Patrizio Dimitri

Summary

The technique of *in situ* hybridization of DNA probes to *Drosophila* chromosomes has been initially applied to the salivary gland polytene chromosomes and is now routinely used for mapping single-copy and repetitive DNA sequences, such as transposable elements, to the euchromatic regions of these chromosomes. However, most of the heterochromatin normally escapes cytogenetic analyses on polytene chromosomes because it is organized in a poorly differentiated cytological structure called the chromocenter. This peculiar organization does not allow a detailed mapping of DNA clones to heterochromatin. Such a limitation can be overcome by the fluorescent *in situ* hybridization (FISH) technique on mitotic chromosomes of *D. melanogaster*, where heterochromatin has been extensively characterized by banding techniques and subdivided into several cytologically diverse regions. Digital images of FISH signals and DAPI staining can be separately recorded by CCD camera, pseudocolored, and merged using specific software for image analysis. The visualization of the signals and DAPI banding pattern on a single chromosome enables the mapping of a given sequence to specific cytological regions of mitotic heterochromatin. This method has initially proven successful in the detection and mapping of transposable element clusters in the heterochromatin of *D. melanogaster* and has been used to study the distribution of repeated and even single-copy sequences.

Key Words: Fluorescent *in situ* hybridization (FISH); mitotic heterochromatin; cytological mapping; heterochromatic sequences; *Drosophila*.

1. Introduction

The technique of *in situ* hybridization of DNA probes to *Drosophila* chromosomes has been initially applied to the salivary gland polytene chromosomes ***(1)*** and is now routinely used for mapping single-copy and repetitive DNA sequences, such as transposable elements to the euchromatic regions of

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

chromosomes (*see* Chapter 2). The heterochromatin, a conspicuous component of the *Drosophila* genome ***(2,3)***, normally escapes cytogenetic analyses on polytene chromosomes because it is positioned in a poorly differentiated cytological structure called the chromocenter. The chromocenter contains two cytological domains ***(4)***: α-heterochromatin, which corresponds to a small compact region located in the middle of the chromocenter and undergoes little if any replication during polytenization ***(5)***, and β-heterochromatin, a diffusely banded mesh-like material that lies between proximal euchromatin and α-heterochromatin and that undergoes extensive DNA replication during polytenization ***(6,7)***. Because of this peculiar organization, *in situ* hybridization to salivary gland chromosomes can allow the localization of repetitive or unique polytenized heterochromatic sequences only to β-heterochromatic regions of chromosome arms without the possibility of performing any further detailed physical mapping. This limitation can be overcome by the *in situ* hybridization on mitotic chromosomes of *D. melanogaster*, where heterochromatin has been extensively characterized by banding techniques and subdivided into several cytologically diverse regions ***(2)***.

This chapter presents methods routinely used for mitotic chromosome preparations and fluorescent *in situ* hybridization (FISH) with transposable element probes to mitotic heterochromatin in *D. melanogaster*.

2. Materials

2.1. Preparation of Mitotic Chromosome Squashes

1. Siliconized glass slides (only as support for droplets of dissection solutions).
2. Siliconized glass coverslips (20 mm × 20 mm or 22 mm × 22 mm).
3. Nonsiliconized glass slides.
4. Petri dishes (35 mm × 10 mm).
5. Microscope for dissection.
6. Dissecting forceps (Dumont no. 5) or needles.
7. Bibulous paper.
8. Razor blade.
9. Saline: 0.7% NaCl in H_2O. Store at 4°C.
10. Hypotonic solution: 0.5% sodium citrate $2H_2O$ in H_2O. Store at 4°C.
11. Fixative: acetic acid, methanol, H_2O (5.5 mL:5.5 mL:1 mL).
12. Acetic acid (45%). Make fresh.
13. Absolute ethanol, chilled at –20°C.
14. Liquid nitrogen or a block of dry ice.

2.2. FISH

1. Coplin jars.
2. Glass coverslips (22 mm × 22 mm or 24 mm × 24 mm).
3. Moist chamber.

4. Dry, dust-free box for holding slides.
5. Rubber cement.
6. Formamide (J. T. Baker). Stored at 4°C.
7. Biotin-nick translation mix (Roche). Stored at –20°C.
8. DIG-nick translation mix (Roche). Stored at –20°C.
9. Rhodamine-nick translation mix (Roche). Stored at –20°C.
10. Fluorescein isothiocyanate (FITC)-conjugated avidin (DCS grade; Vector laboratories) for biotinylated probes. Store at 4°C.
11. Cy3-conjugated avidin (Roche) for biotinylated probes. Store at 4°C.
12. Rhodamine-conjugated anti-digoxigenin (DIG) sheep IgG, Fab fragments (Roche) for digoxigenin-labeled probes. Store at 4°C.
13. Sonicated salmon sperm DNA.
14. 3*M* Sodium acetate, pH 4.5.
15. Ethanol (70%, 90%, and absolute), at room temperature.
16. Ethanol (70%) chilled at –20°C.
17. Ethanol (90% and absolute), chilled at 4°C or on ice.
18. 20X SSC.
19. Tween-20.
20. 4,6-diamino-2-phenylindole-dihydrochloride (DAPI; 0.2 μg/mL), dissolved in 2X SSC. Store at 4°C.
21. Vectashield H-1000 (Vector laboratories). Store at 4°C.

3. Methods

A crucial step for fluorescent *in situ* hybridization to the heterochromatin of diploid cells is the preparation of mitotic chromosome squashes. The quality of chromosome morphology determines the ease of recognizing the heterochromatic landmarks obtained by fluorochromes such as DAPI or Hoecst-33258 that are general indicators of AT-rich regions. The following protocols describe the preparation of mitotic chromosome squashes from both brain and imaginal discs of *D. melanogaster* larvae.

3.1. Preparation of Mitotic Chromosome Squashes From Larval Brains

1. Grow larvae in moderately crowded vials. Select large larvae that are climbing up the sides of the tube (*see* **Note 1**). At this stage, determine the sex of the larvae if chromosomes of a particular sex are required for analysis.
2. Collect and wash the larvae in a 35 mm × 10 mm petri dish with 0.7% saline at room temperature. Transfer three drops of saline (50 μL each) onto a siliconized slide. Place one or two larvae in each drop.
3. Perform dissections in saline solution using sharp forceps (Dumont no. 5) or dissecting needles as follows: Grasp the mouth hooks with one forceps, then grasp the body of the larva midway down with the other forceps. Gently separate the mouth hooks from the rest of the larval body. The brain frequently remains attached to the head portion together with imaginal discs.

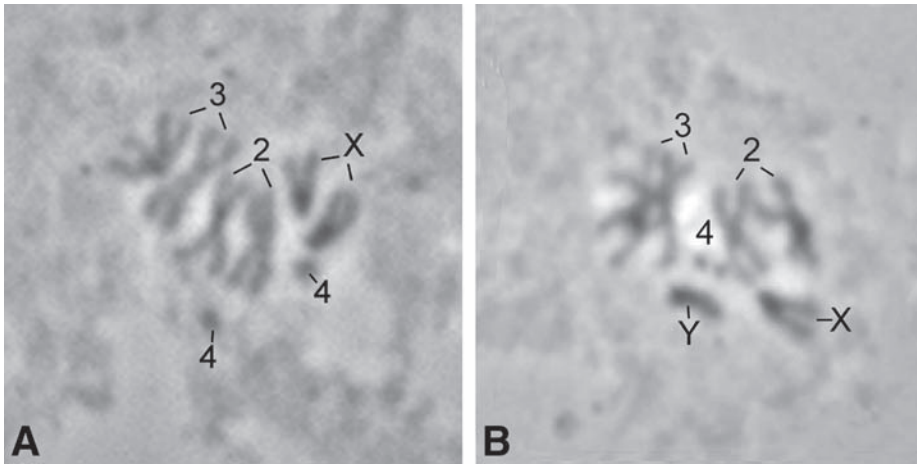


Fig. 1. Examples of unstained chromosomes from mitotic cells of *D. melanogaster* larvae: **(A)** mitotic chromosomes from larval brains. **(B)** Mitotic chromosomes from wing imaginal discs. After removing the coverslip and dehydration in absolute ethanol, slides were examined under phase contrast in an optical microscope.

4. Remove the brain from the mouth hooks, gently detach imaginal discs, and collect the brains in a fresh drop of saline.
5. Transfer the brains into a drop (50 μL) of hypotonic solution placed on a siliconized slide and incubate at room temperature for 10 min.
6. Move brains to a 35 mm × 10 mm petri dish containing a freshly prepared mixture of acetic acid/methanol/H_20 (5.5 mL:5.5 mL:1 mL) for approx 30 s.
7. Transfer a single brain into a small drop (2 μL) of 45% acetic acid placed on a dust-free siliconized coverslip (20 mm × 20 mm or 22 mm × 22 mm). One to four brains can be placed on the same coverslip. Leave the brains in the 45% acetic acid drops for 1–2 min.
8. Pick up coverslip carrying the brains using a dust-free nonsiliconized slide, so that the coverslip will adhere to the slide. Avoid the formation of air bubbles. Flip the slide over and gently press out excessive acetic acid between two sheets of blotting paper, then squash hard using the thumb. During squashing avoid lateral slippage of the coverslip.
9. Freeze slides either in liquid nitrogen or on dry ice for 5 min. By using a sharp razor blade, flip off coverslip with a quick motion and immediately plunge slides in cold (–20°C) absolute ethanol. Let them gradually reach room temperature (it usually takes about 30 min), remove from ethanol, and air-dry. Slides can be stored at 4°C for weeks in a dry, dust-free box.
10. Dried preparations can be checked without a coverslip under a phase contrast microscope. Chromosomes suitable for FISH experiments should appear flat and gray with no refractions (*see* **Fig. 1A**).

3.2. Preparation of Mitotic Chromosome Squashes From Imaginal Discs

1. Select larvae and perform dissections as described in **Subheading 3.1**. The imaginal discs frequently remain attached to the mouth hooks together with the brain.
2. Transfer the discs into a drop (50 μL) of hypotonic solution placed on a siliconized slide and incubate at room temperature for 7 min.
3. Transfer the discs individually into small drops (2 μL) of 45% acetic acid placed on a dust-free siliconized coverslip (20 mm × 20 mm or 22 mm × 22 mm). Leave the discs in the 45% acetic acid drops for 30 s.
4. Squash slides, remove coverslip and check preparations as described in **Subheading 3.1. Figure 1B** shows an example for well-spread imaginal wing disc chromosomes.

3.3. Fluorescent In Situ *Hybridization Mapping of TE Sequences on Mitotic Heterochromatin*

One major disadvantage of using tritiated probes is that the detection of H3-labeled hybridization signals on mitotic chromosomes is time consuming. In addition, either tritiated or biotinylated probes detected by nonfluorescent staining techniques do not allow simultaneous visualization of both the signals and the heterochromatin banding pattern. The fluorescence *in situ* hybridization technique coupled with DAPI staining and digital recording of images solves this problem ***(8)***. For example, digital images of FISH signals and DAPI staining can be separately recorded by charge-coupled device (CCD) camera, pseudocolored, and merged using specific software for image analysis, such as Photoshop®. The visualization of the hybridization signals and DAPI banding pattern on the same chromosome enables the mapping of a given sequence to specific cytological regions within mitotic heterochromatin.

This method can be applied to answer several kinds of questions. For example, it has initially proven successful in the detection and mapping of transposable element clusters located in the heterochromatin of *D. melanogaster* ***(9)***, and it can be used to study the distribution of repeated and even single-copy sequences along the mitotic heterochromatin of *Drosophila* chromosomes ***(7,10–12)*** (*see* **Fig. 2**). FISH mapping of single *P* element insertions along the mitotic heterochromatin (**Fig. 2 E–G**) may be important for genomic studies of *Drosophila* ***(13,14).*** These elements can be assigned to specific heterochromatic bands and can then represent important landmarks for physical mapping of heterochromatin. In addition, if the cytological location of a given *P* element insertion close to the heterochromatic gene of interest is known, insertional alleles or deletions of the gene can be generated by local hopping of the *P* element.

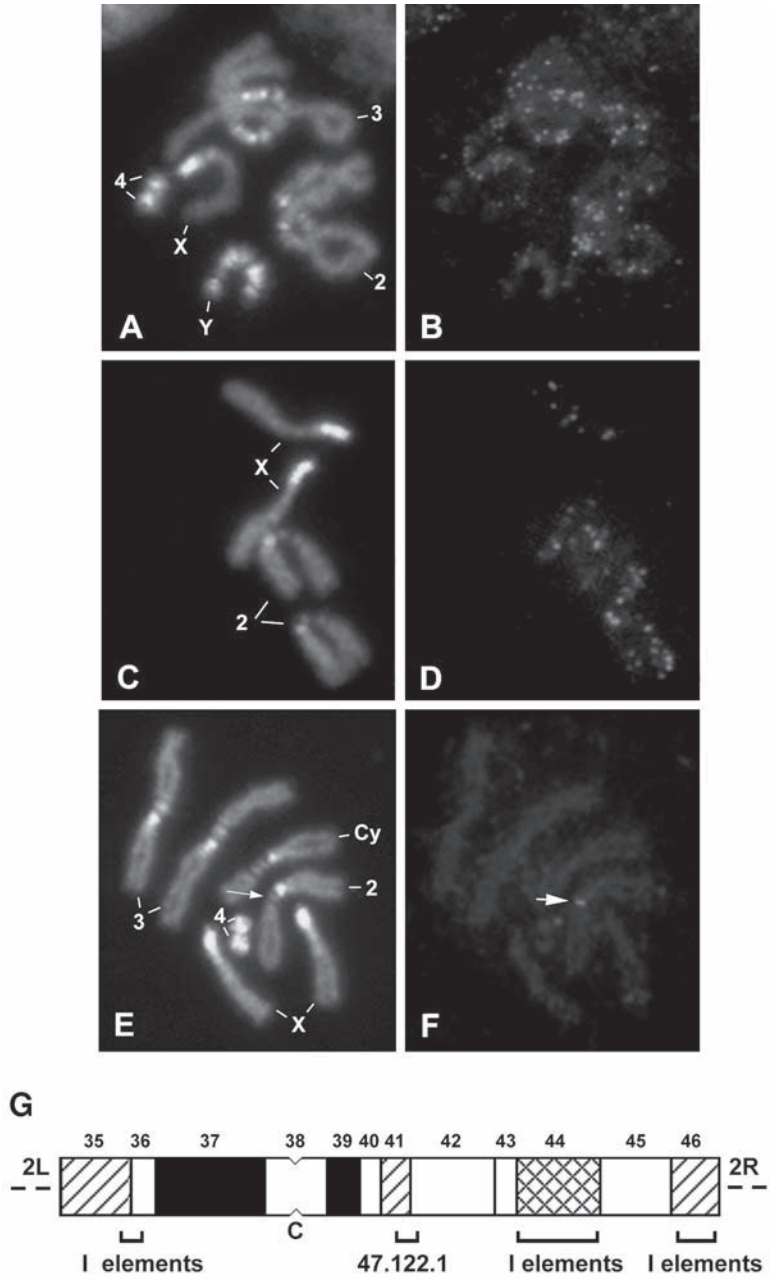

A
B
C
D
E
F
G
3
4
X
Y
2
Cy
2
3
4
X
2
X
35 36 37 38 39 40 41 42 43 44 45 46
2L
2R
C
I elements
47.122.1
I elements
I elements

Fig. 2.

3.3.1. Hybridization

1. Dehydrate 2- to 3-d-old slides (prepared according to **Subheading 3.1.**) by immersion in 70%, 90%, and 100% ethanol (3 min each). Air-dry slides after denaturation at room temperature.
2. Immerse one to three slides in 50 mL of prewarmed denaturation solution (35 mL ultrapure formamide, 5 mL 20X SSC and 10 mL distilled water). Incubate for 2 min in a water bath at 70°C.
3. Quickly transfer slides to 70% ethanol (–20°C), incubate for 3 min, and then dehydrate in ice-cooled 90% and 100% ethanol (3 min each time). Let slides air-dry at room temperature.
4. Label 1 μg of DNA probe (plasmids or PCR fragments) by nick translation using biotin-11-dUTP or digoxigenin-11-dUTP. For DNA labeling, we routinely use biotin-nick translation mix or digoxigenin-nick translation mix (Roche) (*see* **Note 2**).
5. Remove unincorporated nucleotides by ethanol precipitation (*see* **Note 3**) and store the probe at –20°C.
6. Precipitate the labeled DNA (40–80 ng per slide; *see* **Note 4**) by adding sonicated salmon sperm DNA (3 μg per slide), 0.1 volume of 3*M* sodium acetate, pH 4.5, and 2 volumes of cold absolute ethanol (–20°C). Place at –80°C for 15 min and spin at 13,000 rpm for 15 min. Dry the pellet in a Savant centrifuge (*see* **Note 5**).
7. Resuspend DNA in the hybridization mixture (10 μL per slide) by vortexing.
8. Heat the probe solution at 80°C for 8 min. Place tubes on ice for 5 min and centrifuge briefly to bring down any condensation. Keep on ice until used.

Fig. 2. *(previous page)* FISH mapping of *I* elements and a single *P* element insertion (line 47.122.1) to *Drosophila melanogaster* mitotic heterochromatin. **(A)** Oregon-R male prometaphase chromosomes stained with DAPI. **(B)** Hybridization signals detected by the biotinylated *I* element probe. **(C)** Canton-S female partial prometaphase stained with DAPI. **(D)** Hybridization signals detected by the rhodamin-labeled *I* element probe. **(E)** Female prometaphase from the line 47.122.1 stained with DAPI. The 47.122.1 insertion is caused by a single *P* element construct that contains the *miniwhite* eye-color gene, a white enhancer, an *scs* sequence, and a *Fab-7* fragment ***(15)***. **(F)** The hybridization signal corresponding to the 47.122.1 *P* insertion that maps to the distal part of region h41 (*see* arrow). **(G)** Cytological map of chromosome 2 heterochromatin showing the localization of *I* elements and the 47.122.1 *P* insertion. The heterochromatin of chromosome 2 has been subdivided by banding techniques into 13 regions, and numbered h35 to h46 ***(16)***. Filled areas represent the DAPI or Hoechst-33258-bright regions; shaded boxes represent regions of intermediate fluorescence, and open boxes are regions of dull fluorescence. The label 2L indicates the left arm of the chromosome, and 2R is the right arm. C is the centromeric region. Horizontal lines (below) indicate the location of *I* elements and single *P* transposon marked with *miniwhite* gene (47.122.1). **X**, **Y**, and numerals **2–4** indicate their respective chromosomes; Cy is the CyO balancer of chromosome 2.

9. Put 10 μL of probe solution to denatured slides and cover with 24 mm × 24 mm dust-free clean coverslip. Avoid trapping of air bubbles and seal the edges of the coverslip with rubber cement.
10. Put slides in a moist chamber and incubated overnight at 37°C (*see* **Note 6**).
11. Roll off the rubber cement and gently remove the coverslip. If the coverslip does stick to the slide, rinse it once in the washing solution prewarmed to the temperature used for hybridization, and try again (*see* **Note 7**).
12. Wash slides three times (5 min each) in the washing solution (50% formamide, 2X SSC) at 42°C.
13. Wash slides three times (5 min each) in 0.1X SSC at 60°C and remove excess liquid from the slide edges (*see* **Note 8**).
14. Apply 100 μL of blocking solution to each slide. Cover with 24 mm × 24 mm coverslip and incubate at 37°C for 30 min.

3.3.2. Detection of Biotin-Labeled DNA

1. Remove coverslip and blot excess blocking solution from the edges of the slide.
2. Drop onto each slide 50–100 μL of 3.3 μg/mL fluorescein isothiocyanate (FITC)-conjugated avidin (Vector) diluted in 4X SSC, 0.1% bovine serum albumin (BSA), 0.1% Tween-20; cover with a 24 mm × 24 mm coverslip and incubate for 30 min at 37°C in a dark moist chamber.
3. Remove coverslip and wash three times (5 min each) in 4X SSC, 0.1% Tween-20, at 42°C. Remove slides from the washing solutions and let them air-dry at room temperature.
4. Stain with 0.16 μg/mL 4,6-diamino-2-phenylindole-dihydrochloride (DAPI) dissolved in 2X SSC for 5 min at room temperature.
5. Rinse slides once in 2X SSC at room temperature, remove slides from 2X SSC and air dry.
6. Mount slides in 20m*M* Tris-HCl, pH 8, 90% glycerol, containing 2.3% of DABCO [1,4-diazo-bicyclo (2,2,2) octane; Merck] anti-fade (*see* **Note 9**).
7. Seal coverslips with rubber cement and store at 4°C. Slides can be stored for weeks.

3.3.3. Detection of Digoxigenin (DIG)-Labeled DNA

The procedure is identical to that for biotinylated probes described in **Subheading 3.3.2.**, with the exception of **step 2**, which is modified as follows:

2. Drop onto each slide 50–100 μL of 2 μg/mL rhodamine-conjugated anti-digoxigenin sheep IgG, Fab fragments (Roche), diluted in 4X SSC, 1% BSA, 0.1% Tween-20; cover with a 24 mm × 24 mm coverslip and incubate for 30 min at 37°C in a dark moist chamber.

3.3.4. Detection of Rhodamin-Labeled DNA (see ***Note 10***)

After the posthybridization washes (*see* **Subheading 3.3.1.**, **steps 12** and **13**), slides with probes directly labeled with tetramethylrhodamin-6dUTP or other fluorophores should be treated as follows:

1. Wash slides once for 3 min in 2X SSC, 0.1% Tween-20, at room temperature.
2. Stain slides with DAPI and mount as described in **Subheading 3.3.2.**, **steps 4–7**).

3.3.5. Double Labeling

1 For simultaneous *in situ* hybridizations mix the desired amount of biotin- and DIG-labeled probes.
2. Probe preparation: As described in **Subheading 3.3.1.**
3. Hybridization: As described in **Subheading 3.3.1.**
4. Signal detection: Prepare a mixture of 3 μg/mL FITC-conjugated avidin, 2 μg/mL rhodamine-conjugated anti-DIG sheep IgG, Fab fragments diluted in 4X SSC, 1% BSA, 0.1% Tween-20. Apply 80–100 μL per slide and cover with 22 mm × 22 mm or 24 mm × 24 coverslip and incubate at 37°C in the dark, humid chamber.
5. Wash slides, stain and mount preparation as described in **Subheading 3.3.2.**

4. Notes

1. Female larvae frequently have better chromosomes than male larvae.
2. FISH probes can be also labeled directly with fluorophores, usually by incorporation of specifically conjugated nucleotides. Fluorescein-labeled dNTPs (green emission) or Cy3-labeled dUTPs (red emission) are available from several suppliers. I routinely prepare TE probes labeled with tetramethylrhodamin-6 dUTP (red emission) using the rhodamin-nick translation mix from Roche.
3. Labeled DNA may be also recovered by centrifugation with the Microcon centrifugal filter device (Millipore) following the standard protocol of the producer. In the course of *in situ* hybridization experiments aimed to test whether or not a given TE sequence is present within the heterochromatin of mitotic chromosomes, it may be helpful to use a positive control for probe labeling. One option is to check the TE probe on polytene chromosome preparations. If the probe is labeled successfully, multiple euchromatic signals corresponding to the euchromatic copies of the element will be revealed.
4. Use 100–150 ng probe per slide for single *P* element insertions or other single-copy sequences.
5. Alternatively, transfer the desired amount of labeled DNA in an Eppendorf tube, add sonicated salmon sperm DNA (3 μg per slide), and dry in a Savant centrifuge.
6. Temperature used for middle repetitive probes is 37°C. For higher stringent conditions, hybridizations can be performed overnight at 42°C.
7. Keep the slides wet.
8. Lower stringent conditions for washes can be performed in 2X SSC or 4X SSC at 35°C.
9. Commercial anti-fade such as Vectashield H-1000 (Vector laboratories) may also be used.
10. The use of DNA probes directly labeled with tetramethylrhodamin-6 dUTPs or with other fluorescently conjugated nucleotides avoids the blocking and detection steps, and thus it is particularly useful because it reduces the background and shortens the procedure. For mapping of TE clusters, the intensities of hybridiza-

tion signals obtained with tetramethylrhodamin-6 dUTP labeled probes are comparable to those obtained by secondary detection systems. In contrast, signals corresponding to single-copy *P* element insertions of even 7 kb, such as *PZ* elements, are not easily detectable on mitotic heterochromatin with this primary detection method.

Acknowledgments

I wish to thank Nikolaj Junakovic for helpful comments and discussions and Martin Muller for kindly providing the 47.122.1 line.

References

1. Pardue, M. L. and Gall, J. G. (1969) Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc. Natl. Acad. Sci. USA* **64,** 600–604.
2. Gatti, M. and Pimpinelli, S. (1992) Functional elements in *Drosophila melanogaster* heterochromatin. *Annu. Rev. Genet.* **27,** 239–275.
3. Weiler, K. S. and Wakimoto, B. T. (1995) Heterochromatin and gene expression in Drosophila. *Annu. Rev. Genet.* **29,** 577–605.
4. Heitz, F. (1934) Uber alpha-Heterochromatin sowie Konstanz und Bau der Chromomeren bei *Drosophila*. *Biol. Zentralbl.* **45,** 588–609.
5. Gall, G. J., Cohen E. H., and Polan, M. L. (1971) Repetitive DNA sequences in *Drosophila. Chromosoma* **33,** 319–344.
6. Zhang, P. and Spradling, A. C. (1995) The *Drosophila* salivary gland chromocenter contains highly polytenized subdomains of mitotic heterochromatin. *Genetics* **139,** 659–670.
7. Berghella, L. and Dimitri, P. (1996) The heterochromatic *rolled* gene of *Drosophila melanogaster* is extensively polytenized and transcriptionally active in the salivary gland chromocenter. *Genetics* **144,** 117–125.
8. Gatti, M., Bonaccorsi S., and Pimpinelli, S. (1994) Looking at *Drosophila* mitotic chromosomes. *Methods Cell Biol.* **44,** 371–391.
9. Pimpinelli S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., et al. (1995) Transposable elements are stable components of *Drosophila melanogaster* heterochromatin. *Proc. Natl. Acad. Sci. USA* **92,** 3804–3808.
10. Baldrich, E., Dimitri, P., Desset, S., Leblanc, P., Codipietro, D., and Vaury, C. (1997) Genomic distribution of the retrovirus-like element *ZAM* in *Drosophila*. *Genetica* **100,** 131–140.
11. Losada, A., Abad, J. P., and Villasante, A. (1997) Organization of DNA sequences near the centromere of the *Drosophila melanogaster* Y chromosome. *Chromosoma* **106,** 503–512.
12. Makunin, I. V., Pokholkova, G. V., Kholodilov, N. G., Zakharkin, S. O., Bonaccorsi, S., Dimitri, P., et al. (1999) A novel simple satellite DNA colocalizes with the *Stalker* retrotransposon in *Drosophila melanogaster* heterochromatin. *Mol. Gen. Genet.* **261,** 381–387.
13. Yan, C. M., Dobie, K. W., Le, H. D., Konev, A. Y., and Karpen, G. H. (2002) Efficient recovery of centric heterochromatin *P* element insertions in *Drosophila melanogaster*. *Genetics* **161,** 217–229.

14. Corradini, N., Rossi, F., Vernì, F., and Dimitri, P. (2003) FISH analysis of *Drosophila* heterochromatin using BACs and *P*-elements. *Chromosoma* **112,** 26–37.
15. Hagstrom, K., Muller, M., and Schedl, P. (1996) Fab-7 functions as chromatin domain boundary to ensure proper segment specification by the *Drosophila* bithorax complex. *Genes Dev*. **10,** 3202–3215.
16. Pimpinelli, S. and Dimitri, P. (1989) Cytogenetic organization of the *Rsp* (*Responder*) locus in *Drosophila melanogaster*. *Genetics* **121,** 765–772.

4

Southern Blot Analysis of Individual *Drosophila* Flies

Nikolaj Junakovic

Summary

Detection of novel insertions and the ability to explore heterochromatin are two key goals in the study of mobile elements in *Drosophila*. The Southern blot analysis of individual flies can prove useful in both types of investigations, alone or in combination with genetic and cytological approaches. This chapter describes the protocol for carrying out such an analysis and provides some clues for the interpretation of the results.

Key Words: Individual *Drosophila* flies; Southern blot technique; de novo rearrangements; heterochromatin.

1. Introduction

One of the advantages of *Drosophila melanogaster* as model system for the study of transposable elements (TEs) is the comparatively low number of elements per family. This allows investigation of the dynamics of TEs at the genomic level, as opposed to focusing on a phenotypic marker or target sequence that may not be representative of instabilities occurring elsewhere in the genome. A particular TE genomic distribution can be viewed as a phenotypic trait that may provide valuable information on transposition, copy number, insertion sites, and the identity of the stock studied. Since the advent of the transposon display technique, comparable genome-wide surveys can also be carried out in genomes harboring thousands of elements per family. This will likely generate the synergy necessary for unraveling the fundamentals of the TE–host interactions, a subject probably broader than can be handled by any single model system ***(1)***.

This chapter should enable the reader to analyze individual *Drosophila* flies by the Southern blot technique. The presentation is in three parts, beginning with the pieces of equipment that can be fabricated in the laboratory or workshop if so desired. The overall cost should not exceed $100 US, provided

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

some basic equipment for molecular biology is also available. The solutions and buffers used are also listed in this section. The second part describes the protocol for the extraction of DNA from individual flies, the electrophoretic separation, transfer, and hybridization, and the reprobing of filters. Well-established procedures have been revised in an attempt to improve yield, reproducibility, resolution, and the comfort of the experimenter. Finally, clues for the interpretation of TE patterns and some of the questions that can be asked as a result of the technique described are briefly outlined.

2. Materials

1. The pestle: Pestles for 1.5-mL Eppendorf tubes are sold by several companies; however, we were unable to find commercial pestles and tubes that accurately match each other as estimated by immersing the pestle in a droplet of colored liquid (e.g., Bromophenol blue) and finding out how well it fits in the bottom of the tube. Hence, we use a pestle made to measure with dental resin. The help of a vise or clamp may be required to keep the metal rod (5 cm long, 3 mm diameter) in a straight position while the resin is solidifying.
2. The flexible shaft: This shaft transmits rotation from the motor and allows the pestle to be handled like a pencil while homogenizing; it can be purchased from Micro-Mark (www.micromark.com) (*see* **Fig. 1A**) or in a hobby shop. (Hardware shops mostly sell heavy-duty shafts, unsuitable for the purpose.)
3. The motor: This motor should provide a rotation of approx 1000 rpm and can be purchased in a hobby shop; alternatively, a kitchen mixer, set on the turbo option, works fine. A pressure on–off switch is more desirable than a toggle and is therefore recommended. The connection between the shaft and the motor should be accurately aligned to avoid lateral swinging; this and the measurement of the rotation speed may require technical equipment.
4. The homogenization stand: The stand consists of a slab (Plexiglas or wood), approx 20 × 30 cm in size, bearing in the central area an additional piece of Plexiglas, used to oppose the pressure of the pestle. The advantage of this piece is that during homogenization only a light grip of the tube is sufficient, just enough to keep it up and prevent it from rotating. The homogenization is thus easier to perform consistently, particularly when numerous samples are processed.
5. The gel box, combs, and electrodes: One way of increasing the efficiency of transfer is the use of agarose gels made as thin as possible. Gel thickness is problematic with horizontal gels where some of the thickness is contributed by the agarose between the comb and the gel support. Moreover, reduced thickness reduces the depth of the gel slots, in turn reducing the volume of the samples and increasing the chance of overflowing between adjacent slots. We use a home-made vertical gel box that accommodates gel slabs 3 mm thick, and 38 × 19 cm in size (*see* **Fig. 1B**). It is made of glass plates assembled with silicon glue; glass is preferred because agarose slightly sticks to it, thus preventing the upper part from collapsing and squashing the lower part, as experienced with a Plexiglas box. As spacers, rubber strips 3 mm thick and 2.5 cm wide are used. The comb is

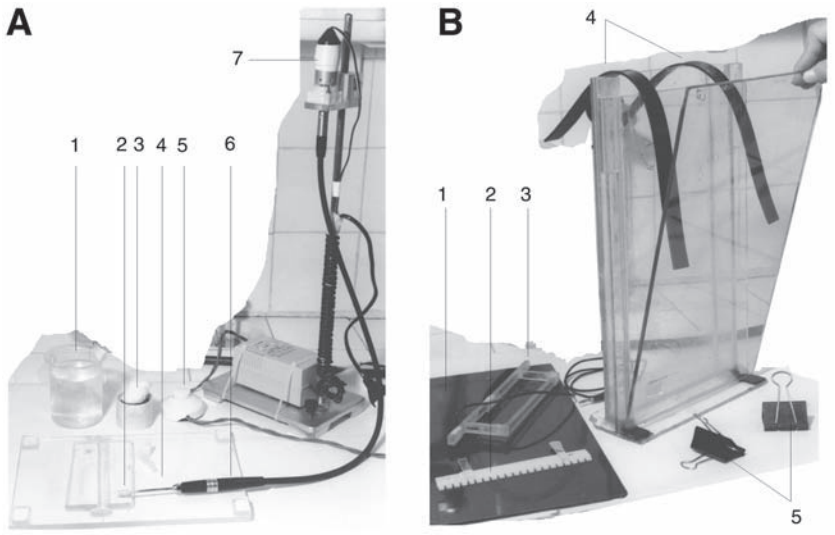


Fig. 1. Equipment for the single-fly Southern blot technique. **(A)** The equipment for the extraction of DNA from individual flies: (1) 250-mL beaker for rinsing the pestle between two homogenizations; (2) the pestle; (3) the paper towel for drying the pestle; (4) the homogenization stand; (5) the pressure on–off switch that commands the motor; (6) the shaft; (7) the motor. **(B)** The gel box is made of glass plates 5 mm thick; the back slab is 20 × 40 cm; the two sides are 3 × 40 cm; the inner front plate is 24 × 38 cm bearing in the upper part 2 × 2 cm protrusions on either side; the basis is 10 × 24 cm. These slabs are permanently assembled with silicone glue for aquariums. To prepare an agarose gel, the rubber spacers (4) are aligned longitudinally on the two sides, the external front plate (24 × 40 cm, tilted in the figure) is placed over the spacers, and the front plate and gel box secured with 6 clamps (5) per side. Note that the front plate rests on two additional spacers that define a 3-mm-wide slot, which is used for pouring approx 50 mL of 2% agarose that seals the lower part of the agarose chamber. Shown are the Plexiglas slab that supports the gel for the picture, black for UV illumination from above (1), the comb (2), and one electrode (3).

made of Plexiglas and requires a workbench for precise cutting; it should be slightly thinner than the spacers, just enough to allow a back-and-forth movement in the gel box; it is 16.5 cm long, accommodating 21 slots, 5 mm wide. This size is suitable for exposure of the filters to the 18-cm-wide X-ray films. In addition, approx 1 cm of unused gel is left on either side where the electric field tends to distort mobility. The electrodes consist simply of a platinum wire supported by some Plexiglas and connected to an electrical wire ending with a powerpack plug.

6. The transfer apparatus: Although in our experience capillary blotting is somewhat more efficient, we use the vacuum blotting transfer (VacuGene XL, Pharmacia) because the transfer can be completed in less than 1 h.

7. The hybridization apparatus: We find that the handling of filters is simpler and the chance of hybridization background reduced with a box placed in a shaking water bath, as compared to cylinders rotating in a rotisserie oven. The box should be large enough to allow some movement of the filters during hybridization.
8. The homogenization buffer: 100 m*M* Tris-HCl, pH 9.0, 100 m*M* EDTA, 1% sodium dodecyl sulfate (SDS).
9. The precipitation solution: 8 *M* potassium acetate.
10. The loading buffer: 20% Ficoll, 100 m*M* ethylenediamine tetraacetic acid (EDTA); the amount of the dye, orange G or bromophenol blue, is determined empirically.
11. The electrophoresis buffer: To make 10X stock solution, dissolve 215 g of Tris, 200 g of NaH_2PO_4, and 9.2 g of EDTA in 5 L of distilled water.
12. The transfer solution: 0.4*N* NaOH.
13. The agarose: High-strength agarose (Roche or BioRad) should be used to prevent the collapsing of the gel.
14. The Church buffer: Add 70 g SDS and 20 mL of 0.5*M* EDTA, pH 8.0 to 500 mL of 1*M* phosphate buffer, pH 7.2; adjust to 1000 mL with distilled water.

3. Methods

3.1. DNA Extraction

1. Place etherized individuals in 1.5-mL Eppendorf tubes containing 100 µL of homogenization buffer. Lift the flies a few millimeters up the tube wall and place tubes on ice (*see* **Note 1**).
2. Insert the pestle into the tube, switch on rotation, and lean the tube against the homogenizing stand. During the first 10 s of homogenization just press the pestle lightly, then start gently moving it up and down by approx 1 mm for an additional 15–20 s; remove most of the lysate still adhering to the rotating pestle by keeping it in contact with the tube wall while slowly withdrawing it. Immerse the pestle in a beaker containing distilled water, and eliminate the clinging droplet of water by lightly touching it to a piece of paper towel (*see* **Note 2**). Switch off rotation. Keep homogenates on ice until a set of 20 flies are completed.
3. Put at 65°C for 1 h; flick tubes individually after approx 30 min.
4. Leave at room temperature to cool for a few minutes, flick once to disperse the sediment, add 14 µL of 8*M* potassium acetate, and keep flicking until the liquid stops foaming.
5. Leave on ice for 30 min.
6. Spin at 10,000 rpm for 10 min in an Eppendorf centrifuge.
7. Transfer the supernatant to fresh tubes (*see* **Note 3**).
8. Add 100 µL of isopropanol, flick, and leave to precipitate 10 min at room temperature (*see* **Note 4**).
9. Spin at 10,000 rpm for 10 min.
10. Remove the supernatant. Add 200 µL of 70% ethanol at room temperature, and flick (*see* **Note 5**).

11. Spin at 13,000 rpm for 3 min.
12. Discard the supernatant and leave open on the bench for 15–30 min (*see* **Note 6**).
13. Add 27 μL of water, leave 10 min to hydrate, and flick. Repeat once (*see* **Note 7**).
14. Add 3 μL of a mix of 10X digestion buffer and restriction enzyme; incubate at appropriate temperature for a few hours to overnight (*see* **Note 8**).
15. Add 6 μL of loading buffer (*see* **Notes 9** and **10**).

3.2. Electrophoresis

1. Assemble the gel box as described in **Fig. 1B** and lean the upper part on a support about 5 cm high, so as to have the gel box inclined at an angle of approx 20°, front plate up.
2. Pour hot 2% agarose in electrophoresis buffer (*see* **Note 11**) into the lower opening of the box filling up to approx 7 cm (that is about 50 mL of agarose); let solidify.
3. Dissolve 250 mL of 0.8% agarose in electrophoresis buffer. Pour a few milliliters with a Pasteur pipet internally, along the spacers, to seal the box.
4. Cool the remaining 250 mL under tap water to approx 70°C; lean the gel box forward and fill with agarose.
5. Replace the gel box in an upright position, let the cooling agarose shrink and air bubbles surface for a few seconds and replenish.
6. As shrinking slows down, place the comb and check that no air bubbles are trapped; stand by the cooling gel for a few minutes ready to add more agarose as shrinking might proceed below the comb.
7. Place the gel box in a tray and position the electrodes, one inside the gel box, the other in the tray. Add electrophoresis buffer to the tray (just enough to ensure contact with lower end of agarose slab), fill in the gel box, and remove the comb. Load samples and marker (*see* **Note 12**). Run at 1.5 V/cm (80 V for the gel box described) for approx 16–17 h.
8. To dismantle the gel, place it horizontally, front plate downwards, on a support that allows removal of the clips from the sides. To facilitate the detachment from the gel box, cut away the excess agarose (above the gel slots) that might be adhering to the inner plate, using a scalpel. While holding down the front plate with the thumb, use four fingers to slowly lift the box. Gravity alone is sufficient to leave the gel on the front plate.
9. Remove any agarose debris that may be released from the slots (*see* **Note 13**).
10. Slide the gel onto a staining support, which may be either a UV-transparent or a black Plexiglas slab (20 × 40 cm) depending on whether a transilluminator or UV illumination from above is used for the exposure (*see* **Note 14**).
11. Stain in ethidium bromide for approx 30 min. Remove the solution with the aid of a suction pump (*see* **Note 15**).
12. Before the picture is taken, place a ruler on the gel for subsequent top–bottom and left–right reference of the area to be transferred (*see* **Note 16**).

3.3. The Transfer

1. Cut and discard the top agarose part bearing the slots (approx 4 cm). Ensure that the high-molecular-weight fragments plus about 1 cm stay in. Cut and discard the other end exceeding 24 cm (*see* **Note 17**).
2. Flood with water the porous support of the transfer apparatus and lay a piece of 3MM paper approx 25 × 30 cm to cover the central area; position the transferring mask.
3. Briefly soak a membrane 20 × 24 cm (Hybond N^+, Amersham) in transfer solution and place it over the transfer window (*see* **Note 18**).
4. Using a dedicated Plexiglas slab, lay the gel over the window in the transferring mask (*see* **Note 19**).
5. Pour the transfer solution over the gel, start transfer at 40 mbar and periodically add more NaOH solution. In approx 45 min to 1 h, 200 mL should go through (*see* **Note 20**).
6. Discard the agarose. Mark on the membrane the contour of the transferring window.
7. Trim the membrane exceeding the transfer area and rinse several times in 2X SSC to remove agarose debris and alkali.
8. Place the membrane, DNA side down, on a transilluminator (a piece of cling film between the two is a good idea). Mark the molecular weight bands (*see* **Note 21**).
9. Mark the filter and store wrapped in cling film at 4°C until used.

3.4. Hybridization and Reprobing of Filters

1. Lightly heat the Church solution until homogeneous and pour 100 mL in the hybridization box (*see* **Note 22**).
2. Add the filter and prehybridize for 15 min (*see* **Note 23**).
3. Remove approx 30 mL to a Falcon tube, add the probe, mix and pour back in the hybridization box (*see* **Note 24**). Hybridize overnight.
4. Discard the hybridization mix (*see* **Note 25**) and wash at the desired stringency, typically 2X SSC, 0.1% SDS at 65°C.
5. Blot the excess washing liquid from the membrane by placing it between two sheets of blotting paper, wrap it between two layers of cling film, and expose to X-ray film at –80°C (*see* **Note 26**).
6. To strip the probe, incubate the filter in 0.2 *N* NaOH at 45°C for 10 min. Wash several times with 2X SSC to remove alkali (*see* **Note 27**).

3.5. Interpreting the TE Patterns

In its simplest form, the pattern revealed with a probe homologous to a TE is formed by DNA fragments that share the TE sequence used as probe and that are heterogeneous in length, because the location of the external restriction sites differs among individual elements that are dispersed in the genome. Thus, a given pattern reflects a particular genomic distribution and is expected to change as elements acquire novel flanking sequences. In practice, however,

there are numerous hypothetical or real flaws that can contribute to a pattern and variations of it in the absence of transposition, or that may hinder the detection of transposition. In addition to contamination (*see* below), these include partial digestion, chromosomal polytenization in somatic tissues and somatic transposition, recombination, structural rearrangements of TEs, and chance comigration of restriction fragments.

3.5.1. Partial Digestion

Partial digestion can be handled along with the control for contamination by reprobing the filter with diverse TEs: if at least one gives rise to homogeneous patterns, it acts also as a control for complete digestion of polymorphic patterns (*see* **Fig. 2**). The filter can also be tested with a probe homologous to a sequence repeated in tandem (e.g., *rDNA, Bari 1*) that on partial digestion gives rise to predictable dimers, trimers etc.

3.5.2. Somatic Tissues

No significant difference has been detected between TE patterns revealed in DNA samples extracted from sperm, embryos, first- and second-instar larvae, brains, and adult flies ***(2)***. Comparisons between additional DNA samples extracted from brains (diploid tissues) and from whole flies (males and females) from a variety of stocks, tested with numerous TE probes, produced the same results (Junakovic, N., unpublished). The only consistent difference is that in DNA extracted from dissected tissues, high-molecular-weight bands tend to be more intense compared to samples from whole flies, most probably because in the latter some mechanical or enzymatic degradation occurs during the extraction. Thus, polytene chromosomes from somatic tissues make no detectable contribution to a pattern of TEs. Note that this is not informative with regard to hypothetical somatic events of instability or amplification that may stay undetected, as they are quantitatively too low to emerge in the average pattern from whole flies.

3.5.3. Recombination

Obviously, recombination occurring anywhere outside the fragments detected by the probe does not affect the number or size of the restriction fragments contributing to a TE pattern. Nor is recombination that is internal to these fragments, at sites that are homozygous for the elements and flanking sequences, expected to affect the TE pattern. The only relevant recombination, probably rare in most lab stocks, is the one occurring at sites that are heterozygous for both the presence of an element and the location of the restriction site in the flanking region. Reprobing can be informative in this case, too, because if homogeneous and heterogeneous patterns are detected on the same filter,

one would have to hypothesize that recombination has affected selectively the elements of the one TE family that is polymorphic (**Fig. 2**).

3.5.4. Structural Rearrangements

Some indirect criteria for distinguishing between instability observed by the Southern blotting technique and structural rearrangements of TEs ***(5)*** can be worked out by comparing the available evidence on the rate of rearrangements and transposition. It has been reported by Petrov and Hartl ***(3)*** that in *D. virilis* deletions of *Helena* elements far outnumber insertions, that the average size of deleted regions is 25 bp, and that the estimated rate of deletions is compatible with the loss of half of an element in 14.3 Myr. As the deletion size of *Helena* elements is comparable in distantly related species, the same value can reasonably be extended to other elements within *D. melanogaster*. Thus, the average, novel, deletion is undetectable by the Southern technique and the deletion rate appears too slow to account for *de novo* bands that may appear over one generation or for the persistence over the years of high polymorphism in small laboratory populations ***(4,25)***.

3.5.5. Chance Comigration of Restriction Fragments

Fragments bearing elements from different genomic locations that happen to exhibit the same mobility may prevent detection of *de novo* insertions by the Southern blot technique. A good electrophoretic separation may help, but as a rule, the lower the number of elements, the higher the chance of detecting a novel insertion; for high-copy families (e.g., *B104*) the transposon display technique may be more appropriate. One way of increasing the chance of detecting new insertions consists of setting up several crosses, and after the females have laid eggs, analyzing the parents and choosing for further analysis the siblings of the one pair that exhibits the lowest number of bands. This "choice" step can be of a more general nature for the selection of the trait of interest, which may be polymorphic in the stock studied: Sublines can be obtained that share a genetic background and that do or do not exhibit a specific band, amplification, or target gene, with and without an insert.

3.6. Why Analyze Individual Flies by the Southern Blot Technique?

The analysis of individual flies by the Southern blot technique may be useful for a broad range of issues, such as the estimate of transposition rate, hybrid dysgenesis, epigenetics, evolution of heterochromatin, telomeric activity, speciation, and aging. These fields share a documented or hypothetical involvement of TEs as well as the methodological requirement to distinguish *de novo* events of instability from drift of pre-existent polymorphism. The technique

can be used in combination with phenotypic variation and *in situ* hybridization technique, or in its own right for the study of the instability occurring in heterochromatin and telomeric regions where genetic markers are scarce or absent, and where chromosomal morphology is not suitable for the cytological analysis of dynamic features.

3.6.1. Transposition

Phenotypic variation correlated with instability of TEs is the most direct approach for the study of transposition. It is, however, limited to the few families that can be mobilized by dysgenic crosses and to a comparatively low number of loci. A broader view can be obtained by *in situ* hybridization to polytene chromosomes, which allows correlation of a particular genomic distribution of the elements of most transposon families, and variations of it, with known chromosomal regions (*see* Chapter 2). A problem with this technique is its coarse resolution, distinguishing only between elements located in different bands. As a result, transposition of those elements that tend to move over short chromosomal distances, within a chromosomal band, stay undetected, leading to an underestimate of the instability ***(6,7)***. In addition, novel insertions are sought via *in situ* technique under the assumption that hybridization signals along the chromosomal arms are primarily due to single elements, i.e., that most chromosomal bands where insertion occurs are devoid of other elements of the same family. This is important because the detection systems based on biotin and digoxigenin are only marginally quantitative: If two or more elements target the same chromosomal band, only the insertion of the first (and excision of the last) would be detected; the intermediate events would be concealed by resident elements.

For a long time the above assumption has been a reasonable one because the number of hybridization sites on polytene chromosomes and the number of genomic elements assessed by the Cot reassociation technique had been initially reported to be comparable ***(8)***. Subsequently, however, TEs have been found to be overabundant in heterochromatin ***(9)***, in turn raising novel questions about the number of genomic elements, the relative proportion within families of elements in euchromatin and heterochromatin, and the number of elements per chromosomal band. The multiple, preferential insertions of *gypsy* elements at the *ovo* locus ***(10)*** and of *P* elements at divisions 35C and 40A ***(11)*** show that some "crowding" at chromosomal bands may take place. This effect is likely due to more accessible genomic regions that are preferential targets of insertion. If so, it is possible that, given a collection of insertion regions that differ in accessibility, only the low- to intermediate-frequency insertions would be detected by *in situ* hybridization, whereas insertions in the most accessible

regions would tend to be concealed by resident elements. In this context, the Southern blot technique might help as it is sensitive to changes in location of a few thousand base pairs, depending on the size of the restriction fragments and the resolution of the electrophoretic separation.

The two approaches have been directly compared by other authors who concluded that "the Southern blotting technique had serious deficiencies as…it revealed less than 30% of the new insertions detected by *in situ* hybridization" ***(26)***. The actual comparison being between *in situ* hybridization on individuals and Southern pictures of pools of flies. This conclusion documents the need for single fly analysis because rare events, likely unique in a batch of 20–25, are obviously undetectable in the average pattern contributed by the majority of the individuals.

As for the requirement that new insertions must be distinguished from drift of pre-existing polymorphism and contamination, let's suppose a laboratory stock has been found highly polymorphic in the distribution of the elements of a transposon family. Formally, three interpretations are possible: (1) elements are currently unstable, (2) in the recent past, elements underwent a burst of transposition that has generated the heterogeneity, and (3) in the recent past, the stock underwent contamination by alien flies. By the Southern blot technique one can set up individual crosses, analyze the parental patterns after the females have laid eggs, choose the pair exhibiting the lowest number of bands, and ask whether in their siblings *de novo* bands are detectable. To check for contamination, one can take advantage of the finding that in most *Drosophila* stocks only a subset of TE families, if any, is found unstable (for an overview *see* **ref. *4***). The same filter can then be reprobed with elements of other families in search of the one that gives rise to homogenous individual patterns. If found, this represents a convincing internal control—on the same individuals—because alien flies are expected to bring in different patterns for all TEs tested (*see* **Fig. 2**). Such a direct approach is not feasible by *in situ* hybridization because polytene chromosomes are taken from larvae that are not yet reproductively mature. One would then have to homogenize the population by repeated sister–sib mating or balancer chromosomes, thus erasing the polymorphism that is the object of the initial interest. In addition, there is evidence for polymorphism in the genomic distribution of TEs persisting (or arising) in stocks that have been subjected to both procedures of homogenization ***(12–14)***. As for contamination, testing pieces of individual salivary glands by *in situ* hybridization with more than one probe, although possible, is a cumbersome procedure. In addition, finding individuals with "new" polymorphic distributions and other individuals of the same population with "old" stable patterns may still reflect a mix of contaminant and contaminated individuals.

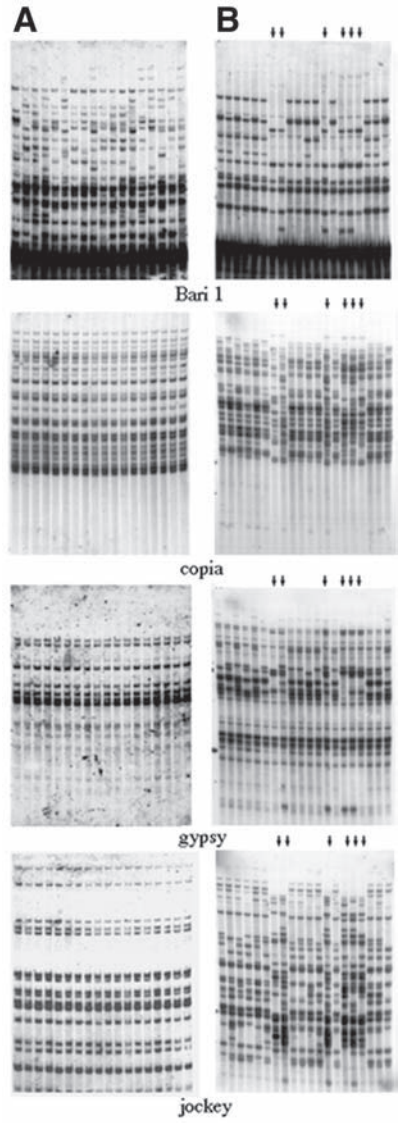


Fig. 2. Reprobing of the filters for the characterization of intrastock polymorphism. **(A)** High intrastock polymorphism is detected on probing with *Bari 1* individuals of the stock A; individuals appear homogeneous on testing the same filter with probes homologous to *copia*, *gypsy*, and *jockey*. Thus, the heterogeneity of *Bari* 1 elements is unlikely to be due to contamination, partial digestion, restriction site polymorphism, recombination, and structural rearrangement of elements (*see* text). **(B)** The stock B has been founded from known parents and tested periodically with *Bari* 1 probe. At generation 23, novel patterns appeared in six individuals (*arrows*); on reprobing, other transposons give rise to novel patterns as well, consistent with contamination by alien flies.

3.6.2. Hybrid Dysgenesis and Induction of Transposition

Relevant questions related to dysgenic systems such as the number of families mobilized and the instability of *P*, *I*, and *hobo* elements in the absence of dysgenic crosses are still poorly understood, primarily as a result of the lack of an experimental design able to distinguish between drift of pre-existing polymorphism and *de novo* instability. Spontaneous instability could be investigated in *D. melanogaster* and *D. virilis* ***(15)*** by the approach described above. The Southern blot patterns of the actual partner flies used for dysgenic (and reciprocal) crosses could be compared to the respective siblings; the analysis could be rapidly extended to numerous transposon families on reprobing of the filters. In addition to the dysgenic systems, there is evidence in *Drosophila* for the induction of mobility by environmental parameters such as heat-shock ***(16)***. Extending further these observations might benefit from the analysis of treated and untreated flies and respective siblings with probes homologous to diverse transposon families.

3.6.3. Evolution of Heterochromatin, Speciation, and Telomeric Activity

A special case of interest is represented by the TEs located in heterochromatin because they are particularly abundant, their dynamics may be difficult to analyze in detail by *in situ* hybridization, and because genes reporting on instability through phenotypic variation are rare in this region. An initial report claiming that DNA fragments originating from heterochromatin are refractory to transfer turned out to be a technical artifact ***(17,18)***. Elements that are located in underreplicated regions can be identified by comparing patterns from whole flies with the patterns obtained from pools of larval brains and salivary glands; elements located on the Y chromosome are easily identified by comparing the male and female patterns.

Speculations about the involvement of TEs in speciation date back to the early 1980s, but experimental evidence consistent with this hypothesis has been reported only recently in *Macropus eugenii* (wallaby) and *Drosophila* ***(19,20)***. The wallaby case shows that a significant change in hybrid species may affect TEs in heterochromatin. Also significant differences have been observed between *D. melanogaster* and *D. simulans* in euchromatic vs heterochromatic location of TEs and TEs on the Y chromosome ***(21)***. To gain information on the dynamics of these changes, the individuals of the two species can be crossed first, analyzed by the Southern blot technique, and then compared to the hybrid sibling flies, both males and females, in search of *de novo* events that may be undetectable as phenotypic variation or by *in situ* hybridization.

Difficulties are encountered in the study of telomeric regions comparable to those of heterochromatin studies. An example of how individual flies analyzed

by the Southern blot technique can provide information on telomeric activity (erosion, transposition, recombination, conversion) is illustrated in **ref**. ***22***.

3.6.4. Epigenetic Variation

Epigenetic variability can be distinguished from TE-induced variability by testing the integrity of the corresponding gene or by making the appropriate genetic crosses. If flies are sterile, the analysis of mutant individuals by Southern blot may be the only way.

3.6.5. Aging

Two reports provide an excellent overview in *Drosophila* ***(23)*** and other model systems ***(24)*** of the experimental evidence and theoretical speculations suggesting a role of TEs in aging. With lifespan as the only phenotypic trait, the Southern blot technique appears as a convenient experimental approach for the analysis of somatic events in individuals of different age. As previously pointed out, however, even if extensive instability of TEs were occurring in old flies, only events taking place in a significant fraction of somatic tissues and affecting recurrent genomic sites could be detected. Individuals should be analyzed because occasional heterogeneity that might still show up in individual patterns could be diluted out in DNA extracted from batches of flies. Finally, the comparison between young and aged flies requires that an age-related difference, if any, be distinguished from a difference due to random sampling. The analysis of individuals can help by studying sibling flies of known parents, which could be used as a third comparison reference.

4. Notes

1. Lifting the flies above the liquid ensures that the side of the pestle is used during homogenization. This is more efficient and reproducible than having the flies at the bottom of the tube.
2. Washing the pestle as described is efficient enough to remove the traces of the previous lysate, as estimated by PCR (Nina Schubert, University of Georgia, personal communication). The towel should be folded several times so as to prevent the rotating pestle from getting wrapped in loose, wet paper.
3. Use a fresh yellow tip for each sample.
4. This is a convenient stopping point. Samples will keep up to two years in our experience.
5. The pellet, albeit small, should be visible under good illumination.
6. Extra care should be taken at this stage to avoid discarding the pellet with the supernatant; the use of a Pasteur pipet with flame-thinned tip (as opposed to yellow tips) has the advantage that, if the pellet is sucked up as well, it might still be seen and replaced into the tube. Quick drying under vacuum may result in difficult resuspension.

7. There is no visual reference to follow resuspension. The ritual suggested here alternates time for hydration with mechanical stirring as experienced with visible pellets.
8. The presence of RNA does not interfere with the restriction digestion, at least with the most common enzymes such as *Eco*RI, *Hin*dIII, *Pst*I, *Bam*HI, *Xho*I, and *Acc*I.
9. Ficoll is osmotically neutral as opposed to glycerol or sucrose, which tend to give rise to "smiling" bands.
10. The expected yield is approx 600 ng per single *D. melanogaster* female and 60% of that per single male. Processing of 40 samples up to the isopropanol precipitation step takes approx 3 h.
11. We use phosphate buffer. Borate and acetate buffers should work equally well provided some adjustment for separation time is made.
12. Have good illumination while loading to be able to see "the skins" (films of agarose formed between the comb and the glass plate) that might bend into the slot and prevent the sample from reaching the bottom of the slot. One way of coping is to load the samples on either side of the slot rather than in the central part; alternatively, after removing the comb, clean the slots by aspiration with the aid of a Pasteur pipet. To make different runs comparable, the migration of the dye is more reliable than electrophoresis time. Typically, a good separation is achieved when Bromophenol Blue has migrated 25–26 cm or Orange G 30–32 cm in phosphate buffer. Recirculation, although not necessary, is useful to keep the pH constant and in case of leakage of the gel box.
13. Agarose debris may be difficult to remove if subsequently pieces slip below the gel.
14. To move the gel from one support to another, incline the donor over the recipient slab, push the bottom side of the gel with the aid of a "pusher" piece of Plexiglas until the top side steps over the recipient slab, and then slowly pull back the donor slab while holding the gel in place with the pusher device.
15. To reduce the volume of Ethidium Bromide solution it is a good idea to have a tray made to measure (45 × 21 × 5 cm).
16. This picture provides information on digestion, amount of DNA per lane, and quality of migration. The same information can be gained by visual inspection using a hand-held UV lamp. In this case the area to be transferred can be defined by placing plastic or paper stripes on the gel.
17. After trimming, the gel size should be slightly larger than the window in the transferring mask (17 × 22 cm). The contour of the window should be marked to make it visible through the membrane and the gel.
18. In our experience presoaking the filter in 0.4*N* NaOH is equivalent to presoaking the gel, and it results in a substantial increase of the hybridization signal, presumably because in the absence of this step, the very first DNA fragments getting in touch with the membrane are still in electrophoresis buffer. Placing the membrane on top of the window allows correction of a possible misalignment between the gel and the transfer window by sliding the membrane along with the gel with the aid of tweezers.

19. The Plexiglas slab, approx 25 × 27 × 0.4 cm should be transparent to allow "aiming" at the transferring window; to facilitate the sliding, the transferring side should be cut at an angle.
20. There are two reasons for adding the transfer solution on top of the gel as opposed to flooding the whole transfer chamber (as in the instruction manual): the first is that less solution is used; the second is to prevent occasional floating of the gel and passage of the liquid between the gel and the membrane. The transfer is more rapid initially and then slows down as gel starts collapsing under vacuum. Time course experiments showed that approximately half of the transfer occurs over the first 10 min.
21. The molecular-weight bands are clearly visible in fluorescence (not so if the gel has been presoaked in NaOH) and can be marked with a pencil or a laundry marker. An accurate mobility reference can thus be obtained by superposing the final autoradiograph and the filter.
22. In our experience, the chance of background is lower with the Church buffer compared with Denhard's solution.
23. Eight filters, possibly more, can be hybridized together in a volume of 200 mL.
24. This step dilutes the probe, which, if added as such, tends to produce a heavy spot of background.
25. The probe can be reused upon redenaturing by heat.
26. With a probe labeled by nick translation (e.g., 40% incorporation of 50 μC ^{32}P dCTP) the expected exposure time at –80°C for a TE pattern or a single-copy gene is overnight to two days.
27. Following this protocol 98% of the signal is washed off as estimated by instantimager quantitation before and after stripping of a prominent band (to be 80-fold more intense than a single-copy sequence). Up to 20 rounds of hybridization can be carried out with some filter-to-filter difference in the rate of DNA loss.

Acknowledgments

I am grateful to C. Biémont for helpful comments and to Livio Lascari and Tullio Riosa, from the workshop of the Department of Genetics and Molecular Biology, who helped with their skill and suggestions in the series of trials and errors that led to the protocol described in this article. The work in my lab is supported by the Consiglio Nazionale delle Ricerche and by the Fondazione Cenci-Bolognetti.

References

1. Kidwell, M. G. and Evgen'ev, M. B. (1999) How valuable are model organisms for transposable elements studies? *Genetica* **107,** 103–111.
2. Di Franco, C., Pisano, C., Dimitri, P., Gigliotti, S., and Junakovic, N. (1989) Genomic distribution of *copia*-like transposable elements in somatic tissues and during development of *D. melanogaster*. *Chromosoma* **98,** 402–410.
3. Petrov, D. A. and Hartl, D. L. (1998) High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. *Mol. Biol. Evol.* **15,** 293–302.

4. Junakovic, N., Di Franco, C., and Terrinoni, A. (1997) Evidence for a role of the host in regulating the activity of transposable elements in *D. melanogaster*: the case of *Bari 1* elements in Charolles stock. *Genetica* **100,** 149–154.
5. Alonso-Gonzalez, L., Dominquez, A., and Albornoz, J. (2003) Structural heterogenity and genomic distribution of Drosophila melanogaster LTR-retrotransposons. *Mol. Biol. Evol.* **20,** 401–409.
6. Tower, J., Karpen, G., Craig, N., and Spradling, A. (1993) Preferential transposition of *Drosophila P* elements to nearby chromosomal sites. *Genetics* **133,** 347–359.
7. Newfeld, S. J. and Takaesu, N. (1999) Local transposition of a hobo element within the decapentaplegic locus of *Drosophila*. *Genetics* **151,** 177–187.
8. Potter, S., Brorein, W. J., Dunsmuir, J. P., and Rubin, G. M. (1979) Transposition of elements of the *412*, *copia* and *297* dispersed repeated gene families in *Drosophila*. *Cell* **17,** 415–427.
9. Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., et al. (1995) Transposable elements are stable components of *Drosophila melanogaster* heterochromatin. *Proc. Natl. Acad. Sci. USA* **92,** 3804–3808.
10. Dej, K. J., Gerasimova, T., Corces, V. G., and Boeke, J. D. (1998) A hotspot for the *Drosophila gypsy* retroelement in the *ovo* locus. *Nucleic Acids Res.* **26,** 4019–4024.
11. Kassis, J. A., Noll, E., VanSickle, E. P., Odenwald, W. F., and Perrimon, N. (1992) Altering the insertional specificity of a *Drosophila* transposable element. *Proc. Natl. Acad. Sci. USA* **89,** 1919–1923.
12. Pasyukova, E. G., Belyaeva, E. S., Ilyinskaya, L. E., and Gvozdev, V. A. (1988) Outcross dependent transposition of copia like mobile genetic elements in chromosomes of an inbred Drosophila melanogaster stock. *Mol. Gen. Genet.* **212,** 281–286.
13. Di Franco, C., Galluppi, D., and Junakovic, N. (1992) Genomic distribution of transposable elements among individuals of an inbred *Drosophila* line. *Genetica* **86,** 1–11.
14. Garcia-Guerrero, M. P. and Biémont, C. (1995) Changes in the chromosomal insertion pattern of the *copia* element during the process of making the chromosomes homogeneous in *Drosophila melanogaster*. *Mol. Gen. Genet*. **246,** 206–211.
15. Petrov, D. A., Schutzman. J. L., Hartl, D. L., and Lozovskaya, E. R. (1995) Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci. USA* **92,** 8050–8054.
16. Vasilyeva, L. A., Bubenshchikova. E. V., and Ratner. V. A. (1999) Heavy heat shock induced retrotransposon transposition in *Drosophila. Genet. Res.* **74,** 111–119.
17. Glaser, R. L. and Spradling, A. C. (1994) Unusual properties of genomic DNA molecules spanning the euchromatic–heterochromatic junction of a *Drosophila* minichromosome. *Nucleic Acids Res*. **22,** 5068–5075.
18. Leach, J. T. and Glaser, R. L. (1998) Quantitative hybridization of genomic DNA fractionated by pulsed-field gel electrophoresis. *Nucleic Acids Res.* **26,** 4787–4789.
19. O'Neill, R. J. W., O'Neill, M. J., and Graves, J. A. M. (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* **393,** 68–72.

20. Labrador, M., Farre, M., Utzet, F., and Fontdevila, A. (1999) Interspecific hybridization increases transposition rates of *Osvaldo*. *Mol. Biol. Evol.* **16,** 931–937.
21. Junakovic, N., Terrinoni, A., Di Franco, C., Vieira, C., and Loevenbruck, C. (1998) Accumulation of transposable elements in the heterochromatin of *D. simulans* and *D. melanogaster*. *J. Mol. Evol.* **46,** 661–668.
22. Fortunati, D. and Junakovic, N. (1999) Evidence for genomic regulation of the telomeric activity in *Drosophila melanogaster*. *Genetica* **107,** 95–102.
23. Woodruff, R. C., Thompson, J. N. Jr., Barker, J. S., and Huai, H. (1999) Transposable DNA elements and life history traits: II. Transposition of *P* DNA elements in somatic cells reduces fitness, mating activity, and locomotion of *Drosophila melanogaster*. *Genetica* **107,** 261–269.
24. Nikitin, A. G. and Shmookler Reis, R. J. (1997) Role of transposable elements in age-related genomic instability. *Genet. Res.* **69,** 183–195.
25. Soriano, S., Fortunati, D., and Junakovic, N. (2002) Evidence for the host contribution in the definition of preferential insertion sites of the elements of *Bari1* transposon family in *Drosophila melanogaster*. *J. Mol. Evol.* **55,** 606–615.
26. Maside, X., Bartolome, C., Assimacopoulos, S., and Charlesworth, B. (2001) Rates of movement and distribution of transposable elements in *Drosophila melanogaster*: *In situ* hybridization vs Southern blotting data. *Genet. Res. Camb.* **78,** 121–136.

5

Computational Analysis of Transposable Element Sequences

I. King Jordan and Nathan J. Bowen

Summary

This chapter provides a simple guide for the computational analysis of transposable element (TE) sequences. Web links are provided for a number of sequence analysis applications, and their potential use in the analysis of TE sequences is briefly described. The level of detail provided is intended to be sufficient for a naive user to begin to analyze TE sequences *in silico*. The emphasis is placed on the identification, retrieval and manipulation of TE sequences. Information is also provided on the evolutionary study of TE sequences including the use phylogenetics programs.

Key Words: Transposable elements; computational analysis; sequence comparison; repeat masker; phylogenetics; alignment.

1. Introduction

The purpose of this chapter is to provide the reader with a simple heuristic guide for the computational analysis of transposable element (TE) sequences (nucleotide and/or amino acid). The current revolution in genomics has produced a wealth of sequence information. These sequence data are of particular relevance to the field of TE biology as mobile elements are ubiquitous and often constitute a substantial fraction of their hosts' genomes. Concurrent with the production of genomic sequence data has been a concerted effort to produce and disseminate the computational tools necessary to analyze and interpret these data. This work has resulted in a vast and potentially confounding array of computational tools and approaches. This chapter attempts to present a beginning framework for the appropriate selection and use of these tools. This work in no way represents a comprehensive or in-depth survey of the conceptual foundations and computational tools available for sequence analysis. Programs and analytical approaches presented here are chosen on the basis

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

of ease of use and the authors' familiarity. An emphasis is placed on brevity, and just enough information is provided for the naive user to get started with a variety of sequence analysis techniques. Only software that is freely available, either on Web servers or as downloadable executable code, is described here. In addition, an attempt is made to provide information relevant to users working under the Windows® (PC), Macintosh®, and/or UNIX® operating systems.

Given that evolution is a unifying theme in biology and that TEs are known to have a major impact on genome evolution and organization, the emphasis here is on sequence analyses that enable detailed evolutionary inferences. However, it is hoped that the tools and approaches described here will prove to be relevant to the computational analysis of TEs in any biological context.

2. Material

URLs (Web addresses) where the programs can be found that are recommended for computational sequence analysis are given, for the most part, in **Tables 1–5** and in **Subheading 2.** The prefix http:// has been omitted from the addresses in the interest of space, but is required when typing the URL.

3. Methods

3.1. Sequence Retrieval and Manipulation

Molecular sequence data are represented as strings of characters (e.g., A, T, C, and G for nucleotides) and are stored with annotation as text files in a variety of different file formats. Some familiarity with a few of the more common sequence formats (e.g., FASTA format; *see* www.ncbi.nlm.nih.gov/BLAST/fasta.html) will be helpful to the user. Sequence analysis programs require that data be entered in specific and often different formats. Thus the use of a file-format converter will likely be inevitable in any sustained effort at sequence analysis. Format converters can input sequences (usually aligned) in a variety of formats and then output them in a different user-defined format. **Table 1** lists some programs that include file-format conversion functions.

The first step in any sequence analysis project is the retrieval of sequence data. A number of databases exist that are designed to store, organize, and disseminate sequence data (*see* **Table 2**). These include very comprehensive databases such as Genbank, more focused databases such as the TIGR microbial genome database (www.tigr.org/tdb/), and even organism-specific databases such as the yeast genome database (genome-www.stanford.edu/Saccharomyces/). Sequence retrieval from Genbank will be described here.

Table 1
Programs That Perform Sequence File-Format Conversion

Program	Format		URL (All addresses require http:// prefix)	Operating system		
	Executable	Server		Windows	Macintosh	UNIX
Seqpup	**X**		iubio.bio.indiana.edu/soft/molbio/java/apps/seqpup/	**X**	**X**	**X**
DAMBE	**X**		web.hku.hk/~xxia/software/software.htm	**X**		
ClustalX	**X**		www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html	**X**	**X**	**X**
GDE	**X**		ftp.bio.indiana.edu/soft/molbio/unix/GDE/			**X**
ReadSeq		**X**	dot.imgen.bcm.tmc.edu/seq-util/readseq.html	**X**	**X**	**X**

Table 2
Molecular Sequence Databases

Database	URL (All addresses require http:// prefix)
Genbank (NCBI)	www.ncbi.nlm.nih.gov/
EMBL Nucleotide sequence database (EBI)	www.ebi.ac.uk/embl/index.html
DDBJ (DNA databank of Japan)	www.ddbj.nig.ac.jp/
SwissPROT (SIB and EBI)	www.ebi.ac.uk/swissprot/

Genbank's Entrez search and retrieval system (www.ncbi.nlm.nih.gov/Entrez/) can be used to do a variety of string searches to identify transposable element sequences of interest. For example, entering the boolean search command 'gypsy AND transposable element' in the Entrez nucleotide search field will retrieve a number of *gypsy*-like retrotransposon sequences. In addition to such string searches, the user will probably want to perform a sequence-similarity search to locate sequences that show some similarity (and thus putative relatedness) to their element of interest. This can be done using any number of different BLAST ***(1,2)*** searches (www.ncbi.nlm.nih.gov/BLAST/). To perform a search using BLAST, the user selects a query sequence (element of interest) to search the database of choice. The BLAST program then retrieves all sequences in the chosen database that have a similarity score at or above a user-defined value. The use of amino-acid query sequences usually results in more sensitive searches than those conducted with nucleotide query sequences, and it can effectively retrieve distantly related sequences. Another way to increase the sensitivity of a sequence search is to incorporate the site-specific information embedded in multiple alignments of related sequences. This can be accomplished using PSI-BLAST ***(2)***. PSI-BLAST generates a multiple sequence alignment based on an initial BLAST search. The site-specific variation derived from this alignment is then used to iteratively re-search the database for more distantly related sequences. At each iteration the user has the option of choosing which sequences to include in the next multiple sequence alignment. The process is repeated until it converges and no new sequences are retrieved. The use of PSI-BLAST has the advantage that it can retrieve very distantly related sequences that may not be detected with a standard BLAST search. However, the increased sensitivity of PSI-BLAST can also result in more false positives, and this approach necessarily involves more input and consideration from the user.

A very powerful and useful tool designed explicitly for the identification of TEs and other repetitive sequences is the Repeat Masker program

Table 3
Multiple Alignment Programs

Program	Format		URL	Operating system		
	Executable	Server		Windows	Macintosh	UNIX
ClustalW		**X**	http://www.ebi.ac.uk/clustalw/	**X**	**X**	**X**
ClustalW	**X**		ftp://ftp.bio.indiana.edu/molbio/align/clustal/	**X**	**X**	**X**
ClustalX	**X**		http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html	**X**	**X**	**X**
SAM		**X**	http://www.cse.ucsc.edu/research/compbio/HMM-apps/tuneup-alignment.html	**X**	**X**	**X**
MultAlin		**X**	http://www.toulouse.inra.fr/multalin.html	**X**	**X**	**X**
DIALIGN	**X**		http://www.gsf.de/biodv/dialign.html			**X**

Table 4
Phylogenetic Analysis Programs

Program	URL (All addresses require http:// prefix)	Operating system		
		Windows	Macintosh	UNIX
PHYLIP	evolution.genetics.washington.edu/phylip.html	**X**	**X**	**X**
MEGA	www.megasoftware.net/	**X**		
ClustalX	www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html	**X**	**X**	**X**
DAMBE	web.hku.hk/~xxia/software/software.htm	**X**		
Treeview	taxonomy.zoology.gla.ac.uk/rod/treeview.html	**X**	**X**	
PAUP*	paup.csit.fsu.edu	**X**	**X**	**X**

Table 5
Sequence Divergence and Polymorphism Programs

Program	URL (All addresses require http:// prefix)	Operating system		
		Windows	Macintosh	UNIX
DnaSP	www.ub.es/dnasp	**X**		
MEGA	www.megasoftware.net/	**X**		
DAMBE	web.hku.hk/~xxia/software/software.htm	**X**		
JaDis	biom3.univ-lyon1.fr/software/jadis.html	**X**	**X**	**X**
PAML	abacus.gene.ucl.ac.uk/software/paml.html	**X**	**X**	**X**

(repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker) (Smit, A. F. A. and Green, P., 2001, unpublished). Repeat Masker takes a FASTA-formatted sequence as input and characterizes any repetitive elements in the sequence that have similarity to the elements in its database. The "masked" regions of the input sequence, i.e., those with similarity to known repetitive DNA, are annotated and returned to the user. This program works best for systems where most if not all of the repetitive DNA elements have been defined and are thus already present in the database.

Once the user has retrieved a number of TE sequences of potential interest, the next step is to align the sequences. Multiple sequence alignment involves the identification and alignment of homologous residues among a group of related sequences, and it is a prerequisite to the extraction of meaningful biological and evolutionary information from the sequences. A number of multiple sequence alignment methods and programs are available (**Table 3**). The most commonly used program is Clustal. Clustal users can choose from two different interfaces that run the same alignment algorithm: ClustalW ***(3)*** has a text interface, and ClustalX ***(4)*** has a more user-friendly Windows interface. Sequences can be input into Clustal in a number of different formats. Prior to multiple sequence alignment, the user has the option of adjusting a number of alignment parameters, including gap penalties and the protein or DNA weight matrix to be used. Once the alignment is complete, the user can re-align selected sequences or a selected residue range. These procedures are often iterated, with variations, a number of times until the best alignment is obtained.

It is important to note that multiple sequence alignment can be quite inexact, especially with distantly related sequences. The user should always visually inspect the output from any multiple alignment program. In obvious cases, the user may choose to manually adjust any misaligned region. Some biological knowledge of the sequence in question as well as familiarity with previous work done on similar or related sequences can greatly aid in manually aligning conserved motifs. A more conservative approach would entail the removal of any ambiguous or poorly aligned region from the alignment. The reliability of any subsequent analysis depends on the accuracy of the alignment. It is therefore critical to ensure that the best possible alignment is obtained, using both the parameters of the program and manual adjustment if necessary, before proceeding with analysis of the data.

3.2. Sequence Analysis

Computational analysis of TE sequences is a comparative endeavor which at its core entails the detection and study of shared patterns among groups of related elements. The patterns that are revealed through these efforts reflect the historical process of evolution. With this relationship between pattern and

process in mind, the sequence analysis section consists of four parts. The first two parts, *Phylogenetic Analysis* and *Aging of Elements*, emphasize the detection of evolutionary patterns among elements. While the final two parts, *Genome-Level Selection* and *Host-Level Selection*, describe the study of some of the forces involved in the process of element evolution.

3.2.1. Phylogenetic Analysis

The evolutionary relationships among a related group of TEs can be discerned using phylogenetic analysis. Phylogenetic analysis can also be used to identify novel families of TEs ***(5)*** or to uncover recombination events between individual elements ***(6)***. Phylogenetic reconstruction begins with a reliably aligned set of sequences (*see* **Subheading 3.1.**). There are three general methods of phylogenetic reconstruction: distance-based, parsimony, and maximum-likelihood ***(7)***. Many programs that implement one or more of these approaches are available (**Table 4** and evolution.genetics.washington.edu/phylip/software.html).

The most commonly used programs are PAUP* ***(8)*** and PHYLIP ***(9)***. PAUP*, as implemented on the Macintosh operating system, is probably the most user-friendly phylogenetic analysis program. However, PAUP* is not freely available. PHYLIP, although not quite as user friendly as PAUP*, is also quite useful and widely employed. PHYLIP can perform all three general methods of phylogenetic reconstruction. The program requires the user to supply an input file of aligned sequences in a format specific to the program. Alignments can be converted into this PHYLIP format using a sequence converter (**Table 1**). In order to execute a program in PHYLIP, it is simplest to put the input file in the same directory as the program to be used and name that file "infile." Use of PHYLIP can be somewhat unwieldy, as it requires each step in the analysis to be performed by a different program. For example, with distance-based phylogenetic reconstruction the user must first calculate a distance matrix using either the DNADIST or PROTDIST programs, depending on the type of sequence being analyzed.

Once a distance matrix has been calculated, it can be used with one of several tree-building programs to reconstruct the phylogeny. The most common measure of support for individual branches of phylogenetic trees is the bootstrap. As with phylogenetic reconstruction, bootstrapping with PHYLIP requires the separate use of several different programs. First, randomized replicate sequence alignments are generated using SEQBOOT. Then, for each alignment a distance matrix is calculated, followed by multiple tree reconstructions using the programs described above. Finally, a consensus phylogeny is built using the CONSENSE program. The percentage of times each branch shows up in all of the phylogenies reconstructed from the bootstrapped align-

ments can be determined from the CONSENSE output; this is taken as the measure of support for that branch. Viewing and graphically manipulating the phylogenies produced by PHYLIP and other programs can be done using the Treeview ***(10)*** program (**Table 4**).

3.2.2. Aging of Elements

The age of elements in the genome may be of interest to the TE researcher. Long terminal repeat (LTR)-containing retrotransposons can be aged in a straightforward way by comparing the sequences of their 5' and 3' LTRs ***(11,12)***. Due to the mechanism of reverse transcription, when an LTR retrotransposon inserts into the genome its LTRs are expected to be identical in sequence. Subsequent to insertion, the LTRs accumulate mutations. Thus the level of sequence divergence between 5' and 3' LTRs of an element can be used to assess approximately how much time has elapsed since it inserted in the genome. For non-LTR elements, ages can be estimated for groups or subfamilies of related elements ***(13)*** as opposed to the aging of individual elements possible with LTR retrotransposons. This technique also relies on the fact that elements accumulate mutations subsequent to their insertion in the genome. In the case of a related group of non-LTR elements, the sequence of a common ancestor can be estimated either by using a consensus sequence, or by using a phylogenetic approach. Once an ancestral sequence is estimated for a given group of elements, the average number of mutations that have accumulated between the ancestral sequence and each extant sequence can be determined. This average level of sequence diversity can be used to estimate the age of a group or subfamily of elements.

3.2.3. Genome-Level Selection

In addition to the patterns of evolution revealed as described above, analysis of TE sequences can also yield information on the process of element evolution. Selection on TE sequences at the level of the genome, or inter-element selection ***(14)***, occurs as a result of differential reproductive success (i.e., transposition rates) among members of a TE family. This type of selection is consistent with the "selfish DNA" hypothesis ***(15,16)*** of TE evolution. The role of inter-element selection can be inferred by comparing paralogous copies of elements within genomes ***(17)***. Comparison of protein encoding nucleotide sequences can yield evidence of inter-element selection. In order to perform such comparisons, it is necessary to align the codons of the protein-encoding nucleotide sequences. This action can be accomplished by first aligning the encoded protein sequences, and then ensuring that the gaps in the corresponding nucleotide sequence alignment match those in the encoded amino-acid sequence alignment. This procedure is implemented in the DAMBE ***(18)***

program (**Table 1**) for PC users or the DNA stacks ***(19)*** program (biology.fullerton.edu/deernisse/dnastacks.html) for Macintosh users.

Once the codons are properly aligned, comparison of synonymous (ds) and nonsynonymous (dn) substitution rates will yield information on the nature of selection that has operated on element sequences within genomes. A number of programs are available that can calculate ds and dn (**Table 5**) as well as other measures of nucleotide variation. Some of these programs also include more sophisticated tests that may reveal subtle effects of selection not detected by a simple ds versus dn comparison. For example, comparison of ds and dn can be done for individual branches of a phylogenetic tree to evaluate different historical episodes of selection ***(20,21)***.

3.2.4. Host-Level Selection

The detection of host-level selection (i.e., between organisms) on element sequences is not as straightforward as the detection of inter-element selection and as yet is less common. The availability of identically located (orthologous) element sequences from the genomes of different but closely related species is critical to this endeavor. Unless the elements have a site-specific insertion mechanism, the presence of orthologous element sequences in related genomes indicates that these elements inserted prior to the evolutionary divergence of the genomes (species). Thus any selection acting on these sequences necessarily occurred after transposition (insertion), or in other words, during the process of species divergence ***(22)***. Orthologous element sequences can be compared in the same way as described for paralogous copies above. For example, ds and dn comparisons can be made to assess whether orthologous element sequences are being conserved between species due to selection acting at the host level. In addition, nonencoding orthologous element sequences can be compared to determine if they may be conserved between species and thus potentially play some regulatory role for the host species. Selection of TE sequences at the host level is not consistent with the "selfish DNA" hypothesis of TE evolution and indicates that the element sequences in question are performing an essential function for their host species ***(23)***.

References

1. Altschul, S. F., Gish, W., Miller, W., Mayers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.
2. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.
3. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence

weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680.
4. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25,** 4876–4882.
5. Bowen, N. J. and McDonald, J. F. (1999) Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements. *Genome Res.* **9,** 924–935.
6. Jordan, I. K. and McDonald, J. F. (1999) Phylogenetic perspective reveals abundant Ty1/Ty2 hybrid elements in the *Saccharomyces cerevisiae* genome. *Mol. Biol. Evol.* **16,** 419–422.
7. Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. Phylogenetic inference. In *Molecular Systematics* (Hillis, D. M., Moritz, C., and Mable, B. K., eds.), Sinauer Associates, Sunderland, MA, 1996, pp. 407–514.
8. Swofford, D. L. (ed.) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).* Sinauer Associates, Sunderland, MA, 1998.
9. Felsenstein, J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266,** 418–427.
10. Page, R. D. M. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12,** 357–358.
11. SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y., and Bennetzen, J. L. (1998) The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20,** 43–45.
12. Promislow, D. E., Jordan, I. K., and McDonald, J. F. (1999) Genomic demography: a life-history analysis of transposable element evolution. *Proc. R. Soc. Lond. B Biol. Sci.* **266,** 1555–1560.
13. Kapitonov, V. and Jurka, J. (1996) The age of Alu subfamilies. *J. Mol. Evol.* **42,** 59–65.
14. McDonald, J. F., Matyunina, L. V., Wilson, S., Jordan, I. K., Bowen, N. J., and Miller, W. J. (1997) LTR retrotransposons and the evolution of eukaryotic enhancers. *Genetica* **100,** 3–13.
15. Doolittle, W. F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284,** 601–603.
16. Orgel, L. E. and Crick, F. H. (1980) Selfish DNA: the ultimate parasite. *Nature* **284,** 604–604.
17. Jordan, I. K. and McDonald, J. F. (1999) Comparative genomics and evolutionary dynamics of Saccharomyces cerevisiae Ty elements. *Genetica* **107,** 3–13.
18. Xia, X. (ed.) *Data Analysis in Molecular Biology and Evolution*. Kluwer Academic Publishers, Norwell, MA, 2000.
19. Eernisse, D. J. (1992) DNA Translator and Aligner: HyperCard utilities to aid phylogenetic analysis of molecules. *Comput. Appl. Biosci.* **8,** 177–184.
20. Messier, W. and Stewart, C. B. (1997) Episodic adaptive evolution of primate lysozymes. *Nature* **385,** 151–154.

21. Zhang, J., Rosenberg, H. F., and Nei, M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95,** 3708–3713.
22. Witherspoon, D. J., Doak, T. G., Williams, K. R., Seegmiller, A., Seger, J., and Herrick, G. (1997) Selection on the protein-coding genes of the TBE1 family of transposable elements in the ciliates *Oxytricha fallax* and *O. trifallax*. *Mol. Biol. Evol.* **14,** 696–706.
23. Miller, W. J., McDonald J. F., and Pinsker, W. (1997) Molecular domestication of mobile elements. *Genetica* **100,** 261–270.

6

Biochemical Analysis of Long Terminal Repeat Retrotransposons

Andreas Bachmair, Karin Garber, Shin Takeda, Kazuhiko Sugimoto, Tetsuji Kakutani, and Hirohiko Hirochika

Summary

A brief survey of the retrotransposition cycle is given with emphasis on intermediary steps and their identification by biochemical analysis. References include some older key publications that may not be easily accessible via electronic databases. Methods for enrichment of viruslike particles from plant material and from yeast are described.

Key Words: Virus-like particles (VLPs); sucrose gradient centrifugation; RNA packaging; reverse transcriptase assay; cDNA detection.

1. Introduction

Retrotransposons can be viewed either as an intrinsic part of the organism where they occur, or, alternatively, as separate entities that display a set of transformations that are usually called the "retrotransposon life cycle." While it is clear that many host factors are required for this life cycle, a core of critical reactions, particularly those that may not always be in the best interest of the host, rely on retrotransposon-encoded proteins. A major goal of retrotransposon activity is to replicate element sequences more often than host DNA replication occurs for the rest of the genome. Parasitic aspects of host–element relationships are enforced by the similarity of long terminal repeat (LTR)-containing retroelements to retroviruses, which are known as nonessential, often disease-causing, contagious entities.

Biochemical investigation of retrotransposons has been largely restricted to the analysis of retroelement-encoded functions, with the notable exception of elements from *Drosophila melanogaster*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*. Many methods of analysis have been developed

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

by adapting methods originally employed for retroviruses. Experiments have demonstrated considerable diversity regarding priming mechanism and protein expression strategy. One major difference between retroviruses and retrotransposons lies in the abundance of products present in an element-containing/infected cell. Whereas retroviruses may kill host cells without risk of extinction, retrotransposon activity has to be much more restricted to ensure host (and thus element) survival.

Low abundance of components to be analyzed is certainly a major limitation to further progress in dissecting and understanding retrotransposon-related reactions. The problem may be circumvented by overexpression and/or by use of a heterologous host. In some cases, in vitro reactions have been used successfully for analysis. Advantages of the use of a homologous host for analysis are that all aspects of element biology can, in principle, be analyzed; all results reflect "natural" behavior. A disadvantage is the low abundance of transposition intermediates. Furthermore, investigations cannot be too manipulative, e.g., a test of altered versions of retrotransposon components is complicated or even compromised by the background of unchanged elements. Heterologous hosts may be closely related (e.g., use of a plant host for a plant element) in order to reflect as much as possible the original biological context, or they may be chosen for reasons of convenience. In the latter case, microbial hosts offer the most advantages.

From the generally accepted life cycle ***(1)***, one may split biochemical analysis into four different steps: (1) particle formation; (2) RNA packaging and maturation; (3) reverse transcription; and (4) target site selection and integration.

1.1. Particle Formation

Particle assembly can be analyzed by enrichment of virus-like particles (VLPs; also called Gag particles) via centrifugation through a sucrose gradient or a sucrose cushion ***(2,3)***. Because all eukaryotic hosts potentially harbor many classes of elements, specific antibodies are necessary to identify and distinguish the object of research from other elements and their products. Proteins and nucleic acids within VLPs are, at least to a limited extent, protected from hydrolytic enzymes contained in crude cell extracts, so that many protocols do not take extensive care for absence of nucleases. However, it has been reported that yeast Ty*1* VLPs are permeable to Ribonuclease A, which has a size of approx 14 kDa ***(4)***.

1.2. RNA Packaging and Maturation

RNA content can be analyzed in fractions enriched for VLPs (*see* **Subheading 3.2.**, **step 4**). Experiments with yeast Ty*1* indicate that in addition to the

expected components (retrotransposon mRNA and priming tRNA), nucleic acids without an obvious role in the life cycle may also be present, such as additional tRNA species or cellular mRNAs ***(5,6)***. Alternatively, in vitro studies may be used to investigate aspects of packaging and folding of RNA ***(7,8)***. In addition to the RNA, protein products of the VLP are also in a dynamic transition even after particle assembly ***(1,4)***. The process is usually called maturation and involves proteolytic cleavage of the particle-forming Gag protein, which can in principle be followed, with the help of antibodies, by pulse-chase experiments. Precise determination of cleavage sites has proven exceptionally tricky in those rare cases where protein processing was studied ***(9,10)***, but may be aided in the future by recently developed mass spectrometric methods.

1.3. Reverse Transcription

A robust proof of reverse transcriptase (RT) activity is the detection of reverse transcripts in VLPs (*see* **Subheading 3.1.**). Direct determination of cDNA synthesis has been possible by supplying a VLP fraction with labeled nucleotides. Variants of the labeling procedure either provide only the nucleotides to probe incorporation based on templates inside VLPs, or provide RNA template and DNA primer as an exogenous nucleic acid source, as well. Experiments have shown that poly(rC) and an oligo(dG) primer give both a high signal and a low background in this type of assay with Ty*1* VLPs ***(11,12)***. This primer–template combination has therefore become the standard tool for detection of retroelement RT activity. An additional tool for analysis is the insertion of the sequence of a presumed reverse transcriptase to replace Ty*1* RT. A broad range of biochemical and genetic assays may then allow detection of this "heterologous" RT activity in yeast ***(13,14)***.

1.4. Integration and Target Site Selection

Investigation of retrotransposon integration has been extended to the biochemical level in at least two instances, namely for Ty*3* and Ty*5* of *S. cerevisiae* ***(15,16)***. In both cases, integrase apparently interacts with specific chromatin proteins, the distribution of which determines the target site specificity. More common, however, is the indirect determination of integration preferences by sequencing of integration sites.

In the following section, two methods for enrichment and analysis of VLPs are presented. The first method, described in **Subheading 3.1.**, uses plant cells as a starting material. This method is based on a protocol originally developed for isolation of Cauliflower Mosaic Virus DNA from plant tissues ***(17)***. It is applied for enrichment of Gag particles of an autonomous tobacco retrotransposon, *Tto1*, from *Arabidopsis* calli carrying highly expressed *Tto1* cop-

ies ***(18–20)***. For simplicity, the protocol is written in terms of enrichment of DNA intermediates. However, it can be adapted for other applications and for plant LTR-retrotransposons other than *Tto1*, if Gag particles or DNA intermediates accumulate to a sufficient extent. Studies of Gag particles of another plant LTR-retrotransposon have recently been described by Jääskeläinen et al. ***(21)***.

The second method, reported in **Subheading 3.2.**, was originally developed for analysis of retrotransposons from baker's yeast, *S. cerevisiae* ***(11)***. Retrotransposons from other species, however, can be analyzed in a similar way after overexpression in *S. cerevisiae*. In our laboratory, the method is currently used to study proteolytic processing and other steps after galactose-induced overexpression of *Tto1* components in *S. cerevisiae*.

2. Materials

2.1. Isolation of VLPs From Plant Tissue

1. Extraction buffer: 0.2*M* Tris-HCl, pH 7.0, 20 m*M* EDTA, 1.5 *M* Urea (use within 2 wk, store at 4°C, check pH at 4°C just before use).
2. 15% Sucrose, 10 m*M* potassium phosphate buffer, pH 7.2.
3. Particle suspension buffer: 0.1*M* Tris-HCl, pH 7.4, 2.5 m*M* $MgCl_2$.
4. DNase I (before use, dilute to 1 mg/mL with particle suspension buffer).
5. 0.25 *M* EDTA, pH 8.0.
6. 10% Sodium dodecyl sulfate (SDS).
7. 3*M* sodium acetate, pH 7.0 (adjust pH with acetic acid).
8. Phenol:chloroform mixture (made from molecular biology grade phenol by addition of an equal amount of chloroform. Equilibrate with 0.1*M* Tris-HCl, pH 7.5).
9. Triton X-100.
10. 100% Ethanol.
11. Sea sand (40 to 80 mesh).
12. Mortar and pestle.
13. SpeedVac evaporator.
14. Tabletop ultracentrifuge (Beckman TL-100), rotor TLA45.

2.2. Materials for Isolation of VLPs From S. cerevisiae

1. Buffer B/EDTA, 20 m*M* HEPES/KOH, pH 7.8, 15 m*M* KCl, 3 m*M* DTT, 10 m*M* EDTA (add DTT freshly from 1 *M* stock).
2. 1*M* dithiothreitol (DTT) stock (store at –20°C).
3. Roche protease inhibitor tablets (Complete Mini, cat. no. 1 836 153).
4. Sucrose for density gradient centrifugation (Merck).
5. Glycerol.
6. Acid-washed glass beads (0.5 mm diameter).

7. Ultracentrifuge (Beckman, L8-70), rotor SW41Ti, translucent (nitrocellulose) centrifuge tubes.
8. Centrifuge (Beckman J2-21), rotor JA-10.

3. Methods

3.1. Isolation of VLPs From Plant Tissue

1. With a prechilled mortar and pestle, vigorously homogenize 0.2 to 0.3 g of plant tissues mixed with 0.5 mL of extraction buffer and 0.5 mL of sea sand on ice. Add 0.5 mL of extraction buffer and 20 µL Triton X-100 (2% final concentration), then mix completely (*see* **Note 1**).
2. Transfer extracts into a 1.5-mL tube and centrifuge for 5 min at 180*g*, 4°C (1500 rpm in a conventional microfuge rotor).
3. Transfer supernatant into a new 1.5-mL tube and centrifuge for 10 min at 180*g*, 4°C.
4. Carefully transfer supernatant into a new 1.5-mL tube and keep on ice. Do not remove any pellet or debris. Quantify soluble protein content, if required.
5. Transfer approx 700 µL of supernant gently onto 250 µL of prechilled 15% sucrose, 10 m*M* potassium phosphate buffer, pH 7.2, in a 1.5-mL tube for ultracentrifugation.
6. Ultracentrifuge for 90 min at 109,000*g*, 4°C (45,000 rpm in a Beckman TLA45 rotor).
7. Discard supernant carefully. If required, pellet can be washed by addition of 200 µL of 15% sucrose, 10 m*M* potassium phosphate buffer, pH 7.2, and ultracentrifuge again for 20 min.
8. Resuspend pellet in 0.5 mL of particle suspension buffer, carefully using pipet tips, and place on ice. (For analysis by Western blot, pellet can be directly dissolved in an appropriate volume of SDS-loading buffer. Detection of *Tto1* Gag polypeptides in the pellet fraction is shown in **Fig. 1A**.)
9. Add 5 µL of 1 mg/mL DNase I (10 µg/mL final), gently mix, and incubate at 37°C for 10 min. DNA in Gag particles should be protected from the digestion.
10. Add 20 µL of 0.25*M* EDTA, 50 µL of 10% SDS, 25 µL of 10 mg/mL proteinase, mix, and incubate at 65°C for 10 min.
11. Purify DNA samples by extraction with 0.5 mL of phenol:chloroform three times, and with 0.5 mL of chloroform once. Transfer last aqueous phase into 1.5-mL tube and add 40 µL of 3*M* sodium acetate (pH 7.0), and 1 mL of cold 100% ethanol to precipitate DNA. Mix and incubate at –80°C for 10 min or more.
12. Centrifuge for 10 min at 18,000*g*, 4°C (15,000 rpm in a microfuge). Pour off supernatant and wash pellet with 0.5 mL of 70% ethanol; centrifuge for 5 min. Pour off supernatant and dry pellet in a SpeedVac evaporator. Resuspend DNA pellet in water or TE buffer. Prepared DNA samples are ready for gel blot analysis with specific probes or for ligation reaction (*see* **Note 2**). Detection of a full-length cDNA intermediate of *Tto1* by DNA gel blot analysis is shown in **Fig. 1B**.

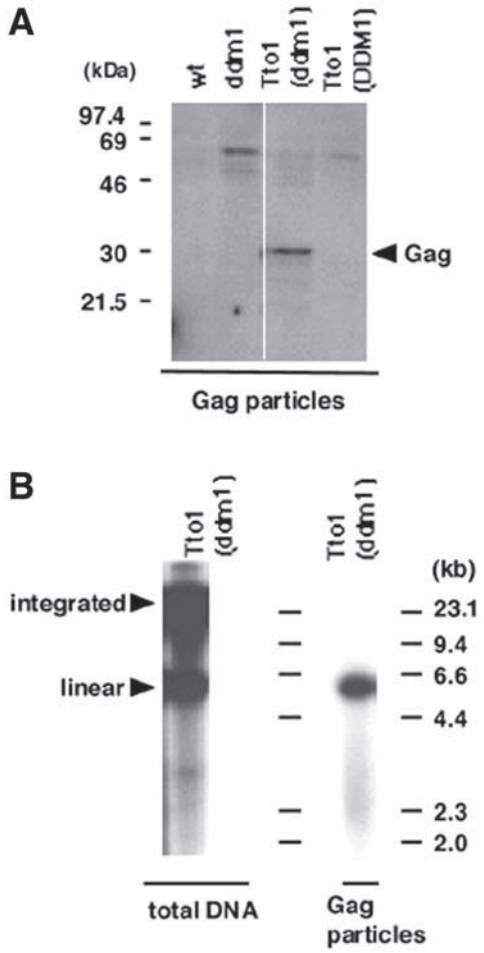


Fig. 1. Enrichment of Gag particles of *Tto1* from transgenic *Arabidopsis* calli. In the transgenic lines, expression of *Tto1* copies is suppressed, but re-activated in *ddm1* mutant background ***(20,28)***. **(A)** *Tto1* Gag polypeptides in the prepared Gag-particle fraction were detected by Western blot analysis using anti-Tto1 Gag antibodies ***(24)***. Gag-particle fractions were prepared from various *Arabidopsis* callus lines indicated: wt, wild type; Tto1 (ddm1), transgenic callus with *ddm1* background; Tto1 (DDM1), transgenic callus with *DDM1* background. **(B)** *Tto1* linear DNA molecules in the prepared Gag-particle fraction. Total DNA (0.5 μg) extracted from a transgenic *ddm1* callus line, and DNA in the Gag-particle fraction of the same line were examined by DNA blot analysis without digestion by restriction enzymes, using ^{32}P-labeled DNA probe for Tto1 ***(24)***.

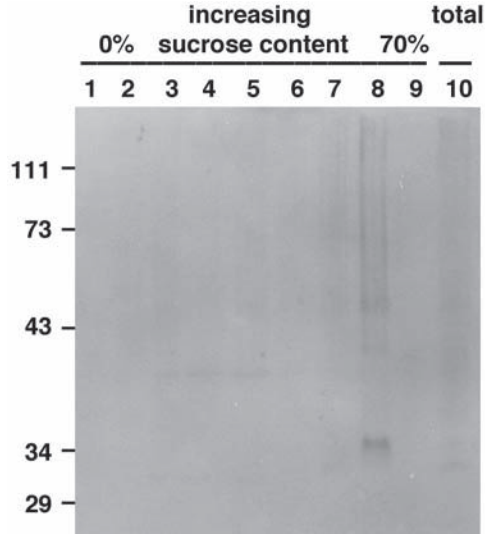


Fig. 2. Detection of *Tto1* proteins in extracts from *S. cerevisiae* fractionated by sucrose step gradient. Antibodies directed towards the Gag part were used in Western blot analysis. The fraction with most detectable material (*lane 8*) is from the border of the 30% and 70% sucrose zones. *Lane 10* indicates total lysate as applied onto the gradient.

3.2. Isolation of VLPs From S. cerevisiae *After Heterologous Expression of a Plant Retrotransposon*

1. An overnight culture of *S. cerevisiae* expressing retrotransposon proteins is grown in 5 mL selective medium with 2% galactose at 30°C. This culture is then used to inoculate 500 mL of selective, galactose-containing medium (incubation for 24 h at 30°C) (*see* **Note 3**).
2. The cells are harvested by centrifugation for 10 min at 1100*g* (2500 rpm, rotor Beckman JA-10), washed with 50 mL dH_2O (precooled at 4°C), and resuspended in 3 mL of cold B/EDTA (optional: add protease inhibitor tablet from Roche). All the following steps are performed at 4°C. Four grams of ice-cold, nitric-acid-washed glass beads are added, and cell walls are broken by vortexing for 5 min (repeat this step if required) (*see* **Note 4**).
3. The cell debris is pelleted for 10 min at 17,000*g* (10,000 rpm, rotor Beckman JA-10), and the supernatant is loaded on a sucrose gradient (2 mL 70% sucrose, 2 mL 30% sucrose, and finally 6 mL 20% sucrose, all made in B/EDTA, are carefully pipetted into a clear (nitrocellulose) centrifugation tube).
4. The gradient is centrifuged for 3 h at 83,500*g* (26,000 rpm in the Beckman SW41Ti rotor).

5. After centrifugation, the gradient is fractionated from the top; usually 1.3 mL fractions are made. These fractions can be used for further analysis (**Fig. 2**) or might be stored at –80°C after addition of 10% glycerol.
6. To concentrate retrotransposon proteins, the peak fractions might be diluted 1:2 with buffer B/EDTA and centrifuged overnight at 15,000*g* (10,000 rpm in rotor SW41Ti). Afterwards, the pellet is dissolved in 100 µL B/EDTA.
7. 5–10 µL of the concentrated fractions are loaded onto a 10–12% SDS gel.

4. Notes

1. Extraction with a buffer containing 10 m*M* HEPES-KOH (pH 7.8), 15 m*M* KCl, 5 m*M* EDTA, 10 m*M* β-mercaptoethanol, and 1 m*M* PMSF, and subsequent ultracentrifuge fractionation using 30% sucrose in the same buffer, resulted in almost the same recovery of *Tto1* Gag products. An attempt to isolate DNA intermediates in Gag particles from frozen tissues was not successful.
2. It should be noted that in wild-type tobacco plants, the amount of *Tto1* DNA intermediates seems to be very low, even in the stress conditions in which expression of *Tto1* is activated ***(18,22,23)***. In this case, the replicated *Tto1* DNA can only be detected by subsequent adaptor ligation and PCR amplification ***(24)***.
3. Production of retrotransposons or their components in yeast can utilize a galactose-inducible promoter for high and controlled expression ***(11,25)***. In the protocol, we assume that yeast cells harbor plasmids with selectable markers for maintenance and galactose-inducible genes for retrotransposon expression. If promoter activity does not depend on galactose, 2% glucose is the carbon source of choice.
4. Using a different cell-disruption method, the VLP enrichment procedure was also applied to retrotransposon proteins expressed in *E. coli* ***(26,27)***.

References

1. Boeke, J. D. and Stoye, J. P. Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., eds.), CSHL Press, Plainview, NY, 1997, pp. 343–435.
2. Shiba, T. and Saigo, K. (1983) Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosophila melanogaster*. *Nature* **302,** 119–124.
3. Eichinger, D. J. and Boeke, J. D. (1988) The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. *Cell* **54,** 955–966.
4. Roth, J.-F. (2000) The yeast Ty virus-like particles. *Yeast* **16,** 785–795.
5. Pochart, P., Agoutin, B., Fix, C., Keith, G., and Heyman, T. (1993) A very poorly expressed tRNA-Ser is highly concentrated together with primer initiator tRNA-Met in the yeast Ty1 virus-like particles. *Nucleic Acids Res.* **21,** 1517–1521.
6. Xu, H. and Boeke, J. D. (1990) Localization of sequences required in cis for yeast Ty*1* element transposition near the long terminal repeats: analysis of mini-Ty*1* elements. *Mol. Cell. Biol.* **10,** 2695–2702.

7. Gabus, C., Ficheux, D., Rau, M., Keith, G., Sandmeyer, S., and Darlix, J.-L. (1998) The yeast Ty3 retrotransposon contains a 5'– 3' bipartite primer-binding site and encodes nucleocapsid protein NCp9 functionally homologous to HIV-1 NCp7. *EMBO J.* **17,** 4873–4880.
8. Friant, S., Heyman, T., Wilhelm, M. L., and Wilhelm, F. X. (1996) Extended interactions between the primer tRNAi-Met and genomic RNA of the yeast Ty1 retrotransposon. *Nucleic Acids Res.* **24,** 441–449.
9. Merkulov, G. V., Swiderek, K. M., Baker Brachmann, C., and Boeke, J. D. (1996) A critical proteolytic cleavage site near the C terminus of the yeast retrotransposon Ty1 Gag protein. *J. Virol*. **70,** 5548–5556.
10. Martin-Rendon, E., Hurd, D. W., Marfany, G., Kingsman, S. M., and Kingsman, A. J. (1996) Identification of proteolytic cleavage sites within the gag-analogue protein of Ty1 virus-like particles. *Molec. Microbiol.* **22,** 1035–1043.
11. Garfinkel, D. J., Boeke, J. D., and Fink, G. R. (1985) Ty element transposition: reverse transcriptase and virus-like particles. *Cell* **42,** 507–517.
12. Mellor, J., Malim, M. H., Gull, K., Tuite, M. F., McCready, S., Dibbayawan, T., et al. (1985) Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. *Nature* **318,** 583–586.
13. Dombroski, B., Feng, Q., Mathias, S. L., Sassaman, D. M., Scott, A. F., Kazazian, H. H., et al. (1994) An *in vivo* assay for the reverse transcriptase of human retrotransposon L1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **14,** 4485–4492.
14. Nissley, D. V., Garfinkel, D. J., and Strathern, J. N. (1996) HIV reverse transcription in yeast. *Nature* **380,** 30.
15. Yieh, L., Kassavetis, G., Geiduschek, E. P., and Sandmeyer, S. B. (2000) The Brf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the Gypsy-like element, Ty3. *J. Biol. Chem.* **275,** 29,800–29,807.
16. Gai, X. and Voytas, D. F. (1998) A single amino acid change in yeast retrotransposon Ty5 abolishes targeting to silent chromatin. *Molec. Cell* **1,** 1051–1055.
17. Gardner, R. C. and Shepherd, R. J. (1980) A procedure for rapid isolation and analysis of cauliflower mosaic virus DNA. *Virology* **106,** 159–161.
18. Hirochika, H. (1993) Activation of tobacco retrotransposons during tissue culture. *EMBO J.* **12,** 2521–2528.
19. Hirochika, H., Otsuki, H., Yoshikawa, M., Otsuki, Y., Sugimoto, K., and Takeda, S. (1996) Autonomous transposition of the tobacco retrotransposon *Tto1* in rice. *Plant Cell* **8,** 725–734.
20. Hirochika, H., Okamoto, H., and Kakutani, T. (2000) Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* **12,** 357–369.
21. Jääskeläinen, M., Mykkänen, A.-H., Arna, T., Vicient, C. M., Suoniemi, A., Kalendar, R., et al. (1999) Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells. *Plant J.* **20,** 413–422.
22. Takeda, S., Sugimoto, K., Otsuki, H., and Hirochika, H. (1998) Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. *Plant Mol. Biol.* **36,** 365–376.

23. Takeda, S., Sugimoto, K., Otsuki, H., and Hirochika, H. (1999) A 13-bp *cis*-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J.* **18,** 383–393.
24. Takeda, S., Sugimoto, K., Kakutani, T., and Hirochika, H., (2001) Linear DNA intermediates of the *Tto1* retrotransposon in Gag particles accumulated in stressed tobacco and *Arabidopsis thaliana*. *Plant J*. **28**, 307–317.
25. Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. (1985) Ty elements transpose through an RNA intermediate. *Cell* **40,** 491–500.
26. Luschnig, C., Hess, M., Pusch, O., Brookman, J., and Bachmair, A. (1995) The *gag* homologue of retrotransposon Ty1 assembles into spherical particles in *Escherichia coli*. *Eur. J. Biochem.* **228,** 739–744.
27. Luschnig, C. and Bachmair, A. (1997) RNA packaging of yeast retrotransposon Ty1 in the heterologous host, *Escherichia coli*. *Biol. Chem.* **378,** 39–46.
28. Jeddeloh, J. A., Bender, J., and Richards, E. J. (1998). The DNA methylation locus *DDM1* is required for maintenance of gene silencing in *Arabidopsis*. *Genes Dev.* **12,** 1714–1725.

7

Tn*5* as a Molecular Genetics Tool

InVitro Transposition and the Coupling of In Vitro Technologies With In Vivo Transposition

William S. Reznikoff, Igor Y. Goryshin, and Jerry J. Jendrisak

Summary

The development of in vitro transposition technologies have provided many powerful tools for the molecular genetics research laboratory. In this chapter we describe some of these tools with a focus on the Tn*5* transposition system. Tn*5* technologies are particularly useful because the Tn*5* transposition system has simple requirements, is efficient, random in target recognition, and robust. In particular we will describe the use of in vitro Tn*5* transposition in transposon tagging and in the generation of nested deletions. We will also describe a unique in vitro/in vivo technology in which Tn*5* inserts can be generated in a wide spectrum of bacterial species through the electroporation of preformed tranposase–transposon DNA complexes.

Key Words: Tn*5*; in vitro transposition; transposon tagging; nested deletions; electroporation tranposase.

1. Introduction

Transposition is a powerful tool for investigating and manipulating genomes. The obvious use for this technology is to create knockout mutations. However, the power of transposon mutagenesis is greatly enhanced by the fact that the DNA internal to the transposon ends can encompass a wide variety of sequences as long as the transposase can be provided by some other source. Thus the internal sequences can include, for example, selectable markers, reporter functions (such as the gene for β-galactosidase), controlling elements (such as a regulated promoter), primer binding sites, an origin of replication, an origin of gene transfer, epitope encoding sequences, and site specific recombination recognition sites.

Transposition technology is undergoing a major change as a result of the development of in vitro transposition systems. Most transposition applications

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

have previously been performed in vivo. There are several limitations imposed by using in vivo approaches. For instance, one must provide for the production of the transposase in the host cells, which requires the construction of specialized transposase expression systems for each target organism. In addition, it is typically desirable to limit transposase presence to a defined time frame, so that transposition events occur, but the products are stable (i.e., no subsequent transposition events occur). This outcome requires the introduction of transposase expression systems on suicide vectors that are not replicated by the target cells; the use of very tightly regulated transposase expression systems; or the transfer of the transposition products from one cell to another. Expression of active transposase in cells can be deleterious to the cells even in the absence of the desired transposition event. All of these limitations are bypassed by in vitro approaches.

A number of efficient in vitro transposition systems have been developed for practical applications. These include those derived from Tn*5* ***(1)***, Ty*1* ***(2)***, Tn*7* ***(3)***, *Mu* ***(4)***, *Mariner* ***(5)***, and Tn*502* ***(6)***. We will discuss Tn*5*-based technologies. The Tn*5* system is simple, requiring but three macromolecular components: transposase, transposon DNA, and target DNA. This system is also efficient, random, and robust ***(1)***. In addition, one of the technologies that we will describe, electroporation of synaptic complexes ***(7,8)*** was first developed for the Tn*5* system.

The basic technology used in in vitro transposition systems involves the introduction of the transposon DNA into the target DNA (*see* **Fig. 1A**). If the target (or the transposon) contains an origin of replication, then the resulting product will be a replicon that can be introduced into target cells by, for instance, electroporation, and the resulting cells containing independently replicating transposon inserts are selected. An example of this approach is the use of in vitro transposition to distribute primer binding sites along the length of a BAC clone ***(9)***. If no origin of replication is present, the single-strand gaps next to the ends of the transposon insert will need to be repaired, and then the transposon inserts can be incorporated into the target cell chromosome by homologous recombination. Examples of this approach for the generation of transposon insert mutations in *Streptomyces coelicolor* and *Vibrio cholerae* have recently been published ***(10,11)***.

Tn*5* will also transpose in an intramolecular fashion ***(12)***. That is, the transposon ends will attack the transposon itself rather than a second DNA (**Fig. 2**). Two different groups of products are generated: One group consists of two deletion circles, and the second group consists of inversion circles. These two groups should be approximately evenly represented in the products because their respective generation is a function of the orientation of the attacking transposon ends onto the target. If the body of the transposon

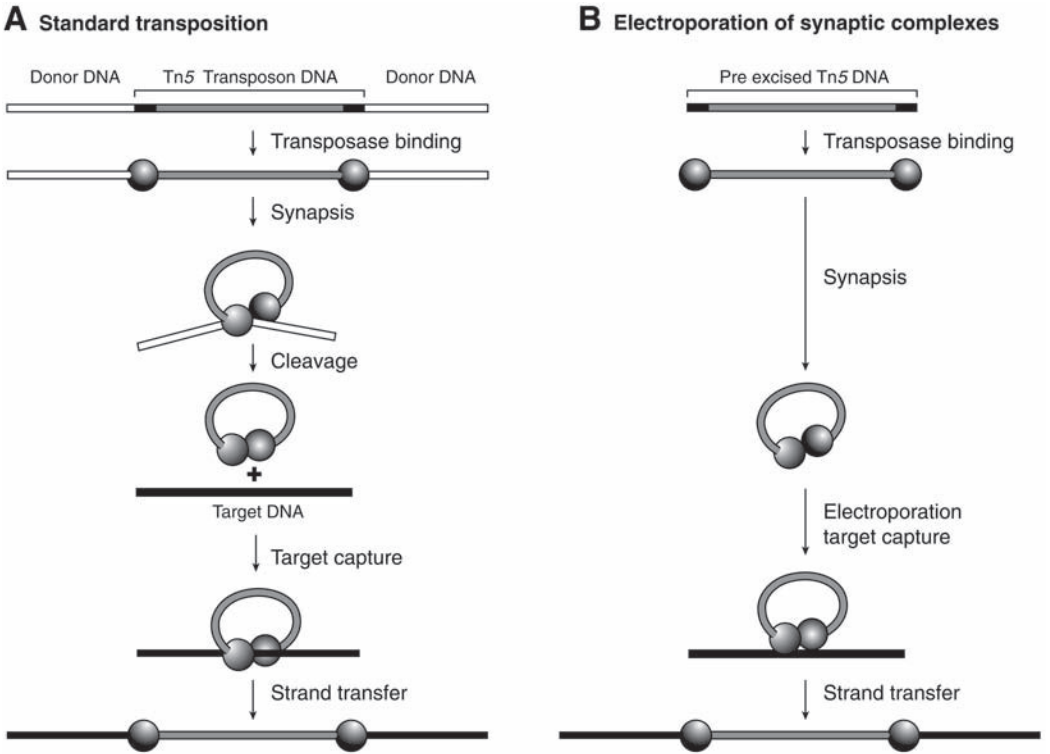


Fig. 1. Steps in Tn*5* transposition. **(A)** Standard transposition. The steps in the Tn*5* transposition process are described in more detail in references ***(13)*** and ***(14)***. **(B)** Electroporation of synaptic complexes. The steps involved in the electroporation/transposition technology bypass some steps normally followed in Tn*5* transposition. Preexcised transposon DNA is used. This allows transposase binding and synapsis without the cleavage steps. Thus these steps are performed in the absence of Mg^{2+}. The complexes are then electroporated into cells in which they encounter Mg^{2+} and target DNA and undergo strand transfer to insert the transposon.

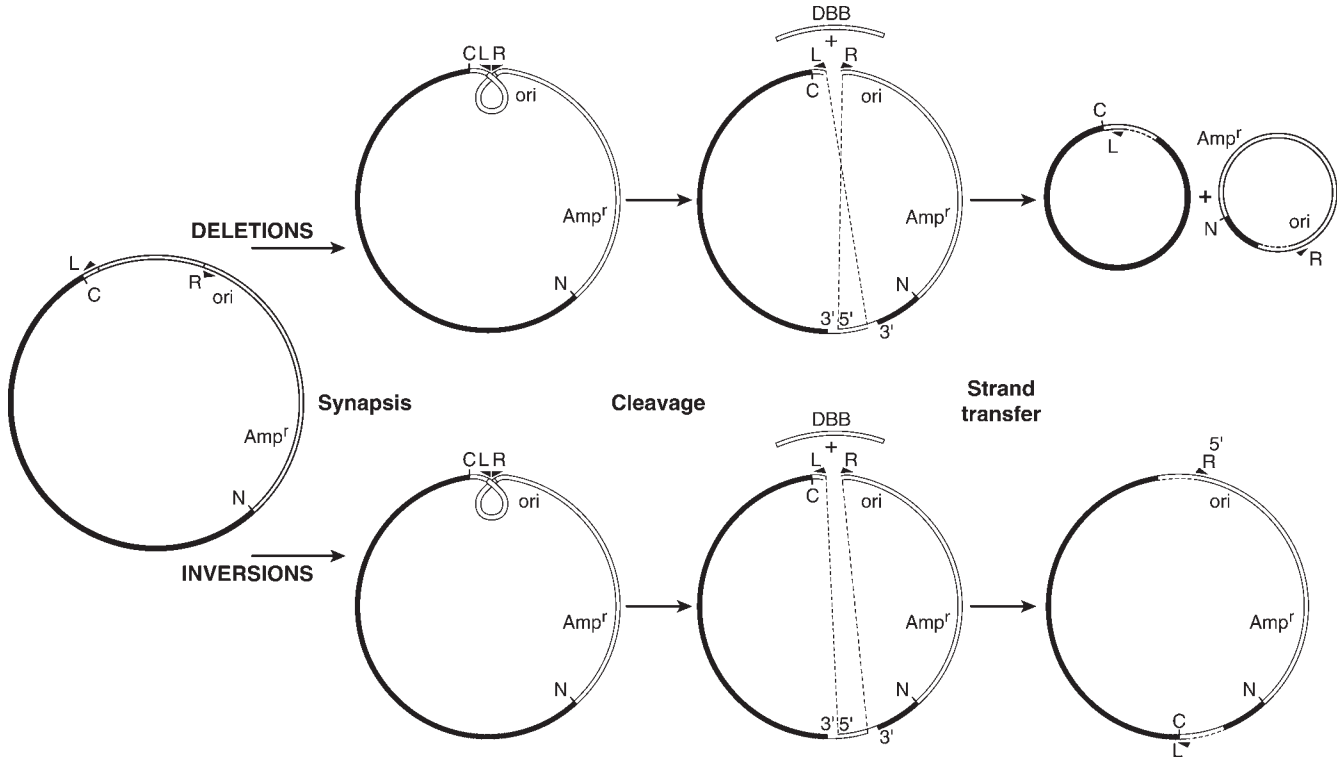


Fig. 2. Intramolecular transposition forms nested deletions and nested inversions. The substrate plasmid contains a transposon defined by two end sequences (indicated by triangles marked with L and R) that contains between the ends an origin of replication (ori), the Amp^r gene, and a target sequence encoding a protein (the N and C termini are indicated). Intramolecular transposition results in the formation of deletion circles (*top*) or inversions (*bottom*). Only one product is shown for each type of event, but in fact a library of products are made with approximately random deletion and inversion points.

substrate contains an origin of replication and a selectable marker as pictured in **Fig. 2**, one of the two types of deletion circles and the inversion circles can be propagated after introduction into a suitable host. Intramolecular transposition is a convenient technology for generating nested deletion families of the target gene.

We have recently developed a third technology that is a combined in vitro and in vivo approach ***(7)***. In essence this technology involves the formation of synaptic complexes in vitro, electroporation of the complexes into the cells, and selection for the products of transposition events. This technology has been successfully applied to many organisms, such as *Escherichia coli* K12, *Salmonella typhimurium*, *Proteus vulgaris*, *Pseudomonas* spp., and *Mycobacterium smegmatis* ***(7,8)***. It should be amenable to any organism that can be transformed via electroporation.

The basic mechanism of Tn*5* transposition and how the electroporation/transposition technology has been designed around the transposition mechanism is presented in **Fig. 1A** and **B** (*see* **refs**. ***7,13,14***). First, the transposase binds to the 19 bp sequences that define the ends of the transposon. Second, the transposase dimerizes through the formation of protein–protein and *trans* protein–DNA contacts to form a synaptic complex. Third, the transposase catalyzes cleavage of the transposon DNA free of the adjoining donor DNA. These cleavage events require the presence of Mg^{2+}. Fourth, the released synaptic complex binds to target DNA. Fifth, the 3'—OH ends of the transposon DNA attacks the target DNA with a 9-bp stagger, inserting the transposon DNA into the target. This insertion event also only occurs in the presence of Mg^{2+}. The optimal form of the electroporation/transposition technology bypasses the third step, transposase-catalyzed cleavage of the transposon DNA from the adjoining donor DNA. Rather, precleaved transposon DNA is used in the first step, and then the end-bound transposase molecules dimerize to form precleaved synaptic complexes. These complexes are formed and are quite stable in the absence of Mg^{2+}. The synaptic complexes are then electroporated into target cells. Inside the cells, the transposase is activated presumably because the cell contains Mg^{2+}; transposition then occurs. Because no transposase is synthesized in the cells, no further transposition occurs.

There are several key aspects of the Tn*5* transposition system that allows the electroporation/transposition technology to be used. First, in vitro experiments have shown that Tn*5* transposition has simple macromolecular requirements: transposase, transposon DNA, and target DNA ***(1)***. In other words, no host functions are required, so Tn*5* transposition can occur in many organisms. Second, we have generated very active forms of the transposase and the transposon end sequences so that the efficiency of transposition is quite high ***(15–17)***. Third, transposase binds to precleaved transposon end sequences in the absence

of donor or target DNA and Mg^{2+}. Fourth, transposase–transposon synaptic complexes are quite stable in the absence of Mg^{2+}. Fifth, transposase–transposon synaptic complexes are activated to transpose in the presence of Mg^{2+}.

Although there are limitations on the electroporation/transposition technology as we will discuss subsequently, it is an extremely powerful tool ***(7)***. In particular, it bypasses species barriers. Moreover, using an optimal *E. coli* K12 system and a 1.8-kb transposon, we have generated viable transposition products in as much as 8% of the cells that survived the electroporation treatment.

2. Materials

1. *Transposase*: The transposase is a hyperactive triple-mutant version of the Tn*5* transposase. The mutations are at residues 54 (E to K), 56 (M to A) and 372 (L to P) ***(1)***. The enzyme can be purchased from Epicentre Technologies (*see* **Note 1**). N-terminal His-tagged and maltose-binding protein-fusion versions of the hyperactive transposase have also been constructed and used successfully (Yigit, H. and Reznikoff, W. S., unpublished) ***(18)***.
2. *Transposon DNA*: The transposon DNA can be defined by either the outside end (OE) sequence, the naturally occurring 19-bp sequence that defines the ends of Tn*5* ***(19)***, or, preferably, by the mosaic end (ME) sequence that is a hyperactive mutant version of the OE ***(17)***. The ME end-defined transposon is described in **Fig. 3**. The DNA between the two ME sequences can carry a variety of antibiotic resistant markers or, alternatively, a multiple-cloning site for construction of the desired transposon (*see* **Note 2**). The latter is available from Epicentre Technologies.
3. *Use of Precleaved Transposon*: Both the in vitro transposition and the electroporation/transposition systems function with transposon DNA still embedded within plasmid DNA, and with prereleased transposon DNA. However, precleavage of the transposon DNA from adjoining sequences increases the electroporation/transposition frequency by over 10-fold. Precleaved transposons can be produced by two means: restriction digestion, or PCR amplification. As shown in **Fig. 3**, the ME–donor DNA sequence can be made cleavable by *Pvu*II or *Psh*AI (*Box*I) to release the transposon from an appropriate plasmid (*see* **Note 3**).

   An ME PCR primer (5' CTGTCTCTTATACACATCT 3') can be used for production of the "precleaved" transposon through PCR amplification. (Note: This single primer will function at both ends of the transposon. In addition, it is not necessary to employ a kinase with this primer or the resulting PCR product in order to make a functional transposon.) Since the percent of GC is low (37%), a 37°C annealing temperature is recommended for thermocycling. *Taq* DNA polymerase will produce untemplated 3'-A additions to a significant portion of the PCR products, which will lower the overall performance of the transposon preparation. PCR in the presence of a proofreading thermophilic DNA polymerase

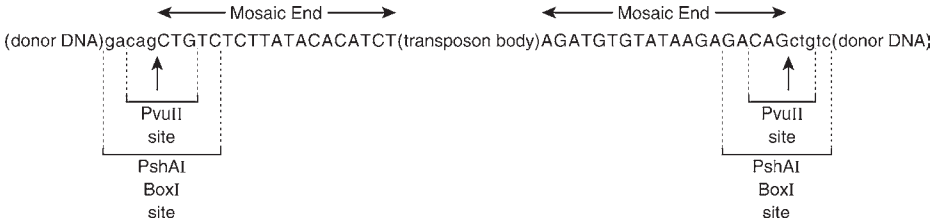


Fig. 3. Basic transposon structure. The optimal transposon for the Tn*5* in vitro system is defined by 19-bp mosaic end sequences. The transposon body can contain any desired sequence. The mosaic end–donor DNA junctions can be defined by *Pvu* II or *Psh*A I (*Box* I) sites if precleavage of the transposon–donor boundaries is desired.

(e.g., *Pfu* or *Pwo*) along with *Taq* DNA polymerase will increase the yield of perfectly blunt-ended molecules. If *Taq* DNA polymerase is solely used for PCR, the transposon DNA can be end-repaired by incubation with *T4* DNA polymerase and dNTPs (after purifying *Taq* DNA polymerase away), to yield blunt-ended DNA.

If PCR is done with primers annealing outside of the ME sequences, the PCR product (purified of *Taq* DNA polymerase) can be digested with *Pvu*II or *Psh*AI (*Box* I) to yield perfectly blunt-ended DNA. If there are internal sites for these restriction enzymes within the amplified DNA, one is left with PCR using the ME primer.

The precleaved transposon DNA can be purified if desired by agarose gel electrophoresis followed by extraction using the QIAquick Gel Extraction Kit, Qiagen. Transposon DNA is stored in 10 m*M* Tris-HCl, pH 7.5, 1 m*M* EDTA.

4. *Transposase Storage Buffer*: The choice of an appropriate buffer for storage of the transposase is important because at high concentrations, transposase will aggregate into an inactive form at low salt conditions. For transposase concentrations equal to or less than 50 µg/mL, we typically use 0.05*M* Tris-HCl, pH 7.5, 0.1 m*M* EDTA, 0.1*M* NaCl, 1 m*M* dithiothreitol (DTT), 50% (v/v) glycerol, and 0.1% Triton X-100. For higher concentrations of transposase, the NaCl concentration is kept at 0.5 *M*.
5. *Synaptic Complex Formation Buffer*: Efficient formation of synaptic complexes has been found to occur in a variety of buffers although it is important that they not contain Mg^{2+}. (Mg^{2+} is a required divalent anion for transposition catalysis ***[1]***.) We typically use 0.05*M* Tris-acetate, pH 7.5, 0.15*M* potassium acetate, 4 m*M* spermidine; and 1 m*M* EDTA. If the resulting synaptic complexes are to be stored long term at –20°C for future use, an equal volume of 100% glycerol is added after the complexes are formed. This solution does not freeze at –20°C.
6. *Transposition Reaction Buffer*: Efficient transposition reactions also occur in a variety of buffers. We routinely use the synaptic complex formation buffer adjusted to contain 10 m*M* $MgCl_2$.

3. Methods

3.1. Intermolecular In Vitro Transposition

The typical way to consider transposition is as an intermolecular process as pictured in **Fig. 1A**; that is, as the movement of a DNA sequence from a site on one replicon to a site on a second distinct replicon. It is this process that an investigator uses to tag DNA molecules with a transposon. The in vitro application of this process is primarily used in conjunction with the sequencing of large DNA stretches, by randomly inserting DNA polymerase priming sites within the target sequence (*see* **Note 4**). A typical reaction is performed in 10 μL of a solution that contains 0.1 μ*M* transposase, 0.01 μ*M* transposon DNA (for a transposase-to-transposon ratio of 10) and 0.01 μ*M* target DNA. The important feature of this reaction mixture is that the molar concentration of the transposon should not exceed that of the target. If higher ratios of transposon-to-target are used, one can generate a significant number of double insertions, thus complicating the downstream analysis (*see* **Note 5**). The stated transposase concentration is well in excess and can be reduced somewhat with no significant impact on the reactions. The mixture is incubated for 2 h at 37°C, and then the reaction is terminated by adding 1 μL of 1% sodium dodecyl sulfate (SDS) and heating at 70°C for 10 min to inactivate the transposase. The mixture can be directly introduced into appropriate cells (by, for instance, electroporation) and transposition clones selected; it can be analyzed by electrophoresis; or it can be stored at –20°C for future analysis.

For experiments that involve the transposon tagging of plasmids to be introduced into *E. coli* K12, the products are electroporated (*see* **Subheading 3.4.** for a suggested protocol) into the appropriate host. Cells that have received tagged DNAs are selected by using an antibiotic for which resistance is encoded by the transposon. The plasmid DNAs from individual isolates can then analyzed regarding the location of the insert.

3.2. Intramolecular In Vitro Transposition (Formation of Nested Deletions and Deletions/Inversions)

Intramolecular transposition reactions are performed exactly as described for intermolecular transposition reactions (*see* **Subheading 3.4.**) except that the second target DNA is omitted. The transposon containing clone is in essence both a transposon donor and the target.

3.3. Formation of Synaptic Complexes

Transposon DNA is incubated with hyperactive transposase typically at a 5:1 molar ratio of transposase to transposon in a 20-μL reaction volume for 1 h at 37°C. The concentrations of transposase and DNAs are typically 0.1 μ*M* and

0.02 μ*M* respectively. For smaller transposons (<2 kb) it is preferable to use larger incubation volumes to decrease the abundance of DNA intermolecular complexes (*see* **Note 6**). For these reactions a volume of 400 μL with a DNA concentration of 2.5 μg/mL can be used. The large-volume reactions need to be subsequently concentrated about 20-fold using a 0.05 μm Millipore VM membrane.

Following incubation a sample can be analyzed through a mobility shift analysis. Samples are electrophoresed on a 1% agarose gel in TAE buffer. A transposon-only control and a molecular-weight marker sample will aid in the identification of the desired intramolecular complexes.

3.4. Electroporation

Electroporation of the synaptic complexes can be effected using published procedures ***(20)***.

1. Electrocompetent cells are generally prepared by growing 100 mL of cells to mid-log (0.5–0.6 $A_{600\ nm}$).
2. Cells are chilled, harvested by centrifugation, and washed with ice-cold 10% glycerol three times before suspending them in 250 μL of 10% glycerol. Cells are stored frozen at –70°C.
3. Electroporation is carried out by thawing cells, adding 1 μL of the synaptic complex reagent to 50 μL of thawed cells, and transferring the mixture to a 2.0-mm gap cuvet (*see* **Note 7**).
4. Cells are electroporated at 2500 volts (tau = 5 ms) using an Eppendorf multiporator. Slightly different conditions may be recommended for other brand electroporators.
5. Following electroporation, cells are diluted to 1 mL with Luria Bertani (LB) and incubated at 37°C with aeration for 1 h.

3.5. Selection of Transposition Events

Transposition events are selected by standard microbiological procedures based upon the nature of the selectable marker encoded by the transposon (*see* **Notes 8** and **9**).

3.6. Conclusion

The Tn*5* transposition technology is powerful, robust and simple. The in vitro intermolecular transposition technology allows the generation of transposon insert libraries into any cloned DNA molecule. The inserts can contain virtually any desired sequence, and they will be found at essentially random locations. The in vitro intramolecular transposition technology allows the formation of random nested deletion and nested deletion/inversion libraries though a simple one-step reaction. The electroporation/transposition methodology has already allowed the expansion of transposition mutagenesis schemes to diverse microorganisms that previously were not amenable to genetic analysis.

4. Notes

1. Transposase toxicity: The hyperactive transposase is toxic to cells even in the absence of its specific DNA recognition sequences. Therefore, growth of strains producing hyperactive transposase can present problems. For instance, mutant forms of the transposase gene may arise that inactivate the protein. It is for this reason that purchase of the active enzyme is recommended.
2. DNA length limitations: There are two possible transposon-length limitations on the electroporation/transposition technology that have not been well studied. First, the ability to form intramolecular complexes should be dependent upon the relative concentration of the two ends, which is in turn dependent upon the chain length between the ends. We have found that molecules less than 5 kb long work effectively, but we have not tested longer molecules.

   The second event that is length dependent is electroporation. It has been shown, for example, that smaller plasmid DNAs (e.g., 2.9 kb) are more efficiently electroporated into *E. coli* DH5α than larger DNAs (approx 50% reduction in efficiency for plasmids twice this size) ***(21)***. It is expected that larger complexes (due to the use of larger transposons) will likewise be subjected to size discrimination in many hosts, with the added complication that a nucleoprotein complex, and not naked DNA, is being electroporated into the cells. To what extent complexes are size biased has not been carefully studied.
3. *Pvu*II star activities: It is most convenient to use prereleased transposon DNA for the electroporation/transposition technology. Frequently the prereleased transposon DNA is generated through the use of *Pvu*II. Use of too high a concentration of *Pvu*II and/or failure to adequately purify substrate DNA can result in extra enzyme cleavage products during the preparation of precleaved transposon DNA.
4. Target choice randomness: The randomness in target choice is an important consideration especially for the strictly in vitro technologies that we have discussed; that is, the intermolecular transposition is primarily used for introducing primer binding sites in conjunction with DNA sequencing, and the transposon must go in with sufficient randomness to permit sequence coverage. Likewise, randomness is an important feature when making nested deletions through the intramolecular transposition protocol. All transposition systems are likely to display some sequence biases in regards to target selection. A saturation transposition target-selection analysis for the Tn*5* system has demonstrated some sequence preferences ***(22)***. However, even in this study, approximately 10% of the possible target sites were hit. In **Fig. 4** we present example results of an in vitro intermolecular transposition experiment in which 55 independent inserts were generated in a 7775 bp target. Fifty-four separate sites were hit (one site was hit twice), and there appears to be no obvious base-composition bias for the insert sites ***(23)***.
5. Multiple insertion events following electroporation of transposition complexes: It is possible, although unlikely, that multiple transposon inserts could be generated by the electroporation of more than one complex into a cell. The number of inserts can be determined by a Southern blot assay. Using this procedure we analyzed 14 independent isolates and found only 1 isolate had more than one insert ***(7)***.

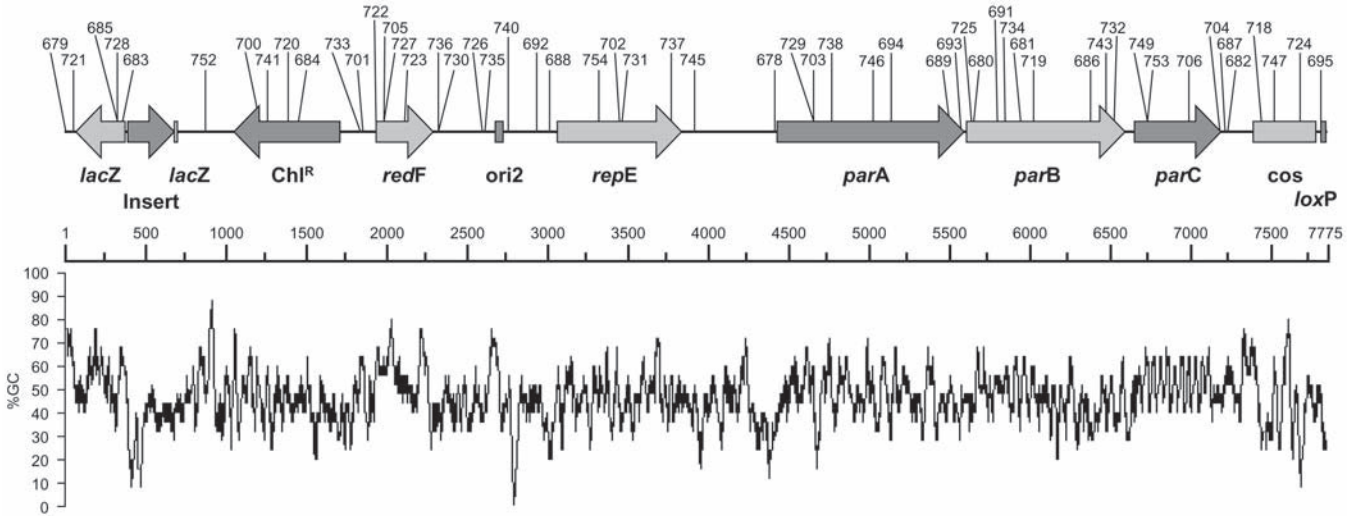


Fig. 4. Insertion data for 55 inserts generated in a 7775-bp target. Inserts were generated through the in vitro intermolecular reaction, the resulting products were introduced into cells, and the insertion sites were determined by DNA sequencing. Only two inserts (685 and 728) landed at the same site. Below the insert map is a GC content map using a 25-bp window. The insert sites demonstrated no obvious base-composition bias. The data was kindly provided by R. Meis and is similar to a figure published previously ***(23)***.

Simple ways to avoid this problem include performing the electroporation experiment at submaximal levels of complex and using more concentrated cells.

6. Excess transposon DNA and the formation of intermolecular synaptic complexes: Scientists frequently add excess reagents to increase the yield of various protocols. With regard to the electroporation/transposition technology, use of high-transposon DNA concentrations can lead to the formation of intermolecular instead of intramolecular transposase–transposon complexes, in which two ends from different transposon molecules are complexed together. While the result of electroporating such intermolecular complexes has not been deliberately studied, it is likely that they will yield double-strand breaks instead of insertion events. The abundance of intermolecular complexes can be estimated by prior agarose gel electrophoresis of the complex mixture. We have found that electroporation/transposition can be performed successfully even in the presence of some intermolecular complexes, as long as intramolecular complexes are also present and evident on gel analysis.
7. Electroporation of greater volume samples: Our recommendation for sample electroporation is to use 1 μL of sample. In order to increase the yield of cells containing transposon inserts, it may be desirable to use larger transposition-complex sample volumes, but this method will lead to arching due to the salt content. To increase the transposition-complex sample volume we recommend dialysis of the sample versus 5 m*M* Tris-HCl, pH 7.9, 10% glycerol. Dialysis can be accomplished by floating a 25-mm Millipore disk (0.05 μm, VM) on the surface of 50 mL of buffer and applying a drop of up to 40 μL sample to the disk for 30 min. Glycerol is necessary if the sample is to be stored at –20°C. If the sample is to be electroporated immediately, no glycerol is necessary.
8. Use of too high a concentration of antibiotic: Many standard media contain antibiotic levels suitable for selection of multicopy antibiotic-resistant genes. The electroporation/transposition technology generates monocopy inserts, and thus standard recipes may contain excess antibiotic. This problem can be addressed by predetermining a minimum inhibitory concentration for the relevant antibiotic and the host strain, and adjusting the selection conditions accordingly.
9. Background plasmid clones after electroporation: Contamination of *Pvu*II or *Psh*AI digestion products with uncut plasmid DNA can result in some of the antibiotic-resistant clones arising from plasmid transformation instead of transposon integration, if the target cells support DNA replication of the plasmid from which the transposon was derived. Overdigestion with restriction enzymes followed by careful gel purification should alleviate this problem.

   For PCR generated transposons, intact plasmid contamination is generally not an issue because only small amounts of plasmid template are used to amplify the transposon sequence. However, digestion of the plasmid DNA in a region outside of the transposon proper can both improve PCR performance and further lower the background. To completely eliminate any potential for background clones, the PCR reaction product can be digested with *Dpn*I which will cleave any N6A-methylated DNA template at GATC, which would be produced in most

E. coli strains because they are Dam+. PCR products, not being methylated, are resistant to digestion.

Acknowledgments

Research in the laboratory of WSR was supported by grants from the National Institute of General Medical Sciences (GM50692) and the University of Wisconsin University–Industry Relations program (1350538). We thank Ron Meis for providing us with data.

References

1. Goryshin, I. Y. and Reznikoff, W. S. (1998) Tn*5 in vitro* transposition. *J. Biol. Chem.* **273,** 7367–7374.
2. Devine, S. E. and Boeke, J. D. (1994) Efficient integration of artificial transposons into plasmid targets in vitro: a useful tool for DNA mapping, sequencing and genetic analysis. *Nucleic Acids Res.* **22,** 3765–3772.
3. Gwinn, M. L., Stellwagen, A. E., Craig, N. L., Tomb, J. F., and Smith, H. O. (1997) *In vitro* Tn*7* mutagenesis of *Haemophilus influenzae* Rd and characterization of the role of *atpA* in transformation. *J. Bact.* **179,** 7315–7320.
4. Haapa, S., Taira, S., Heikkinen, E., and Salvilahti, H. (1999) An efficient and accurate integration of mini-Mu transposons *in vitro*: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res.* **27,** 2777–2784.
5. Akerley, B. J., Rubin, E. J., Camilli, A., Lampe, D. J., Robertson, H. M., and Mekalanos, J. J. (1998) Systematic identification of essential genes by *in vitro mariner* mutagenesis. *Proc. Natl. Acad. Sci. USA* **95,** 8927–8932.
6. Griffin, T. J. IV, Parsons, L., Leschziner, A. E., DeVost, J., Derbyshire, K. M., and Grindley, N. D. F. (1999) *In vitro* transposition of Tn*552*: a tool for DNA sequencing and mutagenesis. *Nucleic Acids Res.* **27,** 3859–3865.
7. Goryshin, I. Y., Jendrisak, J., Hoffman, L. M., Meis, R., and Reznikoff, W. S. (2000) Insertional transposon mutagenesis by electroporation of released Tn*5* transposition complexes. *Nature Biotech.* **18,** 97–100.
8. Hoffman, L. M., Jendrisak, J. J., Meis, R. J., Goryshin, I. Y., and Reznikoff, W. S. (2000) Transposome insertional mutagenesis and direct sequencing of microbial genomics. *Genetica* **108,** 19–24.
9. O'Mullan, P. J. (2000) Direct sequencing of BAC clones without subcloning or primer walking using EZ::TN transposon tools. *Epicentre Forum* **7,** 1–3.
10. Gehring, A. M., Nodwell, J. R., Beverley, S. M., and Losick, R. (2000) Genomewide insertional mutagenesis in *Streptomyces coelicolor* reveals additional genes involved in morphological differentiation. *Proc. Natl. Acad. Sci. USA* **97,** 9642–9647.
11. Judson, N. and Mekalanos, J. J. (2000) TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. *Nature Biotech.* **18,** 740–745.
12. York, D., Welch, J., Goryshin, I. Y., and Reznikoff, W. S. (1998) Simple and efficient generation *in vitro* of nested deletions and inversions: Tn*5* intramolecular transposition. *Nucleic Acids Res.* **26,** 1927–1933.

13. Reznikoff, W. S., Bhasin, A., Davies, D. R., Goryshin, I. Y., Naumann, T., Rayment, I., et al. (1999) Tn*5*: A molecular window on transposition. *Biochem. Biophys. Res. Commun.* **266,** 729–734.
14. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., and Rayment, I. (2000) Three-dimensional structure on the Tn*5* synaptic complex transposition intermediate. *Science* **289,** 77–85.
15. Weinreich, M. D., Gasch, A., and Reznikoff, W. S. (1994) Evidence that the *cis* preference of the Tn*5* transposase is caused by non-productive multimerization. *Genes Dev.* **8,** 2363–2374.
16. Zhou, M. and Reznikoff, W. S. (1997) Tn*5* transposase mutants that alter DNA binding specificity. *J. Mol. Biology* **271,** 362–373.
17. Zhou, M., Bhasin, A., and Reznikoff, W. S. (1998) Molecular genetic analysis of transposase-end sequence recognition: Cooperation of three adjacent base pairs in specific interaction with a mutant Tn*5* transposase. *J. Mol. Biology* **276,** 913–925.
18. Bhasin, A., Goryshin, I. Y., Steiniger-White, M., York, D., and Reznikoff, W. S. (2000) Characterization of a Tn*5* pre-cleavage synaptic complex. *J. Mol. Biology* **302,** 49–63.
19. Johnson, R. C. and Reznikoff, W. S. (1983) DNA sequences at the ends of transposon Tn*5* required for transposition. *Nature* **304,** 280–282.
20. Birren, B., Mancino, V., and Shizuya, H. Bacterial Artificial Chromosomes. In *Genome Analysis: A Laboratory Manual* (Birren, B., Green, E. D., Klapholz, S., Meyers, R. M., Riethman, H., and Roskams, J., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, vol. 3, 1999, pp. 274–279.
21. Siguret, V., Ribba, A. S., Cherel, G., Meyer, D., and Pietu, G. (1994) Effect of plasmid size on transformation efficiency by electroporation of *Escherichia coli* DH5 alpha. *Biotechniques* **16,** 422–426.
22. Goryshin, I. Y., Miller, J. A., Kil, Y. V., Lanzov, V. A., and Reznikoff, W. S. (1998) Tn*5*/IS*50* target recognition. *Proc. Natl. Acad. Sci. USA* **95,** 10716–10721.
23. Meis, R. J. (2000) EZ::TN transposon insertions into target DNA *in vitro* are highly random. *Epicentre Forum* **7,** 5.

8

Gene Targeting in *Drosophila*

Gregory B. Gloor

Summary

DNA double-strand breaks provide a powerful means to modify the genome. This chapter describes how to generate and use these breaks to target specific sequences, or other modifications to the *Drosophila* genome. Both *P* element dependent gene conversion, in which the chromosomal DNA is broken, and the Rong and Golic gene-targeting technique, in which the targeting vector contains the DNA break are explained. The strengths and limitations of both methods are presented so that the user can choose the appropriate method for their particular situation. The efficiency of both methods depends upon the genomic location being modified, although few, if any, genomic locations are refractory to either method. It cannot be emphasized strongly enough that the investigator should be prepared to invest sufficient time into setting up and running these experiments properly.

Key Words: Gene targeting; DNA breaks; genome modification; gene conversion; I-*Sce* endonuclease; *P* element.

1. Introduction

The recent determination of the complete genome of *Drosophila melanogaster* presents us with an unprecedented genetic opportunity ***(1)***. *Drosophila* is an extremely useful model organism for the study of processes in a complex multicellular organism that relate to developmental biology, signal transduction, and genetic regulation ***(2,3)***.

Until recently one of the major impediments facing *Drosophila* biologists has been their inability to target DNA sequences to a specific locus. Such alterations have long been possible in the yeast, *Saccharomyces cerevisiae*, and in the mouse. This situation has now changed for the better. An investigator in *Drosophila* can now choose between two different methods of gene targeting. Both methods utilize the natural tendency of the cell to repair a DNA double-strand bread. Such breaks are lethal to the cell or can cause chromosome rearrangements if left unrepaired or if repaired inappropriately. Both methods

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

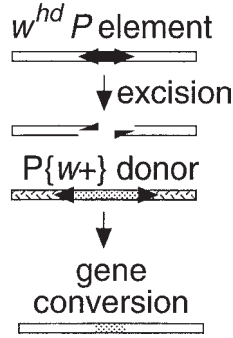


Fig. 1. *P* element induced gene conversion in *Drosophila*. The w^{hd} *P* element in a nonautonomous element inserted in exon 6 of the *white* gene. This insertion results in a white-eyed phenotype, and precise loss of the insertion reverts the eye color to the wild-type red color. The donor element is a *white* gene inserted into a *P* element transformation vector and inserted in the genome at an ectopic site. This *white* gene is modified by single base alterations that change the restriction map without affecting expression of the gene. Excision of the *P* element, induced by addition of P transposase, causes a double strand break that has 17 nt 3' extended ends composed of part of the *P* element inverted termini. The *white* gene sequence at the break site is used to conduct a homology search. When the ends find the homologous *white* gene, they invade it, and serve as primers to initiate DNA synthesis. The result is a gene conversion product in which the *white* gene sequence at the break site is replaced by sequence copied from the ectopic template.

produce targeted genetic alterations at a rate that permits the investigator to identify the events through a phenotypic screen or in some cases by a molecular screen.

The first method requires a *P* element insertion in or near the gene of interest ***(4,5)***. As shown in **Fig. 1**, excision of the *P* element results in a break in the chromosome. The cell senses the double-strand break, and double-strand-break repair factors are recruited to the site. In *S. cerevisiae*, and presumably in other eukaryotes, the ends are processed to leave single-strand 3' extended ends ***(6)***. The single-strand DNA at the broken ends is used to conduct a genome-wide search for a homologous donor sequence. This sequence can be located on the sister chromatid, the homolog, an ectopic site, or on a plasmid injected into the cell ***(7–11)***. The single-strand 3' ends base-pair with their complementary sequences, displacing the identical sequences at the homologous site. This invading 3' end becomes a primer for DNA synthesis. Replication results in complete reconstitution of the sequence that was lost at the break site, without the donor site being modified. Single-base mismatches, insertions of at least 20 kbp, and deletions can be introduced by this process ***(5,12)***. The double-

strand break made by *P* element excision can be up to 10 kbp distant from the desired modification site (Gloor, G. B., unpublished). In general, 0.1% to 1% of the double-strand breaks are repaired by this method, provided an appropriate donor molecule is present. Application of this at the *white* ***(5)***, *yellow* ***(13)***, *dpp* ***(14)***, *bithorax* ***(15)***, *suppressor of forked* ***(16)***, and *forked* ***(17)*** loci has produced specific in vivo modifications. I will refer to this approach as gene conversion for the remainder of this chapter.

The standard ends-in method of gene targeting that has been used with such tremendous effect in *S. cerevisiae* and mice can now be used in *Drosophila*, thanks to a clever in vivo manipulation invented by Rong and Golic ***(18)***. They recognized that the major impediment to classical gene targeting in *Drosophila* was the inability to introduce the appropriate linear DNA fragment into fly cells. Individual cells once removed from the embryo cannot be grown in culture and reintroduced to form the germ line. Therefore, an embryonic stem cell culture does not exist for *Drosophila*. This situation makes it very difficult to introduce DNA into the *Drosophila* germline. All existing protocols for doing so call for microinjection into the syncytial oocyte at the place where the germ cells will take up residence. It is currently not practical to inject a gene-targeting vector into a sufficient number of embryos in this manner. However, the *Drosophila* germline can be transformed by *P* element-mediated transposition from an injected plasmid into a random genomic site ***(19)***. The major insight of Rong and Golic was to realize that *P* element integrants could be altered in vivo into a gene-targeting vector by the expression of the appropriate enzymes. I will refer to this as *gene targeting* for the remainder of this article.

An outline of Rong and Golic's gene-targeting method is presented in **Fig. 2**. In this method, a *white* gene controlled by a truncated hsp70 promoter and flanked by directly repeated FRT sites is integrated into the genome via *P* element-mediated transformation. A fragment of the gene of interest is cloned in this vector between the FRT sites without disrupting *white* gene expression. An *I-Sce*1 endonuclease recognition sequence is inserted into the middle of the sequence of interest. Expression of the FLP recombinase results in the release of a circular DNA molecule, while expression of the *I-Sce*1 endonuclease results in this circular molecule being linearized, thus converting it to an ends-in gene-targeting vector ***(20)***. Targeted insertion of this molecule via double-strand break repair is recognized by linkage of the *white* gene to the appropriate genetic locus. It is not currently possible to produce deletion alleles of the targeted locus because every insertion results in the duplication of the sequence that is in the targeting vector. However, both duplicated sequences could be made nonfunctional if the targeting vector contains a 5' and 3' truncation of the locus of interest. In this case, one of the duplicated segments would have a 5' truncation, and the other would have a 3' truncation. It is also possible

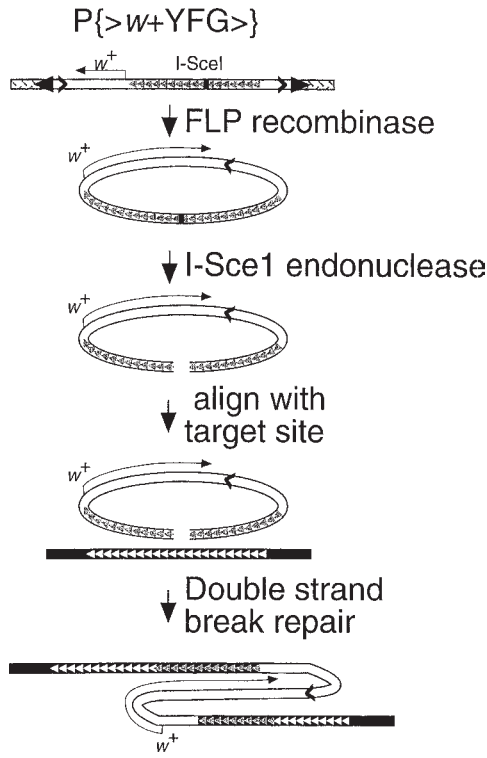


Fig. 2. Gene targeting in *Drosophila*. The gene-targeting vector is a modified P{>w^{hs}>} in which a *Not*1 cloning site has been introduced outside of the *white* gene coding sequence but between the FLP recombinase target sites (FRTs). This vector is referred to as P{>w^{hs}·*N*>}. The gene of interest (YFG) is cloned into this site, and an *I-Sce*1 endonuclease recognition sequence is inserted into the middle of YFG. This recognition sequence should be at least 1 kbp from either end of the YFG sequence. This vector is transformed into the *Drosophila* germline by *P* element-mediated transformation. The sequence between the FRT sites is excised from the chromosome upon induction of the FLP recombinase enzyme. Induction of the *I-Sce*1 endonuclease results in linearization of this excised DNA circle. This DNA fragment is now used for gene targeting via the ends-in targeting method. The end result is a duplication of the target gene with an integrated copy of the *white* gene.

to introduce point mutations that cause frameshift mutations in the 5' and 3' sequences. In this way, each of the duplicated sequences at the target locus could be made nonfunctional ***(18)***.

This chapter will describe how to generate both types of gene-targeting events. Both methods can be used to alter the genome, and each method has its own particular strengths and weaknesses. *P* element-dependent gene conver-

sion occurs at a sufficiently high frequency that physical assays for the products are feasible. In at least some instances, the frequency is high enough to permit gene conversion from DNA injected into embryos ***(10)***. The major limitation of this technique is that the site to be modified must be close to an existing *P* element insertion. Several studies have shown that the probability of a point mutation being introduced by this method varies exponentially with its distance from the site of *P* element excision. This probability (P) can be estimated by the formula $P_n = x^n$, where *n* is the distance of the point mutation from the *P* element excision site, and *x* is equal to 0.99873 ***(7)***. However, this formula underestimates the probability at distant sites. For example, a site 2 kb distant from the *P* element excision site is converted at about 10–15% of the frequency of a site within a few base pairs of the *P* element ***(5)***, but the formula predicts a conversion frequency of approx 8%.

One important additional factor is the mobility of the *P* element. The *P* element must excise frequently to ensure the production of sufficient double-strand breaks. In contrast, the gene-targeting procedure of Rong and Golic should in principle work at any site in the genome ***(18,21)***.

This method has two disadvantages: First, the frequency of gene targets is somewhat lower than is observed with *P* element-dependent gene conversion. It cannot be emphasized too strongly that the investigator should be prepared to invest enough time into setting up the experiment properly. Rong and Golic observed a frequency of one event per 500 gametes (0.2%) with the *X*-linked *yellow* locus ***(18)***, and a frequency of one event per approx 30,000 gametes with an autosomal locus ***(21)***. Second, as shown in **Fig. 3**, this method generates duplications of the targeted gene with the marker gene in between the duplicated region. Thus, this method does not make deletion alleles, but it is possible to generate a null allele by truncating both duplicated segments, or by introducing point mutations into the duplicated sequences. The ability to perform ends-out targeting, which is less efficient, would negate this problem.

The gene-conversion and gene-targeting methods share several common features. First, the regions of homology on each side of the double-strand break should be at least 1 kb. Several studies of the amount of homology required for double-strand break repair in *Drosophila* and other organisms suggest that this size is the prudent lower limit ***(22)***. Near optimal double-strand break repair should work with about 500 bp of homology, but this has only been tested at a small number of loci. Second, the homologous sequence flanking the double-strand break should be isogenic with the target site if possible ***(7)*** with the exception of the desired point mutation. Each individual mismatch between the interacting DNA molecules will decrease the rate of targeting or conversion by a slight amount. Third, both methods will work most conveniently if the site to be modified, the *P* element vector, and the inducer genes (transposase

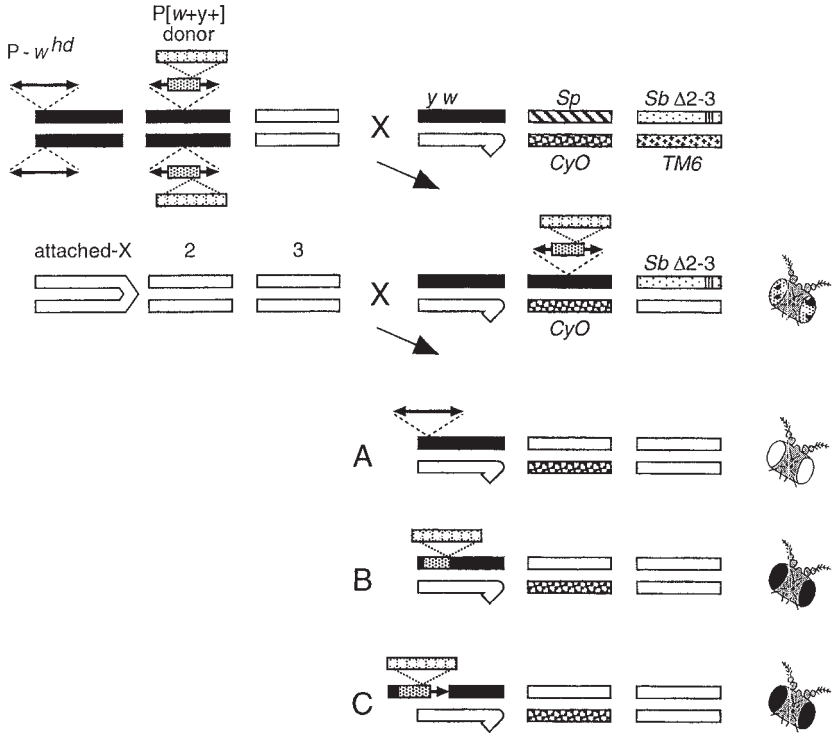


Fig. 3. Genetic screen for gene conversion in *Drosophila*. In the parental mating (top), females homozygous for the w^{hd} *P* element and the P{w+ y+} donor element are mated to males carrying a *CyO* balancer chromosome and chromosome carrying the *P* element transposase source, Δ2–3(99B), that is marked with a dominant *Sb* allele. The F_1 male progeny from this cross contain the w^{hd} *P* element, the P{w+ y+} donor element heterozygous with the *CyO* balancer chromosome, and the transposase-carrying third chromosome. These males show a mosaic eye-color phenotype because of the mobilization of the w^{hd} *P* element and the donor *P* element. These F_1 males are mated individually with 3–5 attached-*X* females, and the F_2 male progeny are collected that contain the *CyO* balancer chromosome and that lack the *Sb* marker. The *X* chromosome in these F_2 males is inherited from their father, and these are the chromosomes in which double-strand break repair has occurred. There are two eye-color phenotypes observed in these flies. The first is a non-wild-type eye color, which can range from white to a dull red. Flies with a white eye color are not gene conversion products and are discarded. Flies with an intermediate eye color most commonly result from a transposition of the P{w+ y+} donor element to a new location in the genome. Flies with this eye-color phenotype can also be discarded. Flies with a wild-type eye color are kept for molecular analysis and are mated individually to attached-*X* virgin females to establish a stock.

source for gene conversion, fragment length polymorphism recombinase and *I-Sce*I endonuclease for gene targeting) are each located on a separate chromosome. I will outline methods for gene conversion to the *white* locus on the *X* chromosome, and for gene targeting to the *Drosophila* homologue of the *Ku80* gene on the second chromosome as examples.

2. Materials

Water for the solutions should be of the best quality available. All chemicals are Life Technologies Industries Ultrapure, Sigma Molecular Biology grade, or equivalent unless noted otherwise.

2.1. Genetic Analysis

1. Dissecting microscope.
2. CO_2 anesthetizing apparatus.
3. Fly food is the standard cornmeal, sucrose, and agar food.
4. Fly strains:
   a. *y* w^{hd} *f* ; P{w+ y+} ; +
   b. *y w* ; *CyO*/*Sp* ; *Sb* Δ2-3(99B)*/TM6*
   c. *C(1)DX, y w f* ; + ; + (attached-*X*)
   d. *FM7c* , w^{a}*/Df(1)N19* ; + ; *P*{*70FLP*} *P*{*70I-Sce*1}*/TM6B, Tb*
   e. *P*{w+ YFG^{Sce}); + ; *TM3, Ser*/*Sb*
   f. *w m f* ; + ; +

2.2. DNA Preparation

1. 1.5-mL polypropylene tubes (Eppendorf).
2. Disposable plastic pestles (Kontes Pellet Pestle, Fisher).
3. Proteinase K (Promega), 20 mg/mL in water is kept as a frozen stock at –20°C.
4. Squashing buffer: 10 m*M* Tris-HCl, pH 7.8, 25 m*M* NaCl, 0.1 m*M* EDTA, 0.1% Triton-X 100. This buffer may be kept at room temperature for several months. To minimize contamination it is important that the squashing buffer be made up with solutions and glassware that have never been in contact with plasmids or PCR products. Proteinase K is diluted 1:50 in squashing buffer just before use.
5. Refrigerator for 4°C incubation.
6. Dry bath incubators at 50°C and 95°C. It is important that the entire wall of the tube be in complete contact with the incubator.
7. 0.1*N* HCl.

2.3. PCR

1. Taq polymerase (Perkin-Elmer, or Amersham Pharmacia Biotech) at 5 units/μL.
2. 10X PCR buffer: 200 m*M* Tris-HCl pH 8.6, 500 m*M* KCl, 0.1% Triton X-100
3. 50 m*M* $MgCl_2$.

4. The four deoxyribonucleotide triphosphates (dNTPs) are purchased as ultrapure solutions from Pharmacia at a concentration of 100 m*M*. Equal amounts of each dNTP are mixed and then frozen as 50 µL aliquots at –70°C. Any dNTP mix not used in 9 mo is discarded. Adding 950 µL of distilled water before use dilutes each dNTP solution. The final concentration is 1.25 m*M* of each dNTP in the mix. This diluted dNTP mix can be stored at –20°C for 4–8 wk before use.
5. Oligonucleotides are purchased as deblocked crude pellets and are used without further purification. They are resuspended at a final concentration of 200 µ*M* in water and kept at –70°C, where they are stable for at least 1 yr. Working stocks of oligonucleotide primers are diluted to a concentration of 20 µ*M* and kept at –20°C until use. They are stable for several months under these conditions. The 10 oligonucleotides required for these experiments are:
   a. GGTTGTCGTACCTCTCATGG ef+
   b. ACAGCGAAAGAGCAACTACG Hi–
   c. GCAGCCTTCCACTGCGAATC P310
   d. GGTTGGCGCGATCTCGCGCTCT I+
   e. AAGAGATAGCGGACGCAGCG CasRt2
   f. GAGTGTCGTATTGAGTCTGAG 20108
   g. GAGAGAGCAATAGTACAGAGA vOutA
   h. TTTACTGTCAAGTAGACCCATA vOutB
   i. CTGCTCAAAGATGCTGGGAA KuA
   j. AAGTAAGTTGGCAACGCGGT KuB
6. Light mineral oil.
7. Eight-well, thin-walled strip tubes (ABGene, Epsom)

2.4. Restriction Endonuclease Digestion

1. 0.1*M* $MgCl_2$, made up in distilled water and autoclaved.
2. 0.25*M* Tris-HCl, pH 7.8, made up in distilled water and autoclaved.
3. 5*M* NaCl, made up in distilled water and autoclaved.
4. 10X agarose gel loading buffer: 10% Ficoll 70, 1% sodium dodecyl sulfate (SDS), 20 m*M* EDTA, 0.1% bromophenol blue.

3. Methods

3.1. Gene Conversion From an Integrated Donor

Figure 4 shows the optimal mating scheme for gene conversion to an *X*-linked site. In this scheme, the *P* element that is excised to make the double-strand break is inserted in a gene on the *X* chromosome. In this instance it is in the *white* locus, and its insertion causes a null phenotype; this *white* allele will be referred to as w^{hd}. Precise loss of the *P* element, which occurs when gene conversion repairs the double-strand break, results in reversion of the phenotype (*see* **Note 1**). The second chromosome carries the donor *P* element that contains white+ and yellow+ alleles. The yellow+ allele is embedded within the

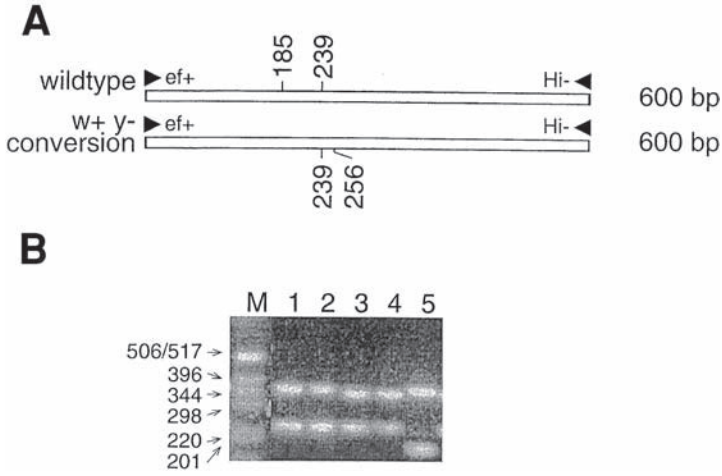


Fig. 4. Confirmation of gene conversion. The P{w+ y+} donor element contains single-base alterations that permit identification of gene conversion at the *white* locus. A portion of the *white* locus that flanks the site of the DNA break is amplified as in **Subheading 3.4.** and subsequently digested with *Hae*III. **(A)** shows the *Hae*III restriction map of the PCR product amplified from a wild-type *white* gene, and the *Hae*III restriction map of the same region amplified from a *white* gene in which this sequence was replaced by gene conversion by copying the corresponding sequence from the P{w+ y+} donor. In this case *Hae*III digestion of the wild-type sequence would produce fragments of 361, 185, and 54 base pairs. Digestion of the converted sequence would produce fragments of 344, 239, and 17 base pairs. It is also possible to replace only part of the wild-type sequence by gene conversion. In such cases restriction digestion fragments of 361 and 239 base pairs or fragments of 344, 185, 54, and 17 base pairs are obtained. **(B)** shows the restriction digestion patterns of the two such gene conversion events and of the wild-type *white* gene sequence. In this example the *white* gene sequence was amplified by PCR with the ef+/Hi– primers, digested with *Hae*III, and run on a 3% agarose gel. *Lane M* contains a molecular weight marker. *Lanes 1* and *2* contain samples in which the *white* gene sequence is converted for the *Hae*III site at position 185, but not position 256 (fragment sizes of 361 and 239 base pairs). *Lanes 3* and *4* contain samples converted for both *Hae*III sites (fragment sizes of 344 and 239 base pairs). *Lane 5* contains a sample of wild-type *white* gene sequence (fragment sizes of 361 and 185 base pairs).

white gene with >4 kb of homology to the *white* gene on one side, and 763 bp of homology to the *white* gene on the other. The transposase source, Δ2-3(99B) is located on the third chromosome.

The parental cross mates a female fly of the genotype *y* w^{hd} ; P{w+ y+}/*CyO* to a male fly of the genotype *w* ; *CyO*/*Sp* ; *Sb* Δ2-3(99B)/*TM6B*. The male progeny of this cross with the genotype w^{hd}; P{w+ y+}/*CyO* ; *Sb* Δ2-3/+ are

crossed individually to 3–5 attached-*X*, *y w f* females. It is important that this cross be done with individual male flies, as many of the desired gene conversion events occur in the germline before meiosis. Many sibling flies with identical gene-conversion products are found in the progeny of an individual male because of this. We routinely set up from 50 to 100 individual matings to ensure enough progeny flies for analysis.

The F_2 progeny potentially contain the gene conversion events. Only male F_2 progeny are examined because the *X* chromosome is inherited from the father in this cross. In this particular cross gene conversions may be recognized because of reversion of the w^{hd} allele to wild type (*see* **Note 2**). The F_2 progeny are further sorted by the presence of the *CyO* and *Sb* markers. Only those progeny that are Sb+ are kept, ensuring that the transposase source is absent in these flies, and therefore preventing further transposition of the w^{hd} *P* element or of the P{w+ y+} donor element. We further keep only those F_2 progeny that contain the *CyO* chromosome. The presence of the *CyO* chromosome in the F_2 progeny indicates that the chromosome that originally carried the P{w+ y+} donor element has been lost by segregation.

These F_2 progeny are then individually mated to 3–5 attached-*X*, *y w f* females. The F_3 progeny are examined for the segregation of the w+ phenotype or the w+ y+ phenotypes. Only those lines that show linkage between the *X* chromosome and the w+ or w+ y+ phenotypes are kept for further analysis (*see* **Note 3**).

Several of the F_3 male progeny are mated to attached-*X*, *y w f* females to generate a continuing line of flies.

3.2. DNA Preparation

DNA is prepared from four of the remaining F_3 flies for analysis by PCR. We use a slight modification of the protocol published previously. The major modification is that the flies are macerated in a larger volume of squashing buffer with a plastic pestle. Three main points must be emphasized: First, it is important to minimize the time that the DNA sample is held at 95°C, because the DNA sample rapidly deteriorates at this temperature. I find that incubation at this temperature for 3 min in a thick-walled Eppendorf tube is optimal. Second, it is imperative that the entire tube be raised to 95°C to minimize the carryover of Proteinase K activity. We accomplish this by plunging the entire tube in a sand bath kept at 95°C. Third, the plastic pestles are washed for at least 12 h in 0.1*N* HCl after use. They can be used after they are rinsed in H_2O.

1. Place individual male flies in a 1.5-mL polypropylene tube. Place the tubes on ice or at –20°C until the flies stop moving.
2. Add 100 µL of squashing buffer to the tube.
3. Macerate the flies with an acid-washed, disposable pestle until all the fly parts can be taken up with a yellow pipet tip (200 µL vol tip).

4. Transfer the macerated fly to a 0.5-mL Eppendorf tube.
5. Incubate at 50°C for at least 30 min, then at 95°C for 3 min.
6. Briefly centrifuge the tube to bring down any condensate. These preparations can be stored for several months at 4°C. Do not freeze. One μL of sample contains sufficient DNA to produce a strong signal in a polymerase chain reaction (PCR) reaction after 20–25 amplification cycles.

3.3. PCR Screen for Donor P *Element Ends*

This PCR screen tests for the presence of donor *P* element ends. Lines carrying a putative gene conversion can contain donor *P* element ends for three reasons: First, because a gene conversion and a transposition of the donor *P* element to the *X* chromosome occurred at the same time. These elements are genetically invisible because the w+ phenotype conferred by the presence of the *P*{w+ y+} element is masked by wild-type expression of the endogenous *white* gene that was repaired by gene conversion. Second, the donor *P* element could have transposed to the *X* chromosome and inserted into a site at which the w+ allele was expressed at or near wild type levels. Third, one of the donor *P* element ends can be copied by gene conversion into the *white* locus. The presence of donor *P* element ends is carried out by two separate PCR reactions: The left-end PCR reaction is carried out with the 20108/CaRt2 primer combination; and the right-end PCR reaction is carried out with the I+/3645 primer combination. It is important to analyze several flies from each line, we routinely analyze four flies per line.

PCR reactions are performed in 20-μL reactions in 8 × 400 μL thin-walled strip tubes. Use of these tubes allows us to use a multichannel pipettor to set up the reactions. PCR reactions are set up as bipartite reactions. Multiply the volumes of each component by the number of sample to be assessed.

1. Assemble Reaction Mix A as follows:

   4.3 μL H_2O.
   3.7 μL 5 m*M* dNTP mix.
   0.7 μL 50 m*M* $MgCl_2$.
   1.16 μL primer 20108.
   1.16 μL primer CaRt2.
   Label as Mix A and aliquot 8 samples of equal volume into an 8 × 400-μL thin-walled strip tube and hold on ice until ready for use.

2. Label a new tube as B and add the following components for each sample to be assessed:

   8.9 μL H_2O.
   2.4 μL 10X PCR Buffer.
   0.24 μL *Taq* DNA polymerase.
   Aliquot 8 samples of equal volume into an 8 × 400 μL thin-walled strip tube and hold on ice until ready for use.

3. Aliquot 30 μL of light mineral oil into each well of the 8 × 400 μL strip tubes with a repeating pipet.
4. Add 1 μL of DNA sample, prepared as in **Subheading 3.2.**, under the oil layer.
5. Add 10 μL of Reaction Mix A to each sample tube using an 8-channel multichannel pipettor.
6. Place the tubes prepared as in **step 5** into the PCR machine with the block preheated to 85°C. Incubate for at least 1 min.
7. Add 10 μL of Reaction Mix B into each sample tube using an 8-channel multichannel pipettor.
8. The first PCR cycle is carried out with the following temperature profile:

   Denature at 95°C for 1 min.
   Anneal at 65°C for 1 min.
   Extend at 72°C for 1 min.

   The annealing temperature in each subsequent cycle is reduced by 1°C until an annealing temperature of 55°C is reached.

   Amplify for a further 25 cycles with the following temperature profile:

   Denature at 95°C for 1 min.
   Anneal at 55°C for 1 min.
   Extend at 72°C for 1 min.
9. Move the completed PCR reactions to another room for all subsequent analysis.
10. Analyze the products by running 10 μL of each PCR reaction on a 1.4% agarose gel. In this case, we are looking for the absence of an amplified product. We always include at least two reactions in each set of genomic DNA containing the donor *P* element as positive controls.

3.4. PCR Screen for Gene Conversion

The P{w+ y+} elements that are the donors in the gene-conversion reaction carry a DNA sequence modification that results in the loss of a *Hae*III site at the site of *P* element excision. We use a restriction digest of a third PCR reaction as a screen for the loss of this *Hae*III site as an indicator of gene conversion. The primers for this reaction are ef+/Hi– if the gene conversion has a y– phenotype. We amplify with wRR/yR if the gene conversion has a y+ phenotype (*see* **Note 4**). DNA samples that did not amplify in the previous section are used to set up these PCR reactions.

1. The reactions are assembled identically to those in **Subheading 3.3.**, except that the different oligonucleotide primers are used.
2. The reaction conditions are identical to those in the previous section.
3. Move the completed PCR reactions to another room for all subsequent analysis.
4. A digestion mix is assembled with the following additions for each sample to be analyzed:
   a. 1.5 μL of 100 m*M* $MgCl_2$.
   b. 1.5 μL of 250 m*M* Tris-HCl, pH 7.8.

c. 1 U of *Hae*III.
d. Water to 5 µL.
e. Add 5 µL of the digestion mix under the oil to each sample to be digested. Mix by pipetting up and down several times.

5. Incubate at 37°C for at least 1 h.
6. Analyze by running 15 µL of the digested sample on a 3% agarose gel.

3.5. Gene Targeting From an Integrated Donor

We are interested in disrupting the *Drosophila* homolog of the *Ku80* gene ***(1)*** (*see* **Note 5**). A targeting vector was made that will result in a duplication structure that has a deletion of the first putative exon and a substantial portion of the last exon. The construct has 958 bp and 660 bp of *Ku80* sequence flanking the *I-Sce*1 cut site. This construct will serve as a good test of the targeting method, because it has sequence homology at or near the lower end of the amount required. We cannot introduce more flanking sequence without regenerating the putative *Ku80* gene in at least one of the duplicated copies (*see* **Note 6**).

Figure 5 shows a convenient mating scheme that can be used to target a locus on the second chromosome. The purpose of this mating scheme is to incorporate the targeting vector (outlined in **Fig. 2**), the FLP recombinase gene, and the *I-Sce*1 endonuclease gene into the same genome. Rong and Golic have placed the FLP recombinase and the *I-Sce*1 endonuclease genes on the same chromosome, and second- and third-linked versions of these constructs are available. The gene targeting method, described by Rong and Golic at the *X*-linked *yellow* locus, worked much more efficiently in the female germline than in the male germline. It is currently unknown if gene targeting to an autosomal locus will also be more efficient in the female germline. The prudent course of action is to perform gene targeting only in female flies until this effect is explored further. Therefore, it is imperative to ensure that the chromosome containing the FLP recombinase and the *I-Sce*1 endonuclease are heterozygous with a balancer chromosome at all times. In this mating, the targeting vector is located on the *X* chromosome, the targeted locus is on the second chromosome, and the recombinase and endonuclease genes are on the third chromosome.

In the parental cross shown in **Fig. 5**, female flies carrying the *FM7c X* chromosome balancer and a chromosome with the P{*70FLP*} and P{*70I-Sce*1} transgenes, which are heterozygous with a third chromosome balancer, are mated to male flies carrying an insertion of the targeting vector and a different third chromosome balancer, *TM3, Ser*. It is most convenient to conduct this mating with three female flies and several male flies in each vial. Vial matings are preferred because an efficient heat shock is essential to successful targeting. Sufficient vials should be set up to ensure collecting at least 500 virgin

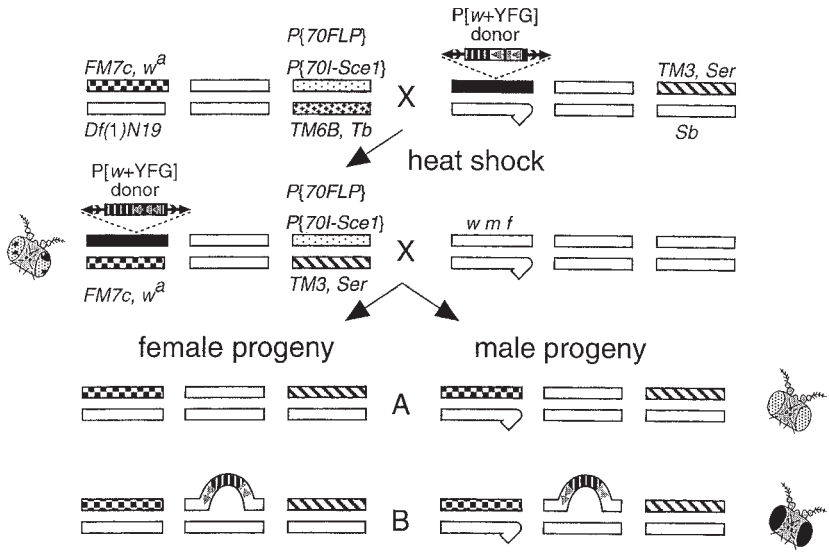


Fig. 5. Genetic screen for gene targeting in *Drosophila*. In the parental mating (top), females containing the *FM7c X* chromosome balancer and that are heterozygous for the *TM6B* third-chromosome balancer and a chromosome carrying the targeting inducer genes (the FLP recombinase and the *I-Sce*1 endonuclease genes on *P* element vectors) are crossed to males that carry the targeting vector and a different third chromosome balancer, *TM3, Ser*. The FLP recombinase and *I-Sce*1 genes are under the control of heat-shock gene promoter elements, and a heat shock is carried out at 0–3 d after mating. Virgin females from this cross that contain both the targeting vector heterozygous with the *FM7c* balancer chromosome and the inducer genes heterozygous with the *TM6B* balancer chromosome are collected, and mated to males that contain a null allele of the *white* gene. These matings can be set up en masse. It is very important to ensure that chromosomes carrying the targeting vector and the targeting inducer genes are heterozygous with a balancer chromosome to prevent meiotic recombination and the subsequent random assortment of these markers into the F_2 progeny. Male and female F_2 progeny that contain the *FM7c* chromosome and the *TM3*, *Ser* chromosome are collected and examined. The F_2 progeny of this mating will be of two eye phenotypes: orange and dark red. The orange-eye phenotype is contributed by the *white*^{apricot} allele carried on the *FM7c* balancer chromosome. Flies with this eye color may be discarded. The F_2 progeny that have a dark red eye color represent gene events in which the *white* gene in the targeting vector has moved to a new genomic locus. The most common cause of such movement will be gene targeting to the desired locus. These flies should be kept for further genetic analysis to ensure linkage of the red-eye phenotype with the desired chromosome and for molecular analysis.

female progeny of the appropriate genotype. The number of required matings will depend on the viability of the stocks, but in general 50 matings should be more than sufficient. The female flies are allowed to lay eggs in the vial for 3 d, and then are transferred to a new vial. The vials are heat-shocked at 38°C for 1 h at 3 d postmating. This heat shock induces expression of the FLP recombinase and the *I-Sce*1 endonuclease.

Virgin female F_1 progeny are collected that contain the chromosome carrying the targeting vector heterozygous with the *FM7c* balancer chromosome and the P{*70FLP*} and P{*70I-Sce*1} transgenes heterozygous with the *TM3, Ser* balancer chromosome. These females will have a mosaic eye color or a white eye color because the *white* gene in the targeting vector is excised from the most of the chromosomes upon site-specific recombination. These F_1 females are mated *en masse* to male flies that have a null white allele; the frequency of targeting is sufficiently low that most targeting events recovered from a single bottle containing 10–15 F_1 females should be independent. The male and female F_2 progeny that contain the *FM7c* and *TM3, Ser* chromosomes are collected and examined for red-eyed flies. In these red-eyed F_2 flies, the original gene targeting vector and the P{*70FLP*} and P{*70I-Sce*1} transgenes have been segregated away from the second chromosome, which contains the targeted locus. It is most convenient to proceed with the male progeny for the genetic analysis because it is much simpler to map the location of the red-eyed phenotype with males than with females. The red-eyed F_2 progeny male flies should be mated to attached-*X*, *y w f* female flies to assess linkage of the white+ phenotype to the second chromosome.

A mating scheme similar to the one outlined above could be performed if the investigator wishes to target a locus on the third chromosome. In this case, the P{*70FLP*} and P{*70I-Sce*1} transgenes should be located on the second chromosome and should be balanced with suitable second chromosome balancers. Rong and Golic outline the mating scheme for an *X*-linked target locus ***(18)*** (*see* **Notes 7** and **8**).

3.6. Molecular Screen for Gene Targeting

Putative gene targeting events from the screen in **Subheading 3.5.** can be identified by linkage of the w+ phenotype to the second chromosome. We collect at least 30 potential gene-targeting events and test for the production of a partial duplication of the *Ku80* gene with the duplicated segments separated by the *white* gene. PCR amplification of the junctions between the duplicated *Ku80* gene segments and the *white* gene will produce diagnostic PCR amplification products. The KuA and KuB oligonucleotides are complementary to sequences just outside the targeted sequences, and the vOutA and vOutB oligonucleotides are just inside the *white* gene sequences in the targeting vector.

PCR amplification with KuA and vOutA will produce an 1189-bp amplification product, and PCR amplification with KuB and vOutB will produce a 799-bp amplification product. PCR reactions with KuA and KuB will produce a 1784-bp product with DNA prepared from flies heterozygous for the targeted insertion or in flies lacking a targeted insertion, but will produce no product with DNA prepared from flies homozygous for the targeted insertion.

1. The reactions are assembled identically to those in **Subheading 3.3.**, except that the different oligonucleotide primers are used.
2. The reaction conditions are identical to those in the previous section.
3. Move the completed PCR reactions to another room for all subsequent analysis.
4. Analyze by running 15 μL of the amplification product on a 1% agarose gel.

4. Notes

1. It is possible to screen for the loss of a *P* element-linked phenotype as an indicator of *P* element excision. If for example the *P* element vector carried a rosy+ allele, the investigator could analyze all the progeny for gene conversion in which the rosy+ phenotype had been lost from the *X* chromosome. This approach has been applied successfully at the UBX and *dpp* loci.
2. In some instances successful gene conversion may be recognized by a change in phenotype other than reversion to wild type. For example, Dr. C. T. Wu has collected many gene conversion events that remove individual *yellow* gene promoter elements by examining the F_2 flies for a particular yellow mutant phenotype that is predicted based on prior knowledge of the tissue specificity of the *yellow* gene promoter elements.
3. Some of the F_3 progeny will have a gene conversion and a transposition of the P{w+ y+} donor element. If the transposition occurs to an autosome these cases can be recognized easily because some of the female F_3 progeny are also w+ and or y+. When this happens we take 8 F_3 male progeny and mate them individually to *C(1)DX, y w f* females and examine the F_4 progeny. We select lines in which only the males exhibit the w+ or w+ y+ phenotypes for further analysis.
4. The wRR primer is located outside of the *white* gene sequence that is found in the *P*{w+ y+} donor element. Therefore, when the gene conversion has a yellow+ phenotype, successful amplification of the presumed product with wRR and yR is sufficient proof of gene conversion.
5. The University of Utah distributes a gene-targeting kit on behalf of Rong and Golic. Some components of this kit are described in their *Science* paper ***(18)***, and others are patented. A material transfer agreement (MTA) must be signed prior to receiving this kit. The MTA can be obtained from: Brent Brown, Technology Manager, The University of Utah, Technology Transfer Office, 615 Arapeen Drive, Suite 110, Salt Lake City, UT, 84108, or on the web at http://www.tto.utah.edu.
6. An alternative strategy that could be employed would be to introduce a frameshift or stop codon in the 3' and 5' to the *I-Sce*1 cut site. The introduction of two

such point mutations would be expected to lower the targeting frequency by only a marginal amount, provided that the rest of the sequence was isogenic ***(7)***.

7. One major drawback to the gene-targeting method as described here is the high frequency of red-eyed flies because of incomplete excision of the *white* gene from the targeting vector. This background frequency can be greater than 100-fold higher than the frequency of targeting. Rong and Golic's second paper ***(21)*** presents one method to reduce this frequency. The F_1 females, in **Subheading 3.5.**, are mated to white-eyed males that carry a homozygous P{*70FLP*}. The progeny of this mating are heat-shocked briefly on day 3 after mating. The result is that all the false-positive progeny contain mosaic eyes, and those that are true targeting events have solid red eyes.
8. I would like to acknowledge the generous help of Dr. Kent Golic in preparing the gene targeting section.

References

1. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287,** 2185–2195.
2. Rubin, G. M. and Lewis, E. B. (2000) A brief history of *Drosophila*'s contributions to genome research. *Science* **287,** 2216–2218.
3. Kornberg, T. B. and Krasnow, M. A. (2000) The *Drosophila* genome sequence: implications for biology and medicine. *Science* **287,** 2218–2220.
4. Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R., and Engels, W. R. (1991) Targeted gene replacement in *Drosophila* via *P* element-induced gap repair. *Science* **253,** 1110–1117.
5. Nassif, N., Penney, J., Pal, S., Engels, W. R., and Gloor, G. B. (1994) Efficient copying of nonhomologous sequences from ectopic sites via *P*-element-induced gap repair. *Trends Biochem. Sci.* **14,** 1613–1625.
6. Paques, F. and Haber, J. E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63,** 349–404.
7. Nassif, N. and Engels, W. (1993) DNA homology requirements for mitotic gap repair in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **90,** 1262–1266.
8. Johnson-Schlitz, D. M. and Engels, W. R. (1993) *P*-element-induced interallelic gene conversion of insertions and deletions in *Drosophila melanogaster*. *Mol. Cell. Biol.* **13,** 7006–7018.
9. Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B., and Sved, J. (1990) High-frequency *P* element loss in *Drosophila* is homolog dependent. *Cell* **62,** 515–525.
10. Keeler, K. J., Dray, T., Penney, J. E., and Gloor, G. B. (1996) Gene targeting of a plasmid-borne sequence to a double-strand DNA break in *Drosophila melanogaster*. *Mol. Cell. Biol.* **16,** 522–528.
11. Engels, W. R., Preston, C. R., and Johnson-Schlitz, D. M. (1994) Long-range cis preference in DNA homology search over the length of a *Drosophila* chromosome. *Science* **263,** 1623–1625.

12. Keeler, K. J. and Gloor, G. B. (1997) Efficient gap repair in *Drosophila melanogaster* requires a maximum of 31 nucleotides of homologous sequence at the searching ends. *Mol. Cell. Biol.* **17,** 627–634.
13. Morris, J. R., Geyer, P. K., and Wu, C. T. (1999) Core promoter elements can regulate transcription on a separate chromosome in trans. *Genes Dev.* **13,** 253–258.
14. Merli, C., Bergstrom, D. E., Cygan, J. A., and Blackman, R. K. (1996) Promoter specificity mediates the independent regulation of neighboring genes. *Genes Dev.* **10,** 1260–1270.
15. McCall, K. and Bender, W. (1996) Probes of chromatin accessibility in the *Drosophila* bithorax complex respond differently to Polycomb-mediated repression. *EMBO J.* **15,** 569–580.
16. Williams, C. J. and O'Hare, K. (1996) Elimination of introns at the *Drosophila* suppressor-of-forked locus by P-element-mediated gene conversion shows that an RNA lacking a stop codon is dispensable. *Genetics* **143,** 345–351.
17. Lankenau, D. H., Corces, V. G., and Engels, W. R. (1996) Comparison of targeted-gene replacement frequencies in *Drosophila melanogaster* at the forked and white loci. *Mol. Cell. Biol.* **16,** 3535–3544.
18. Rong, Y. S. and Golic, K. G. (2000) Gene targeting by homologous recombination in *Drosophila* [see comments]. *Science* **288,** 2013–2018.
19. Spradling, A. C. and Rubin, G. M. (1982) Transposition of cloned *P* elements into *Drosophila* germ line chromosomes. *Science* **218,** 341–347.
20. Hastings, P. J., McGill, C., Shafer, B., and Strathern, J. N. (1993) Ends-in vs. ends-out recombination in yeast. *Genetics* **135,** 973–980.
21. Rong, Y. and Golic, K. (2001) A targeted gene knockout in *Drosophila*. *Genetics* **157,** 1307–1312.
22. Dray, T. and Gloor, G. B. (1997) Homology requirements for targeting heterologous sequences during *P*-induced gap repair in *Drosophila melanogaster*. *Genetics* **147,** 689–699.

9

TE-Based Mutagenesis Systems in Plants

A Gene Family Approach

Michiel Vandenbussche and Tom Gerats

Summary

Insertions in specific genes belonging to large and homogeneous gene families often do not cause a visible phenotype due to genetic redundancy. Therefore, several single-insertion mutants may have to be combined into double or even triple mutants in order to obtain a loss-of-function phenotype. It is therefore most useful to shift from single-gene insertion selection toward selection at the gene family level. Here, we present an alternative screening methodology that is highly suited for the functional analysis of gene families. Labeled primers designed from conserved regions present in the gene family and a transposon derived primer are combined in an optimized polymerase chain reaction (PCR) to screen a three-dimensionally pooled insertion library in a single step. PCR products are sized by polyacrylamide gel electrophoresis (PAGE), and putative insertion fragments are isolated from the gel, reamplified, and sequenced directly. Because the identification of insertions is sequence-based, insertions into highly homologous genes can easily be distinguished. Taking advantage of the presence of conserved domains in gene families, this approach allows simultaneous screening for insertion events in different family members, and it has the additional advantage of identifying yet unknown family members through their corresponding insertion mutant.

Key Words: Petunia; insertion mutagenesis; transposon; transposable elements; *dTph1*; reverse genetics; gene family; genetic redundancy.

1. Introduction

One approach to identifying the function of a gene is the selection of insertion mutants by reverse genetics. The inserting DNA can be either a transposable element or a T-DNA. Insertion elements can be either endogenous or heterologous, and huge populations are available, especially for *Arabidopsis*, maize, and snapdragon *(reviewed in* **refs. *1*,*2***). With the molecular isolation of the *dTph1* transposable element ***(3)***, insertion mutagenesis has become feasible for *Petunia hybrida* ***(4)***. The approach is based on the initially described

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

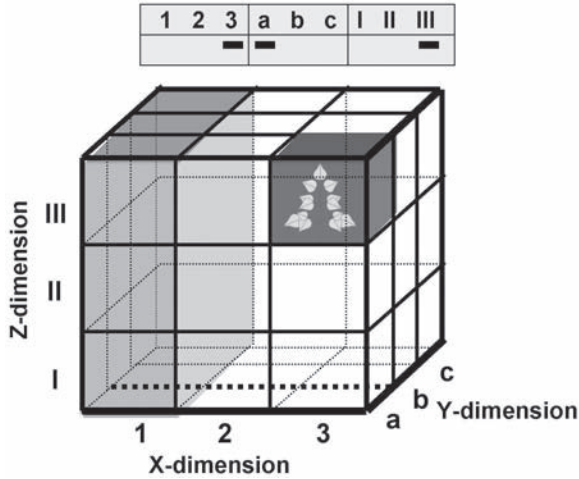


Fig. 1. Schematic representation of a three-dimensional pooling system. Leaves from all plants in one plane (e.g., X1 plane) are pooled for DNA extraction.

strategies using *P* elements in *Drosophila melanogaster* ***(5,6)***. Adopting a particular sample pooling system (*see* **Fig. 1**), each individual plant is sampled three times and identified by three coordinates, a method which facilitates the screening of large populations enormously. For example, selecting an individual plant carrying a specific insertion from a population of 3840 (15 × 16 × 16) plants requires only 47 polymerase chain reactions (PCRs), combining a gene-specific primer with a T-DNA/transposon-specific primer in a PCR. Classically, PCR products are sized by agarose gel electrophoresis, and gene-specific products are identified by hybridization to a gene-specific probe. The autoradiograph will reveal the coordinates of a plant that carries an insertion allele (*see* **Fig. 2**).

Insertion mutagenesis as a reverse-genetics tool may seem a straightforward approach. However, insertions in specific genes belonging to large gene families often do not cause a visible phenotype, simply due to redundancy ***(7–9)***. Therefore, several single-insertion mutants may have to be combined into double or even triple mutants in order to obtain a loss-of-function phenotype. It therefore would be most useful to shift from single-gene insertion selection towards selection at the gene family level.

When screening for insertions into genes belonging to large and highly homologous gene families, cross-hybridization between different family members can seriously complicate downstream analysis. Choosing primers and probes in nonconserved regions might circumvent this problem, but then repeated screenings, targeting each individual family member, are required,

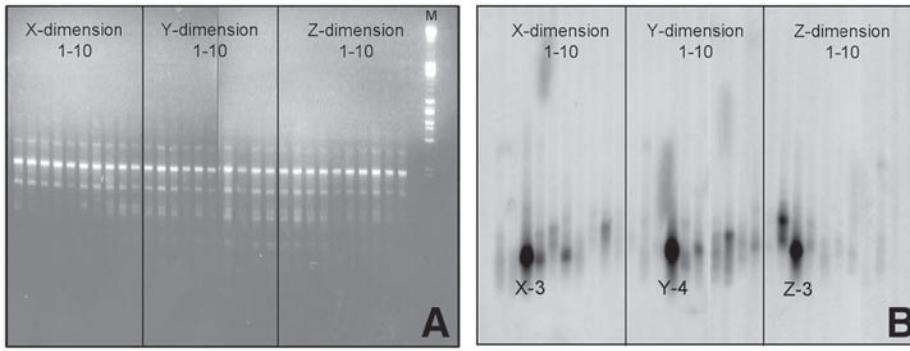


Fig. 2. An example of a classical screening experiment. **(A)** One thousand Petunia W138 plants (10 × 10 × 10) were screened by PCR and sized on a 1% agarose gel (M: lambda PstI ladder). **(B)** The resulting blot hybridized with a gene-specific probe revealing an insertion (in plant [3,4,3]).

making the whole screening labor-intensive and time-consuming. Moreover, this approach would imply that the sequence of "all" family members is available.

To detect insertions in multiple-gene copies simultaneously and to create a more efficient screening system, we developed a method that omits the hybridization step. Because the specificity and sensitivity of the detection step normally is determined by hybridization with a gene-specific probe, these features have to be accomplished differently. First of all, the gene-specific or family-specific primer(s) to be used in the PCR screening are designed in such a way that they preferentially recognize members of the gene family that is screened for. We will call them Family Signature (FS) primers. We design FS primers as follows: A primer sequence (typically a 26–29 mere) is identified in a conserved domain of the family. Since such small conserved regions could be shared by other gene families, we test whether the proposed primer sequence is truly family-specific by performing a homology search against the whole nucleotide database. If homologies are predominantly found in genes belonging to the gene family to be screened, the primer is considered a good FS primer (*see* **Fig. 3** and **Note 1**).

Second, PCR products are visualized by using a labeled FS primer and sized by polyacrylamide gel electrophoresis (*see* **Fig. 4**). After exposure, positive fragments can be directly cut out from the gel and reamplified. Positives are analyzed by direct sequencing of the reamplified fragments. Because the identification is sequence-based, even highly homologous genes can be distinguished, while the exact insertion position is also determined (*see* **Fig. 5**). This approach allows simultaneous identification of insertion events in different

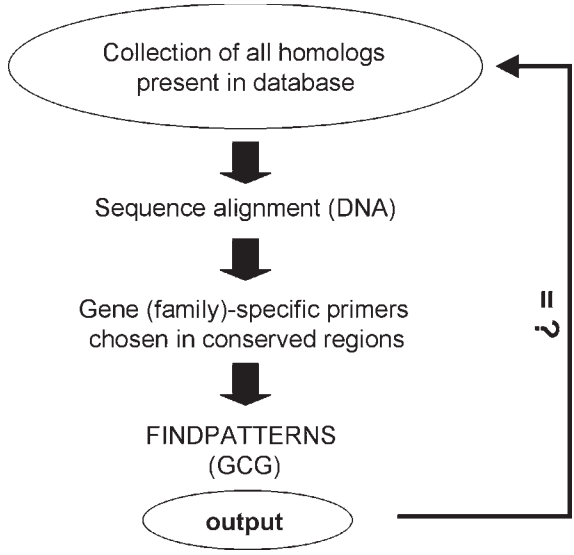


Fig. 3. Family Signature primer design.

family members. It has the additional advantage of identifying yet unknown family members through their corresponding insertion mutant.

We will briefly discuss the efficiency of the *Petunia* TE system as described here. Variations in the number of *dTph1* elements in each plant (approx 100–200) and the variable transposition rate make it difficult to calculate the probability to tag a specific gene ***(10)***. Therefore we prefer to give a practical example of a gene family screening. We designed FS primers to screen for insertions into the *Petunia* MADS-box gene family ***(11,12)***. Screening 12,700 plants, we found 32 insertions into 20 different *Petunia* MADS-box genes ***(12a)***.

2. Materials

2.1. Plant Material and DNA Template Preparation

Large populations of W138 *Petunia* plants (1000–4000 plants, depending on the available space, composed of small families of 20–25 plants) were grown under normal greenhouse conditions. DNA was extracted from pooled leaf material and harvested according to a three-dimensional matrix (*see* **Fig. 1**) as described in **ref. *4***. All plants were selfed, and seeds of every plant were harvested individually. Because the seeds and DNAs can be stored for several years, such a population can be screened many times over a prolonged period. *Petunia* insertion libraries have been set up in Amsterdam and Wageningen

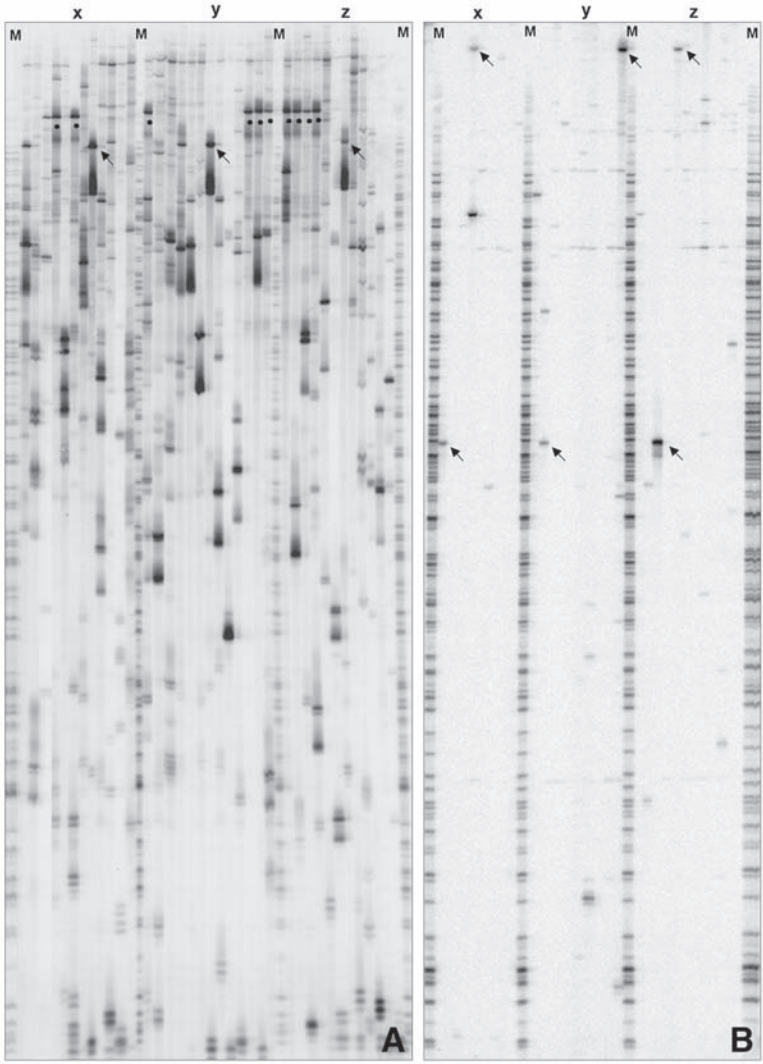


Fig. 4. Two examples of the screening procedure as described in this chapter. The FS primer is ^{33}P-labeled and PCR products are sized on a 4.5% denaturing acrylamide gel. The dried gels have been exposed to a phosphorimage screen (Molecular Dynamics Inc.) for about 3 h. The *X*, *Y*, and *Z* dimensions are separated by a marker M (leftover of AFLP reactions). **(A)** A population of 1872 W138 plants (12 × 13 × 12) screened with a MADS-box gene FS primer. Fragments depicted by black dots represent a segregating insertion in *pmads14* (*see* **Note 11**). Arrows indicate an insertion in pmads12. Although a lot of fragments are visible on the gel (in contrast with **B**), the majority do not correspond to both selection criteria as stated in **Note 11**, and thus should be ignored. **(B)** A population of 1080 W138 plants (9 × 10 × 12) screened with another MADS-box FS primer, showing two insertions in fbp20.

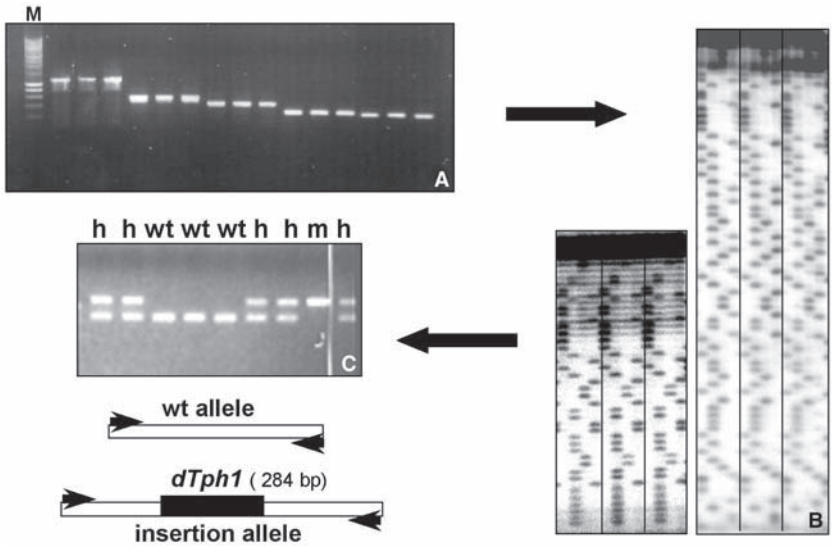


Fig. 5. **(A)** Reamplification products of fragments cut out from acrylamide gels sized on a 1% agarose gel (M = 1 kb, Smartladder, Eurogentec). **(B)** Typical sequencing results of insertion fragments. In this case the labeled FS primer was used as a sequencing primer. Sequences of two sets of 3 coordinates are shown, originating from insertions very close to the FS primer site (24 bp and 53 bp respectively) in two *Petunia* MADS-box genes. The last 18 bp at the end of the sequence represent the *dTph1* IR transposon primer. **(C)** Once the exact insertion position is known, a straightforward segregation analysis can be performed by PCR using two gene-specific primers flanking the insertion site. PCR products are sized on a 1.2% agarose gel. Insertion alleles are 284 bp bigger than wild-type alleles (h = heterozygote, wt = wild type, m = homozygous mutant).

(The Netherlands) and in Ghent (Belgium), and they are being screened co-operatively.

Researchers who wish to screen the *Petunia* insertion libraries are kindly invited to contact R. Koes (e-mail: koes@bio.vu.nl), G. Angenent (e-mail: G.C. Angenent@plant.wag-ur.nl) or T. Gerats (e-mail: toger@gengenp.rug.ac.be).

2.2. Primer Design

1. Primer analysis software (e.g., OLIGO Primer Analysis Software (National Biosciences, Inc.).
2. The findpatterns function (GCG software package, Wisconsin Package Version 10.0, Genetics Computer Group [GCG]).
3. Multiple alignment tools, e.g., the pileup function in the GCG software package, ClustalW ***(13)***, or BlockMaker ***(14)***.

2.3. Primer Labeling

1. FS primer (10 μ*M*).
2. T4 polynucleotide kinase (PNK) (10 units/μL).
3. 10X T4 PNK A buffer: 500 m*M* Tris-HCl, pH 7.6, 100 m*M* $MgCl_2$, 50 m*M* dithiothreitol (DTT), 1 m*M* spermidine-HCl, 1 m*M* EDTA (Eurogentec).
4. (γ-^{33}P)-ATP (10 mCi/mL).
5. 2 heat blocks.

2.4. PCR Screening

(For additional details, *see* **Note 2**.)

1. Template DNA (± 20 ng/μL).
2. Labeled FS primer.
3. Inverted repeat (IR) primer (5'-GAATTCGCTCCGCCCCTG-3'): complementary to the terminal inverted repeat of *dTph1* and containing an *EcoRI* site at the 5' end, (1 pmol/μL).
4. Platinum *Taq* DNA polymerase (GibcoBRL) (*see* **Note 3**).
5. 10 m*M* dNTPs.
6. Deionized H_20 (dH_2O).
7. 10X PCR reaction buffer (with 15 m*M* $MgCl_2$).
8. Thermal cycler: PerkinElmer 9600 or equivalent (*see* **Note 4**).
9. Formamide loading dye: 98% formamide, 10 m*M* EDTA, 0.06% Bromo phenol blue, and 0.06% xylen cyanol as tracking dyes.

2.5. PAGE Analysis

1. Biorad sequencing gel system (50 × 38 × 0.4 cm) or equivalent system.
2. Acrylamide stock solution (40%, acrylamide:methylenbisacrylamide, 19:1).
3. Urea.
4. 10X TBE: 1*M* Tris, 1*M* boric acid, 20 m*M* EDTA; refresh every 2 wk.
5. TEMED (tetramethyl-ethylenediamine, Pharmacia Biotech).
6. 10% ammonium persulfate (APS); store at 4°C only for a maximum of 2 to 3 wk.

2.6. Reamplification and Direct Sequencing of Selected Fragments

1. Kodak BioMax films (Kodak).
2. Platinum *Taq* DNA polymerase (GibcoBRL).
3. 10 m*M* deoxyribonucleotid triphosphates (dNTPs).
4. Deionized H_20 (dH_2O).
5. 10X PCR reaction buffer (with 15 m*M* $MgCl_2$).
6. FS primer (10 μ*M*) and IR primer (10 μ*M*).
7. Thermosequenase cycle-sequencing kit (Amersham Life Science) or equivalent system.
8. ^{33}P-labeled FS or IR primer (*see* **Subheadings 2.3.** and **3.2.** and **Note 5**).

3. Methods

3.1. Primer Design

1. Align nucleotide sequences of all available family members and identify conserved regions.
2. Choose a primer sequence (*see* **Note 6**) in a conserved region according to the following requirements:
   a. The 3' end of the primer is in the most conserved region (DNA level).
   b. Length is approx 25–29 nucleotides.
   c. The GC content is 35–50%.
   d. The T_m value is 70–74°C (*see* **Note 7**).
   e. A minimum of two and maximum of four Gs or Cs in the last six nucleotides at the 3' end, two evenly distributed (*see* **Note 8**).
   f. No internal loop formation occurs with a T_m higher than 45°C.
3. Use the primer sequence in the GCG program findpatterns and screen the plant database allowing three to seven mismatches.
4. Change the primer sequence until the majority of the sequences found in the findpatterns output file correspond to the gene family that will be screened for insertions.

3.2. Primer Labeling

1. For 50 PCR reactions, mix together the following components in a 1.5-mL microcentrifuge tube placed on ice (*see* **Note 9**):

FS primer (10 pmol/µL)	5 µL
T4 PNK 10X buffer	5 µL
H_2O	32.5 µL
T4 PNK	2.5 µL
(γ-^{33}P)-ATP	5 µL

   The total volume should be 50 µL.
2. Incubate the mixture for 30 min at 37°C.
3. Inactivate the enzyme by incubating the tube for 10 min at 80°C.
4. Centrifuge the mixture briefly to collect any condensate that has formed on the top of the tube. The labeled primers can be used immediately or stored for a maximum of 1 wk at –20°C.

3.3. PCR Screening

1. Pipet 5 µL template (± 100 ng) into each PCR tube.
2. For *n* PCR reactions, assemble *n* + 3 times the following PCR master mix in a 2-mL tube:

10X PCR buffer	2.5 µL
dNTPs (10 m*M*)	0.5 µL
IR primer (1 pmol/µL)	2 µL
Platinum *Taq*	0.12 µL

dH_2O	14.88 µL
Labeled FS primer	1 µL

The total volume should be 20 µL.

3. Vortex the mix for a few seconds and spin briefly.
4. Add 20 µL mix to each tube and spin down in a PCR plate centrifuge.
5. Run the following PCR program (*see* **Note 10**):
   a. 14 cycles at 94°C for 15 s; 71°C for 30 s, –1°C/cycle; 72°C for 30–60 s.
   b. 50 cycles at 94°C for 15 s; 56°C for 30 s; 72°C for 30–60 s.
   c. 1 cycle at 4°C as long as desired.
6. Add 20 µL formamide loading dye to the PCR samples, seal the PCR tubes, and vortex briefly.
7. Denature the samples for 5 min at 94°C, then quickly cool down on ice before loading or store at –20°C (maximum 2 wk).

3.4. PAGE Analysis

Two to three microliters of the reaction products are analyzed on a 4.5% denaturing polyacrylamide gel (*see* **Fig. 4**). This gel is in principle a normal sequencing gel, with the exception that a lower percentage of polyacrylamide is used. The gel is cast 2 h before use, by mixing 100 mL of 5%, 19:1 acrylamide:bisacrylamide, 7.5 *M* Urea, 1X TBE, with 500 µL of 10% ammonium persulfate and 100 µL of TEMED. The gel is run using 1X TBE as a running buffer at constant power (40–50 V/cm). After running, the gel is lifted from the glass plate with Whatman 3MM paper and dried on a standard slab gel drier for sequence gels. Gels are exposed to Molecular Dynamics phosphorimage screens for 2–3 h and visualized using a Molecular Dynamics phosphorimage analysis system 445 SI (Molecular Dynamics Inc.). **Note 11** explains how to select candidate bands for further analysis.

3.5. Reamplification and Direct Sequencing of Selected Fragments

1. Expose the dried gel to a film overnight (*see* **Note 12**).
2. Cut the candidate bands from the gel and put the piece of gel (including the Whatman paper) in 200 µL dH_2O (*see* **Note 13**).
3. Allow the DNA to elute for 45–60 min at room temperature and vortex occasionally.
4. Centrifuge the tubes for 1 min to spin down the paper and gel particles. The water phase contains the template DNA.
5. Pipet 5 µL template into each PCR tube.
6. For *n* PCR reamplifications, assemble *n* + 3 times the following PCR master mix in a 2-mL tube:

10X PCR buffer	3.5 µL
dNTPs (10 m*M*)	0.7 µL
IR primer (10 pmol/µL)	1 µL
FS primer (10 pmol/µL)	1 µL

Platinum *Taq*	0.12 µL
dH_2O	23.68 µL

The total volume should be 30 µL.

7. Vortex the mix for a few seconds and spin briefly.
8. Add 30 µL master mix to each tube.
9. Run the following PCR program:
   a. 8 cycles at 94°C for 15 s; 65°C for 30 s, –1°C/cycle; 72°C for 30–60 s.
   b. 30 cycles at 94°C for 15 s; 56°C for 30 s; 72°C for 30–60 s.
10. Load 7–10 µL of the PCR products on an agarose gel to check the purity of the bands. Single bands should be visible without any background (*see* **Fig. 5** and **Note 14**).
11. Use 3–4 µL of reamplified PCR product as a template and 1 µL of labeled primer (IR or FS primer, *see* **Note 5**) to set up the sequencing reactions as described in the radiolabeled primer cycle sequencing protocol accompanying the Thermosequenase cycle-sequencing kit.
12. Use the following cycling program:
    a. 50 cycles at 95°C for 20s; 56°C for 30 s; 72° C for 60 s.
13. Add 4 µL stop solution (supplied with the kit) to the sequencing reactions and heat the samples to 94°C for 5 min.
14. Analyze the sequencing reactions on a denaturing 6% polyacrylamide gel.

4. Notes

1. In a standard PCR with a forward and a reverse gene-specific primer pair, the possibility that a wrong target will be amplified is minimal, because the statistical chance that two gene-specific primers would be able to efficiently anneal at the same time on a wrong target is very low. For this reason, primer pairs can be chosen at any position in a gene, as long as they meet the standard primer design criteria. In contrast, in an insertion screening with standard PCR, only one of the primers contributes to the specificity of the amplification (the transposon or T-DNA primer is obviously not specific for the gene screened for). As a consequence, in classical screening approaches, a hybridization step is necessary to detect the positive PCR products among the background amplification. Here we present a primer design and PCR procedure that enhances specificity of the amplification in such a way that hybridization is not needed anymore.
2. In our lab, we use some extra equipment and disposables that facilitate the handling of large numbers of PCR reactions and that enhance reliability. The stock of template DNA is stored in a 96-well plate, and the DNA is diluted to 5 µL template/PCR reaction (± 100 ng) to allow easy pipetting with an automatic multichannel pipettor. We use 96-well PCR plates and seal them with reusable rubber lids (e.g., MicroAmp™ full-plate cover, PerkinElmer™; must be used in combination with a thermal cycler that has a screw-down or clip-down hot lid applying pressure to the top of the plate). After addition of the formamide loading dye, we seal the plates with disposable aluminium foil lids (Seal and Sample,

Biomek™). Finally, we use a "serial" pipet to distribute the PCR master mix in all PCR tubes.

3. We highly recommend using a hot-start PCR enzyme for the screening. In a hot-start PCR, DNA synthesis is prevented during the warming up to 94°C in the first cycle ***(15)***. This strategy provides a tighter control over the conditions that allow primer–template annealing. We found a major reduction of background amplification using hot-start PCR with, as a result, a much higher sensitivity. There are several methods and different enzymes to obtain a hot start. In our hands, Platinum *Taq* DNA polymerase turned out to be the most efficient one.
4. A thermal cycler is needed that allows programming of touchdown PCR profiles (*see also* **Note 10**) and linking of different programs.
5. Normally, we use the IR primer as a sequencing primer. When the FS primer is used, then separate labeling reactions are required for each primer used in the screening. Moreover, in our experience, some FS primers (approx 10%) gave bad sequencing results for reasons we could not explain.
6. If insertion mutants in "all" family members are of interest, then two options to reduce the total number of screening reactions are available. First of all, one can design a small set of degenerate primers that covers the whole family. The second option is to develop a set of "gene-specific" primers (e.g., one for each subfamily). By choosing the 3' end of each primer in a highly conserved region (subfamily level), most "gene-specific" primers will be able to recognize more than one family member. We tested both approaches, and in our hands the second option was far more successful.

   Working with degenerate primers, we encountered the following problems: It turns out that during synthesis of degenerate primers, there is often a bias for certain positions in the primer to incorporate only one of the nucleotide possibilities. Second, because only a small portion of a degenerate primer pool has a perfect match with each target, amplification occurs at low efficiency. As a result, we obtained only very weak signals, and often one or two of the coordinates could not be identified.
7. T_m value calculated according to the GC percent method.
8. In our experience, we obtained more false positives with primers having a very strong 3' clamp (e.g., 5'-X_NATGCGC-3').
9. Use filter tips to prevent pipettor contamination.
10. The first part of the PCR profile is a touchdown (TD) PCR ***(16)***. A high annealing temperature (T_{ann}) is used during the first cycles to enhance specificity of the amplification, followed by a gradual decrease in the T_{ann} to increase yield.
11. Template DNA of each individual plant is present three times (*x*, *y*, and *z* coordinates) in the pooled samples of the DNA library (samples may harbor DNA from up to 250 individuals). The implication is that if a plant contains an insertion into the gene screened for, then fragments should be selected that appear three times, once in the *x* dimension, once in the *y* dimension, and once in the *z* dimension, all exhibiting exactly the same size. To facilitate the selection of fragments according to these rules, a marker can be loaded to separate the different dimensions

from each other. In case the gel does not run perfectly horizontal, the marker also helps to adjust the image by drawing lines in between corresponding bands in the marker lanes. Fragments on the gel appearing in only one or two dimensions of the population can be caused by somatic insertion events or PCR artifacts. These fragments should be ignored.

Sometimes bands appear more than three times, or even in all lanes. Normally these fragments are also ignored, but there is one exception to this rule. A *Petunia* population generally consists of small families of 20–25 plants, originating from a single seed capsule (self). In case an insertion event took place in the parental plant (or even earlier), the insertion will segregate (or may even be homozygous) in that family. As a result, more than three lanes will exhibit the same positive. In such cases, the coordinates of the bands are checked to see whether they correspond to a single family. An example of a segregating insertion event in the *pmads14* gene is shown in **Fig. 4**.

12. Put the film on the gel in such a way that you can exactly position the film on the gel later on. For example, keep film and gel together by punching staples through it in several places.
13. The gel can be re-exposed to check whether bands have been properly cut out. Try to cut out the fragments (dried gel + paper) as small and as accurately as possible, because bigger slices could result in bad reamplification.
14. Since unpurified PCR product is used for direct sequencing, it is important that primers are used up completely. Leftover primer could compete with the labeled primer in the sequencing reaction, resulting in weak signals. If reamplification was not very efficient, try to increase the elution time or the cycle number in the reamplification reaction.

Acknowledgments

Part of this work was supported by a grant from EC BIOTECH program Bio4-CT97-2217. We thank R. Koes and G. Angenent for kindly providing DNA samples of their *Petunia* insertion libraries.

References

1. Walbot, V. (1992) Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. *Ann. Rev. Plant Physiol. Plant Mol. Biol*. **43,** 69–82.
2. Riechmann, J. L. and Ratcliffe, O. J. (2000) A genomic perspective on plant transcription factors. *Curr. Opin. Plant Biol*. **3,** 423–434.
3. Gerats, A. G. M., Huits, H., Vrijlandt, E., Marana, C., Souer, E., and Beld, M. (1990) Molecular characterization of a nonautonomous transposable element (*dTph1*) of *Petunia*. *Plant Cell* **2,** 1121–1128.
4. Koes, R., Souer, E., Van, H. A., Mur, L., Spelt, C., Quattrocchio, F., et al. (1995) Targeted gene inactivation in *Petunia* by PCR-based selection of transposon insertion mutants. *Proc. Natl. Acad. Sci. USA* **92,** 8149–8153.

5. Ballinger, D. G. and Benzer, S. (1989) Targeted gene mutations in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **86,** 9402–9406.
6. Kaiser, K. and Goodwin, S. F. (1990) "Site-selected" transposon mutagenesis of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **87,** 1686–1690.
7. Bouchez, D. and Hofte, H. (1998) Functional genomics in plants. *Plant Physiol*. **118,** 725–732.
8. Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000) B and C floral organ identity functions require Sepallata MADS-box genes. *Nature* **405,** 200–203.
9. Liljegren, S., Ditta G. S, Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000) Shatterproof MADS-box genes control seed dispersal in Arabidopsis. *Nature* **404,** 766–770.
10. De Keukeleire, P., Maes, T., Sauer, M., Zethof, J., Van Montagu, M., and Gerats, T. (2001) Display of the behavior of the dTph1 element family during ontogeny and inbreeding of Petunia hybrida. *Mol. Genet. Genomics* **265,** 72–81.
11. Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990) Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **250,** 931–936.
12. Theissen, G., Kim, J. T., and Saedler, H. (1996) Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* **43,** 484–516.

12a. Vandenbussche, M., Zethof, S., Souer, E., et al. (2003) Toward the analysis of the Petunia MADS box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require. SEPALLATA-like MADS box genes in Petunia. *Plant Cell* **15,** 2680–2693.

13. Thompson, J. D., Higgens, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680.
14. Henikoff, S., Henikoff, J. G., Alford, W. J., and Pietrokovski, S. (1995) Automated construction and graphical presentation of protein blocks from unaligned sequences. *Gene* **163,** 17–26.
15. D'Aquila, R. T., Bechtel, L. J., Viteler, J. A., Eron, J. J., Gorczyca, P., and Kaplan, J. C. (1991) Maximizing sensitivity and specificity of PCR by pre-amplification heating. *Nucleic Acids Res.* **19,** 3749.
16. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., and Mattick, J. S. (1991) "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19,** 4008.

10

The Use of Double-Stranded RNA to Knock Down Specific Gene Activity

Mary K. Montgomery

Summary

In many eukaryotes, the introduction of double-stranded RNA (dsRNA) into cells triggers the degradation of cognate mRNAs through a posttranscriptional gene silencing mechanism. This phenomenon has been called RNA interference or RNAi. Several methods for delivering dsRNA into the model organism *C. elegans* are described; these methods include (1) microinjecting dsRNA synthesized in vitro into the body cavity of the worm, (2) soaking worms in a solution of dsRNA, (3) feeding worms dsRNA-expressing bacteria, and (4) engineering transgenic worm strains to express dsRNA in vivo. Variations of these methods may be used to perform RNAi in other species as well. The choice of which delivery method to use, along with other options (region to target, length of dsRNA) are also discussed.

Key Words: RNA interference (RNAi); RNA silencing; posttranscriptional gene silencing (PTGS); short interfering RNA (siRNA).

1. Introduction

RNA interference (RNAi) refers to the use of double-stranded RNA (dsRNA) to silence genes in a sequence-specific manner. The phenomenon was first described for the nematode *Caenorhabditis elegans* ***(1)***, but has since been used as a reverse genetic technique to knock down expression of specific genes in several other animal species, including *Drosophila* and other insects ***(2–4)***, planaria ***(5)***, leech ***(6)***, and hydra ***(7)*** among others. RNAi turns out to be a form of posttranscriptional gene silencing (PTGS), as transcription of the targeted gene appears unaffected, but the mRNA it encodes is rapidly degraded, resulting in little to no protein synthesis ***(8)***. PTGS was first described by plant molecular biologists as a subset of cosuppression ***(9)*** events, in which an introduced transgene designed to be overexpressed would instead silence itself as well as endogenous copies with high sequence similarity (*reviewed in* **ref. *10***).

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

A related phenomenon called "quelling" has been described in the fungus *Neurospora crassa* ***(11)***. In all of these cases, PTGS leads to the rapid degradation of cognate mRNAs. The trigger is dsRNA in the case of RNAi and in some, if not all, cases of quelling and plant cosuppression. In cases of cosuppression, dsRNA may be generated because of the way in which multiple transgenes might insert into the genome (e.g., next to cryptic promoters or as inverted repeats leading to synthesis of both sense and antisense RNAs) ***(12)***. Antisense RNAs that bind to sense mRNAs may also be synthesized under certain circumstances by RNA-directed RNA polymerases found to mediate certain aspects of PTGS in plants, fungi, and animals (*reviewed in* **ref. *13***).

The intriguing task of uncovering the mechanism responsible for dsRNA-induced PTGS has attracted the efforts of numerous researchers working with a variety of organisms. The model emerging from biochemical and genetic studies is that many eukaryotes respond to the presence of dsRNA by activating an evolutionarily conserved mechanism designed to suppress viral replication and transposon expression. Genetic screens in the fungus *Neurospora*, the flowering plant *Arabidopsis*, and the nematode *C. elegans* have identified homologous genes required for PTGS ***(14–16)***; some PTGS-defective mutants are also more susceptible to viral infection or less able to suppress transposon hopping compared to wild type ***(17–19)***. Whereas many viruses and transposons produce RNAs with extended double-stranded structure, the host eukaryotic cell in general does not. It appears then that dsRNAs are interpreted by the host cell as "nonself" and potentially harmful. Indeed, in vertebrates detection of dsRNA elicits a global panic response that triggers the interferon response in neighboring cells and leads to a general shutdown of translation in the exposed cell ***(20)***. This more severe and sequence-nonspecific response had initially limited RNAi's usefulness as an effective tool for specific knockdowns in vertebrates. However, the introduction of dsRNA into mammalian oocytes and early embryos (i.e., prior to development of the immune response) has resulted in sequence-specific interference ***(21)***. Furthermore, the introduction of short "preprocessed" 23-nucleotide dsRNAs, called short interfering RNAs (siRNAs) causes sequence-specific degradation of targeted mRNA in cultured mammalian cells without eliciting the nonspecific lethality seen with longer dsRNA sequences ***(22)***.

How does dsRNA lead to sequence-specific mRNA degradation? Biochemical studies indicate that the original dsRNA material is cleaved into approx 23-bp siRNAs by an enzyme with RNaseIII-type activity ***(23)***. This enzyme, named *dicer*, contains RNaseIII and helicase domains and was originally identified from *Drosophila* extracts ***(24)***; putative homologs have been identified in *C. elegans*, plants, fungi, and mammals. The siRNAs are then thought to act as guides that bring a larger enzyme complex, called RNAi induced silencing complex (RISC), to mRNAs with complementary sequence;

presumably, the siRNAs are unwound by a helicase component in RISC that allows base-pairing between one strand of the siRNA and a complementary sequence on the target mRNA. RISC then cleaves the target mRNA, which destabilizes the mRNA, resulting in further degradation by the mRNA surveillance machinery. Potentially, the same activated RISC may be used repeatedly to target dozens or hundreds of mRNA molecules. Such a mechanism would explain earlier observations concerning the ability of dsRNA to function substoichiometrically ***(1)***, as a single dsRNA molecule may produce several siRNA molecules, each of which may activate a separate RISC. Thus, the number of mRNA molecules degraded is far in excess of the number of dsRNA molecules introduced into a cell.

Although the mechanism governing the RNAi/PTGS cellular response is not yet completely understood, researchers have rapidly harnessed RNAi as a tool to analyze gene function, essentially tricking the organism into attacking the expression of one of its own genes by introducing dsRNA equivalent in sequence to the gene of interest. Various methods for delivering the dsRNA have been developed. The most direct method is to microinject dsRNA that has been synthesized in vitro. This technique has been used extensively to generate knockdown phenotypes of thousands of genes in *C. elegans*, including 96% of the 2300 predicted open reading frames on chromosome III ***(25)***. It has also been the means for introducing dsRNA into eggs of *Drosophila* ***(2)***, *Xenopus* ***(26)***, and mice ***(21)***.

Microinjection requires relatively expensive equipment and some expertise. Therefore, efforts to develop alternative methods for dsRNA delivery have resulted in technically less demanding protocols. For organisms such as *C. elegans* that have the ability to transport dsRNA across cell boundaries, alternative methods include simply soaking the worms in a solution containing dsRNA ***(27)*** and feeding the worms dsRNA-expressing bacteria ***(28)***. These methods in general do not result in as efficient a knockdown as microinjection; however, recent improvements using a strain of *E. coli* deficient for RNaseIII and engineered to produce high quantities of specific dsRNAs when fed to *C. elegans* resulted in knockdown phenotypes comparable in severity to genetic loss-of-function phenotypes for a variety of targeted genes ***(29)***. Finally, a fourth method for dsRNA delivery is the use of transgenes to make dsRNA in vivo; this approach has been used successfully in a broad range of species, including the protozoan *Trypanosoma* ***(30,31)***, flowering plants such as tobacco, rice ***(32)***, and *Arabidopsis* ***(33)***, and invertebrates including *C. elegans* ***(34)*** and *Drosophila* ***(35)***. One advantage of the transgenic approach is that "mutant lines" may be maintained over multiple generations, whereas microinjection of dsRNA results in transient and ultimately reversible knockdowns. Additional methods for delivery of dsRNA are transfection via electroporation ***(7,30)***. Essentially, any method used to deliver DNA into an organism can most likely also be used to deliver dsRNA.

2. Materials

2.1. DNA Template Cleanup

(For additional details, *see* **Notes 1** and **2**.)

1. 1X Stop solution: 1*M* ammonium acetate, 10 m*M* EDTA, 0.2% SDS. Filter-sterilize and store at room temperature.
2. Glycogen, molecular biology grade, at 20 mg/mL (Roche Molecular Biochemicals). Store at –20°C.
3. Tris-EDTA (TE) buffer: 10 m*M* Tris-HCl, pH 7.4, 1 m*M* EDTA. Autoclave and store at room temperature.
4. Chloroform (light sensitive; store at room temperature in the dark).
5. Phenol (light sensitive; store at 4°C).

Keep stock chemicals nuclease-free by never placing any item such as a spatula in the stock dry chemical. Measure out chemicals by gently tapping on container; always discard excess chemicals—do not pour back into container.

2.2. In Vitro Synthesis of dsRNA

1. 5X Transcription buffer: 200 m*M* Tris-HCl, pH 7.9, 30 m*M* $MgCl_2$, 10 m*M* spermidine, 50 m*M* NaCl (Promega Corp.).
2. 100 m*M* dithiothreitol (DTT) (Promega Corp.).
3. 5 m*M* nucleotide triphosphates (NTPs): To make nucleotide mixture: Add 5 μL each 100-m*M* adenosine triphosphate (ATP), guanine triphosphate (GTP), cystine triphosphate (CTP), and uridine triphosphate (UTP) to 80 μL RNase-free TE buffer. (Stock NTPs from Pharmacia Biotech Inc.)
4. 1 mg/mL linearized plasmid DNA or polymerase chain reaction (PCR) template DNA in TE buffer (*see* **Notes 1** and **2**).
5. T3 and T7 RNA polymerases at 15–20 U/μL (Promega Corp.). These enzymes are very labile. Use of a –20°C labtop cooler is recommended to minimize the effects of removal from the freezer.
6. RNasin (ribonuclease inhibitor) at 40 U/μL (Promega Corp.).
7. DNase I, RNase-free at 10–50 U/μL (Roche Molecular Biochemicals).
8. 3X Injection Buffer: 60 m*M* phosphate buffer, pH 7.5, 9 m*M* potassium citrate, pH 7.5, 6% PEG 6000.
9. 5 m*M* Tris-HCl, pH 7.5, 0.5 m*M* EDTA.

All solutions should be made up in filter-sterilized, double-distilled water (ddH_2O); it is not necessary to use DEPC-treated water. Store all of the above reagents at –20°C.

2.3. dsRNA Delivery

2.3.1. Delivery of dsRNA by Injection

1. Recovery buffer: 3 m*M* HEPES, pH 7.2, with salmon sperm DNA (1 mg/mL), 2.4 m*M* KCl, 66 m*M* NaCl, 3 m*M* $MgCl_2$, 3 m*M* $CaCl_2$, and 4% glucose.

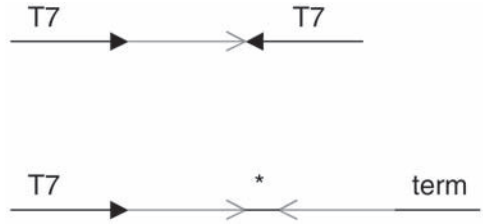


Fig. 1. Two plasmid configurations for generating double-stranded RNA (dsRNA) in the bacterial strain HT115(DE3) to be used in delivery of dsRNA by feeding (*see* **refs. *28*** and ***29***). The top figure represents a configuration that utilizes two T7 promoters flanking a single copy of a segment from a target coding sequence. Both sense and antisense RNAs are made when production of T7 polymerase is induced, and the two strands anneal in vivo to produce dsRNA. The bottom figure represents a configuration in which a single T7 promoter drives expression of RNA with a hairpin structure. The sense and antisense versions of the target (represented by gray arrows) are generated from an inverted repeat. Sense and antisense sequences within the RNA will basepair with each other to form the stem while a non-homologous spacer sequence of approx 60 bp (represented by *) forms the loop of the hairpin. At the 3' end a T7 terminator sequence is included to terminate transcription. Similar configurations can be used to drive expression of dsRNA within transformed experimental organisms or cell lines by replacing T7 with a different promoter. (T7, T7 RNA polymerase binding site; *, short spacer sequence; term, T7 transcription terminator sequence.)

2.3.2. Delivery of dsRNA by Soaking

1. M9 buffer: To make mix 3 g KH_2PO_4, 6 g Na_2HPO_4, and 5 g NaCl in 1L ddH_2O. Autoclave for 20 min. Allow to cool to 60°C. Add 1 mL of sterile 1*M* $MgSO_4$. Aliquot into sterile culture flasks if desired and store at 4°C.

2.3.3. Delivery of dsRNA by Feeding

1. DH5αF' bacteria or other standard cloning strain (Life Technologies).
2. Bacterial strain HT115(DE3) ***(29)***. The bacterial strain HT115:W3110, *rnc*14::ΔTn10 ***(36,37)*** carries a null mutation in the dsRNA-specific endonuclease RNaseIII. HT115(DE3) is an RNaseIII mutant strain that in addition harbors a λDE3 lysogen, which serves as a source of T7 polymerase. This strain will express and accumulate dsRNA when transformed with the appropriate construct. The strain is also tetracycline resistant. It can be obtained from the *Caenorhabditis* Genetics Stock Center at the University of Minnesota (*see* http://biosci.umn.edu/CGC/CGChomepage.htm and search under Strain List).
3. A plasmid construct designed to generate dsRNA from a T7 promoter, as described by Timmons and Fire ***(28)***. The plasmid may contain (1) the target sequence flanked on both sides by T7 promoters or (2) a single T7 promoter driving an inverted repeat structure interrupted by a short nonhomologous spacer DNA (*see* **Fig. 1**).

4. 2X YT culture medium: 1.6% tryptone, 1% yeast extract, 0.5% NaCl. Autoclave.
5. Isopropyl-β-D-thiogalactopyranoside (IPTG).
6. Antibiotics: ampicillin, tetracycline.
7. NGM/agar medium: Add 3.1 g tryptone, 0.55 g Tris-HCl, 0.24 g Tris-base, 2.0 g NaCl, and 17 g agar. Fill up to 1 L with ddH_2O. Add 1.0 mL cholesterol (5 mg/mL). Autoclave. Let cool to 45°C and supplement with antibiotics and IPTG where protocol indicates. Pour 60-mm or 100-mm plates.

3. Methods

3.1. DNA Template Clean-Up

1. Digest 5–10 μg plasmid DNA in 100 μL reaction mix with the appropriate restriction enzyme. Alternatively, PCR products that have been amplified using primers that contain T3 and T7 polymerase promoter sequences can be used (*see* **Notes 1** and **2**).
2. Add 300 μL 1X STOP and 0.3 μL glycogen as carrier.
3. Add 200 μL phenol:chloroform (1:1), mix vigorously and spin at top speed (14,000*g*) for 2 min.
4. Transfer upper aqueous phase to a new tube with 200 μL phenol:chloroform (1:1), mix vigorously and spin at top speed for 2 min.
5. Transfer upper aqueous phase to a new tube with 200 μL chloroform, mix vigorously and spin at top speed (14,000*g*) for 1 min.
6. Transfer upper aqueous phase to a new tube with 1 mL 100% ethanol.
7. Invert tube to mix contents and spin at top speed for 10 min. Carefully dump the supernatant.
8. Add 1 mL 100% ethanol to wash the pellet. Spin at top speed for 2 min. Carefully dump the supernatant.
9. Dry the pellet. Resuspend in 10 μL RNase-free TE.
10. The final concentration of the DNA should be 0.5–1 μg/μL and can be checked by running 1 μL on a standard agarose gel.

3.2. In Vitro Synthesis of dsRNA

1. In a sterile nuclease-free microfuge tube add the following reagents in order as listed (*see* **Note 3**):

   RNase-free ddH_2O 9.0 μL.
   5X transcription buffer 4.0 μL.
   100 m*M* DTT 2.0 μL.
   template DNA (0.5–1.0 μg) (*see* **Notes 1** and **2**) 1.0 μL.
   5 m*M* NTPs 2.0 μL.
   RNasin 1.0 μL.
   T3 or T7 RNA polymerase 1.0 μL.
2. Mix reagents together by tapping tube and then briefly spin in a microcentrifuge.
3. Incubate for 1.5 h at 37°C.
4. To terminate transcription, add 1.0 μL RNase-free DNase, mix, and spin. Incubate for 15 min at 37°C.

5. Add 380 µL 1X stop solution and 0.5 µL glycogen as carrier to the reaction mix (*see* **Note 4**).
6. Add 200 µL phenol:chloroform (1:1), mix vigorously and spin at top speed for 2 min.
7. Transfer upper aqueous phase to a new tube with 200 µL phenol:chloroform (1:1), mix vigorously and spin at top speed for 2 min.
8. Transfer upper aqueous phase to a new tube with 200 µL chloroform, mix vigorously and spin at top speed for 1 min.
9. Transfer upper aqueous phase to a new tube with 1 mL 100% ethanol.
10. Invert tube to mix contents and spin at top speed for 15 min. Carefully dump the supernatant.
11. Add 1 mL 100% ethanol to wash the pellet. Spin at top speed for 2 min. Carefully dump the supernatant.
12. Dry the pellet. Resuspend in 10 µL RNase-free 5 m*M* Tris, pH 7.5, 0.5 m*M* EDTA.
13. The concentration of ssRNA can be estimated by running 1 µL on a standard agarose gel.
14. To make dsRNA, mix 5 µL sense RNA with 5 µL antisense RNA and incubate at 80°C in a heating block for 3 min; this step should denature the two strands to allow subsequent annealing. Remove from the heating block and add 5 µL 3X injection buffer, mix, spin, and incubate at 37°C for 30 min. Store the dsRNA at –20°C if it will be used within several weeks, or at –80°C for longer-term storage.
15. Confirm that the RNA is predominantly double stranded by testing mobility on a standard (nondenaturing) agarose gel. Gel mobility of the dsRNA will be shifted relative to the individual ssRNA preparations.

3.3. dsRNA Delivery

The following protocols have been optimized for delivery of dsRNA into *C. elegans* and other nematodes within the *Caenorhabditis* group. Many of these protocols, however, can be modified for use with other organisms. The choice of which delivery system to use will depend on the organism, temporal and spatial expression of the targeted gene, available equipment and expertise, and the goals of the experiment (e.g., transient vs sustained knockdown).

3.3.1. Microinjection of dsRNA

Because microinjection systems tend to be specialized for each model organism, a detailed description of such apparati is beyond the scope of this chapter and is already available through other sources (for example, *see* **ref. *38***). The protocol below assumes the reader has knowledge of and experience with microinjection, and only points out critical steps for RNAi and/or steps that differ significantly from microinjection used for DNA transformation (*see* **Note 5**).

1. Open fresh package of capillary needles and wear gloves to avoid contaminating needles with RNases. Pull needles and store in dust-free container.
2. Backfill needles with small volume of dsRNA injection mix.
3. Inject dsRNA into body cavity (typically gut or gonad) of L4 larvae or adult hermaphrodites immobilized on an injection pad. A smaller volume than is typically required for DNA transformation can be injected.
4. Allow injected worms to recover in recovery buffer for 1–2 h before transferring to a seeded plate with a thin, small centrally placed bacterial lawn. Such plates can usually be obtained by seeding plates approx 24 h earlier with a small drop of stationary phase *E. coli* OP50. The small thin lawn makes locating worms and embryos easier.
5. Transfer worms to fresh plates approx 6 h following injection to allow the adults to "purge" (i.e., lay) embryos that had been fertilized prior to injection and that will thus not show any RNAi phenotype.
6. Transfer injected worms to new plates every 12 to 24 h. Worms may be individually cloned on separate plates or pooled together 10 per plate. Cloning individual injected worms, however, will help to separate poorly injected worms from those that were properly injected; the latter typically provide more robust RNAi results. The phenotypes of embryos laid on the second set of plates (i.e., between 6 and 36 h following injection) are scored at an appropriate stage or time. The terminal phenotypes of embryonic lethals can be scored 24 h after removal of the adults from the plates. The phenotypes of genes that affect larval or adult stages cannot be scored until 1–3 d following removal of injected adults.

3.3.2. Delivery of dsRNA by Soaking

The original soaking method was first described by Tabara et al. ***(27)***. The following protocol was developed by Subramaniam and Seydoux ***(39)*** and it has been optimized for targeting maternally expressed genes.

1. Collect up to 100 L4 stage larvae in approx 25 µL of M9 in a 1.5-mL microfuge tube.
2. Add 25 µL of dsRNA to the tube. The final concentration of RNA in the 50-µL volume should be approx 150 ng/µL or higher.
3. Place the tube on a nutator overnight. The motion of the nutator will not mix the worms very much but will prevent them from compacting.
4. The following day, remove the tube from the nutator and allow the worms to settle to the bottom of the tube. Without disturbing the worm pellet, remove as much of the solution as possible to a fresh tube and save by storing at –20°C or –80°C. The dsRNA solution can be used at least twice more.
5. Using a clean glass pipet, transfer the worms in the remaining solution to a fresh NGM agar plate seeded with *E. coli* OP50.
6. Allow the worms to recover for 6–12 h. Transfer the adult hermaphrodites to new plates every 12 h and monitor their progeny for RNAi-induced phenotypes. The RNAi effects typically persist in embryos laid over the period 12 to 36 h following removal of the adults from the dsRNA solution.

3.3.3 Delivery of dsRNA by Feeding

The following protocol was developed by Timmons and Fire ***(29)***. Ingestion of dsRNA-expressing bacteria may be more like a "slow-drip IV" rather than the single large dose of dsRNA received via microinjection. Feeding works well to target genes expressed at most stages of development, with the exception of late embryonic stages when the nonfeeding embryo is isolated within the egg shell. Males in general and the nervous system of hermaphrodites also appear less susceptible to RNAi by feeding.

1. Perform all DNA manipulations involved in constructing the desired dsRNA-expressing plasmid in DH5α or other standard cloning bacterial strains (*see* **Fig. 1**). Transform the bacterial host HT115(DE3) with the final plasmid construct using standard $CaCl_2$ transformation protocols and plate on LB-agar plates containing tetracycline at 12.5 μg/mL and any other appropriate antibiotic (e.g., ampicillin if plasmid carries ampicillin-resistance gene).
2. Inoculate a 2-mL 2X YT culture containing tetracycline at 12.5 μg/mL with a single colony of HT115(DE3) cells carrying the plasmid. Culture overnight at 37°C.
3. Dilute the culture 100-fold or more and allow to grow to OD_{600} equal to 0.4.
4. To induce T7 polymerase, add IPTG to a final concentration of 0.4 m*M* and incubate the culture with shaking for 2–4 h at 37°C.
5. Supplement with additional tetracycline, IPTG, and additional appropriate antibiotic (e.g., ampicillin if plasmid carries ampicillin-resistance gene). Apply cells directly, or first concentrate by gentle centrifugation, onto NGM/agar plates supplemented with 12.5 μg/mL tetracycline, 0.4 m*M* IPTG, and 50–100 μg/mL ampicillin (if appropriate).
6. Add one to several L4 or young adult hermaphrodites per freshly seeded plate. Incubate between 16°C and 25°C (*see* **Note 6**). Monitor the phenotypes of the F_1 progeny of the plated animals over the next several days. Plates should contain sufficient quantities of bacteria to support nematode growth during the course of the experiment. Do not let the worms deplete the food source. Transfer to freshly seeded plates if necessary.

3.3.4. Producing dsRNA In Vivo

For organisms in which DNA transformation technology has been developed, transgenes can be constructed that when placed into the organism will express dsRNA in vivo. This method has been used extensively for a broad range of hosts ***(31–35)***, rivaling that of direct dsRNA microinjection as a favored form of dsRNA delivery. The technique entails construction of transgenes in which a single promoter drives expression of an inverted repeat, or two promoters drive expression of the sense and antisense strands corresponding to the targeted gene. Such constructs are similar in configuration to those placed into bacterial strains to deliver dsRNA by feeding (*see* **Fig. 1**),

except that promoters other than T7 are used to drive expression. In general it appears that hairpin structures are more effective at producing a robust RNAi response, as sense and antisense strands that are transcribed separately may not anneal efficiently in the environment of the nucleus. One variation is to use an intronic sequence for the spacer DNA; this results in the loop being spliced out of the hairpin RNA leaving just the dsRNA component.

The choice of promoter will depend on the species being transformed and to what extent the researcher wants to manipulate the timing and spatial expression of the dsRNA. Expression may be driven by ubiquitous promoters or by conditional or inducible promoters (e.g., heat shock, GAL4). The inclusion of *cis*-acting control elements that serve to tightly control expression may have special utility in *Drosophila* and similar organisms in which dsRNA does not appear to cross cell boundaries; the use of such elements may allow the experimenter to selectively knock down gene activity in just a subset of cells. But even in organisms such as *C. elegans*, in which dsRNA made in one tissue spreads to others, conditional or stage-specific promoters can be used to drive expression of dsRNAs at specific developmental stages, thereby allowing the researcher to distinguish between the contributions of maternal versus larval gene expression, for example.

3.4. Analysis of RNAi Results

Because of the varying responses elicited by dsRNA, researchers using RNAi to generate knockdown phenotypes should take the following into consideration: (1) Has the gene product been depleted? and (2) Is the observed phenotype specifically due to the loss of the targeted mRNA? The first question can be answered by looking for either loss of the targeted mRNA or the protein product it encodes; a variety of methods are possible, including *in situ* hybridization, Northern or Western blot, quantitative reverse transcriptase-polyacrylamide gel electrophoresis (RT-PCR), RNase protection, or immunocytochemistry if antibodies specific to the affected protein are available. RNAi may result in phenotypes that resemble null mutations or only partial reduction-of-function phenotypes. Absence of a phenotype may be due to ineffective RNAi or may be the true effect of loss of the target. If follow-up with one of the above techniques indicates that targeted mRNAs and the proteins they encode (*see* **Note 7**) have been effectively eliminated, and still no phenotype is observed, the gene under study may be functionally redundant with others or may have subtle loss-of-function phenotypes not uncovered under the laboratory or assay conditions used.

Appropriate controls for determining the specificity of any RNAi-induced phenotype include testing of a "nonsense" dsRNA designed not to target any gene in the organism. The control dsRNA should be prepared in parallel with

the test dsRNA. The organism and/or its progeny should remain unaffected after introduction of the negative control dsRNA. Such a result indicates that neither the dsRNA itself nor the delivery medium is either toxic to the organism or elicits a nonspecific global panic response. Sequences corresponding to reporters such as *green fluorescent protein* (*GFP*) have been frequently chosen to serve as such negative controls. Such reporters, when targeted in transgenic organisms, can also serve as positive controls confirming that production of dsRNA and the method of delivery is sufficient to elicit an RNAi response. For example, introduction of *GFP* dsRNA into a *GFP*-expressing transgenic organism should result in the specific loss of GFP expression, but an otherwise healthy experimental subject. Another strategy worth considering is to use dsRNAs that target separate portions of the gene under study and to determine whether consistent phenocopies are obtained. Determination that the expression of closely related but untargeted genes remain wild type following dsRNA treatment is also a good indication that any observed phenotypes are caused by the specific loss or reduction in gene activity of the target. If a range of phenotypes are observed, can the severity of the phenotype be correlated with the degree of loss of the target? Unlike with traditional antisense approaches, and because of the efficiency of RNAi, dsRNA treatment in general cannot be rescued by injection of translatable mRNAs; therefore in general this control is not useful.

4. Notes

1. The first step in any RNAi experiment is to choose the appropriate region(s) in the gene of interest to target. Ideally, the dsRNA should target between 200 and 2000 bp of the mature mRNA. Shorter sequences may work, but the knockdown may be less potent. We have found that dsRNAs as short as 60 bp will target some genes effectively but others not at all. Because RNAi targets and degrades mRNAs, the dsRNA sequence should correspond to sequences in the mature message. Therefore, derive template DNA for dsRNA synthesis from a cDNA if available. Subcloning genomic sequences can work also, particularly if segments with long exons interrupted with few and short introns are used; dsRNAs containing a few intronic sequences can still be effective, although the intronic sequences in general do not contribute to interference. Typically sequences anywhere within the mRNA (i.e., 5'UTR, coding, 3'UTR) can serve as targets.

   The ability of dsRNAs to effectively interfere with their targets depends on several factors, including extent of complementarity, stability of the target mRNA, and availability of the RNAi machinery. Secondary structure of the targeted region, as well as the presence of proteins that may naturally bind to and "mask" the targeted region may also affect the ability of a dsRNA to cause interference. If possible, target a unique portion of the sequence to avoid inadvertently affecting the expression of closely related genes. In *C. elegans* the ability of

a dsRNA to interfere with mRNAs depends on a combination of overall sequence similarity and the longest stretch of 100% identity between the dsRNA and potential targets; a dsRNA may interfere with targets sharing 100% identity over a short region (approx 25 bp) if embedded in an overall sequence that shares approximately 80% or higher overall sequence identity (Montgomery, M. K., unpublished). Short sequences in the context of a longer sequence with much lower "homology" are probably not effectively targeted because the RNAi machinery is saturable ***(40)***. Double and triple "mutants" can be generated simply by introducing dsRNAs that target two or three different genes; however, attempting to target a larger number of genes simultaneously may overwhelm the RNAi machinery and result in less effective interference ***(40)***.

2. Template DNA for in vitro dsRNA synthesis may be constructed by amplifying a target region using PCR and cloning into a plasmid vector such as Bluescript (pBS) that contains T3 and T7 promoters that flank the multicloning site. PCR primers can be designed to contain linkers to facilitate cloning into such a vector. Alternatively, PCR primers may contain a T3 or T7 promoter sequence and the cleaned up PCR product used directly as the template DNA for the in vitro reaction. (T3 sequence is ATTAACCCTCACTAAAGGGA and T7 is TAATACGACTCACTATAGGG.) If this second method is chosen, avoid using the same promoter sequence on both the upstream and downstream primers; use of different promoters on the sense and antisense primers will aid sequencing of the PCR product (e.g., use T3 on the sense primer and T7 on the antisense primer). Even if cloned products appear to be the right size, it is highly recommended that all cloned products be sequenced before use in an RNAi experiment. Bring the PCR mixture or purified PCR product up to a total of 100 μL with TE and proceed to **step 2** of the protocol.
3. Alternatively, you can add T3 and T7 RNA polymerases to the same microfuge tube. This results in synthesis of both sense and antisense RNAs and annealing of the strands in a single step. We have used this shortcut frequently and have found that annealing works efficiently under reaction conditions. It is still a good idea to run the dsRNAs against previously made ssRNA preparations and look for a shift in mobility, as otherwise it would be difficult to determine if either the T3 or T7 RNA polymerase is working suboptimally.
4. The 20-μL reaction can be scaled up to synthesize larger amounts of RNA. The 1X Stop can be used at a lower concentration, comprising as little as one-half the total aqueous volume (e.g., if scaled up to 200 μL in vitro transcription reaction, add 200 μL 1X Stop for total volume of 400 μL aqueous prior to phenol:chloroform extraction). The final concentration of ammonium acetate is still sufficient for ethanol precipitation. All steps can be carried out at room temperature; newly synthesized RNAs but few unincorporated nucleotides remain in the final pellet.
5. In *C. elegans* adult hermaphrodites, injection into essentially any tissue of the adult will result in dsRNA transport to other tissues, including the gonad and newly fertilized embryos. Microinjection results in the delivery of a bolus of dsRNA that is then distributed throughout the tissues of the worm and incorpo-

rated into newly fertilized eggs. This method of delivery works particularly well for targeting genes that are active in the early embryo, such as maternally and zygotically expressed genes. Genes expressed in larval and adult stages can also be targeted, but typically less efficiently as the original dsRNA material becomes more and more diluted with each cell division. In *Drosophila*, injection must be into syncytial-stage embryos, as dsRNA does not have the same ability to cross cell boundaries as has been observed in *C. elegans*. The amount of dsRNA that needs to be injected depends upon several factors, including size (in bp) and concentration of the dsRNA, volume of the embryo or organism being injected, concentration and turnover rate of the target mRNA, and whether the target mRNA is present at the time of injection or is transcribed at a later stage of development. Typical amounts of injected dsRNAs range from one to hundreds of nanograms. If using this technique for the first time, a range of dilutions should be tested to determine the optimal concentration and volume of dsRNA to inject.

6. Maintaining *C. elegans* at the lower end of its optimal growth temperature range (e.g., 16°C) following dsRNA microinjection or during dsRNA feeding experiments may enhance RNAi phenotypes. It is not yet known if this applies to other nematode species.
7. RNAi and similar PTGS processes affect stability of mRNAs. Proteins that are exceptionally stable (i.e., long-lived) and present at the time of dsRNA treatment may remain functional over the course of an RNAi experiment. Thus, absence of an observed phenotype may be due to perdurance of the protein. Therefore, when possible, determine to what extent the targeted protein levels have been reduced using antibody staining or Western blot techniques.

References

1. Fire, A., Xu, S., Montgomery, M. K., Kostas, S., Driver, S., and Mello, C. (1998) Potent and specific genetic interference mediated by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391,** 806–811.
2. Kennerdell, J. R. and Carthew, R. W. (1998) Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the Wingless pathway. *Cell* **95,** 1017–1026.
3. Brown, S. J., Mahaffey, J. P., Lorenzen, M. D., Denell, R. E., and Mahaffey, J. W. (1999) Using RNAi to investigate orthologous homeotic gene function during development of distantly related insects. *Evol. Dev.* **1,** 11–15.
4. Hughes, C. L. and Kaufman, T. C. (2000). RNAi analysis of *Deformed, proboscipedia* and *Sex combs reduced* in the milkweed bug *Oncopeltus fasciatus*: novel roles for Hox genes in the Hemipteran head. *Development* **127,** 3683–3694.
5. Sanchez Alvarado, A. and Newmark, P. A. (1999) Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc. Natl. Acad. Sci. USA* **96,** 5049–5054.
6. Baker, M. W. and Macagno, E. R. (2000) RNAi of the receptor tyrosine phosphatase *HmLAR2* in a single cell of an intact leech embryo leads to growth-cone collapse. *Current Biology* **10,** 1071–1074.

7. Lohmann, J. U., Endl, I., and Bosch, T. C. G. (1999) Silencing of developmental genes in *Hydra*. *Developmental Biology* **214,** 211–214.
8. Montgomery, M. K., Xu, S., and Fire, A. (1998) RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **95,** 15,502–15,507.
9. Jorgensen, R. (1990) Altered gene expression in plants due to trans interactions between homologous genes. *Trends Biotechnol.* **8,** 340–344.
10. Vaucheret, H., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., et al. (1998) Transgene-induced gene silencing in plants. *Plant J.* **16,** 651–659.
11. Romano, N. and Macino, G. (1992) Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol. Microbiol.* **6,** 3343–3353.
12. Montgomery, M. K. and Fire, A. (1998) Double-stranded RNA as a mediator in sequence-specific gene silencing and co-suppression. *Trends in Genetics* **14,** 255–258.
13. Maine, E. M. (2000) A conserved mechanism for post-transcriptional gene silencing? *Genome Biol.* **1,** 10–18.
14. Cogoni, C. and Macino, G. (1999) Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399,** 166–169.
15. Fagard, M., Boutet, S., Morel, J. B., Bellini, C., and Vaucheret, H. (2000) AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl. Acad. Sci. USA* **97,** 11,650–11,654.
16. Smardon, A., Spoerke, J. M., Stacey, S. C., Klein, M. E., Mackin, N., and Maine, E. M. (2000) EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. *Curr. Biol.* **10,** 169–178.
17. Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., and Plasterk, R. H. (1999) *Mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99,** 133–141.
18. Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., et al. (1999) The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99,** 123–132.
19. Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., et al. (2000) *Arabidopsis SGS2* and *SGS3* genes are required for post-transcriptional gene silencing and natural virus resistance. *Cell* **101,** 533–542.
20. Kaufman, R. J. (1999) Double-stranded RNA-activated protein kinase mediates virus-induced apoptosis: a new role for an old actor. *Proc. Natl. Acad. Sci. USA* **96,** 11,693–11,695.
21. Wianny, F. and Zernicka-Goetz, M. (2000) Specific interference with gene function by double-stranded RNA in early mouse development. *Nat. Cell Biol.* **2,** 70–75.
22. Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411,** 494–498.

23. Bass, B. L. (2000) Double-stranded RNA as a template for gene silencing. *Cell* **101,** 235–238.
24. Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409,** 363–366.
25. Gonczy, P., Echeverri, G., Oegema, K., Coulson, A., Jones, S. J., Copley, R. R., et al. (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408,** 331–336.
26. Nakano, H., Amemiya, S., Shiokawa, K., and Taira, M. (2000). RNA interference for the organizer-specific gene *Xlim-1* in *Xenopus* embryos. *Biochem. Biophys. Res. Commun.* **274,** 434–439.
27. Tabara, H., Grishok, A., and Mello, C. C. (1998) RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282,** 430–431.
28. Timmons, L. and Fire, A. (1998). Specific interference by ingested dsRNA. *Nature* **395,** 854.
29. Timmons, L., Court, D. L., and Fire, A. (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263,** 103–112.
30. Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. USA* **95,** 14,687–14,692.
31. Bastin, P., Ellis, K., Kohl, L., and Gull, K. (2000) Flagellum ontogeny in trypanosomes studied via an inherited and regulated RNA interference system. *J. Cell Sci.* **113,** 3321–3328.
32. Waterhouse, P. M., Graham, M. W., and Wang, M. (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl. Acad. Sci. USA* **95,** 13,959–13,964.
33. Chuang, C. F. and Meyerowitz, E. M. (2000) Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **97,** 4985–4990.
34. Tavernarakis, N., Wang, S. L., Dorovkov, M., Ryazanov, A., and Driscoll, M. (2000) Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. *Nat. Genet.* **24,** 180–183.
35. Kennerdell, J. R. and Carthew, R. W. (2000) Heritable gene silencing in *Drosophila* using double stranded RNA. *Nat. Biotechnol.* **18,** 896–898.
36. Takiff, H. E., Chen, S-M., and Court, D. L. (1989) Genetic analysis of the *rnc* operon of *E. coli. J. Bacteriol.* **171,** 2581–2590.
37. Dasgupta, S., Fernandez, L., Kameyama, L., Inada, T., Nakamura, Y., Pappas, A., et al. (1998) Genetic uncoupling of the dsRNA-binding and RNA cleavage activities of the *Escherichia coli* endoribonuclease RNaseIII-effect of dsRNA binding on gene expression. *Molec. Microbiol.* **28,** 629–640.
38. Mello, C. C. and Fire, A. (1995) DNA transformation. In *Caenorhabditis elegans*: *Modern Biological Analysis of an Organism* (Epstein, H. R. and Shakes, D. C., eds). *Methods Cell Biol.* **48,** 451–482.

39. Subramaniam, K. and Seydoux, G. (1999) *nos-1* and *nos-2*, two genes related to *Drosophila nanos*, regulate primordial germ cell development and survival in *Caenorhabditis elegans*. *Development* **126,** 4861–4871.
40. Parrish, S., Fleenor, J., Xu, S., Mellow, C., and Fire, A. (2000) Functional anatomy of a dsRNA trigger: differential requirements for the two trigger strands in RNA interference. *Molec. Cell* **6,** 1077–1087.

11

The Application of LTR Retrotransposons as Molecular Markers in Plants

Alan H. Schulman, Andrew J. Flavell, and T. H. Noel Ellis

Summary

Retrotransposons are major, dispersed components of most eukaryotic genomes. They replicate by a cycle of transcription, reverse transcription, and integration of new copies, without excising from the genome in the process. Because they represent a major share of the genome, cause easily detectable genetic changes having known ancestral and derived states, and contain conserved regions for which polymerase chain reaction (PCR) primers may be designed, retrotransposon insertions can be exploited as powerful molecular marker systems. Here, we describe the background and strategies, as well as give detailed laboratory protocols, for four key retrotransposon-based methods: SSAP, IRAP, REMAP, and RBIP. The SSAP, IRAP, and REMAP methods are multiplex and generate anonymous marker bands; RBIP scores individual loci, much as microsatellite-based marker systems do. The methods are variously suited to marker detection on agarose and polyacrylamide slab gels, slab and capillary sequencing devices, and arrays on solid supports. The different strengths and weaknesses of these approaches and their performance relative to conventional marker methods are discussed, together with their applicability to marker-assisted breeding, phylogenetic analyses, biodiversity determinations, and evolutionary studies.

Key Words: Retrotransposon; molecular marker; biodiversity; marker-assisted breeding; molecular evolution; genetic fingerprinting; pedigree analysis; polymorphism detection; transposable element.

1. Introduction

Markers, entities which are heritable as simple Mendelian traits and are easy to score, have long been important in studies of inheritance and variability, in the construction of linkage maps, and in the diagnosis of individuals or lines carrying certain linked genes. Phenotypic and biochemical (enzyme) markers tend to have the disadvantages of a low degree of polymorphism, limiting their ability to be mapped in crosses; relatively few loci, limiting the density of maps

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

which can be produced; and environmentally variable expression, complicating scoring and the determination of genotype. These marker types have been superceded by DNA-based methods that generate "fingerprints" or "molecular markers," distinctive patterns of DNA fragments resolved by electrophoresis in agarose or acrylamide gels and detected by staining or labeling. A molecular marker is in essence a nucleotide sequence corresponding to a particular physical location in the genome. Its sequence needs to be polymorphic enough to allow its pattern of inheritance to be easily followed.

1.1. Molecular Markers

Restriction fragment length polymorphism (RFLP) was the first DNA-based molecular marker technique and was an outgrowth of the development of gene cloning and filter hybridization methods in the 1970s. The polymorphisms it exploits are the presence or absence of restriction sites in genomic sequences for which a cloned hybridization probe exists. Originally, RFLP analysis required Southern blotting and hybridization ***(1)***. The RFLP method is still used to generate widely shared "anchor" markers, which are those used by many researchers to combine segregation data from different experiments onto recombinational maps, although it suffers from laboriousness as well as from a paucity of alleles and loci. The advent of the polymerase chain reaction (PCR) made possible the detection of variation in randomly amplified polymorphic DNAs (RAPDs) ***(2)***. The RAPDs are indeed rapid, being independent of the need for sequence data, but they suffer from low polymorphism information content (PIC), poor correlation with other marker data, and problems in reproducibility resulting from the low annealing temperatures in the reactions.

Around 1990, methods that detect variability in the number of simple sequence repeats (SSRs) in microsatellites ***(3)***, or that measure variability in the occurrence of two microsatellites close to one another ***(4)***, were developed for plants. In the mid-1990s, the amplified fragment length polymorphism method (AFLP®) was introduced. The AFLP approach is a conceptual hybrid between RFLP and the PCR methods because, whereas the method is PCR-based, its polymorphism is derived from variations in restriction site occurrence or digestibility ***(5)***.

The polymorphism detected by the foregoing methods for generating molecular markers are primarily those of small sequence variations. The RFLP and AFLP methods detect polymorphisms in restriction sites, typically comprising 4 to 6 bp. Although insertions or deletions within a restriction fragment would also generate an RFLP or AFLP polymorphism, the resolution limits of gel electrophoresis restricts insertions that can be scored to several kilobases in length. Polymorphisms in RAPDs primarily affect the ability of the 9 or 10 nt primers to anneal efficiently under the reaction conditions of particular experi-

ments. Microsatellite alleles are generated by the gain or loss of repeat units of only a few base pairs. These changes are, furthermore, bidirectional in the sense that further mutations can restore a restriction site or primer binding site. This bidirectionality reduces the usefulness of these marker systems in resolving phylogenies and pedigrees.

An ideal molecular marker technique would exploit large physical changes in a genome to visualize genetic diversity. The loci scored by the method should be spread throughout the genome at high frequency, enabling dense and well-distributed recombinational maps to be generated. Such a method should be universal in its application, with low investment required for marker development in any particular species. Generation of the marker pattern should be robust and reproducible, and detection should be inexpensive and technically straightforward. Retrotransposons, described below, meet many of these requirements and have been recently developed as molecular marker systems. After providing an introduction to retrotransposons as biological phenomena, the main marker techniques currently applied to retrotransposons will be described in detail.

1.2. LTR Retrotransposons

Retrotransposons are an abundant class of mobile genetic elements ***(6)***. They have little in common with the Class II transposable elements, which are DNA transposons such as *Ac*, *En/Spm*, or *Mutator,* or with MITEs such as *Stowaway*. Unlike the DNA transposons, the retrotransposons do not excise as a concomitant part of their invasion of new loci in the genome, but instead enter new loci as copies of the mother element, which remains fixed in the genome. Retrotransposons fall into two clearly separated groups: the long terminal repeat (LTR)-containing elements ***(7,8)***; those lacking LTRs, the long interspersed elements (LINEs); and short interspersed elements (SINEs) ***(9)***.

Retrotransposons share many similarities with the retroviruses in their organization, in the gene products they encode, and in the steps of their life cycles. Both retrotransposons ***(6)*** and retroviruses ***(10)*** propagate through cycles of successive transcription, reverse transcription, and genomic integration. Their extensive similarities suggest that present-day retrotransposons and retroviruses are derived from a common ancestor ***(11,12)***. The extant retroviruses can be distinguished by their possession of an *envelope* (*env*) domain encoding a glycoprotein necessary for infective passage from cell to cell through the plasma membrane. The related defective elements in humans are called human endogenous retroviruses (HERVs) rather than retrotransposons ***(13)***. Among the LTR retrotransposons (**Fig. 1**), the *gypsy*-like group is most similar in sequence and organization to retroviruses, whereas the *copia*-like elements share the same coding domains but in a different linear order. The *gypsy* family

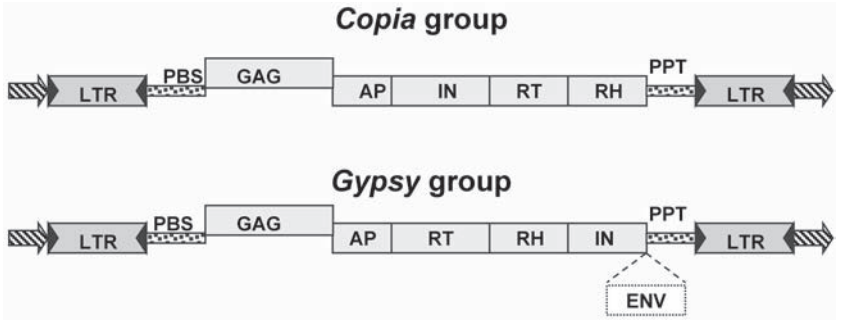


Fig. 1. Organization of LTR retrotransposons. The elements can be classified into two major groups, the *copia*-like and the *gypsy*-like elements, named after the type members of *Drosophila melanogaster*. The elements are flanked by 5-bp direct repeats in the host DNA (hatched arrows), formed by the integration of the element. The retrotransposons consist of two LTRs that contain short inverted repeats at their edges (dark triangles) and that bound the internal domain. Adjacent to the 5' and 3' LTRs are the primer binding site (PBS) and the polypurine tract (PPT), which respectively are responsible for priming the synthesis of the (–)-strand and (+)-strand of the cDNA. The protein coding domain is generally synthesized as a polyprotein, and it contains the Gag domain, which encodes the protein forming the capsid of the virus-like particle; the aspartic proteinase (AP), which cleaves the polyprotein into functional units; the integrase (IN), which inserts the cDNA copy into the genome; and the reverse transcriptase (RT) and RNase H (RH), which synthesize the cDNA from the RNA transcript. The Gag protein may be expressed in some elements in a different reading frame, and it is shown shifted upwards to reflect this. *Gypsy*-like elements differ from *copia*-like elements in the position of the IN domain. Some retrotransposons may contain, as do retroviruses, an envelope (ENV) domain generally expressed in a separate reading frame. The figure is not drawn to scale.

of elements from *Drosophila melanogaster* is in fact transitional between retroviruses and retrotransposons and can be infective under experimental conditions ***(14)***.

1.3. Retrotransposons and the Genome

Retrotransposon transcripts each have the formal potential to be reintegrated into the genome as cDNA copies, which can then serve as further sources of transcripts. The newly integrated retrotransposon copies can be inherited if they are present in cells ultimately giving rise to gametes. In view of the many somatic cell divisions that take place prior to the differentiation of germ cells in plants, it is not totally surprising that retrotransposons have succeeded in becoming major genomic components. In plants with large genomes, retrotransposons are the major class of repetitive DNA and can comprise 40–60% of the

genome as a whole ***(15–17)***. The major families of retrotransposons are, with a few exceptions, dispersed throughout the chromosomes in the plant species examined ***(18–20)***. In some cases, retrotransposon copy number increases appear to have been a major factor in genome size growth in the plants ***(21,22)***. Not only are retrotransposons highly prevalent within the genomes of plants, but also both *copia*-like ***(23,24)*** and *gypsy*-like ***(25)*** retrotransposons are ubiquitous throughout the plant kingdom.

1.4. Retrotransposons as Molecular Markers

The ubiquitous nature of retrotransposons and their activity in creating genomic diversity by stably integrating large DNA segments into dispersed chromosomal loci make these elements ideal for development as molecular markers. Integration sites shared between germplasm accessions are highly likely to have been present in their last common ancestor. Therefore, retrotransposon insertional polymorphisms can help establish pedigrees and phylogenies ***(26,27)*** as well as serve as biodiversity indicators.

In recent years, several molecular marker methods based on retrotransposons have been developed ***(28)*** and they are presented in detail below (**Fig. 2**). All rely on the principle that a joint is formed, during retrotransposon integration, between genomic DNA and the retrotransposon. These joints may be detected by amplification between a primer corresponding to the retrotransposon and a primer matching a nearby motif in the genome. The methods have been named according to the particular motif that provides the second priming site. The Sequence-Specific Amplified Polymorphism (SSAP) method (**Figs. 2A** and **3**), the first retrotransposon-based method to be described, amplifies products between a retrotransposon integration site and a restriction site to which an adapter has been ligated. In Inter-Retrotransposon Amplified Polymorphism (IRAP) (*see* **Figs. 2B** and **4**), segments between two nearby retrotransposons or LTRs are amplified. The Retrotransposon-Microsatellite Amplified Polymorphism (REMAP) (*see* **Figs. 2C** and **5**) technique detects retrotransposons integrated near a microsatellite or stretch of SSRs. The Retrotransposon-Based Insertional Polymorphism (RBIP) (*see* **Figs. 2D** and **6**) marker system, in contrast to the others, detects a given locus in both alternative states, namely, empty and occupied by a retrotransposon, by using both flanking primers and a retrotransposon primer.

Although these methods are presented here as examples with primers specific to a particular family of retrotransposons, it is important to note that retrotransposon marker methods are generic. Any organism in which retrotransposons are dispersed components of the genome, and in which they have been active over a timescale relevant to the question being asked, can be examined with retrotransposon markers. A couple of direct comparisons of

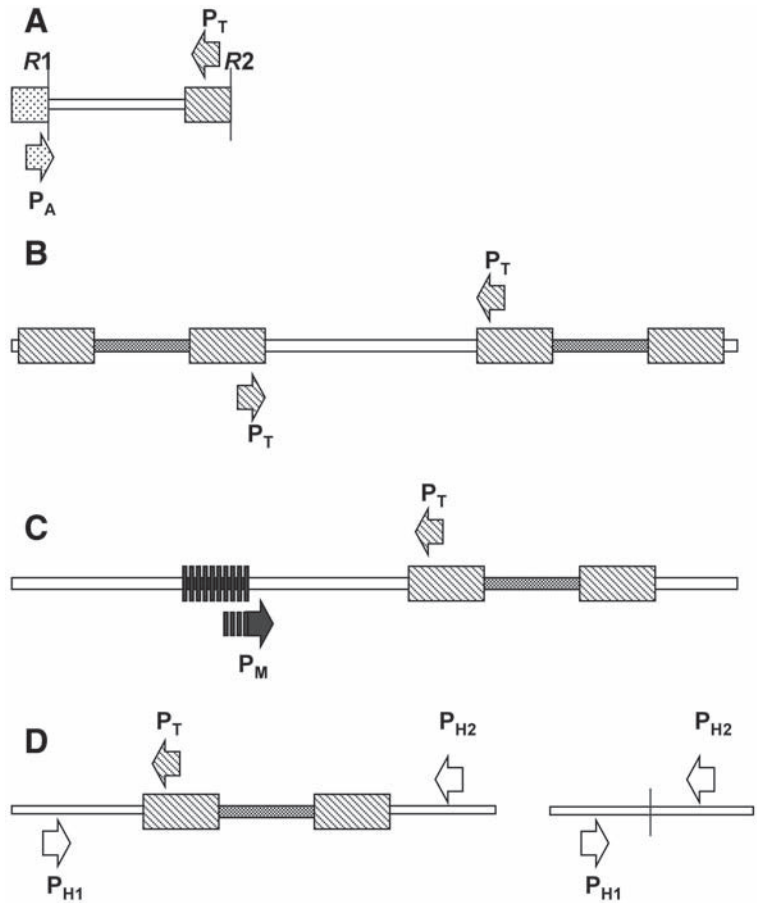


Fig. 2. Marker methods based on LTR Retrotransposons. **(A)** Sequence-specific amplified polymorphism (SSAP). The DNA template is digested by two restriction enzymes (*R1*, *R2*), an adapter ligated (stippled box), and fragments sharing both a retrotransposon region and restriction site *R*1 amplified by PCR with adapter primers (P_A) and retrotransposon primers (P_T). **(B)** Inter-retrotransposon amplified polymorphism (IRAP). Regions of the genome flanked by two retrotransposons are amplified by PCR using either two identical or two different retrotransposon primers (P_T). **(C)** Retrotransposon-microsatellite amplified polymorphism (REMAP). Regions of the genome flanked by a microsatellite domain (left) and a retrotransposon are amplified by PCR using primers containing simple sequence repeats with 3' anchor nucleotides (P_M) and retrotransposon primers (P_T). **(D)** Retrotransposon-based insertional polymorphism (RBIP). Individual sites that are polymorphic for retrotransposon insertion can be detected by PCR in both allelic states, full (left) and empty (right). To detect the presence of the retrotransposon, primers specific to the host DNA on one side of the integrated element (P_{H1}) are used together with a retrotransposon primer (P_T). To detect the empty site, primers to the two host flanks are combined (P_{H1}, P_{H2}).

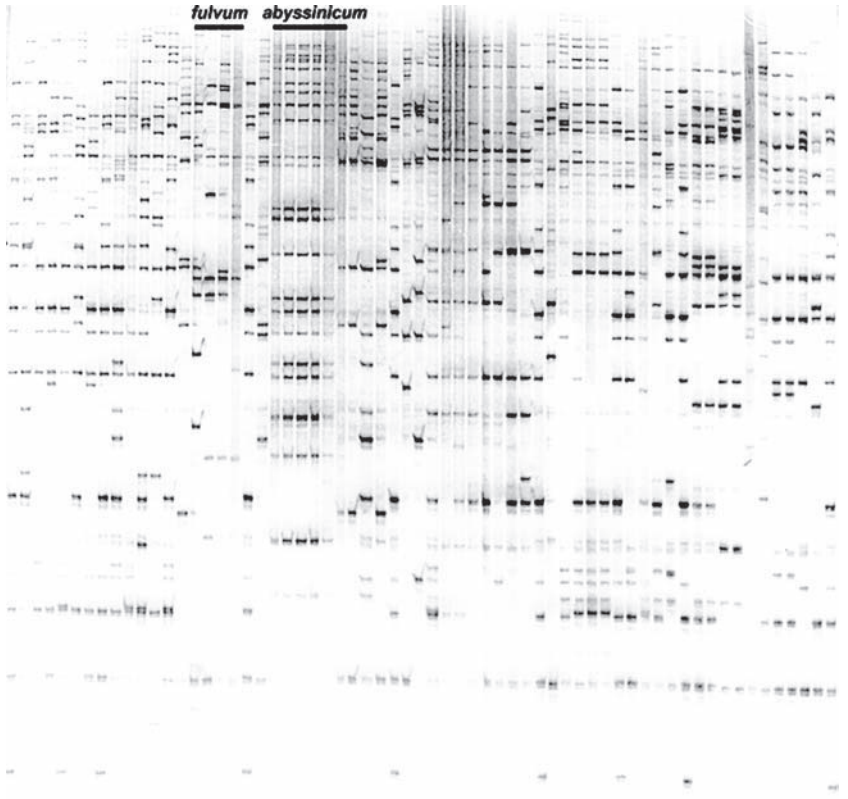


Fig. 3. SSAP Analysis. The figure shows an autoradiograph of a sequencing gel resolving SSAP products. Products were generated from a set of *Pisum* accessions (lanes) using a ^{33}P-labeled PCR primer specific to the PPT of the *Pisum* retrotransposon *PDR*1 and a primer, with selective bases TT, matching a *Taq* I restriction site adapter. The first set of lanes are *P. sativum* accessions, the set labeled ***fulvum*** are *P. fulvum*, and the set labeled ***abyssinicum***, *P. abyssinicum*. The other lanes contain accessions of various *Pisum* species. From **ref. *34*** with permission.

retrotransposon marker methods with AFLP indicate that the retrotransposon markers are some 25% more polymorphic ***(29,30)***. In principle, retrotransposon-based or retrovirus-based molecular markers could prove highly useful in animals, including mammals and birds.

1.5. Sequence-Specific Amplified Polymorphism (SSAP)

Sequence-specific amplified polymorphism (SSAP) was described by Waugh and coworkers in 1997 ***(30)***, but it has several origins and embodiments ***(31–34)***. The SSAP method can be considered to be a modification of AFLP ***(5)***, or as a variant of anchored PCR ***(32)***. The method described by

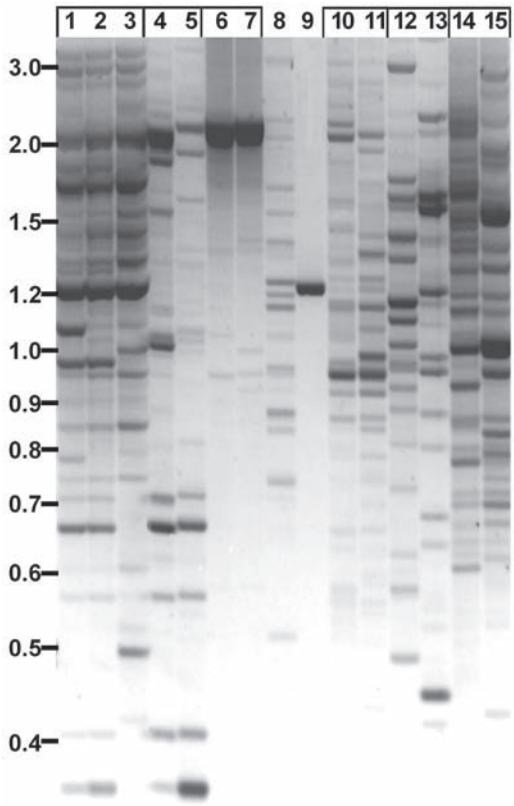


Fig. 4. IRAP analysis. IRAP amplification products from various cereals using *BARE*-1 primers is displayed. The gel has been ethidium bromide-stained, and the fluorescence detected with UV light; a negative image is shown. This example shows that even heterologous primers can produce useful IRAP amplification. The samples represented are: bread wheat (*Triticum aestivum*) var. Tjalve (*lane 1*), bread wheat var. Mahti (*lane 2*), durum wheat (*T. durum*) (*lane 3*), *Aegilops tauschii* line 1691 (*lane 4*), *A. tauschii* line 1704 (*lane 5*), rye (*Secale cereale*) line P105 (*lane 6*), rye line P87 (*lane 7*), oat var. (*Avena sativa*) Veli (*lane 8*), rice var. (*Oryza sativa*) IRRI52886 (*lane 9*), timothy (*Phleum pratense*) line 22 (*lane 10*), timothy line 16 (*lane 11*), *Spartina maritima* (*lane 12*), *S. alterniflora* (*lane 13*), *Leymus arenarius* (*lane 14*), *Leymus mollis* (*Lane 15*). Marker sizes in kb are indicated on the left axis. From **ref. *40*** with permission.

Waugh and colleagues ***(30)*** has many similarities to AFLP, especially in that two different enzymes are used to generate the template for the specific primer PCR, and in that selective bases are used in the adapter primer.

Two implementations of SSAP (**Fig. 2A**) are described below. The first (**Fig. 3**) was designed for use with a retrotransposon found in relatively few copies

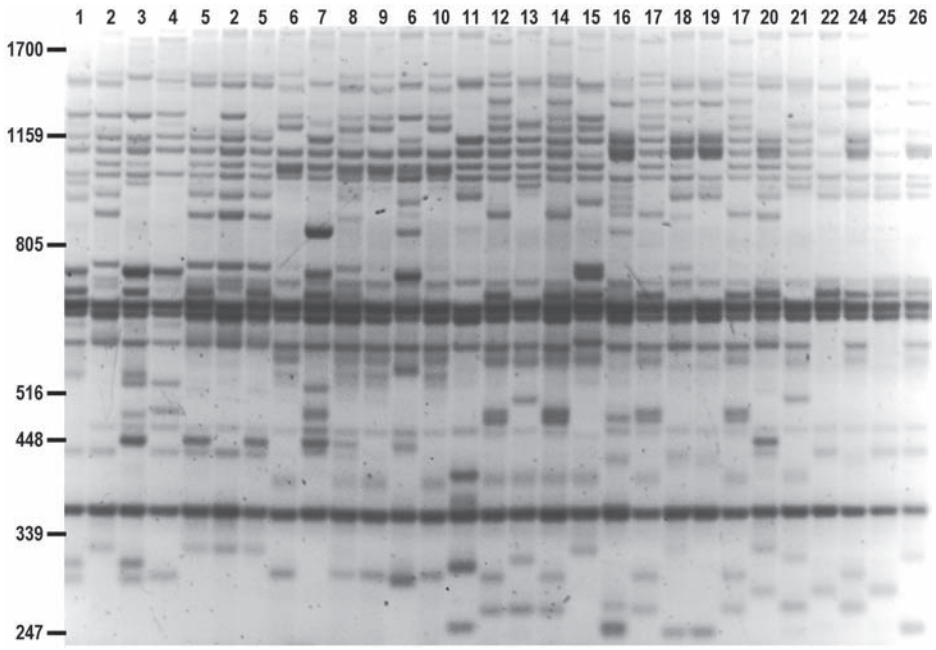


Fig. 5. REMAP analysis. A gel is shown of REMAP amplification products from *Hordeum spontaneum* using *BARE*-1 primers. The 26 genotypes shown (gel lanes) can be distinguished by their *BARE*-1 insertion patterns. The REMAP system is useful for such population studies as well as for cultivar distinction. The banding pattern has been detected as in Fig. 4. Size markers in bp derive from a bacteriophage λ *Pst*I digest. From **ref. *47*** with permission.

(*see* **Subheadings 2.1.** and **3.1.**). In this procedure it is important to maximize the sequence complexity of the template for the specific primer amplification, so a single enzyme digestion is used ***(34)***. As with the method described for *BARE*-1 ***(30)***, the adapter primer is selective. This is a matter of convenience, and nonselective primers could be substituted when the enzyme used for digestion has a larger recognition sequence, or when the copy number is lower. In general, LTR ends are convenient for the design of SSAP primers ***(35)***. However, in the case of *PDR1,* the LTR is exceptionally short at 156 bp ***(36)***, so a GC-rich primer could be designed corresponding to the polypurine tract (PPT) which is found internal to the 3' LTR in retrotransposons. The second implementation is for *BARE*-1 in barley (*see* **Subheadings 2.2.** and **3.2.**), based on the published method ***(30)***. For *BARE*-1 and other high-copy-number families, the number of selective bases may be increased compared to the first version of the protocol. Furthermore, *BARE*-1 and most other retrotransposons

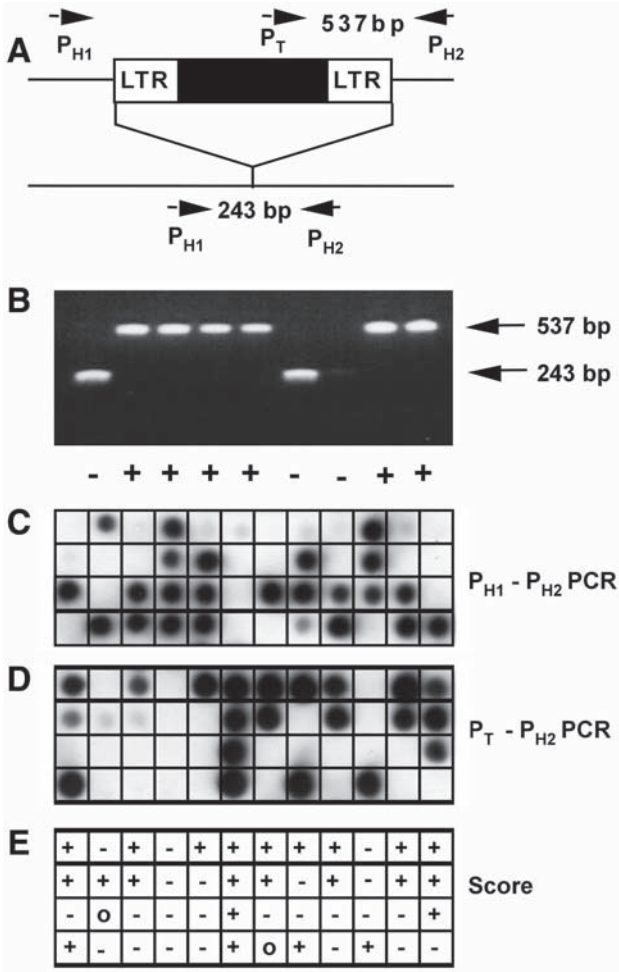


Fig. 6. RBIP Analysis. **(A)** Agarose gel electrophoresis products of RBIP PCRs containing two host-specific primers (P_{H1} and P_{H2}) and a retrotransposon-specific primer P_T. Only one of the two possible products is produced per sample, and the size indicates the product and hence the state of each locus. **(B)** Gel of PCR reaction of genotypes polymorphic for the retrotransposon insertion diagramed in **(A)**. **+** refers to the occupied (retrotransposon-containing) allele, **–** to the unoccupied (retrotransposon-lacking) allele. **(C)** RBIP dot analysis. Altogether 48 DNA samples were subjected to PCR with host specific primers P_{H1} and P_{H2}, then dotted onto nylon membrane and hybridized to a radiolabeled P_{H1}–P_{H2} probe. Samples producing a signal correspond to an unoccupied locus. **(D)** The same samples were subjected to PCR with the transposon-specific primer P_T and host-specific primer P_{H2}, then treated as above. Samples producing a signal correspond to an occupied locus. **(E)** The deduced scores for the 48 scored at accessions at one locus. **+** indicates the occupied site, **–** indicates the unoccupied site, and **o**, no score (failed PCR or absence of both alleles).

have long LTRs, necessitating an anchor primer in the LTR near to the external terminus.

Several features of the first protocol are specific to *PDR1*, but could be used with other retrotransposons of similar structure and copy number. The main feature of the procedure that should be modified for other situations is the location of the sequence-specific primer ***(35)***. The choice of this primer is critical, and it can be modified according to need. For example, internal primer sites have been exploited to describe structural variation within retrotransposons ***(37)***, and the primers can be applied to defined sequences other than the LTR or PPT.

1.6. Inter-Retrotransposon Amplified Polymorphism (IRAP)

IRAP (**Fig. 2B** and **Fig. 4**) detects two retrotransposons or LTRs sufficiently close to one another in the genome to permit PCR amplification of the intervening region. Unlike AFLP or SSAP, the method requires only intact genomic DNA as the template and PCR reagents and apparatus for amplification. There are no restriction enzyme digestion or adapter ligation steps. The amplification products are generally resolved by electrophoresis in wide-resolution agarose gels, but if labeled primers are used, sequencing gel systems may be employed. The amplified fragments range from under 100 bp to over several kilobase pairs, with the minimum size depending on the placement of the PCR priming sites with respect to the ends of the retrotransposon.

The IRAP method ***(38)*** has found applications in gene mapping in barley ***(39)*** and wheat ***(39a)*** and in studies of genome evolution in the grasses ***(40)***. We have adapted it as well for maize and soybean (Schulman, A. H., et al., unpublished). Even given a large genome and a highly prevalent retrotransposon family, one would not expect the IRAP method to produce very many resolvable PCR products. Taking the *BARE*-1 elements in barley as an example, the genome is approx 4.7×10^9 bp in size ***(41)***, and the retrotransposon family is present in approx 1.5×10^4 full-length copies in addition to 1.7×10^5 solo LTRs ***(22)***. The full-length *BARE*-1 is 8932 bp and the LTRs are 1809 bp, comprising a total of 4.4×10^8 bp in the genome and leaving 4.3×10^9 bp of the genome *not BARE*-1. Were insertions to be random within the genome, they would be expected to follow a Poisson distribution. If the total of 1.85×10^5 intact *BARE*-1 elements and solo LTRs were equidistantly dispersed within the remaining, non-*BARE* part of the genome, they would be situated on average roughly 23 kb apart, with most insertions too far from another for PCR and beyond the resolution of conventional agarose gel electrophoresis. The method, however, does produce a range of subkilobase fragments, in part because barley ***(17,42)*** and at least other grass genomes are organized into gene-rich islands surrounded by seas of repetitive DNA. The retrotransposons, which comprise large portions of the repeat seas, tend to be nested, one inserted into another,

in barley, maize, and other grasses ***(16,17,40)***. The IRAP amplification products can derive, therefore, variously both from nearby solo LTRs and full-length elements interspersed with nonretrotransposon DNA and from nested retrotransposons.

The example given below is for the *BARE*-1 element in barley (*see* **Subheadings 3.2.** and **3.3.**). However, the method is applicable at least to any grass ***(40)***, functions as well in other monocots ***(42a)***, and may be useful in other genomes with structures similar to that of the grasses.

1.7. Retrotransposon-Microsatellite Amplified Polymorphism (REMAP)

REMAP (**Fig. 2C** and **Fig. 5**) is conceptually similar to IRAP, but it differs in that it detects polymorphisms in the presence of retrotransposons or LTR derivatives sufficiently near simple sequence repeats (SSRs), often referred to as microsatellites, to allow PCR amplification. Microsatellites are ubiquitous features of eukaryotic genomes, and they have served directly to generate molecular markers in many plants ***(4,43,44)***. For this reason, we were interested in determining whether retrotransposons were associated with microsatellites in the genome, and to what extent such associational polymorphism could serve as molecular markers. We found ***(38)*** that indeed, for *BARE*-1 in barley, retrotransposon insertions near microsatellites are considerably polymorphic. This result was later confirmed by others ***(45)***.

The REMAP method combines outward-facing LTR primers of the sort used in IRAP with SSR primers containing a set of repeats and one or more nonrepeat nucleotides at the 3' end to serve as an anchor. The anchor is necessary to provide specificity to the PCR amplification; otherwise, the repetitive structure of the primer might cause it to anneal in multiple positions in any given microsatellite. Both IRAP and REMAP consist of PCR carried out on undigested template DNA and resolve the products on agarose gels. Following the initial publication of the technique by us ***(38)***, and almost simultaneously by others ***(46)*** under the guise "*copia-SSR*," REMAP has been used to examine genome evolution in wild barley ***(47)***, to map a major resistance gene in barley ***(39)***, and as a sensitive method for detecting genomic copies of retrotransposons amidst retrotransposon cDNAs ***(48)***. The implementation in **Subheadings 2.4.** and **3.4.** is for *BARE*-1, which is useful in a range of cereals and grasses, but the method is generic and could be applied to other plants as well.

1.8. Retrotransposon-Based Insertional Polymorphism (RBIP)

1.8.1. Overview

RBIP (**Fig. 2D** and **Fig. 6**) is in essence the simple PCR-based detection of retrotransposon insertions using PCR between primers flanking the insertion site and primers from the insertion itself. A complementary reaction using

primers from the surrounding DNA alone detects the unoccupied site (**Fig. 6A**). Because retrotransposon insertions are thousands of bases in length, the "unoccupied site PCR" produces no product from an occupied site. The particular feature of RBIP that distinguishes it from the other retrotransposon-based marker methods described in this chapter is that it is a single locus, co-dominant technique.

RBIP is a robust technique. For low numbers of samples, the products are detected by normal agarose gel electrophoresis (**Fig. 6B**). Both reactions are carried out in the same tube, and the size of the PCR product indicates which allele (occupied or unoccupied) has been amplified. The technical problems with this basic RBIP method are all associated with the acquisition of the sequence information for the flanking primers. This situation is closely analogous to the collection of new flanking sequences for microsatellite or SSR markers. Sequence data for new RBIP markers may be obtained from sequence analysis of genomic clones. Alternatively, SSAP markers can be converted into RBIP markers.

The basic RBIP method can be automated by adopting a dot-based assay (**Fig. 6C–E**) to replace gel electrophoresis ***(49)***. In this case, the occupied-site and unoccupied-site PCRs are carried out separately, and the products are dotted onto nylon membrane and probed with a locus-specific probe. This approach avoids a size-separation step, and can be scaled up to handle many thousands of plant samples by robotic spotting. Production of the raw marker data (hybridization signals) is independent of sample number, and data capture and processing, using the technology developed for scoring microarrays ***(50)***, allow the full automation of the process. Such modifications have been developed ***(50a)***.

1.8.2. Converting Other Retrotransposon Markers Into RBIP Markers

In principle, a marker from any of the systems discussed above (SSAP, IRAP, REMAP) can be converted into a corresponding RBIP marker and *vice versa*. Markers from the former set of techniques are very easy to obtain, and they can be rapidly prescreened for their potential informativeness before investing in the effort of developing a corresponding RBIP marker. An SSAP electrophoresis band represents one side of the insertion. It is easy to cut out these bands from a gel, amplify the fragments by PCR, and sequence them to obtain the sequence of one side of the insert. This process, however, is insufficient to allow the detection of the unoccupied site, which is a major disadvantage because much of the strength of the RBIP technique lies in the very high accuracy of a double (or codominant) assay method. A description of standard methods for obtaining the sequence corresponding to the other side of the insertion is given in **Note 1**.

1.8.3. RBIP Compared to the Other Retrotransposon-Based Marker Systems

Retrotransposon-based SSAP, REMAP, and IRAP are well suited to deal with tens to hundreds of samples ***(30,34,38,39)***. RBIP is more useful for far larger numbers of samples because it can, in principal, be completely automated. The RBIP method is also very well suited to phylogeny and biodiversity assessments because it is a codominant marker system, and retrotransposon insertions are quite stable, with a known ground state, namely absence of the insertion ***(49)***. A strategy akin to RBIP was used successfully to determine the distant phylogenetic relationships between whales and ungulates ***(27)***.

2. Materials

2.1. SSAP for PDR1 in Pea (Pisum sativum)

1. RL buffer: 10 m*M* Tris-acetate pH 7.5, 10 m*M* magnesium acetate, 50 m*M* potassium acetate, 5 m*M* dithiothreitol (DTT), 5 ng/μL bovine serum albumin (BSA).
2. Primers: For PDR1, the PPT primer is 5' ATTCACCAGCTTGAGGGGAG 3'
3. Stop solution: 0.25% w/v Bromophenol blue and xylene cyanol in 98% formamide, 10 m*M* EDTA, pH 8.0
4. Resolution of the SSAP products: Acrylamide gel solution of 4.5% for the casting of polyacrylamide gels, either homemade according to standard protocols or commercially prepared.

2.2. SSAP for BARE-1 in Barley

1. RL buffer: as in **Subheading 2.1.**
2. Preparation of adaptors: These should *not* be phosphorylated when synthesized or subsequently treated with kinase.

   *Mse*I: 25 μg 5' GACGATGAGTCCTGAG 3'
   25 μg 3' TACTCAGGACTCAT 5'

   Make up to 100 μL with water; incubate at 65°C for 10 min, then place on ice, add 1 μL 1*M* MgOAc. Bring to 37°C for 10 min, then 25°C for 10 min; place on ice (store at –20°C).

   *Pst*I**:** 25 μg 5' CTCGTAGACTGCGTACATGCA 3'
   25 μg 3' CATCTGACGCATGT 5'

   Treat as for *Mse*I adaptors.
3. Preparation of primers:

   BARE-1 primer: 5' CTAGGGCATAATTCCAACAA 3'
   *Mse*I primer: 5' GATGAGTCCTGAGTAA 3'
   *Pst*I primer: 5' GACTGCGTACATGCAG 3'

   Selective primers are derived from the basic nonselective *Mse*I and *Pst*I primers above, referred to as M(0) and P(0) respectively. The selective *Mse*I primers are: M(C); M(AC); M(ACA). The selective *Pst*I primers are: P(C); P(CG); P(CGA).
4. T0.1E: 10 m*M* Tris-HCl pH 7.5, 0.1 m*M* EDTA

2.3. IRAP for BARE-1 in Cereals

1. Preparation of template DNA: DNA prepared by most standard methods or commercial kits is suitable. Inhibitors of PCR reactions such as polyphenols ***(51)*** or other pigments that may be present in the template preparation will interfere with PCR in IRAP as well.
2. Primers: Primers are made in unphosphorylated, unlabeled form. For separation on sequencing systems, fluorescein or Cy5-labeled primers may be used, but the reaction conditions should be reoptimized as these dyes affect primer annealing to the template.

   Direct *BARE*-1 primer: 5' CTACATCAACCGCGTTTATT 3'

   This sequence corresponds to the LTR at nt 1993 to 2012 of accession Z17327, situated 105–124 nt from the right, 3' end of the LTR.

   Inverse *BARE*-1 primer: 5' GCCTCTAGGGCATAATTCCAAC 3'

   This primer will hybridize to LTR templates 1 nt from the left edge of the LTR, nt 310 to 331 in accession Z17327. This primer is complementary to the coding strand, and therefore faces outward, as does the direct primer, from the element toward the flanking DNA.
3. PCR buffer: The 10X stock contains 750 m*M* Tris-HCl, pH 8.8, 200 m*M* $(NH_4)_2SO_4$, 15 m*M* $MgCl_2$, 0.1% Tween-20.
4. Thermostable DNA polymerases: We have tried a range of thermostable polymerases including *Taq* polymerase from suppliers including, but not limited to, Promega (M1861, storage buffer "A"), Epicentre (Masteramp™ Q82100), Solis BioDyne (Tartu, Estonia, *FIRE*Pol), Finnzymes (Espoo, Finland, DyNAzyme™) and PE Applied Biosystems (Amplitaq®), and have not found differences in the results.
5. Thermocyclers: We have used either a Mastercycler Gradient (Eppendorf-Netheler-Hinz GmbH, Germany) or a PCT-225 DNA Engine Tetrad (MJ Research, Waltham, MA, USA) but have not extensively surveyed others. When using primers in cross-species experiments, it is best to consider possible differences in ramping time for various thermocycler and tube combinations, and to optimize these variables.
6. Agarose: High resolution over a wide range of fragment sizes is important. We have used RESolute™ Wide Range Agarose (Product 337100, BIOzymTC bv, Landgraaf, The Netherlands). Alternatively, 3:1 Nusieve® agarose (50090, FMC Bioproducts, Rockland, ME, USA) may give good results.

2.4. REMAP for BARE-1 in Cereals

Materials for REMAP are the same as described for IRAP in **Subheading 2.3.** with the exception of the primers.

1. *BARE*-1 reverse primer: 5' CATTGCCTCTAGGGCATAATTCCAACA 3'

   This primer is equivalent to LTR-B, described previously ***(39)***, and is complementary to nt 309–335 of the *BARE*-1a sequence (accession Z17327), extending to the left terminus of the LTR.

Table 1
SSR Primers for REMAP

	Hybridization temperature for PCR, °C	
SSR	*BARE*-1 reverse	*BARE*-1 direct
$(GA)_9C$	56	56
$(GT)_9C$	56	56
$(CA)_{10}G$	57	57
$(CT)_9G$	56	56
$(AC)_9C$	56	56
$(AC)_9G$	56	56
$(AC)_9T$	56	56
$(AG)_9C$	56	56
$(TG)_9A$	56	56
$(TG)_9C$	56	56
$(AGC)_6C$	60	60
$(AGC)_6G$	60	60
$(AGC)_6T$	60	60
$(CAC)_7A$	60	60
$(CAC)_7G$	60	60
$(CAC)_7T$	60	60
$(ACC)_6C$	60	60
$(ACC)_6G$	60	60
$(ACC)_6T$	60	60
$(CTC)_6A$	60	60
$(CTC)_6G$	60	60
$(GAG)_6C$	60	60
$(GCT)_6A$	60	60
$(GCT)_6C$	60	60
$(GTG)_7A$	60	60
$(GTG)_7C$	60	60
$(TCG)_6G$	60	60
$(TGC)_6A$	60	60
$(TGC)_6C$	60	60

2. *BARE*-1 forward primer: 5' CTACATCAACCGCGTTTATT 3'

   This matches nucleotides 1993–2012 of *BARE*-1a, extending to 105 bp from the 3' terminus of the LTR.

   A range of SSR primers can be used in combination with either the forward or the reverse retrotransposon primer. These primers are given in **Table 1**, together with the hybridization temperature to be used in PCR:

2.5. RBIP

1. DNA: High DNA quality is not important for the success of RBIP. Miniprep plant DNA, with large amounts of contaminating RNA and polysaccharides, do not affect the success rate of the technique.
2. Reagents: Standard proprietary PCR reagents are used. As in all PCR, success is more likely with hot-start *Taq* enzyme.

3. Methods

3.1. SSAP for PDR1 in Pea

1. DNA digestion: Digest approx 0.5 μg genomic DNA in RL buffer with 5 units restriction endonuclease *Taq* I in a total volume of 40 μL. Incubate 65°C for 2 to 3 h (*see* **Note 2**).
2. Adapter ligation: To the 40 μL digest from **step 1** add 12.5 pmol *Taq* adapter (from 50 pmol/μL stock). Make up to 1 m*M* ATP, and add 1 unit T4 DNA ligase, adjust the total volume to 50 μL in 1X RL. Incubate at 37°C overnight.
3. Template preparation and storage: Dilute the ligated SSAP template DNA from **step 2** by addition of 100 μL TE, pH 8.0, and store at –20°C. (Use 3 μL of this diluted template for a 10-μL PCR volume.)
4. Labeling reaction: Kinase-label the sequence-specific primer in bulk and later dispense the labeled primer among the reactions. The quantity depends on the number of reactions required; the example shown is designed for 30 reactions. The label used here, ^{33}P, is safer and more convenient than ^{32}P, but ensure that appropriate shielding, transport, and disposal procedures are followed (*see* **Note 3**).

   Labeling mix (total volume 20.0 μL):

Primer (100 ng/μL)	4.5 μL
[γ-^{33}P]ATP	2.0 μL (370 kBq/μL)
10X T4 polynucleotide kinase buffer	2.0 μL
Water	11.0 μL
5 units T4 polynucleotide kinase (10 U/μL)	0.5 μL

   Incubate at 37°C for at least 1 h.

   Assemble the reaction components, except for the [γ-^{33}P]ATP, together in a clearly marked screwcapped 1.5-mL Eppendorf tube; dispense the [γ-^{33}P]ATP in a laboratory appropriately equipped for work with radioactivity according to local safety guidelines. Incubate the labeling reaction at 37°C in a heating block designated for radioactive work.
5. Labeled PCR: Assemble as follows for 30 reactions of 10 μL. Each reaction will use 3 μL of template, so 7 μL of the reaction mix must be added to each. Therefore, in this example 210 μL reaction mix must be prepared for aliquoting.

Labeled primer	20 μL (from 4). Primers should be equimolar
Adapter primer (7.5 ng/μL)	60 μL
10X PCR buffer	30 μL
1 m*M* dNTP	60 μL (200 μ*M* each final concentration)
Taq DNA Polymerase	6 U

   Add water (sterile) for a final volume of 210 μL.

Dispense 7 μL to each 3 μL template sample and set up the PCR according to Vos and coworkers ***(5)***:

a. 10 cycles (94°C for 30 s; 55°C [reducing by 1°C per cycle] for 30 s; 72°C for 60 s).
b. 20 cycles (94°C for 30 s; 45°C for 30 s; 72°C for 60 s).
c. A final extension step at 72°C for 7 min.

Check the PCR machine with the Geiger counter before and after use

6. Stopping the reaction: Add 10 μL of Stop solution to each 10 μL PCR; denature by heating to 95°C for 3 min, and cool on ice. Store the reactions at –20°C until ready to load onto a gel. Use care; formamide is a mutagen.
7. Setting up of the polyacrylamide gels: Prepare the sequencing gel apparatus and cast the gel according to standard procedures suited for your specific apparatus.
8. Running and processing the gel (*see* **Note 4**): Mount the gel/glass plate assembly on the electrophoresis unit; add TBE buffer to top and bottom trays; and clean out the wells with buffer using a syringe and needle. Connect up to a power pack and prerun the gel for approx 30 min at 1500–1600V, to warm up. Disconnect the electrophoresis unit, flush out the wells with buffer, and load the denatured samples into the wells (1 μL of sample is generally enough). Continue running the gel for the desired time at 1500–1600V (2 h). Discard the buffer into a drain designated for disposal of low-grade radioactive liquid waste. When the plates have cooled down, remove one of the side spacers. Pry the plates apart, using a thin spatula placed in the gap between the plates at a corner. This is a hazardous procedure as glass fragments may break off or plates may crack and shatter. The gel should remain attached to the nonsilanized plate and can be transferred onto 3MM paper with an extra sheet for backing; trim the excess paper close to the gel. Place a piece of cling film over the gel to protect the gel drier cover from contamination. Dry for 1–2 h at 80°C in the vacuum gel drier. Expose the dried gels to an X-ray film or Phosphoimager plates. An example SSAP gel for *Pisum* is shown in **Fig. 3.**

3.2. SSAP for BARE-*1 in Barley*

1. DNA Digestion: Total genomic DNA from the plant of interest is completely digested using two restriction enzymes: one a rare cutter, the other a frequent cutter. The rationale for this approach is explained by Vos and coworkers ***(5)*** and is summarized below.

   The frequent cutter will generate small DNA fragments, which will amplify well by PCR and are in the correct size range for separation on a denaturing or sequencing gel. The number of fragments amplified can be reduced by using a combination of rare- and frequent-cutting restriction enzymes, allowing amplification of fragments with a rare-cutter site at one end and a frequent-cutter site at the other, to the exclusion of the other fragments. Presumably, it also decreases the chance of a fragment ligating to itself. In this example, we used *Mse*I and

*Pst*I, as these had been previously used in barley ***(30)***, although any combination of rare and frequent cutting enzymes could be tried.

*Mse*I cuts: T TAA
AAT T
*Pst*I cuts: CTGCA G
G ACGTC

Prepare a digest as follows:

Total genomic DNA	1.0 μg
*Mse*I	5 U
*Pst*I	5 U
10X RL buffer	2 μL

Add H_2O for a final volume of 20 μL.

Digest at 37°C for at least 1 h

2. *Mse*I/*Pst*I Adaptor Ligation: Take digested DNA (1 μg in 20 μL) and add the following:

*Mse*I adaptors (40 pmol)	1.0 μL
*Pst*I adaptor (20 pmol)	0.5 μL
10 m*M* ATP	1.0 μL
RL buffer	0.4 μL
T4 ligase	0.5 μL

Incubate at 37°C for 3 h, then store template DNA (at a final concentration of 40 ng/μL) at –20°C.

3. Preamplification PCR (*see* **Note 5**): This procedure is useful when working with large genome sizes, to reduce the restriction fragments to a manageable number. The PCR conditions are the preferred ones for our Techne Genius PCR machine, and should be adjusted as appropriate to others.

10X PCR buffer	2.5 μL
dNTPs	4 μL (1.25 m*M*)
*Mse*I primer	75 ng
*Pst*I primer	75 ng
Template	0.75 μL (approx 30 ng)
Taq DNA polymerase	1 U (0.2 μL)

Add H_2O for a final volume of 25 μL.

We use the following PCR program:

a. 1 min 95°C warmup.
b. 30 cycles (1 min 94°C denaturing; 1 min 60°C annealing; 1 min 72°C extension).
c. 7 min 72°C final extension.

After the reaction is complete, add 55 μL T0.1E and store at –20°C.

2. End-labeling of the *BARE*-1 oligo: This oligo complements the start of the *BARE*-1 5' LTR. The final A on this primer is a selective base, designed to anneal to and amplify only the fraction of fragments in which the first nucleotide of the flanking

sequence is an A. Also, this A is one of two nucleotides which cause mismatches to the 3' LTR, thus reducing the chance of priming into the retrotransposon from this LTR. A total of 1 μL of labeled oligo is made per PCR reaction. We have mainly used [γ-^{32}P]ATP, but ^{33}P label may be used.

Per PCR reaction:

[γ-^{32}P]ATP	1 μCi (3000 Ci/mMol)
BARE-1 oligo (50 ng/μL)	0.13 μL
10X kinase buffer	0.1 μL
T4 polynucleotide kinase	0.25 U (0.025 μL)
Add H_2O for a final volume of 1 μL.	

Incubate at 37°C for at least 30 min. Denature kinase at 70°C for 10 min, then place on ice immediately. Spin at 15,000*g* for 15 s on desktop microcentrifuge. Store at –20°C.

5. Labeled SSAP PCR reaction (*see* **Note 6**):

   Add the following per PCR reaction:

[γ-^{32}P]ATP-labeled *BARE*-1 oligo	1 μL
Unlabeled *BARE*-1 oligo (50 ng/μL)	0.5 μL
Selective *Mse*I or *Pst*I primer (50 ng/μL) (*see* **Subheading 2.2.**)	0.6 μL
10X PCR buffer	2 μL
dNTPs	3.2 μL (1.25 m*M*)
Preamplified DNA (from **step 3**)	2 μL
Taq DNA polymerase	0.5 U (0.1 μL)
Add H_2O for a final volume of 20 μL.	

   The PCR program is as follows, 36 cycles in total:

   a. 94°C, 1 min.
   b. 13 cycles (65°C for 1 min, imposing a –0.7°C decrease per cycle ["touch-down PCR"]; 72°C for 1 min; 94°C for 1 min).
   c. 22 cycles (56°C for 1 min; 72°C for 1 min; 94°C for 1 min).
   d. A final extension at 72°C for 7 min.

6. Running samples on a denaturing gel: Gels are set up as in **step 7** of **Subheading 3.1.** Add 20 μL of sequencing Stop buffer to each sample, mix well. Denature by incubation at 90°C for 5 min, then place on ice immediately. Load each sample onto a 6% denaturing polyacrylamide gel. Load an amount appropriate to the size of combs you are using. We use "shark-tooth" combs, but larger well-forming combs can be used. Samples usually take 1.75–2.0 h to run. It is also useful to run a marker along side. Fix gel if necessary. Gels are exposed with X-ray film for one to five days. Do not use an intensifying screen for ^{32}P gels. If the procedure is working reasonably well, you should get a visible result in a day or two. An alternative is to use a phosphoimager and imaging plates rather than X-ray film.

3.3. IRAP for BARE-1 in Cereals

The technique is presented as developed for barley (*see* **Note 7**).

1. Set up the PCR reaction: The reaction here is designed for 20-μL tubes, but it can be scaled down for use in microtiter plates.
   Each reaction contains:

10X PCR buffer	2 μL
Template DNA (10 ng/μL)	20 ng
PCR primers (one, the other, or both)	200 n*M* each final concentration
dNTPs	0.2 m*M* final concentration
1 U *Taq* DNA polymerase	
Add H_2O for a final volume of 20 μL.	

2. Carry out the PCR as follows:
   a. 94°C for 2 min.
   b. 30 cycles (94°C for 20 s; 60°C for 20 s; 72°C for 2 min).
   c. A final extension at 72°C for 10 min.
   d. Maintenance at 4°C (*see* **Note 8**).
3. Electrophoretic resolution of the PCR products: Take one-fifth of the PCR reaction, mix with loading buffer, and analyze on a wide-resolution agarose gel. We have used 2% RESolute™ agarose, but 1.2–1.5% Seakem 3:1 NuSieve® agarose is expected to work as well. Carry out the electrophoresis in a 20-cm-long gel for 7 h at 100 V in a Pharmacia GNA-200, 20 × 20 cm format, in standard Tris-borate (0.5X TBE) buffer, and visualize by staining with ethidium bromide (*see* **Note 9**).

3.4. REMAP for BARE-1 in Cereals

The example given is for barley.

1. Set up the PCR reaction: The reaction here is designed for 20-μL tubes, but it can be scaled down for use in microtiter plates.
   Each reaction contains:

10X PCR buffer (as for IRAP)	2 μL
Template DNA (10 ng/μL)	20 ng
PCR primers (one, the other, or both)	200 n*M* each final concentration
dNTPs	0.2 m*M* final concentration
1 U *Taq* DNA polymerase	
Add H_2O for a final volume of 20 μL.	

2. Carry out the PCR as follows:
   a. 94°C for 2 min.
   b. 28–32 cycles (94°C for 20 s; 56–60°C [according to primer pair; *see* **Subheading 2.4.**] for 20 s; 72°C for 2 min).
   c. A final extension at 72°C for 10 min.
   d. Maintenance at 4°C (*see* **Note 10**).

3. Electrophoretic resolution of the PCR products: As for IRAP; *see* **step 3** in **Subheading 3.3.** An example REMAP gel is shown in **Fig. 5**.

3.5. RBIP

1. Set up the PCR reaction: The amount of template DNA here is based on use with pea (*see* **Note 11**).

   Each reaction contains:

10X PCR buffer (Promega)	2 μL
Template DNA	10 ng
PCR primers	40 ng each
dNTPs	3.2 μL (1.25 m*M*)
1 U *Taq* DNA polymerase	
Add H_2O for a final volume of 20 μL.	

2. Carry out the PCR: This program was constructed for a Techne Genius machine but can be adapted to others. It consists of 95°C for 1 min; 35 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72°C for 1 min; a final extension at 72°C for 7 min; maintenance at 4°C.
3. Analyze the RBIP products: For gel-based analysis, the products are electrophoretically separated on 1.5% agarose gels containing ethidium bromide in TBE buffer (**Fig. 6B**). For dot-based analysis, 0.5 μL of each PCR product is spotted onto Biodyne A nylon membrane (*see* **Note 12**). The samples are left to dry at room temperature for 5 min, then placed on a sheet of Whatman 3MM chromatography paper wetted with 0.5*M* NaOH, 1*M* NaCl, for 1 min (*see* **Note 13**). After this, they are transferred to another sheet of 3MM paper soaked in 1*M* Tris-HCl, pH 7.5, 3*M* NaCl, then immersed in 2X SSC for 5 min. Finally, the denatured DNA is cross-linked to the filter.

   The filters are hybridized to the probe using standard methods for Southern blot analysis. Prehybridization in 4X SSC, 0.1% SDS at 60°C for at least 30 min is followed by hybridization to the probe in the same solution overnight. Probes are ^{32}P-labeled by the oligonucleotide-primed method. Posthybridization washes use 2X SSC, 60°C final stringency (*see* **Note 14**).

   Radiographic detection is best carried out with a phosphoimager, to allow rapid and accurate quantification of the hybridized signal per dot. We use a Fuji Model BAS-1500 machine, with 1–3 d exposure. The scanned data are quantified by superimposing a square sector matrix using MacBAS software (Fuji; version 2.5). The data are then output to a Microsoft Excel spreadsheet, which subtracts background signals from each scanned sector, calculates occupied/unoccupied signal ratios for each line, and scores the ratios. Ratios of occupied to unoccupied signals of 10 or greater are scored as an occupied site; ratios lower than 0.2 are scored as an unoccupied site; and ratios between these values are scored as heterozygotes (both occupied and unoccupied). Finally, if the summed total signal for both occupied and unoccupied sectors fall below a preset threshold, the locus is scored as a failure (no score; *see* **Note 15**).

3.6. Prospects

Retrotransposons are highly useful as molecular markers, in the analysis of genome structure, and as tools for the reverse-genetic characterization of gene function *(28)*. The protocols presented here have been built around specific retrotransposon families and particular plants. However, retrotransposons throughout the eukaryotes share common structures and life cycles, permitting adaptation to a wide range of research materials. Key considerations for adaptation of the method to the plant of interest are the LTR length, copy number of the retrotransposon family for which the PCR primers are designed, and the genome structure of the plant. Long LTRs necessitate primers near the termini, whereas LTRs of only several hundred base pairs allow more flexibility in this regard. Retrotransposons in high copy number may produce too many bands for efficient amplification or gel resolution in all methods except RBIP. This problem can be overcome by increasing the number of selective bases in SSAP or by designing the retrotransposon primer in IRAP or REMAP to bridge the joint between the LTR and the flanking region and to carry selective bases at its 3' end. Genome organization, particularly the nesting of retrotransposons insertion sites and the proximity of microsatellites to retrotransposons, affects the relative efficacy of IRAP, REMAP, and SSAP.

A valuable aspect of retrotransposon marker systems is that the phylogenetic resolution is dependent on the activity of any particular retrotransposon family. The more active the family, the better the resolution in closely related germplasm. The many examples of explosions in retrotransposon copy number in particular clades of plants *(21,22)* show that certain retrotransposon families can be phylogenetically diagnostic as well. To take advantage of this feature, one must employ a general method for the isolation of new retrotransposon families. The internal domains of retrotransposons contain conserved motifs necessary for carrying out the life cycle. In particular, the RNase domain in the case of *copia*-like elements, and the integrase domain for *gypsy*-like retrotransposons, are sufficiently close to the 3' LTR to permit an SSAP or genome-walking method to be used, employing a PCR primer anchored in either of these regions, to isolate the 5' termini of LTRs of almost any retrotransposon from most eukaryotes *(35)*. In this way, novel elements can be applied to IRAP, REMAP, and SSAP and then in turn the integration sites developed for RBIP.

The RBIP method itself has recently been adapted for more efficient, high-throughput analyses *(50a)*. The improved methodology is based on the use of fluorescent primers where different fluorochromes allow the multiplexing of the PCR reactions as well as on the use of arrays allowing simultaneous analysis of thousands of samples. We therefore expect that retrotransposon marker

systems will find increasing use in the near future for phylogenetic studies, fingerprinting, and germplasm characterization.

4. Notes

1. Rapid ways exist for obtaining the other side of any given retrotransposon insertion. The first of these relies upon the fact that retrotransposons generate a duplication of host insertion site sequence when they insert. For Ty*1-copia* group retrotransposons, this site is a random 5-bp sequence that can be obtained from sequencing the SSAP, IRAP, or REMAP band. This same 5-bp sequence is present at the other side of the insertion, and these sequences can be used as selective bases at the 3' end of a primer specific for the other (unsequenced) end of the insertion. The SSAP, IRAP, or REMAP amplification with this primer on accessions containing the particular insertion and accessions lacking it (as indicated by the marker data) usually yields a very small number of candidate bands corresponding to the other side of the insertion. The correct band can be chosen by its cosegregation with the original marker in a set of samples that are polymorphic for the band. This band can then be sequenced to give the other side of the insertion, and that is all that is needed for the RBIP marker.

   Alternatively, the GenomeWalker™ kit (BD Biosciences Clonetech) or similar products can be used. This procedure is similar to SSAP in principle, but uses a specific primer derived from the host DNA flanking the insertion rather than from the retrotransposon itself, oriented for synthesis toward the insertion site. Sequence analysis of the fragments obtained from accessions lacking the insertion reveals the sequence at the other side of the insertion.
2. DNA digestion: On occasion, the digestion step does not run to completion, presumably as a consequence of some contaminant in the DNA prep. The result is a track with extra bands on the final gel, so that the sample appears exceptional in element number and also distantly related to the other samples (because many bands are not shared). The presence of incomplete digestion can be checked by digesting some of the final sample to be run on the gel: Bands will disappear, revealing the presence of amplification products with internal *Taq* I sites. Alternatively, a specific enzyme digestion buffer can be used and changed for the ligation step; however, this is a little tedious and does not often appear to be necessary. Enzymes other than *Taq* I, or two enzymes, could be used in this step.

   This type of behavior can be exploited in studies of DNA methylation. For example *Sau*3A will not cut C-methylated sites, but *Mbo*I will ***(52)*** so the comparison of *Sau*3A and *Mbo*I SSAPs is informative. Some enzymes are blocked by C-methylation; this blockage may not occur at a symmetric sequence, and there may be no convenient isoschizomer control (e.g., *Hin*d III). In such cases the comparison of the SSAP products with *Hin*d III digested SSAPs can be a useful alternative.
3. ^{33}P poses a hazard mainly as a consequence of contamination. The β-particle emission is low energy compared to ^{32}P. Follow safety guidelines appropriate for handling of radioactive materials.

4. Gradient gels ***(53)*** or high-salt bottom buffers can be used to compress the banding towards the bottom of the gel, maximizing the information content yield from each run.
5. The primers in this step carry no selective bases. The adapter/primer configuration are as described in **Subheading 2.2.**
6. The selective primers used here gave us the most polymorphism with the *BARE*-1 primer and a manageable number of strong bands with the least background on the film. The number of selective bases has to be optimized for each retrotransposon family in a given species. It should be remembered that for any given combination of restriction enzymes (in this example, *Pst*I/*Mse*I) and selective primers, only a subset of the retrotransposon family is amplified. Although this is an inevitable consequence of the limits of PCR amplification and gel-based fragment resolution, additional combinations of digests, adapters, and primers allow analysis of other subsets of the potential integration sites.
7. If the primer is not fully complementary to the template retrotransposon (as would be the case in unconserved regions of a retrotransposon or in divergent families of elements), the PCR buffer, in particular the salt and pH, but not the polymerase, may influence the results.
8. The number of reaction cycles, template quantity, primer concentration, and enzyme quantity may need to be optimized for specific retrotransposon families and plant species. We use up to 1.2 units of enzyme and up to 35 cycles in some cases. The annealing temperature must be adjusted to match the primers used.
9. The IRAP reaction generates a complex mixture of fragments of wide size range. Slow electrophoresis as described improves the fragment resolution, as does longer separation distances and high-quality agarose. We routinely use a 20 × 20 cm Pharmacia gel box (GNA-200) and combs having 1 mm thickness. An example IRAP gel is shown in **Fig. 4**.
10. If there is high background in the lanes, the amount of template can be reduced to 10 ng.
11. Several different proprietary PCR buffers (PE, Promega, Qiagen) have been tried and all have worked. Primers should follow the normal rules for good primer design. In particular, they should be carefully screened against the possibility of primer–dimer artifacts, and we have always been careful to keep the T_m of all primers used in a single reaction to within 2°C of one another. Typically, we use primers of around 20 bases with 40–50% G/C content.
12. Early versions of this protocol used manual spotting, but this method has been superceded by robotic spotting, using a Robbins Hydra with Automated Plate Positioning (APP). This allows 96-well or 384-well PCR plate formats to be spotted automatically at a density of 384 or 1536 dots per 12 cm × 8 cm sheet.
13. These soaking steps are similar to the Southern blot steps, in that they denature the DNA for efficient binding to the membrane and subsequent hybridization. These steps can be omitted and the spotted DNA immediately cross-linked to the membrane, but the signal from the hybridized probe(s) falls by at least a factor of fivefold.

14. Be careful to avoid too stringent washing conditions. The short length of the probes (typically 100–300 bp) and frequent high A/T content from nongenic DNA indicate that the conditions quoted here are usually close to the T_m of the hybrid.
15. Failed PCRs generate low or nonexistent signals in both occupied and unoccupied sectors. Visual screening of scanned hybridized filters gives a very good idea of whether this has happened for any given sample, and the scanned signals for several chosen failures can be used to set the failure threshold signal for all the samples. Typical failure rates are between 3% and 5% in our experience.

Acknowledgments

Development of the methods described in this chapter was funded by contracts BIO-4-CT-960508 and QLK5-CT-2000-01502 to the Commission of the European Communities. We are grateful to Ruslan Kalendar, Maggie Knox, and Steven Pearce for material contributions to the methods presented here. Nigel Kilby and Caroline Stuart-Rogers are thanked for proofreading of the text.

References

1. Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980) Construction of a genetic linkage map using restriction fragment length polymorphisms. *Am. J. Human Genet.* **32,** 314–331.
2. Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A., and Tingey, S. V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18,** 6531–6535.
3. Zhao, X. and Kochert, G. (1993) Phylogenetic distribution and genetic mapping of a $(GCG)_n$ microsatellite from rice (*Oryza sativa*). *Plant Mol. Biol.* **21,** 607–614.
4. Zietkiewicz, E., Rafalski, A., and Labuda, D. (1989) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20,** 176–183.
5. Vos, P., Hogers, R., Bleeker, M., Reijans, M., van der Lee, T., Hornes, M., et al. (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* **21,** 4407–4414.
6. Kumar, A. and Bennetzen, J. (1999) Plant retrotransposons. *Annu. Rev. Genet.* **33,** 479–532.
7. Grandbastien, M.-A. (1992) Retroelements in higher plants. *Trends Genet.* **8,** 103–108.
8. Bennetzen, J. L. (1996) The contributions of retroelements to plant genome organization, function and evolution. *Trends Microbiol.* **4,** 347–353.
9. Schmidt, T. (1999) LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant Mol. Biol.* **40,** 903–910.
10. Frankel, A. D. and Young, J. A. (1998) HIV-1: Fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67,** 1–25.

11. Xiong, Y. and Eickbush, T. H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9,** 3353–3362.
12. Doolittle, R. F., Feng, D. F., McClure, M. A., and Johnson, M. S. (1990) Retrovirus phylogeny and evolution. *Curr. Top. Microbiol. Immunol.* **157,** 1–18.
13. Sverdlov, E. D. (2000) Retroviruses and primate evolution. *BioEssays* **22,** 161–171.
14. Sandmeyer, S. B. and Menees, T. M. (1996) Morphogenesis at the retrotransposon-retrovirus interface: *gypsy* and *copia* families in yeast and *Drosophila*. *Curr. Top. Microbiol. Immunol.* **2124,** 261–296.
15. Pearce, S. R., Harrison, G., Li, D., Heslop-Harrison, J. S., Kumar, A., and Flavell, A. J. (1996) The Ty*1-copia* group of retrotransposons in *Vicia* species: Copy number, sequence heterogeneity and chromosomal localisation. *Mol. Gen. Genet.* **205,** 305–315.
16. SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., et al. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274,** 765–768.
17. Shirasu, K., Schulman, A. H., Lahaye, T., and Schulze-Lefert, P. (2000) A contiguous 66 kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res.* **10,** 908–915.
18. Suoniemi, A., Anamthawat-Jónsson, K., Arna, T., and Schulman, A. H. (1996) Retrotransposon *BARE*-1 is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. *Plant Mol. Biol.* **30,** 1321–1329.
19. Kumar, A., Pearce, S. R., McLean, K., Harrison, G., Heslop-Harrison, J. S., Waugh, R., et al. (1997) The Ty*1-copia* group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. *Genetica* **100,** 205–217.
20. Heslop-Harrison, J. S., Brandes, A., Taketa, S., Schmidt, T., Vershinin, A. V., Alkhimova, E. G., et al. (1997) The chromosomal distributions of Ty*1-copia* group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* **100,** 197–204.
21. SanMiguel, P. and Bennetzen, J. L. (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.* **82,** 37–44.
22. Vicient, C. M., Suoniemi, A., Anamthawat-Jónsson, K., Tanskanen, J., Beharav, A., Nevo, E., et al. (1999) Retrotransposon *BARE*-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11,** 1769–1784.
23. Voytas, D. F., Cummings, M. P., Konieczny, A. K., Ausubel, F. M., and Rodermel, S. R. (1992) *Copia*-like retrotransposons are ubiquitous among plants. *Proc. Natl. Acad. Sci. USA* **89,** 7124–7128.
24. Flavell, A. J., Dunbar, E., Anderson, R., Pearce, S. R., Hartley, R., and Kumar, A. (1992) Ty*1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Res.* **20,** 3639–3644.
25. Suoniemi, A., Tanskanen, J., and Schulman, A. H. (1998) *Gypsy*-like retrotransposons are widespread in the plant kingdom. *Plant J.* **13,** 699–705.

26. Rowold, D. J. and Herrara, R. J. (2000) *Alu* elements and the human genome. *Genetica* **108,** 57–72.
27. Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., et al. (1997) Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* **388,** 666–670.
28. Kumar, A. and Hirochika, H. (2001) Applications of retrotransposons as genetic tools in plant biology. *Trends Plant Sci.* **6,** 127–134.
29. Yu, G.-X. and Wise, R. P. (2000) An anchored AFLP- and retrotransposon-based map of diploid *Avena*. *Genome* **43,** 736–749.
30. Waugh, R., McLean, K., Flavell, A. J., Pearce, S. R., Kumar, A., Thomas, B. B. T., et al. (1997) Genetic distribution of *BARE*-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (SSAP). *Mol. Gen. Genet.* **253,** 687–694.
31. Vogel, J. M. and Morgante, M. A microsatellite-based, multiplexed genome assay. In *Plant Genome III Conference*, San Diego, CA, 1992.
32. Korswagen, H. C., Smits, M. T., Durbin, R. M., and Plasterk, R. H. A. (1996) Transposon *Tc1*-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. *Proc. Natl. Acad. Sci. USA* **93,** 14,680–14,685.
33. Van den Broeck, D., Maes, T., Sauer, M., Zethof, J., De Keukeleire, P., D'Hauw, M., et al. (1998) *Transposon Display* identifies individual transposable elements in high copy number lines. *Plant Jour.* **13,** 121–129.
34. Ellis, T. H. N., Poyser, S. J., Knox, M. R., Vershinin, A. V., and Ambrose, M. J. (1998) Polymorphism of insertion sites of Ty*1-copia* class retrotransposons and its use for linkage and diversity analysis in pea. *Mol. Gen. Genet.* **260,** 9–19.
35. Pearce, S. R., Stuart-Rogers, C., Knox, M. R., Kumar, A., Noel Ellis, T. H., and Flavell, A. J. (1999) Rapid isolation of plant Ty*1-copia* group retrotransposon LTR sequences for molecular marker studies. *Plant J.* **19,** 711–717.
36. Lee, D., Ellis, T. H. N., Turner, L., Hellens, R. P., and Cleary, W. G. (1990) A *copia*-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic analysis. *Plant Mol. Biol.* **15,** 707–722.
37. Vershinin, A. V. and Ellis, T. H. (1999) Heterogeneity of the internal structure of PDR1, a family of Ty*1/copia*-like retrotransposons in pea. *Mol. Gen. Genet.* **262,** 703–713.
38. Kalendar, R., Grob, T., Regina, M., Suoniemi, A., and Schulman, A. H. (1999) IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.* **98,** 704–711.
39. Manninen, O., Kalendar, R., Robinson, J., and Schulman, A. H. (2000) Application of *BARE*-1 retrotransposon markers to map a major resistance gene for net blotch in barley. *Mol. Gen. Genet.* **264,** 325–334.
39a. Boyko, E., Kalendar, R., Korzun, V., et al. (2002) A high density cytogenetic map of the *Aegilops tauschii* genome incorporating retrotransposons and defense related genes: insight into cereal chromosome structure and function. *Plant Mol. Biol.* **48,** 767–790.
40. Vicient, C. M., Jääskeläinen, M., Kalendar, R., and Schulman, A. H. (2001) Active retrotransposons are a common feature of grass genomes. *Plant Physiol.* **125,** 1283–1292.

41. Kankaanpää, J., Mannonen, L., and Schulman, A. H. (1996) The genome sizes of *Hordeum* species show considerable variation. *Genome* **39,** 730–735.
42. Panstruga, R., Büschges, R., Piffanelli, P., and Schulze-Lefert, P. (1998) A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. *Nucleic Acids Res.* **26,** 1056–1062.
42a. Price, Z., Schulman, A. H., and Mayes, S. (2003) Development of new marker methods—an example from oil palm. *Plant Genet. Resour.* (in press).
43. McCouch, S. R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y. G., et al. (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Mol. Biol.* **35,** 89–99.
44. Saghai Maroof, M. A., Biyashev, R. M., Yang, G. P., Zhang, Q., and Allard, R. W. (1994) Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. *Proc. Natl. Acad. Sci. USA* **91,** 5466–5470.
45. Ramsay, L., Macaulay, M., Carle, L., Morgante, M., degli Ivanissevich, S., Maestri, E., et al. (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.* **17,** 415–425.
46. Provan, J., Thomas, W. T. B., Forester, B. P., and Powell, W. (1999) *Copia*-SSR: A simple marker technique which can be used on total genomic DNA. *Genome* **42,** 363–366.
47. Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., and Schulman, A. H. (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA* **97,** 6603–6607.
48. Jääskeläinen, M., Mykkänen, A.-H., Arna, T., Vicient, C., Suoniemi, A., Kalendar, R., et al. (1999) Retrotransposon *BARE*-1: Expression of encoded proteins and formation of virus-like particles in barley cells. *Plant J.* **20,** 413–422.
49. Flavell, A. J., Knox, M. R., Pearce, S. R., and Ellis, T. H. N. (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J.* **16,** 643–650.
50. DeRisi, J. L., Vishwanath, R. I., and Brown, P. O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278,** 680–686.
50a. Flavell, A. J., Bolshakov, V. N., Booth, A., et al. (2003) A microarray-based high throughput molecular marker genotyping method: the tagged microarray marker approach. *Nucl. Acids Res.* **31,** e115.
51. Singh, R. P., Singh, M., and King, R. R. (1998) Use of citric acid for neutralizing polymerase chain reaction inhibition by chlorogenic acid in potato extracts. *J. Virol. Methods* **74,** 231–235.
52. McClelland, M., Nelson, M., and Raschke, E. (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res.* **22,** 3640–3659.
53. Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983) Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.

12

MITE Display

Alexandra M. Casa, Alexander Nagel, and Susan R. Wessler

Summary

Genome size differences among crop plants are largely due to unequal accumulation of repetitive DNA sequences, mainly transposable elements (TEs). Over the past decade, many families of miniature inverted-repeat transposable elements (MITEs) have been identified and characterized in a variety of organisms including animals and plants. MITEs are characterized by short terminal inverted repeats (TIRs) (10–15 bp), small size (approx 100 to 500 bp), high-copy-number (approx 1000 to 15,000 per haploid genome), and a preference for insertion into 2-bp to 3-bp targets that are rich in A and T residues. In this chapter, we present a modified transposon display procedure based on the maize MITE family *Heartbreaker* (*Hbr*). This technique is similar to AFLP in which AFLP adaptors are ligated to compatible ends of digested genomic DNA. Subsets of *Hbr*-containing fragments are then amplified using one AFLP primer and another primer complementary to an internal sequence of the *Hbr* element. Like AFLP, the *Hbr* display method permits the simultaneous analysis of numerous DNA fragments. Given the plethora of available marker systems, the major advantage of *Hbr* markers, and perhaps most MITE-based markers, is a preference for insertion in or near transcriptionally active genomic regions. This feature may be especially valuable in the large genomes of agriculturally important plants like maize, wheat, and barley where gene-rich islands are thought to exist in a sea of retrotransposons. Having a class of markers that are enriched in genic regions, coupled with the ease of isolating MITE markers, could expedite chromosome walks and map-based cloning protocols in these organisms.

Key Words: Transposon display; Heartbreaker; molecular markers; MITEs.

1. Introduction

Two decades ago the repertoire of plant genetic markers was limited, thus restricting the construction of detailed genetic maps. The few available phenotypic markers were not optimal because of difficulties involved in constructing multiply marked lines, and because of the large amount of labor required to generate and use these markers ***(1)***. The advent of DNA markers has facilitated a variety of genetic and genome studies, including map-based gene cloning

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

(2,3), and marker-assisted selection *(4,5)*. This chapter describes a strategy for exploiting the unique properties of a group of transposable elements (TEs) called miniature inverted repeat transposable elements (MITEs) to create a new class of molecular markers in maize *(6)* and rice.

MITEs were first discovered in association with the genes of several grass species including maize *(7–9)*, sorghum *(10)*, rice *(11)*, and barley *(7,8,12)*. They are also present in the genomes of dicotyledenous plants such as green pepper *(13)*, *Arabidopsis (14,15)* and *Medicago (16)*. MITEs are not restricted to plant genomes, having recently been described in fungi *(17)* and in several animal genomes including *C. elegans (18,19)*, insects *(20–22)*, *Xenopus (23)*, humans *(24,25)* and zebrafish *(26)*. MITEs are characterized by short terminal inverted repeats (TIRs) generally ranging from 10 to 15 bp, small size (approx 100 to 500 bp), high-copy-number (approx 1000 to 15,000 per haploid genome), and a preference for insertion into 2-bp to 3-bp targets that are rich in A and T residues *(7)*. A recent study described the characterization of a maize MITE family called *Heartbreaker* (*Hbr*) *(27)*. Unlike previously reported MITEs from plants *(7,8)*, most of the 3000–4000 members of the *Hbr* family display over 90% sequence identity.

In this chapter we present a modified transposon display procedure *(28)* for use with the *Hbr* family. We call this procedure MITE display *(6)* since it has been successfully applied in genetic mapping of many MITE families, in addition to *Hbr*, in maize and rice. This technique comprises four major steps: (1) digestion of DNA with restriction enzymes and ligation to double-stranded adapters, (2) preselective amplification, (3) selective amplification, and (4) detection of PCR products on denaturing polyacrylamide gels (*see* **Fig. 1A** and **1B**).

1.1. Applications

MITE display has been used both for genetic mapping in maize *(6)* and rice (Nagel, A. and Wessler, S. R., unpublished) and for fingerprint analysis in maize *(29)*. Because they are highly polymorphic, *Hbr* and other MITE-derived markers may be used to characterize intra- and interspecific diversity, and thus to provide important insights into their role in shaping genome structure and evolution.

Fig. 1. *(opposite page)* Schematic of the *Hbr*-display protocol. **(A)** P1 and P2 depict the location of internal primers used in the preselective and selective amplification reactions (for details *see* **Subheadings 3.6.2.** and **3.6.3.**). P2 is labeled either with ^{33}P or with a fluorescent tag. **(B)** ^{33}P-labeled PCR products obtained with primer combination *Mse*I+A/*Hbr*Int5-F. L, size standard; V, *Hbr*-containing vector; B and M represent parental lines B73 and Mo17, respectively. Numbers to the right of parent pair denote the progeny number derived from the B X M cross. Figure not to scale.

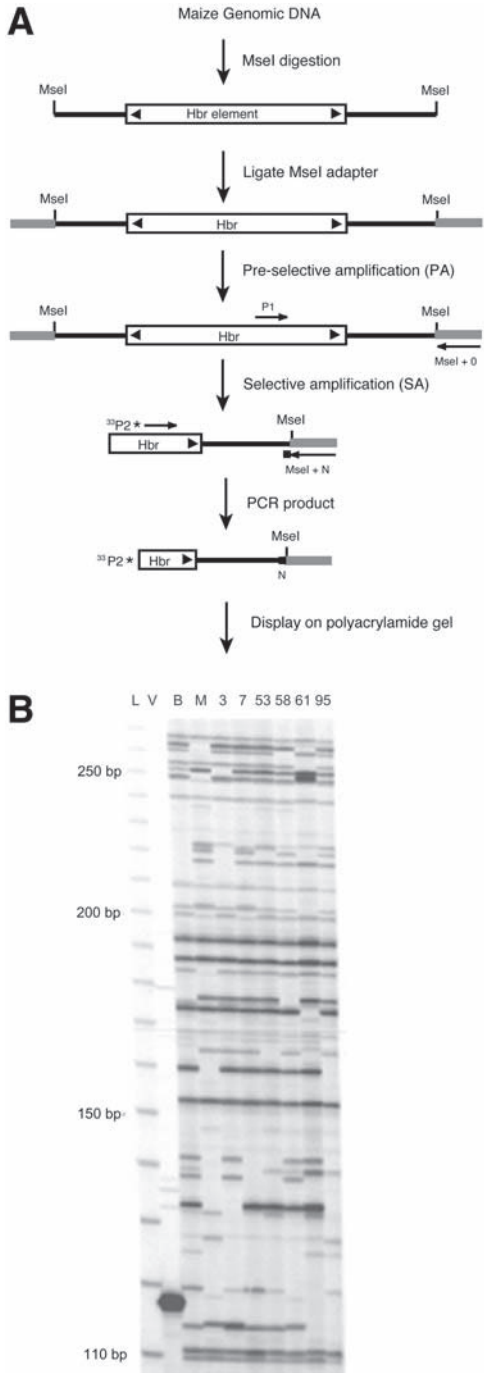

A
Maize Genomic DNA
MseI digestion
MseI
Hbr element
MseI
Ligate MseI adapter
MseI
Hbr
MseI
Pre-selective amplification (PA)
MseI
P1
Hbr
MseI
MseI + 0
Selective amplification (SA)
33P2*
Hbr
MseI
MseI + N
PCR product
MseI
33P2*
Hbr
N
Display on polyacrylamide gel
B
L V B M 3 7 53 58 61 95
250 bp
200 bp
150 bp
110 bp

Fig. 1.

Given the plethora of available marker systems, the major advantage of *Hbr* markers, and perhaps most MITE markers, is a preference for genic regions ***(11,27)***. This feature may be especially valuable in the large genomes of agriculturally important plants like maize, wheat, and barley where gene-rich islands are thought to exist in a sea of retrotransposons ***(30)***. Having a class of markers that are enriched in genic regions, coupled with the ease of isolating MITE markers, should expedite chromosome walks and map-based cloning protocols in these organisms. In addition, the presence of MITEs in several insect species including the mosquito ***(20,21)*** may, in conjunction with the display protocol, provide a means to monitor population dynamics of important disease vectors ***(31)***. Finally, MITE display of parents and their progeny should facilitate the identification of strains with actively transposing MITEs ***(32)***.

The display protocol described in this chapter should also be applicable to short interspersed nuclear elements (SINEs), which share some features with MITEs. Like MITEs, SINEs are short (100–500 bp), dispersed throughout the genome, have attained very high copy number (up to a million per haploid genome for *Alu*) ***(33)***, and populate the genomes of both plants and animals ***(34)***. Unlike Class 2 DNA elements (including MITEs), Class 1 retroelements (including SINEs) do not excise, and thus they permit unambiguous assignment of ancestry. The use of SINEs in phylogenetic analysis is discussed in Chapter 13.

2. Materials

2.1. Materials

1. Thermocycler (MJ Research Inc.).
2. Radioactive detection: power supply and electrophoresis chamber, glass plates, filter paper, Biomax MR-1 film, X-ray cassettes, and needles (Fisher Scientific).
3. Fluorescent detection: glass plates, gel cassettes and 377 DNA sequencer (Applied Biosystem) or equivalent.

2.2. Buffers, Solutions, and Other Supplies

1. 12 m*M* Adenosine triphosphate (ATP): In 8 mL of sterile water dissolve 66 mg of ATP (FW 551) (Sigma-Aldrich). Adjust pH to 7.0 with 0.1*N* NaOH (Sigma-Aldrich). Adjust the volume to 10 mL with distilled water. Dispense solution into small aliquots and store at –70°C.
2. 0.5*M* Dithiothreitol (DTT): In 20 mL of 0.01*M* sodium acetate, pH 5.2 (Sigma-Aldrich) dissolve 1.55 g of DTT (Sigma-Aldrich). Sterilize by filtration (do not autoclave) in 1-mL aliquots and store at –20°C.

3. TE buffer (1X): Add 0.12 g of Tris-HCl base (FW 121.1) (Bio-Rad Laboratories) and 0.037 g of EDTA (FW 372.2) (Bio-Rad Laboratories) to 70 mL of water; adjust pH to 8.0 (by adding HCl) and add water to 100 mL. Autoclave for 20 min at 15 psi on liquid cycle. Store at room temperature.
4. TBE-buffer (10X): Add 108 g of Tris-HCl base (FW 121.1), 55 g of boric acid (FW 61.83) (Bio-Rad Laboratories), and 8.3 g EDTA (FW 372.2) to 800 mL of water. After dissolved, add water to 1 L (*see* **Note 1**).
5. Loading-denaturing buffer for radioactive gel: To 10 mL of deionized formamide (98%) (Amresco Inc.), add 2.5 mg of xylene cyanole FF (0.025%) (Sigma-Aldrich), 2.5 mg of Bromphenol Blue (0.025%) (Sigma-Aldrich), and 3.8 mg of EDTA (FW 372.2).
6. Loading-denaturing buffer for fluorescent gel: Mix 1 part of loading buffer (provided with fluorescent standard) with 4 parts of deionized formamide.
7. *Mse*I/*Bfa*I restriction enzymes (New England BioLabs Inc.) with bovine serum albumin (BSA) 10 mg/mL (provided with restriction enzyme).
8. One-phor-all buffer (OPA) (APBiotech).
9. T4 DNA ligase (Life Technologies).
10. *Taq* DNA polymerase (Promega). 10X PCR buffer and $MgCl_2$ provided with polymerase.
11. Deoxynucleotide triphosphates (dNTPs) (Promega).
12. Adapters: 5' GACGATGAGTCCTGAG and 5' TACTCAGGACTCAT
13. Primers:
    Preselective amplification (*see* **Subheading 3.6.2.**)

    *Hbr*Int5-E: 5' GATTCTCCCCACAGCCAGATTC, *and*
    *Mse*I: 5' GACGATGAGTCCTGAGTAA *or*
    *Bfa*I: 5' GACGATGAGTCCTGAGTAG

    Selective amplification (*see* **Subheading 3.6.3.**):

    *Hbr*Int5-F: 5' GAGCCAGATTTTCAGAAAAGCTG, *and*
    *Mse*I: 5' GACGATGAGTCCTGAGTAA+N *or*
    *Bfa*I: 5' GACGATGAGTCCTGAGTAG+N
14. 40% Acrylamide:Bisacrylamide (19:1) (Bio-Rad Laboratories).
15. Agarose (Bio-Rad Laboratories).
16. Genescan 500 XL [TAMRA] (Applied Biosystem) (loading buffer provided with the size standard for fluorescence detection).
17. [γ ^{33}P] ATP for radioactivity detection (NEN Life Science Products, Inc.).
18. 30–330 bp Amplified fragment length polymorphism (AFLP) DNA ladder (Life Technologies).
19. T4 polynucleotide kinase (Life Technologies) supplied with 5X Reaction Exchange Buffer.
20. TA cloning kit (Invitrogen).
21. QIAquick columns (QIAGEN).

2.3. Software

The following programs are available free on the Web.

1. Identification of MITE sequences/database search: FINDMITE (http://jaketu.biochem.vt.edu/); BlastN (http://www.ncbi.nlm.nih.gov/); MAK (http://perl.idmb.tamu.edu/mak.htm).
2. Multiple sequence alignment: ClustalW (http://www.ebi.ac.uk/clustalw).
3. Restriction site map: Restriction Analysis (v.1.01) (http://molecularworkshop.com/pl/restr102.pl).
4. Primer design: Primer3 (http://www.broad.mit.edu/cgi-bin/primer/primer3).

3. Methods

3.1. Identification of MITEs Suitable for Display

The two most important features in the selection of a MITE family for transposon display are high sequence identity among family members and high copy number. High sequence identity is usually the hallmark of a family that is still active or one that has been active in the recent past. Because recent activity leads to significant levels of polymorphism (defined here as the presence vs the absence of an element at a locus within members of the same species), MITE families with high sequence identity are most likely to produce higher number of markers. Furthermore, high sequence identity allows the design of primers with little or no degeneracy, thus permitting the use of stringent PCR conditions and leading to reproducible results.

MITEs have been identified both experimentally and through database searches ***(7–10,35)***. After a single MITE has been isolated, additional family members can be isolated from genomic libraries ***(27)*** or by computer searches ***(10,19,36)***. Sequence comparison of family members is necessary to derive a consensus sequence for primer design and to identify restriction enzyme sites within the MITE. Such sites will identify enzymes that should be avoided when digesting genomic DNA prior to amplification.

3.2. Selection of MITE Primers

To increase the specificity of the display assay, nested PCR was used with two sets of primers that recognize conserved regions adjacent to the TIRs. However, for certain MITE families, clear and reproducible banding patterns have been obtained when the same MITE primer was used in both the preselective and the selective amplification reactions.

Although the TIR sequence is usually the most conserved among members of a MITE family, these sequences should be avoided in the design of primers for MITE display. It has been observed that primers complementary to the TIRs may result in nonspecific amplification because different MITE families

often have related TIRs ***(6)***. Another consideration is that internal primers allow one to verify that amplified fragments are indeed anchored in the desired MITE family (*see* **Subheading 3.7.**).

As with any PCR-based protocol, primer design is critical to the ultimate success of MITE display. Because of space limitations, we cannot summarize the features of good primers and how they are designed and selected. For this information, the reader is referred to some good books and reviews ***(35,36)***.

3.3. Adapters

Adapters are formed by annealing two oligonucleotides with complementary sequences (core sequence) and an enzyme specific sequence (*see* **Note 2**). There are universal adapter sequences ***(39)***, but they can be altered according to necessity (use of a different restriction site or to increment the melting temperature of the adapter primer). The sequence of the adapter (*see* **Note 3**) and the adjacent restriction site serve as primer binding sites for subsequent amplification of the restriction fragments. In general, one to three selective bases are added to the primer complementary to the adapter. The number of selective bases will vary according to the copy number of the MITE family.

3.4. Polymerase Chain Reaction Conditions

The touchdown protocol ***(40)***, used for *Hbr* display, increases the specificity of amplification when the primers have disparate melting temperatures (T_m > 5°C). However, for any given primer pair, a general PCR program can be selected based on the GC content and length of the primers and the length of the expected PCR product ***(37,38)***.

3.5. Plant Material

Genomic DNA for MITE display can be isolated by using a CTAB procedure ***(41)***. To ensure specific amplification of MITE sequences, it is important to include genomic DNA from plants lacking the MITE family as a negative control.

3.6. Hbr Display

3.6.1. DNA Restriction and Ligation of Adapters

1. Digest total genomic DNA (200–500 ng maize, 50 ng rice) (*see* **Note 4**) to completion for 3 h at 37°C in 40 µL containing 2 units *Mse*I or *Bfa*I, 5 m*M* DTT, 5 µg BSA and 1X OPA buffer.
2. Ligate adapters (*see* **Note 5**) to the digested DNAs by adding 10 µL of a mix containing 1X OPA buffer, 1.2 m*M* ATP, 5 m*M* DTT, 5 µg BSA, 50 pmol adapters, and 1 Weiss unit T4 DNA ligase, and incubate for 3 h at 37°C.

3. Check the quality of the DNA digestion by running 15 μL of the restriction/ligation reactions on 0.8% agarose gels. A smear of DNA fragments ranging in size from approx 100 bp to approx 1000 bp should be visualized.
4. Dilute the remaining restriction and ligation reactions fourfold with 0.1X TE.

3.6.2. Preselective Amplification

PCRs are performed using a primer complementary to the adapters (*Mse*I or *Bfa*I) and another primer (*Hbr*Int5-E) complementary to an internal *Hbr* element sequence (*see* **Subheading 2.2.**).

1. For radioactive detection, PCR amplifications are done in 50-μL volumes and include 5 μL of the diluted, digested, and ligated genomic DNA, 12 pmol of each primer (one complementary to the adapter [either *Mse*I or *Bfa*I], and the internal MITE primer, *Hbr*Int5-E), 1X PCR buffer, 0.2 m*M* dNTPs, 2.5 m*M* $MgCl_2$, and 1 U *Taq* DNA polymerase.
2. For detection in fluorescent format, PCR amplifications are in 20 μL containing 3 μL of the diluted restriction and ligation reaction, 8 pmol of each primer, 1X PCR buffer, 0.2 m*M* dNTPs, 1.5 m*M* $MgCl_2$, and 0.4 U *Taq* polymerase.
3. Conduct amplification using the following cycling parameters:
   a. 94°C for 5 min.
   b. 24 cycles (94°C for 30 s; 59°C for 30 s; 72°C for 1 min).
   c. A final cycle of 72°C for 5 min.
4. Check results of preselective amplification reactions by running 10 μL of each PCR on 1.2% agarose gels stained with ethidium bromide. A smear similar to that observed with the digestion/ligation reactions should be observed.
5. Dilute the remaining volume 20-fold with 0.1X TE.

3.6.3. Selective Amplification

1. For radioactive detection, set up selective amplification reaction in a 20-μL volume containing 5 μL of the diluted preselective amplification products, 8 pmol of selective primer *Mse*I+N, 1.25 pmol ^{33}P-labeled *Hbr*Int5-F (*see* **Subheading 3.6.4.**), 1X PCR buffer, 0.2 m*M* dNTPs, 2.5 m*M* $MgCl_2$, and 0.4 units *Taq* DNA polymerase.
2. For the fluorescence assay, PCRs are set up as above, except that either *Mse*I+N or *Bfa*I+N primers (4 pmol) are used in combination with 4 pmol of the *Hbr*Int5-F primer labeled with 6-FAM (Applied Biosystems) (*see* **Note 6**), and the $MgCl_2$ concentration is reduced to 1.5 m*M*.
3. Use the following "touchdown" protocol:
   a. 94°C for 5 min.
   b. 10 cycles (94°C for 30 s; 70°C for 30 s, –1°C per cycle; 72°C for 1 min).
   c. 27 cycles (94°C for 30 s; 61°C for 30 s; 72°C for 1 min).
   d. A final cycle at 72°C for 5 min.

3.6.4. Primer and DNA Ladder Labeling

Primer (for 20 reactions):

1. In a 1.5-mL Eppendorf tube add 2.5 μL of *Hbr* primer at 10 pmol/μL, 5 μL of [γ ^{33}P] ATP (10 Ci/μL), 1 μL T4 polynucleotide kinase (10 units/μL), 1.25 μL 10X OPA buffer, and water for a final volume of 12.5 μL (*see* **Note 7**).
2. Incubate the reaction for 30 min at 37°C, then for 10 min at 70°C.
3. Spin the tube down briefly and add the contents to the PCR mix.

DNA ladder:

1. Place a 0.5-mL tube on ice and pipet into the tube 2 μL of the 30–330 bp AFLP DNA ladder, 1 μL of Exchange Reaction Buffer, 1 μL of [γ ^{33}P] ATP (10 Ci/μL), and 1 μL of T4 polynucleotide kinase (10 units/μL).
2. Mix the contents thoroughly, centrifuge the tube briefly, and incubate the mixture for 10 min at 37°C.
3. Stop the reaction by heating the tube for 15 min at 65°C.
4. To the reaction mix add an equal volume of TE buffer (1X), and 25 μL of the loading denaturing buffer.

3.6.5. Gel Preparation

General procedures for preparing polyacrylamide gels for both fluorescence and radioactivity detection are similar. See the protocol described in the GeneScan Reference Guide (ABI 373 and ABI Prism 377 DNA Sequencers, Applied Biosystems).

3.6.6. Sample Preparation and Gel Electrophoresis

3.6.6.1. Radioactivity Detection

1. Add 20 μL of loading-denaturing buffer (*see* **Subheading 2.2.–2.5.**) to the PCR reactions.
2. Denature the samples and the 30–330 bp AFLP DNA ladder at 95°C for 5 min, place them on ice, and immediately load 3 μL of the mixture (*see* **Note 8**) onto a 6% denaturing (7.5 *M* urea) acrylamide:bisacrylamide (19:1) gel in 1X TBE buffer.
3. Conduct electrophoresis for 2 h at 35 mA constant (the slower-migrating dye front—xylene cyanole—should have migrated about 25 cm from the origin).
4. Transfer the gel to filter paper, dry, and expose to X-ray film for 24 h (**Fig. 1B**) (exposure time will depend on signal intensity).

3.6.6.2. Fluorescence Detection

1. In a small tube (0.2–0.5 mL) add 0.3–0.5 μL of the PCR products, 0.1 μL GeneScan 500 XL [TAMRA] (*see* **Note 9**) internal lane size standard, and 1.6 μL of loading buffer (*see* **Subheading 2.2.–2.6.**).

2. Denature samples at 95°C for 5 min, place them on ice, and load 0.8 μL of the mixture onto a 5% denaturing (6 *M* urea) acrylamide:bisacrylamide (19:1) gel in 1X TBE.
3. Perform electrophoresis on an automated DNA sequencer (Model 377, Perkin-Elmer/ABI) at 3000 V for 3 h at 51°C in 1X TBE. If the fluorescent signal is too strong (average peak height > 4000), PCR products should be diluted accordingly.

3.7. Recovery of Bands From Acrylamide Gels

The protocol described below has been modified slightly from that previously described ***(42)***.

1. Align the X-ray film and the filter paper and cut a window through the film exposing the fragment of interest.
2. Scratch the gel with a fine needle or pipet tip (*see* **Note 10**) and place in PCR tubes containing 20 μL reaction mix for about 1 min before removing, discarding, and starting PCR. All amplification parameters are as described (*see* **Subheading 3.6.2.**), except that the internal selective primer *Hbr*Int5-F is used, and the number of cycles is increased from 24 to 30.
3. Run PCR products on 1–1.5% agarose gels, excise, purify (QIAquick, QIAGEN), and clone (TA cloning kit, INVITROGEN) fragments of correct size according to the manufacturer's instructions.

3.8. Isolating Genomic DNA Flanking MITE Markers

Marker bands that are isolated from a gel (*see* **Subheading 3.7.**), reamplified, and sequenced contain genomic DNA flanking only one MITE terminus. For some procedures, however, the genomic sequence flanking both termini is required (for example, to convert a MITE marker into a marker that can be used in other populations or members of subspecies, or to obtain the whole sequence of the MITE at a locus for phylogenetic analysis).

For organisms with large DNA sequence databases (e.g., *Arabidopsis*, rice, *C. elegans*), computer searches could be conducted to obtain the genomic sequence flanking the other MITE terminus. Alternatively, one can use a procedure similar to transposon display, but replacing the MITE primers with nested primers derived from the available flanking DNA and oriented toward the MITE. To this end, genomic DNA is digested with a restriction enzyme that does not cut within the MITE and ligated to adapters. The ligation mix is then amplified as described for the preselective and selective amplification reactions (*see* **Subheadings 3.6.2.** and **3.6.3.**) except that the two flanking primers are used in conjunction with the adapter primers.

There are two methods to distinguish the MITE anchored fragments from the vast majority of PCR products that will be anchored only in adapter primers. The first is to radioactively label the flanking primer used in selective

amplification and to isolate the labeled fragment from the acrylamide gel (*see* **Subheading 3.7.**). The second method does not involve radioactivity but requires altering the PCR conditions to reduce the background of fragments not anchored in MITEs. Some alterations include: (1) using adapter primers with one selective base in both preselective and selective amplification, (2) reducing the concentration of only the adapter primers (to approx 10% of the amount used in transposon display), and (3) designing flanking primers with higher annealing temperatures than adapter primers and performing PCR at the higher temperatures. PCR products are then resolved on agarose gels and the band of interest is excised and cloned.

The TAIL-PCR protocol could also be used to specifically amplify the MITE-anchored fragment of interest without using radioactivity ***(43)***. Several methods have been applied successfully. It should be noted that negative controls are very important with each procedure.

4. Notes

1. Precipitates form when concentrated solutions of TBE are stored for long periods of time. To avoid this problem, store the 10X solution in glass bottles at room temperature or make a less concentrated (5X) solution. Discard buffers that precipitate.
2. Adapters should be designed so that restriction enzyme recognition sites are not regenerated after ligation to restricted genomic DNA.
3. The oligonucleotides comprising the double-stranded adapter are not 5'-phosphorylated. This leads to the ligation of only one strand to the ends of the restriction fragments. The complementary strand is filled in by the *Taq* polymerase during the heating step or in the process of assembling the reaction mixture ***(37)***.
4. In our lab, rice genomic DNA ranging from 50 ng to 750 ng resulted in identical banding patterns with similar band intensities.
5. Since *Mse*I and *Bfa*I generate identical 3' overhangs, the same adapters can be used for both ligations.
6. Other fluorescent dye labels can be used, such as TET, HEX, NED, and JOE (Applied BioSystems).
7. We have also successfully used the Life Technologies protocol accompanying the T4 kinase.
8. Loading more of the sample than indicated significantly increases background signal. Therefore, enhance weak radioactive band intensities by increasing exposure time.
9. If the TE-specific primer is labeled with an alternative fluorescent dye, make sure the size standard is compatible with the dye/virtual filter set employed. For example, a TAMRA-labeled size standard is used with 6FAM, TET, and HEX labels (set C dyes), and ROX-labeled standards are used with 6FAM, HEX, and NED (set D dyes).
10. A fine needle is used to isolate fragments from the gel because cutting bands out of the gel may lead to the recovery of neighboring fragments (± 1 base pair).

Acknowledgments

We thank Xiaoyu Zhang and Cedric Feschotte for providing helpful comments and suggestions. This work was supported by funding from Pioneer HiBred International (to AMC and SRW), NSF RTG in Mechanisms of Plant Evolution (to AN), and NIH (to SRW).

References

1. Rafalski, J. A., Vogel, J. M., Morgante, M., Powell, W., Andre, C., and Tingey, S. V. Generating and using DNA markers in plants. In *Non-Mammalian Genomic Analysis: A Practical Guide* (Birren, B. and Lai, E., eds.), Academic Press, New York, NY, 1997, pp. 75–134.
2. Martin, G. B., Brommonscenkel, S. H., Chunwongse, J., Frary, A., Ganal, M. W., Spivey, R., et al. (1993) Map-based cloning of a protein-kinase gene conferring disease resistance in tomato. *Science* **262,** 1432–1436.
3. Ron, M., Kantety, R., Martin, G. B., Avidan, N., Eshed, Y., Zamir, D., et al. (2000) High-resolution linkage analysis and physical characterization of the EIX-responding locus in tomato. *Theor. Appl. Genet*. **100,** 184–189.
4. Miklas, P. N., Larsen, R. C., Riley, R., and Kelly, J. D. (2000) Potential marker-assisted selection for bc-1(2) resistance to bean common mosaic polyvirus in common bean. *Euphytica* **116,** 211–219.
5. Sanchez, A. C., Brar, D. S., Huang, N., Li, Z., and Khush, G. S. (2000) Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. *Crop Sci*. **40,** 792–797.
6. Casa, A. M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S., et al. (2000) The MITE family *Heartbreaker* (*Hbr*): molecular markers in maize. *Proc. Natl. Acad. Sci. USA* **97,** 10,083–10,090.
7. Bureau, T. E. and Wessler, S. R. (1992) *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* **4,** 1283–1294.
8. Bureau, T. E. and Wessler, S. R. (1994) Mobile inverted repeat elements of the *Tourist* family are associated with the genes of many cereal grasses. *Proc. Natl. Acad. Sci. USA* **91,** 1411–1415.
9. Rio, A., Puigdomenech, P., and Casacuberta, J. M. (1996) *Mrs*, a new subfamily of *Tourist* transposable elements. *Plant Mol. Biol*. **32,** 1221–1226.
10. Bureau, T. E. and Wessler, S. R. (1994) *Stowaway*: A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* **6,** 907–916.
11. Bureau, T. E., Ronald, P. C., and Wessler, S. R. (1996) A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild type rice genes. *Proc. Natl. Acad. Sci. USA* **93,** 8524–8529.
12. Shirasu, K., Schulman, A. H., Lahaye, T., and Schulze-Lefert, P. (2000) A contiguous 66Kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res.* **10,** 908–915.

13. Pozueta-Romero, J., Houlne, G., and Schantz, R. (1996) Nonautonomous inverted repeat *Alien* transposable elements are associated with genes of both mono and dicotyledoneous plants. *Gene* **171,** 147–153.
14. Casacuberta, E., Casacuberta, J. M., Puigdomenech, P., and Monfort, A. (1998) Presence of miniature inverted repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the *Emigrant* family of elements. *Plant J.* **16,** 79–85.
15. Surzycki, S. A. and Belknap, W. R. (1999) Characterization of repetitive DNA elements in *Arabidopsis*. *J. Mol. Evol.* **48,** 684–691.
16. Charrier, B., Foucher, F., Kondorosi, E., d'Aubenton-Carafa, Y., Thermes, C., Kondorosi, A., et al. (1999) *Bigfoot*: a new family of MITE elements characterized from the *Medicago* genus. *Plant J.* **18,** 431–441.
17. Yeadon, P. J. and Catcheside, D. E. A. (1995) *Guest*: a 98 bp inverted repeat transposable element in *Neurospora crassa*. *Mol. Gen. Genet*. **247,** 105–109.
18. Oosumi, T., Garlick, B., and Belknap, W. R. (1995) Identification and characterization of putative transposable DNA elements in solanaceous plants and *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **92,** 8886–8890.
19. Surzycki, S. A. and Belknap, W. R. (2000) Repetitive DNA elements are similarly distributed on *Caenorhabditis elegans* autosomes. *Proc. Natl. Acad. Sci. USA* **97,** 245–249.
20. Tu, Z. (1997) Three novel families of miniature inverted repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **94,** 7475–7480.
21. Feschotte, C. and Mouches, C. (2000) Recent amplification of miniature inverted-repeat transposable elements in the vector mosquito *Culex pipiens*: characterization of the *Mimo* family. *Gene* **250,** 109–116.
22. Braquart, C., Royer, V., and Bouhin, H. (1999) DEC: a new miniature inverted-repeat transposable element from the genome of the beetle *Tenebrio molitor*. *Insect Mol. Biol.* **8,** 571–574.
23. Ünsal, K. and Morgan, G. T. (1995) A novel group of families of short interspersed repetitive elements (SINEs) in *Xenopus*: Evidence of a specific target site for DNA-mediated transposition of inverted-repeat SINEs. *J. Mol. Biol.* **248,** 812–823.
24. Oosumi, T., Belknap, W. R., and Garlick, B. (1995) *Mariner* transposons in humans. *Nature* **378,** 672.
25. Morgan, G. T. (1995) Identification in the human genome of mobile elements spread by DNA-mediated transposition. *J. Mol. Biol*. **254,** 1–5.
26. Izsvak, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H., and Hackett, P. B. (1999) Short inverted repeat transposable elements in teleost fish and implications for a mechanism of their amplification. *J. Mol. Evol.* **48,** 13–21.
27. Zhang, Q., Arbuckle, J., and Wessler, S. R. (2000) Recent, extensive, and preferential insertion of members of the family *Heartbreaker* into genic regions of maize. *Proc. Natl. Acad. Sci. USA* **97,** 1160–1165.
28. Van den Broeck, D., Maes, T., Sauer, M., Zethof, J., Keukeleire, P., D'Hauw, M., et al. (1998) Transposon display identifies individual transposable elements in high copy number lines. *Plant J*. **13,** 121–129.

29. Casa, A. M., Mitchell, S. E., Smith, O. S., et al. (2002) Evaluation of Hbr (MITE) markers for assessment of genetic relationships among maize (*Zea mays L.*) inbre lines. *Theor. Appl. Genet.* **104,** 104–110.
30. San Miguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z., and Bennetzen, J. L. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274,** 765–768.
31. Costantini, C. and Besansky, N. J. (2000) Molecular entomology in Africa. *Parasitol. Today* **16,** 6–7.
32. Jiang, N., Bao, Z., Zhang, X., et al. (2003) An active DNA transposon family in rice. *Nature* **421,** 163–167.
33. Rowold, D. J. and Herrera, R. J. (2000) *Alu* elements and the human genome. *Genetica* **108,** 57–72.
34. Shedlock, A. M. and Okada, N. (2000) SINE insertions: powerful tools for molecular systematics. *BioEssays* **22,** 148–160.
35. Tu, Z. (2001) Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **98,** 1699–1704.
36. Le, Q. H., Wright, S., Yu, Z., and Bureau, T. (2000) Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **97,** 7376–7381.
37. Innis, M. A. and Gelfand, D. H. Optimization of PCRs. In *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. eds.), Academic Press, San Diego, CA, 1990, pp. 3–12.
38. Saiki, R. K. Amplification of genomic DNA. In *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. eds.), Academic Press, San Diego, CA, 1990, pp. 13–20.
39. Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., et al. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23,** 4407–4414.
40. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., and Mattick, J. S. (1991) Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res*. **19,** 4008.
41. Stewart, C. N. and Via, L. E. (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques* **14,** 748–750.
42. Stumm, G. B., Vedder, H., and Schlegel, J. (1997) A simple method for isolation of PCR fragments from silver-stained polyacrylamide gels by scratching with a fine needle. *Technical Tips Online* 1:17:T01115. (http://research.bmn.com/tto/).
43. Liu, Y. G. and Whittier, R. F. (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25,** 674–681.

13

Retroposon Mapping in Molecular Systematics

Norihiro Okada, Andrew M. Shedlock, and Masato Nikaido

Summary

Advances in genome sciences are demonstrating the dynamic nature of noncoding DNA regions, which are comprised largely of repetitive elements with no apparent function. Retroposons are one class of mobile genetic elements that amplify and move about the genome via a copy-and-paste mechanism that employs an RNA intermediate. Short and long interspersed elements (SINEs and LINEs, respectively) are types of retroposons of particular interest because of their active role in shaping the architecture of genomes and their diagnostic value as evolutionary markers for studies of phylogeny and population biology. Although the use of SINEs and LINEs for molecular systematic studies is proliferating, a comprehensive laboratory protocol that explicitly outlines how to isolate and characterize retroposons for systematic studies in a detailed, step-by-step fashion has been lacking. The present chapter addresses this gap in the literature by focusing on the strategy for isolating new SINEs from a genomic library, the screening process, the sequencing and characterization of clones into subfamilies, quantification of copy number in host taxa, and the critical diagnosis of phylogenetically informative SINE and LINE insertion patterns. Practical limits to the method are discussed in relation to sampling design, systematic character theory, and the empirical distribution of elements observed in eukaryotic lineages. Major steps in the experimental process are illustrated with case examples from a diversity of taxonomic groups and by published results in the molecular biology and systematics literature.

Key Words: DNA repeat; mobile DNA; interspersed element; retroposon; SINE; LINE; retrotransposition; reverse transcriptase; cDNA; RNA; systematics; genomics; eukaryote.

1 Introduction

1.1. General Introduction

The automation of DNA sequencing and molecular cloning techniques continues to dramatically expand access to the genomes of organisms. Furthermore, the flood of digital genomic information is drawing many comparative biologists into a computerized informatics cooperative. Results of comprehensive efforts, such as the Human Genome Project, have demonstrated

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

in fine detail what was made apparent almost 30 yr ago by nucleic acid renaturation studies; that is, the great majority of the eukaryotic genome does not encode for specific protein products, but is instead composed of repetitive elements with no apparent function ***(1)***. In the wake of genomics, this so-called "junkyard" is gaining new appreciation as a dynamic molecular "jungle," filled with active elements capable of moving about chromosomes, competing for critical enzymes, and parasitizing each other to gain new functionality. Retroposons are one such group of ubiquitous repetitive elements that are attracting attention as important evolutionary agents that can be readily exploited as systematic tools for phylogeny reconstruction and population analysis ***(2,3)***. As such, their diagnosis provides an important bridge between related subfields of evolutionary biology both above and below the organismal level.

The term retroposon refers to the way in which such elements move between a parent and target locus within the genome via an RNA intermediate ***(4)***. This copy-and-paste process creates a reverse flow of genetic information from RNA back into chromosomal DNA ***(5,6)*** and distinguishes retroposons from transposons, such as *mariner* and *P*, which jump about chromosomes directly via a cut-and-paste mechanism that leaves no original copies behind at parent loci ***(7,8)***. An important feature that is used to categorize retroposons is their ability to encode for reverse transcriptase (RTase) a critical enzyme for self-amplification ***(9–11)***. Of those retroposons that do not self-amplify, short interspersed elements, or SINEs, are the most numerous in the genome and are straightforward to diagnose in eukaryotes for systematic studies. SINEs range from 70 to 500 bp in size and can be grouped into two categories, one derived from tRNA and the other derived from 7SL RNA. The 7SL RNA-derived SINEs include only a few families of SINEs, namely, the primate *Alu* family, the rodent *B1* family, and the tree shrew *Tu* type I,II ***(12)***. All other SINEs characterized to date have been shown to be derived from tRNAs.

LINEs, which encode RTase for their amplification, can also be grouped into two categories, namely, the mammalian L1, which appears not to require specific sequences for amplification but rather requires a simple Poly-A stretch, and another category, comprising most LINEs, which requires strict sequence motifs at their 3' end to be recognized by RTase for amplification.

Most nonmammalian SINEs and LINEs share the same 3' end tail sequence, by the presence of which SINEs can be amplified by using RTase encoded by LINEs ***(13–17)***. In the case of mammals, most including those derived from tRNAs as well as 7SL RNA, appear to be amplified by the help of RTase encoded by mammalian L1. Therefore, in this case, the conserved sequence motifs are not observed at the 3'-end tail of SINEs. A diagram for the retrotransposition of SINEs that share the 3' end of partner LINEs is pictured in **Fig. 1**.

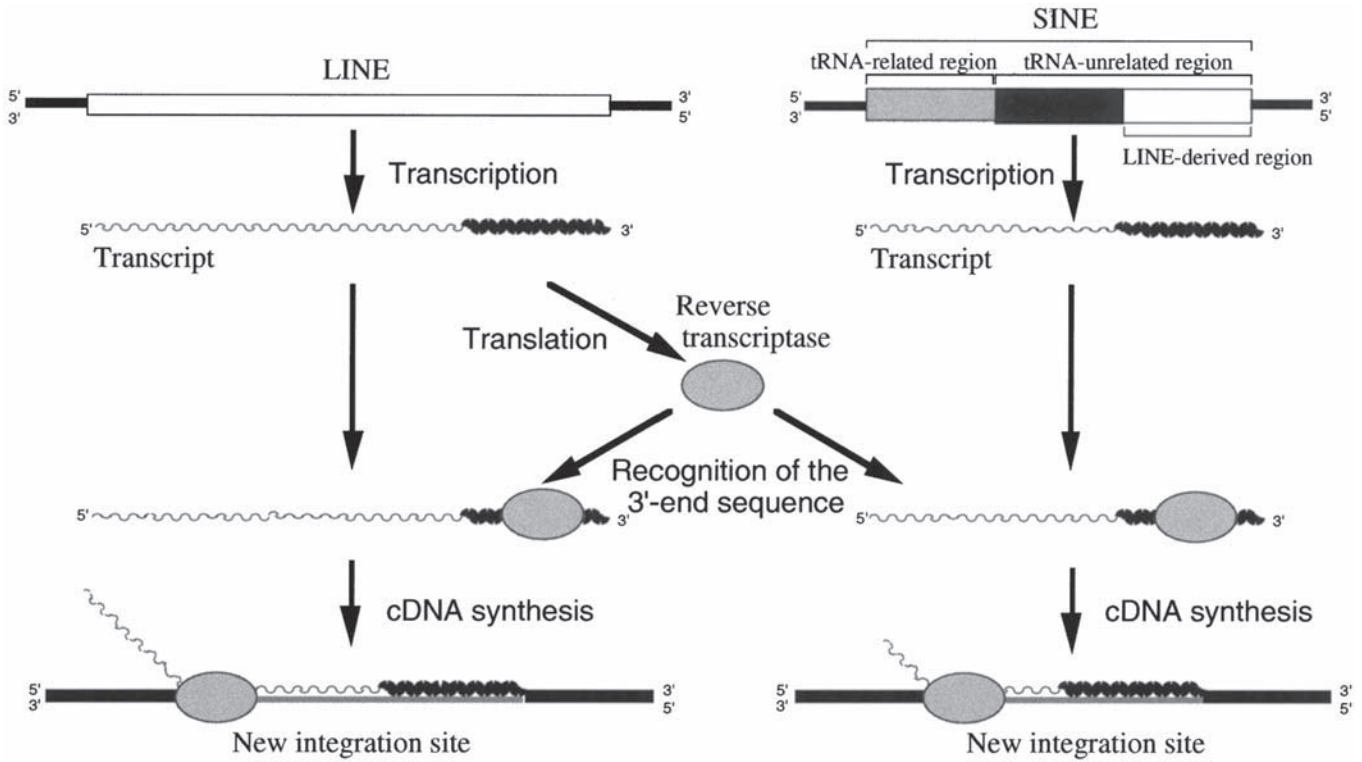


Fig. 1. A model for a retroposition process of SINEs. Enzymes required for retroposition of SINEs can be provided by a partner LINE that shares the same 3' tail sequence. The SINE transcript is reverse-transcribed into cDNA, and cDNA is then integrated into the host genome by using the mechanism called TPRT (target DNA-primed reverse transcription) in a similar manner adopted by LINEs.

The polarity of this process and the common occurrence of these elements in $>10^4$ copies throughout the host genome distinguish SINE insertion analysis as a powerful new approach to molecular systematics that may complement analyses of other forms of character data, most notably DNA sequences and morphology (e.g., **refs. *18–21***). Full reviews of SINE evolution and their importance as systematic tools are available elsewhere ***(3,5,6,22–26)***. In light of their broadening practical value, this chapter focuses on a detailed protocol for how isolation and characterization of SINEs in the laboratory can be completed for diagnosing common ancestry among eukaryotes.

1.2. SINE Evolution

Application of SINEs for systematic inference depends critically on an understanding of how SINEs evolve. SINEs can be categorized into families based on sequence similarity, and into subfamilies based on the presence of diagnostic nucleotides and/or deletions, as described in detail later in this review. The fate of a given SINE element will depend on numerous factors in the chromosomal environment ***(27)*** and on the accumulation of deleterious mutations that could preclude successful amplification of an element. Furthermore, because SINEs parasitize partner LINEs for access to the RTase necessary for their successful amplification, the death of a LINE automatically dictates the extinction of its corresponding SINEs in the same organism ***(15)***.

With regard to the long-term amplification profiles evident for SINEs in the genome, two contrasting models of SINE evolution have emerged from available empirical evidence: the master gene model, and the multiple-source gene model. In the master gene model (**Fig. 2A**) only a single or a few "master" loci give rise to all offspring copies, which do not have the capacity to replicate on their own. In this scenario, the amplification rate over time completely depends on the condition and activity of the master gene(s). Alternatively, the multiple-source gene model (**Fig. 2B**) includes offspring that can propagate in the same manner as parent copies, thus serving as "multiple-source genes" over evolutionary time. In this latter model, amplification rate is a function of the differential increase or decrease of total copy number derived from all source genes. The master gene model was largely developed in response to empirical evidence from the ID SINEs of rodents ***(28)*** and early studies of human *Alu* repeats ***(29)***, although comparative data from a variety of other taxonomic groups, including additional *Alu* work, has suggested that the multiple-source gene model is likely to be the most typical mode of evolution for most SINE families ***(3,27,30–33)***. In the practical context, the subfamily characterized by the presence of diagnostic nucleotides and/or deletions represents each source gene.

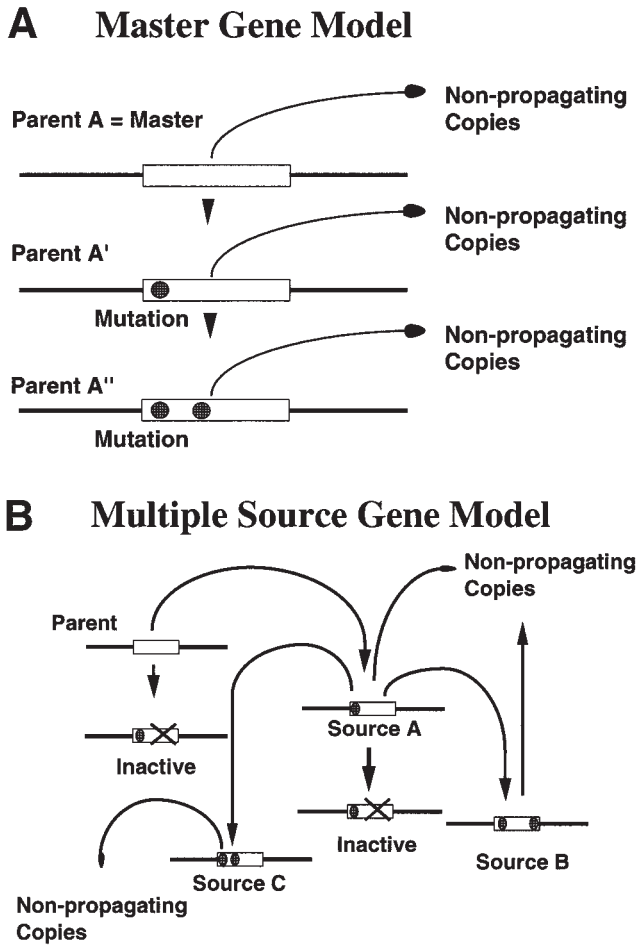


Fig. 2. Two alternative models to explain expansion of SINEs in the genome during evolution. **(A)** Master Gene Model, in which a single parent SINE, A, and its derived subfamilies A' and A", give rise only to nonpropagating copies. **(B)** Multiple-Source Gene Model, in which some offsprings of a parent SINE become inactive and others can propagate and serve as multiple sources (A, B, C) for new SINE copies over evolutionary time.

The birth and death of SINEs will define the active lifespan of these elements in the host genome, which is of direct relevance to their use as systematic tools. If the average sequence divergence of members of a SINE subfamily is small, it is reasonable to expect the subfamily is fairly young and still actively proliferating in the host. If the average divergence among members of the subfamily is large, then this subfamily of SINEs is probably relatively old and

may already be inactive or dead. Thus both living and dead SINEs are detectable in the genome. However, diagnosis of common ancestry among host taxa using SINE insertions is only possible within the active lifespan of any given subfamily. This basic tenet of the SINE method is further illuminated by the information contained in **Subheading 3.**

1.3. SINE Insertion Dynamics and Character Theory

The key to why SINEs can be used as powerful systematic tools lies in their irreversible, independent insertion into the host genome ***(3,18)***. Because there is no known mechanism that specifically removes SINEs from the genome, and because the likelihood of two elements inserting in exactly the same locus or being precisely excised at the same locus is negligible, we can consider the presence of a SINE at the same locus in two different taxa to be a polarized, derived phenotype in the genome. This defines a clade, or monophyletic group, in the strict sense of the methods of Hennig ***(34)***, which employ only synapomorphies, or shared, derived characters, to construct phylogenetic hypotheses. In this cladistic context, the known ancestral condition, or lack of a SINE insertion at a given locus, defines an outgroup unambiguously, without the need for establishing character polarity via comparative methods that produce well-known artifacts ***(35)***.

These basic assumptions of SINE insertion analysis are met only if the SINEs examined are fixed at loci in the host genomes in question, and if we can reasonably dismiss the probability of incongruence between gene trees and organismal trees because of ancestral polymorphism and incomplete lineage sorting (Shedlock, A. W., et al., unpublished). These latter issues raise statistical concerns that warrant the evaluation of SINE insertion results within the context of population genetics theory, but they are not prohibitive problems for the application of the method to a large array of important systematic problems above and below the species level. Full discussions of these theoretical issues and exceptional cases where the basic assumptions of the SINE method may not apply are available elsewhere in the literature ***(36)*** (Shedlock, A. W., et al., unpublished).

1.4. Working With SINEs in the Laboratory

Although the SINE method is rapidly proliferating in the literature as a new approach to phylogenetic inference, a comprehensive laboratory protocol that explicitly outlines how to isolate and characterize SINEs for molecular systematic studies in a detailed, step-by-step fashion has been lacking. This chapter attempts to close this gap in the literature by focusing on the strategy for isolating new SINEs from a genomic library, the screening process, the sequencing and characterization of clones into subfamilies, quantification of copy num-

ber in representative host taxa, and the critical diagnosis of phylogenetically informative SINE insertion patterns. Practical limits to the method are discussed in detail with respect to sampling limitations inherent both within and among genomes, and for cases where the taxonomic scope of a problem may not be realistic to address with SINE data.

2. Materials

2.1. In Vitro Transcription of Total Genomic DNA in a HeLa Cell Extract

1. A HeLa cell extract, prepared as described by Manley et al. ***(37)*** with a slight modification ***(38)*** to make the extraction of RNA polymerases more efficient. The extract can be stored at –80°C for more than one year.
2. In our analysis of the transcripts, we used 8% nondenaturing polyacrylamid gel.
3. Total genomic DNA transcription is performed with Manley's buffer, containing 46 m*M* $MgCl_2$, 160 m*M* $(NH_4)_2SO_4$, 1.2 m*M* EDTA, 9 m*M* dithiothreitol (DTT), and 40% (v/v) glycerol.
4. When α-^{32}P-guanosine triphosphate (GTP) is used as a precursor, transcription is performed in the presence of 3 m*M* adenosine triphosphate (ATP), 3 m*M* cytidine triphosphate (CTP), 3 m*M* uridine triphosphate (UTP), and 125 μ*M* GTP.
5. Genomic DNAs are extracted as described by Blin and Stafford (**ref. *39*** and *see* **Note 1**). Genomic DNA should be stored at 4°C.
6. There are many types of polyacrylamide gel electrophoresis (PAGE) apparatus available commercially, and the arrangements of these apparatuses differ from one manufacturer to another. This experiment can be performed on any PAGE apparatus.
7. TBE buffer is used for the electrophoresis.

2.2. Construction and Screening of Genomic Library

1. According to the experiment, probes can be labeled by radioisotopes in different ways. A ^{32}P-labeled RNA is prepared with total genomic DNA transcription as mentioned above.
2. PCR product is labeled with α-^{32}P dCTP by primer extension using BcaBest DNA polymerase (Takara).
3. Oligonucleotides are labeled with γ-^{32}P ATP by kination using T4 polynucleotide kinase (Nippon Gene).
4. The genomic library of each animal is constructed by the ligation of DNA (digested by an appropriate restriction enzyme) and a vector. We usually use plasmid (pUC 18 or 19) or bacteriophage λgt10 arms (Stratagene) to construct a library.
5. Two different concentration of Sucrose (40% and 10% (w/v)) should be prepared for making a sucrose density gradient. The sucrose solutions are made in a buffer containing 1*M* NaCl, 20 m*M* Tris-HCl, pH 8.0, and 5 m*M* EDTA, pH 8.0.
6. To make the sucrose density gradient, we use a gradient maker, centrifuge tubes (Ultra-Clear™, Beckman), and flexible plastic tubes. The centrifugation is performed in an SW41 rotor using an L8-70M Ultracentrifuge (Beckman).

7. The following solutions are used for the Southern hybridization: 20X SSC, 10% SDS; 50X Denhardt's solution; and herring sperm DNA (10 mg/mL). They are diluted to the optimal concentration (*see* **Subheading 3.2.**).
8. We usually use XL1 blue, JM 105, and DH5α strains of *E. coli* for transformation to clone the DNA.
9. To perform a Southern hybridization, we use a nylon membrane (GeneScreen Plus, NEN Research Products).

2.3. Sequencing

The sequence is performed by the dideoxy chain-termination method. We use two different types of sequencers. The DNA sequencer (LI-COR) requires preparation of sequencing gels, and the sequence reaction is performed with SequiTherm EXCEL™ Long-Read™ DNA Sequencing Kits-LC (Epicentre Technologies), which enable us to sequence about 1.2 Kbp for each clone. The ABI 3100 sequencer (Applied Biosystems) uses a capillary type, and sequence reaction is performed with BigDye Terminator Cycle Sequencing FS Ready Reaction Kit (Applied Biosystems), by which we can sequence about 700 bp for each clone. We choose the appropriate sequencing machine depending on the clone whose sequence we wanted to determine. For sequencing, we usually use the primers that match to the plasmid sequence or to the particular SINE sequence to obtain the SINE flanking sequences.

2.4. Analyzing the PCR Products

1. Generally, the bands obtained by SINE flanking PCR range from 300 bp to 800 bp. So, we prepare 3% (w/v) SeaKem GTG agarose gel (FMC BioProducts) for the electrophoresis. It is carried out with TAE buffer, followed by staining with ethidium bromide and visualization of bands of DNA under UV irradiation.
2. We use a horizontal slab gel for agarose gel electrophoresis, which is carried out in a normal tank. To confirm the presence or absence of SINE, we also determine the sequences of PCR bands.
3. If necessary, the DNA is recovered from the agarose gel by using a QIAquick™ Gel Extraction Kit (QIAGEN) and is ligated to the vector.
4. We use pT7Blue T-Vector (Novagen) and pGEM-T Vector (Promega) for cloning the PCR bands. The ligation kit ver. II (Takara) is used for the ligation of foreign DNA and plasmid DNA.

2.5. Software

1. To analyze the sequences collected by the experiments, we use GENETYX-MAC (Software Development Co., Ltd). This software provides many useful options (e.g., homology search, multiple alignment, constructing the NJ tree, and so on) and is useful for analyzing the SINE containing sequences.
2. To find the sequence similarity between our sequences and those deposited in databases (DDBJ, EMBL, GenBank), we use the BLAST program ***(40)***.

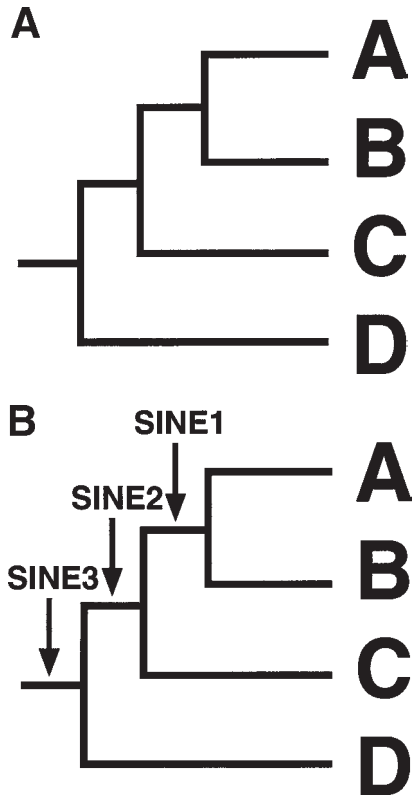


Fig. 3. Schematic representation of the strategy for the cladogram construction by using the SINE method. **(A)** The real phylogeny of the species A, B, C, and D. **(B)** Three SINE loci inserted in an ancestor of all the four species A, B, C, and D. (*Continued on the next page*.)

3. To design the oligonucleotide primers for the experiment, the CPrimer program is used in our laboratory. This program tells us the T_m temperature and the possibility of dimers and hairpin structures (*see* **Note 2**).

3. Methods

3.1. How to Select a Species for Creating a Genomic Library

The SINE method involves a number of basics steps: (1) making a genomic library from a selected species; (2) isolating clones that contain a SINE locus; (3) determining the DNA sequence of clones; (4) designing the polymerase chain reaction (PCR) primers in the flanking sequence of the SINE locus; and (5) PCR diagnosis of SINE presence or absence among related species in question (**Fig. 3C**). Although the SINE method provides conclusive results,

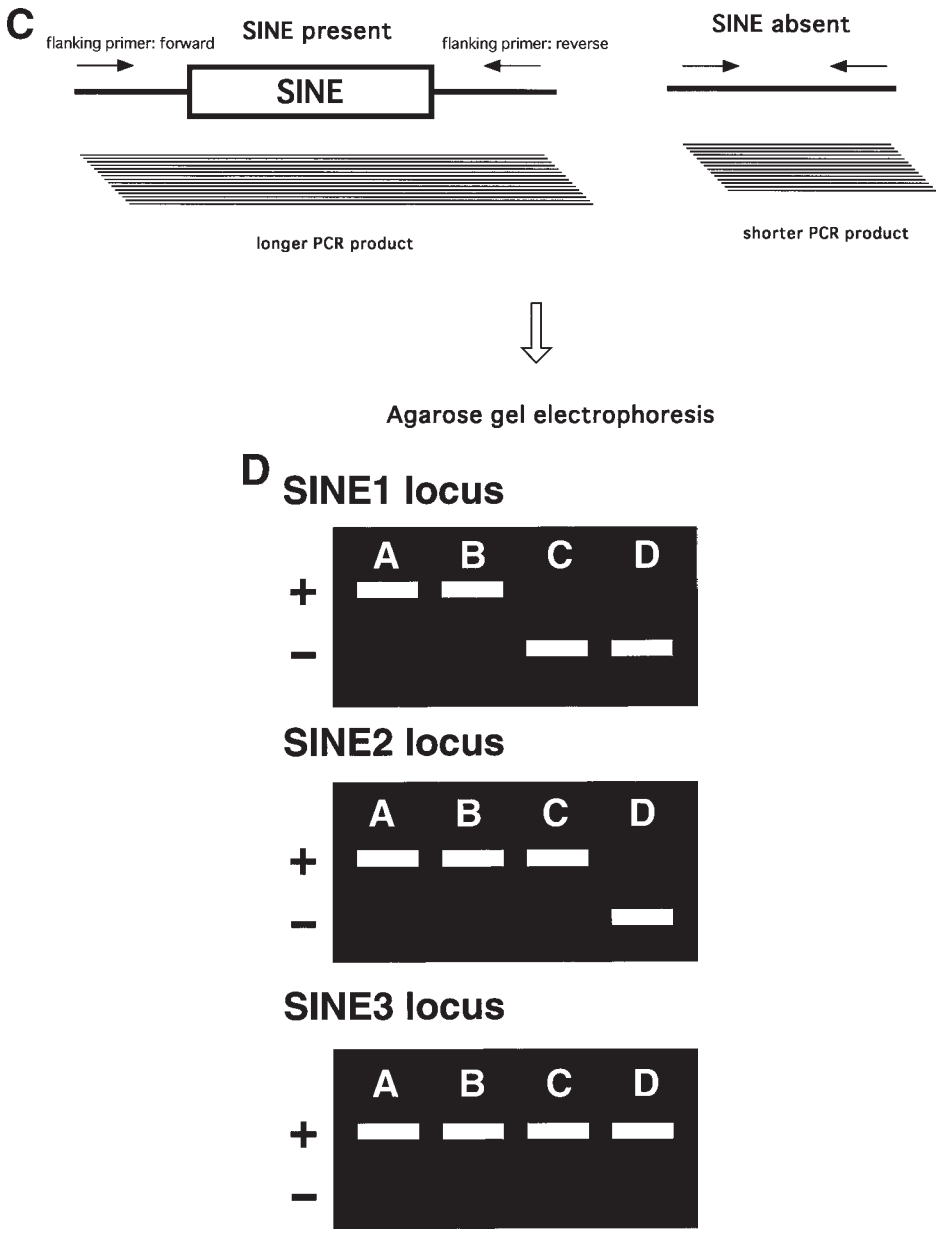


Fig. 3. *(continued)* **(C)** Forward- and reverse-strand primers are designed to anneal at sites flanking a SINE element inserted at a specific locus. **(D)** The PCR patterns of the gel electrophoresis of three SINE loci in which presence (+) or absence (–) of SINE insertions are assayed.

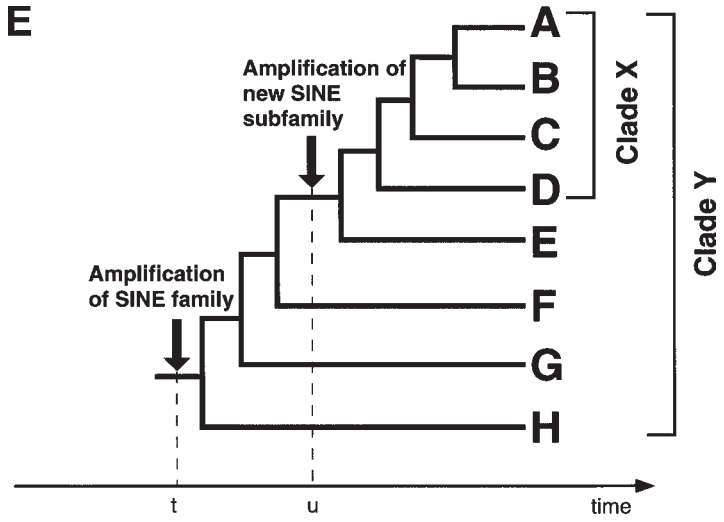


Fig. 3. *(continued)* **(E)** Phylogenetic tree showing the timing of the amplification of new SINE family and subfamily.

it requires substantial time and effort in the laboratory, and successful experiments must be carefully designed.

Consider the case for determining the phylogenetic relationships among closely related species A, B, C, and D. Suppose that the real phylogeny of these species is as described in **Fig. 3A**. In this case, if species D is chosen as the host genome from which to isolate SINEs, it will not yield any phylogenetically informative SINE loci because species D contains SINE loci that were inserted in a common ancestor of all four taxa in question, as well as loci of more ancient origin. Such loci inserted in an ancestor of all 4 species are indicated as SINE 3 in **Fig. 3B**, and their PCR pattern for insertion presence (+) or absence (–) would look like the gel shown at the bottom of **Fig. 3D**. To isolate informative SINEs, one must select species A or B, which can possibly provide all three possible SINE loci indicated in **Fig. 3B**, namely, SINE 1, SINE 2, and SINE 3. Each of these loci defines a clade, or monophyletic group that shares a common ancestor, within the evolutionary history of these species.

Ideally, it would be best to make genomic libraries from all four of the species, in order to optimize sampling of informative loci and to reduce the experimental bias created by ascertaining clones from only a single individual or subset of individuals per species ***(41)*** (Shedlock, A. M., et al., unpublished) (*see* **Note 3**).

3.2. How to Isolate a New SINE Family From a Species

When no SINE families are known in a particular species, for example, species A in **Fig. 3**, it is necessary to newly characterize them from the genome. There are two possible methods to isolate a new SINE family from a selected species. One involves total genomic DNA transcription in vitro ***(42)***, and the other involves sequencing more than 60 kbp of genomic DNA facilitated by new high-throughput automated DNA sequencing methods.

3.2.1. Total Genomic DNA Transcription In Vitro

Most SINEs are known to have been derived from tRNAs, so they have internal promoters for RNA polymerase III. Although it is known that SINEs are very rarely transcribed in vivo, they can easily be transcribed in vitro from naked DNA ***(42)***. In vitro transcription is performed as follows:

1. Combine the following in a single tube in the order listed:
   a. 5 μL of 10X Manley buffer.
   b. 10 μL of 5X NTP.
   c. 3 μg of naked DNA.
   d. 1 μL of α-^{32}P GTP.
   e. 20 μL of HeLa extract.
   f. Add distilled water for a final volume of 50 μL.
2. Put the tube in a water bath at 29°C for one hour.
3. Treat the reaction mixture with phenol:chloroform (1:1), followed by ethanol precipitation. Dissolve the pellet in 10 μL of distilled water.
4. Perform an electrophoresis using 8% polyacrylamide gel at 800 V for 7 h.
5. Remove only one of the glass plates and cover the gel with Saran wrap.
6. Obtain an autoradiogram by exposing the gel for approx 24 h to X-ray film at –80°C with an intensifying screen.

In several cases, one to several discrete bands is detected by in vitro transcription of total genomic DNA ***(42,43)***. This radio-labeled RNA can be used as a probe to screen a genomic library from a species of interest. When the transcript forms a clear band in a gel, we can be confident that it represents a real SINE family, because we expect all identical transcripts from each locus of a given SINE family to collectively form a discrete band. Even in situations producing smeared transcripts from genomic DNA, we can use these transcripts as a probe for screening. However, in this latter case it is possible that the transcripts may represent multiple SINE families (*see* **Note 4**). **Figure 4** shows several examples of the pattern of transcripts from total genomic DNA from selected animal species (modified from Fig. 1 in **ref. *42***; used with permission).

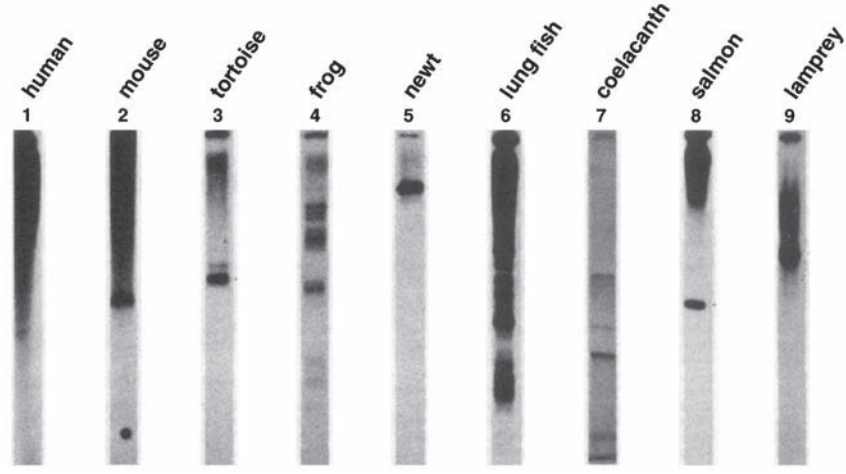


Fig. 4. Examples of the pattern of transcripts from total genomic DNA of several animal species. (Reprinted with permission from **ref. *42***.)

3.2.2. Screening of a Genomic Library and Sequencing of SINE Flanking Regions

The screening of a genomic library is indispensable for obtaining the sequences of SINEs and their flanking regions from the genomes of animals. If we already know the SINE family in certain animal species, the SINE sequence (oligonucleotide or PCR product) is used as probe for screening. In the case that no SINE families are characterized to date, total genomic DNA transcript is used as probe for screening. We usually perform a colony hybridization method for screening a plasmid library, so this technique will be focused in this section. The construction of genomic library is performed in the following procedures:

1. Digestion of 100 μg of genomic DNA by a restriction enzyme. (The choice of restriction enzyme does not affect the efficiency of the genomic library. If there is a restriction site in the particular SINE sequence of interest, however, we have to exchange the enzyme with an appropriate one.) The volume of this reaction mixture is adjusted to 200 μL in total. The condition of digestion is checked by electrophoresis in 1% agarose gel using an aliquot of 5 μL of the reaction mixture.
2. Prepare the gradient containing 1*M* NaCl, 20 m*M* Tris-HCl, pH 8.0, and 5 m*M* EDTA, pH 8.0, in 14 × 89 mm ultracentrifuge tubes (Ultra-Clear™, Beckman).
3. Load the digested DNA (about 200 μL) on the top of the gradient.
4. Centrifuge the gradients at 22,000*g* for 15 h at 15°C in a Beckman SW41-Ti rotor using an L8-70M Ultracentrifuge.

5. Collect the fractions through flexible plastic tube with a 50-μL capillary inserted into the bottom of the centrifuge tube.
6. Check the fractionated DNA by electrophoresis in 1% agarose gel.
7. Collect the appropriate fractions, which contain approx 1.5 to 2.5 kbp DNA fragments. Purify the fractions by ethanol precipitation. Add an equal volume of 100% ethanol and 1/10 volume of 3*M* sodium acetate. Centrifuge the samples at 12,000*g* for 15 min at 4°C. Rinse the pellets with 70% Ethanol.
8. Dissolve the pellets in 15 μL of distilled water.
9. Prepare the ligation reaction as follows: Mix the 1 μL of vector DNA (200 ng/μL), an equimolar amount of fractionated DNA (usually, it is about 1 μL), 12 μL of ligation buffer (Takara ligation kit ver. II sol. A), and 2 μL of ligase (Takara ligation kit ver. II sol. B) in a sterile tube.
10. Incubate the reaction mixture for more than 1 h at 17°C.

Transformation of the genomic library with bacteria is performed as follows:

1. Add the genomic library mixture to 400 μL of competent cells, and swirl the tubes gently several times. Put the tube on ice for 30 min.
2. Transfer the tube to water bath heated at 42°C and wait 30 s. Put back on ice and leave to chill bacteria for 1–2 min.
3. Add 1.2 mL of Luria Bertani (LB) medium to the mixture. Warm the cultures for 10 min at 37°C, shaking several times.
4. Transfer the 200 μL of transformed competent cell onto agar LB medium containing ampicillin (50 μg/mL). Spread the transformed cells over the surface of the agar plate. In total, eight plates will be prepared.
5. Incubate the plates at 37°C. It takes approx 12 to 15 h to detect the colonies.

Colonies are then screened with SINE probes:

1. Place a nylon membrane (Colony/Plaque Screen™, NEN™ Life Science Products), and wait 1 h at 37°C or room temperature.
2. Mark the nylon membrane in three or more asymmetric locations by stabbing through it and into the agar.
3. Remove the membranes and put them to denaturing solutions (0.4*M* NaOH, 0.6*M* NaCl) and wait 3 min. The solutions should be poured into an appropriate plastic box.
4. Transfer the membranes to neutralizing solution (1*M* NaCl, 0.5*M* Tris-HCl, pH 7.0) and wait 3 min.
5. Lay the membranes on the dry sheet, allowing them to dry at room temperature for 1 h.
6. Put the agar plate into a 37°C incubator for 1–2 h to allow the colonies regrow. After checking the regrowth of the colonies, store the plates at 4°C.
7. Put the membranes into the hybridization plastic bag. Different hybridization solutions should be prepared for different kind of probes. The content of the solution for PCR products or total genomic DNA transcripts contains 50% formamide,

1% SDS, 6X SSC, 2X Denhardt's solution, and herring sperm DNA (100 μg/mL). For the oligonucleotide probes, formamide is not added to this solution.
8. Add the hybridization solution to the bag, and squeeze the bubble. Seal the open end of the bag with the heat sealer.
9. Incubate the bag for 1–2 h at 42°C.
10. Open the bag by cutting the corner. Add the probe to the solution, and then squeeze as much air as possible from the bag. Reseal the bag, and put it into a water bath set at 42°C for an appropriate period (in our case, about 12 h).
11. Remove the bag from the water bath and cut off the corner. Pour the hybridization solution into a container for disposal. Remove the membranes and submerge them in a tray containing wash solution (2X SSC and 1% SDS) and incubate for 10 min at appropriate temperature (approx 42°C). Dump the wash solution into the disposal container, and repeat this procedure three times.
12. Place the washed membranes on the used X-ray film marked with radioactive ink, and cover it with plastic film. Put the membrane sheet, X-ray film, and intensifying screen into a film folder in a darkroom. Expose the membranes to X-ray film for appropriate period (approx 16 h, empirically) at –80°C.
13. After developing the X-ray film, adjust the position of membranes and check the corresponding colonies on the agar plate. Next, we isolate plasmids from the colonies and sequence the SINE flanking region. Minipreparation is performed to isolate plasmids from the cultured positive bacteria.

To obtain the sequences of unknown SINEs, the sequencing of the positive clones selected by the total DNA transcript is needed. In this case, the sequencing is performed by the primers that match the plasmid sequences of both side of the insert, and then SINE-like structure should be searched (*see* **Subheading 3.3.**). If the structure of a certain SINE family has already been characterized, the primers for sequencing are designated in the internal region of the SINE (this region should be conserved among the SINE family). For determining the sequence of the downstream region of the SINE, the primer should be designated from the 5' to 3' end of the SINE. For the upstream region, the primer should be designated from the 3' to 5' of the opposite strand.

3.2.3. Sequencing of Genomic DNA of More Than 60 kbp by an Automatic Sequencer

Our empirical studies indicate that the copy number of a SINE family in a genome usually exceeds ten thousand. Suppose the total genome is 3×10^9 bp in length, and the size of one SINE is 300 bp. Such a SINE family would thus represent 0.1% of the genome (e.g., $300 \times 10^4 = 3 \times 10^6$). One can reliably find two independent SINE sequences in randomly isolated DNA fragments of this species by sequencing more than 60 kbp (e.g., $600 \times 100/0.1 = 6 \times 10^6$). This outcome is becoming an increasingly straightforward and easily attainable goal in the lab with access to newer models of high-throughput automatic DNA

sequencers. For example, a new SINE family was recently characterized from the elephant genome, and it was shown that this new SINE is distributed among all species of Afrotheria ***(44)***. This method can be applied to the genomes of all mammals and possibly most vertebrates, but it may be difficult to develop for invertebrate genomes without more baseline information about the existence of SINE families in this group of eukaryotes (*see* **Subheading 5.3.**).

3.3. How to Accurately Identify a SINE Family and Deduce Its tRNA-Like Structure

If one determines sequences of multiple copies of repetitive units according to the methods described above, it is possible to align them and deduce a consensus sequence of a repetitive family. Because there are many repetitive sequences in the genome besides SINE elements, it is essential to diagnose the sequence properly. Most SINEs are known to have been derived from tRNA, so they contain promoters for RNA polymerase III. RNA polymerase III promoters are conserved sequence blocks, with the profiles of the first and second promoters separated from each other in the genome. The second promoter is highly conserved and can be easily recognized empirically.

Consider the example of CHR-2 SINEs. The tRNA-like structure of these elements can be established as follows (**Fig. 5**):

1. First, build a consensus sequence of CHR-2 SINEs from an alignment of several sequences of CHR-2 (*see* **Subheading 3.4.1.**).
2. Visually search for the consensus sequence of the second promoter for RNA polymerase III. The sequence is 5'-GT(orA)TCG(orA)-3'. There can be no exceptions to this exact motif when screening for this intact promoter. When this motif has been found, make a stem-and-loop structure that includes this second promoter sequence. The number of bases in the loop is seven, and that of base pairs in the stem is five. Even if all bases in the stem region do not make base pairs, put these bases in the appropriate positions in a tRNA context, as indicated in **Fig. 5A**.
3. The five bases in the 5' upstream direction from the stem region are considered as a unit because in cytoplasmic class I tRNAs the extra loop region consists of five bases. In the case of the CHR-2 SINE, the sequence of this unit of five bases is 3'-CAGGG-5'. The sequence of 3'-PyPyPuPuPu-5' is typical of the extra loop in several tRNAs ***(45)***, and this finding adds confidence in deducing the tRNA origin of this SINE (**Fig. 5B**).
4. The next five bases are taken as another unit, which forms the aminoacyl-stem region of the tRNA-like structure. In this case, the sequence is 3'-GACGT-5' (**Fig. 5C**).
5. The next seven bases are taken as yet another unit, which forms the anticodon-loop region of the tRNA-like structure. In this case, the sequence is 3'-AACCGTC-5'. The AA residues at the 3' end and the 3'-TC-5' residues at the 5' end are a good indication of the tRNA origin of this SINE, because these bases are highly conserved in most tRNAs ***(46)***. This further supports the tRNA origin of this SINE (**Fig. 5D**).

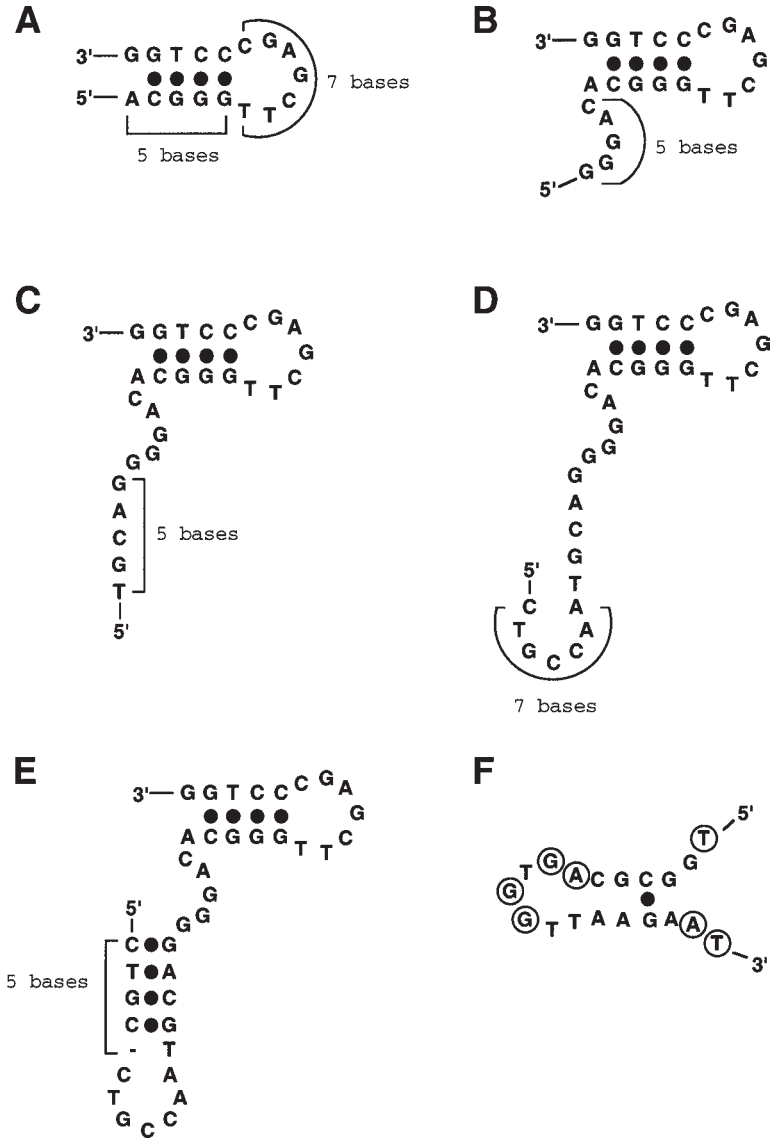


Fig. 5. A procedure of steps for deducing the tRNA-like structure of a SINE. For details *see* **Subheading 3.3.** (*Continued on the next page*.)

6. Usually, the next five bases are taken as another unit, and they should form base pairs with the five bases assigned to the anticodon-stem region (**Fig. 5C**). In the present case, the sequence is 3'-CGTCT-5', and only the first 4 bases are well matched with the other half of the anticodon stem. To accurately align this unit, a

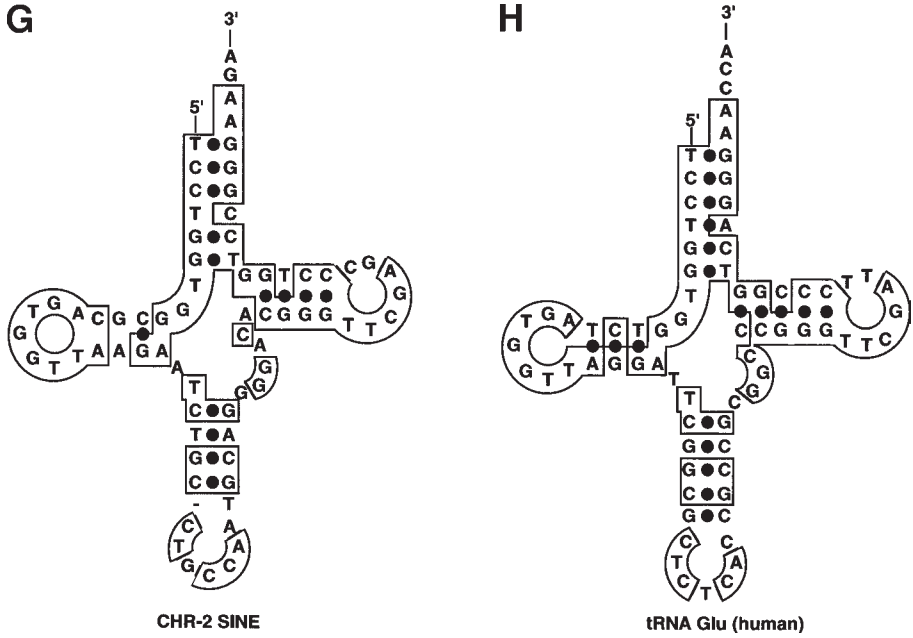


Fig. 5. *(continued)* A procedure of steps for deducing the tRNA-like structure of a SINE. For details *see* **Subheading 3.3.**

deletion is placed at the position of the first base on the 3' side of the anticodon stem (**Fig. 5E**).

7. Next, construct a stem and loop structure for the D-region of this tRNA-like structure. The base numbers of the stem and loop are usually four and eight, respectively, but may vary by 1–2 bases, especially in the tRNA-like structure of SINEs. Apparently, the next several bases of CHR-2 do not form a significant secondary structure. In this case, focus attention on the first promoter region. The most prominent feature in the first promoter region is the presence of two Gs in this region. Other features are G at position 15 and A at position 14 in the loop. This A at position 14 is the first base in the loop on the 5' side. Therefore, place these bases into the corresponding positions of the loop of the tRNA-like structure (**Fig. 5F**). It should be noted that the T base, namely the first base in **Fig. 5F**, is well known to be highly conserved in all tRNA molecules.
8. The tRNA-like structure of a CHR-2 SINEs may be deduced by combining the sequence shown in **Fig. 5F** with other sequences for this family (**Fig. 5G**).
9. Next, search for similarities between the CHR-2 SINE and actual tRNAs by using the BLASTN program in the GenBank DNA database ***(40)***. In this example, tRNAGlu is the most similar to the sequence of CHR-2. **Figure 5H** shows the secondary structure of human tRNAGlu.

3.4. How to Characterize a SINE Family Into Subfamilies

Suppose that a SINE family has been characterized in the genome of species A in the phylogeny shown in **Fig. 3A**, and the time at which this SINE family was first generated during evolution is not known. Also, suppose that this SINE family was first generated in an ancient common ancestor of all taxa of clade X in **Fig. 3E**. In this case, SINE copies present in the genome of species A include old SINEs amplified at the time *t* in **Fig. 3E**, as well as younger SINEs amplified at the time *u*. When a genomic library of species A is screened with the consensus sequence of this SINE family, both old SINEs as well as new SINEs may be isolated. Since only the phylogeny of species A, B, C, and D is sought, it would be inefficient to examine the times of all the amplification events of the SINEs isolated. Rather, it is far more efficient to try to isolate SINE loci that amplified at times near the divergence of species D and that span the divergences of all 4 taxa comprising clade X in **Fig. 3E**.

As briefly described in the Introduction, SINEs as well as LINEs are believed to amplify according to the multiple-source gene models ***(27,47)***. If a certain source gene was subject to mutations and was successfully amplified during evolution, this mutated source gene can be recognized as a subfamily within its respective SINE family ***(48,49)***. Subfamilies are amplified at certain stages of evolution. Therefore, if a subfamily can be characterized that was only amplified in a common ancestor of species A, B, C, and D, copies of this subfamily can be effectively used for determination of the phylogenetic relationships of these four taxa.

The consensus sequence of the SINE family in species A is established as part of the procedure described above, and the procedure allows for the design of one PCR primer at the 5' end of the SINE and another primer at the conserved region near the 3' end, encompassing almost the entire SINE sequence. This primer set can be used to amplify many copies of SINEs by PCR using genomic DNA from species A. The PCR product of this reaction can be cloned in an appropriate vector DNA and sequenced. At this point, sequence determination of 100 copies of SINEs from the PCR product is not a difficult task. By aligning these sequences, diagnostic nucleotides or possible deletions can be identified that represent subfamilies of this SINE family (**Fig. 6**). Diagnostic nucleotides are defined as those that changed cooperatively in more than one nucleotide position in a certain subfamily, and they can be distinguished from neutral mutations that were accumulated randomly in the SINE sequence during evolution. After successful characterization of subfamilies based on the presence of diagnostic nucleotides, and sometimes of specific deletions, it may be possible to examine the taxonomic distribution of a given subfamily by dot-blot hybridization or PCR (*see* **Subheading 3.5.1.** and **Fig. 9**).

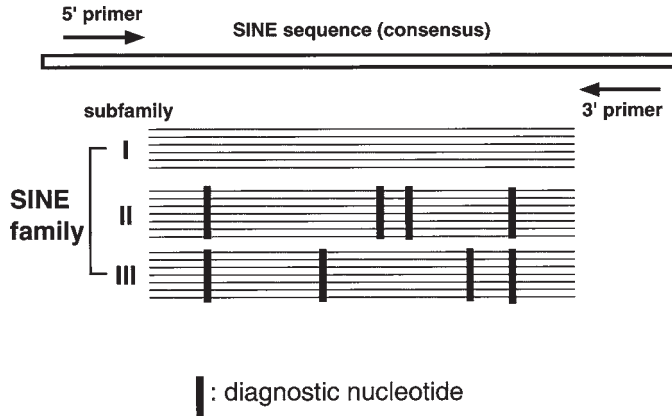


Fig. 6. Schematic representation to explain the presence of diagnostic nucleotides, which are used to divide a SINE family into subfamilies. Thick black bars represent diagnostic nucleotides shared between or among subfamilies.

3.4.1. An Empirical Study

Figure 7 shows an alignment of copies of the CHR-2 SINE family, which was originally characterized as being present in genomes of Cetaceans, Hippopotamuses, and Ruminants ***(19)***. It is easily recognized that there are six subfamilies in CHR-2 SINEs ***(50)***. Because of the presence of deletions, full length (FL), middle deletion I (MDI), middle deletion II (MDII), and the shortest group are characterized. Then, the shortest group can be separated into subfamilies of deletion type (DT), cetacean deletion type (CD), and cetacean deletion odontoceti specific (CDO). **Figure 8** shows an alignment of consensus sequences of these subfamilies. **Figure 9** shows a dot-hybridization experiment using probes specific to CD, CDO, and other subfamilies, respectively. The result clearly shows that the CD subfamily is specific to the genome of Cetaceans (toothed whales plus baleen whales), and that the CDO subfamily is specific to the genome of toothed whales. Therefore, SINEs belonging to the CD subfamily are useful for inferring the phylogeny of Cetaceans, especially the baleen whales, whereas those belonging to the CDO subfamily will be useful for inferring relationships of the toothed whales ***(51)***. The distribution of copies of the CDO subfamily also suggests the monophyly of odontocetes, including sperm whales, which has been one of the most contentious issues in mammalian systematic biology.

3.5. Flanking SINE PCR

After isolating SINE loci from species A that belong to a subfamily generated in a common ancestor of clade X (**Fig. 3E**), and determining their

sequences, PCR experiments may be performed to diagnose the presence or absence of insertion at a given locus. Referring to the flanking sequences, primer sequences can be selected. Primer design should take precautions against formation of secondary structural folds and tandem annealing between upstream and downstream primers. This situation can be easily checked with a variety of standard software programs written to facilitate PCR primer design that are available commercially or over the internet. We typically choose the melting temperature of the oligonucleotide primers at approx 55°C, so the annealing temperature for PCR should be based on this temperature when optimizing amplification of the orthologous locus from species B, C, and D (*see* **Note 5**).

3.5.1. Empirical Studies

Figure 10 shows an example of PCR results together with hybridization experiments completed with a single filter using two different probes. **Figure 10A** shows the pattern of PCR, providing evidence that ocean dolphins are monophyletic. PCR products from ocean dolphins have the expected size of a fragment containing an inserted element, whereas those from other toothed whales have the expected size of a fragment lacking insertion at the Mago 19 locus. **Figure 10B** shows the hybridization experiment using the SINE probe, whereas **Fig. 10C** shows the hybridization experiment using the flanking DNA of this locus. This latter experiment was performed to demonstrate that the orthologous locus was faithfully amplified by PCR in species other than the short-finned pilot whale, from which this locus was originally isolated and characterized.

3.6. Interpretation of PCR Data

In cases where relatively recently diverged species are investigated, flanking sequences at orthologous SINE loci are fairly conserved and typically do not cause prohibitive problems for PCR diagnostics. However, when investigating taxa with relatively old divergences, failed PCR may become more frequent and make interpretation of experimental results less straightforward. Failed PCR should not be confused with SINE-minus data, that is, the successful PCR amplification at a given locus that demonstrates the absence of SINE insertion. Failed PCR is a form of missing data and should be coded as such (e.g., with "?") when performing parsimony analysis of a SINE character matrix encoded for patterns of presence or absence of insertion. This important methodological issue is discussed in detail in a recent review of SINE analysis and its implications for understanding the origin of whales (*see* **ref. *24*** and **Note 6**).

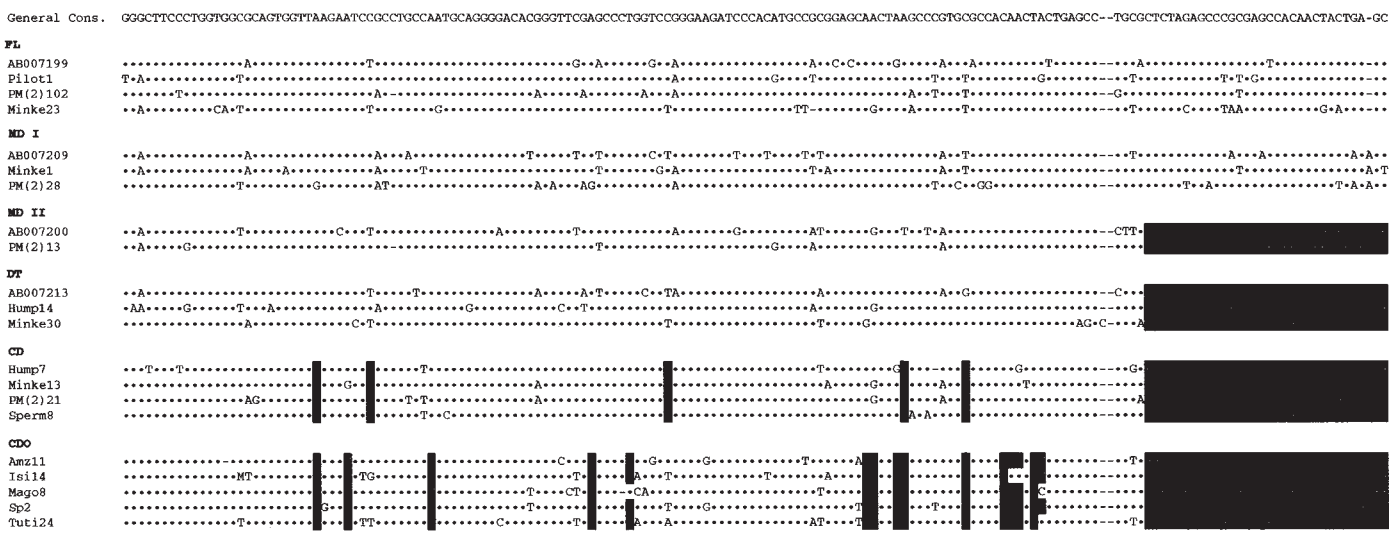


Fig. 7. An alignment of sequences of CHR-2 SINEs. The correlations between names of clones and animals are as follows: Hump, humpback whale; Minke, minke whale; PMs, Sperm, and Sp, sperm whale; Amz, Amazon river dolphin; Isi, Dall's porpoise; Mago, short-finned pilot whale; Tuti, Baird's beaked whale; ABs, hippopotamus. Diagnostic nucleotides and deletions are highlighted.

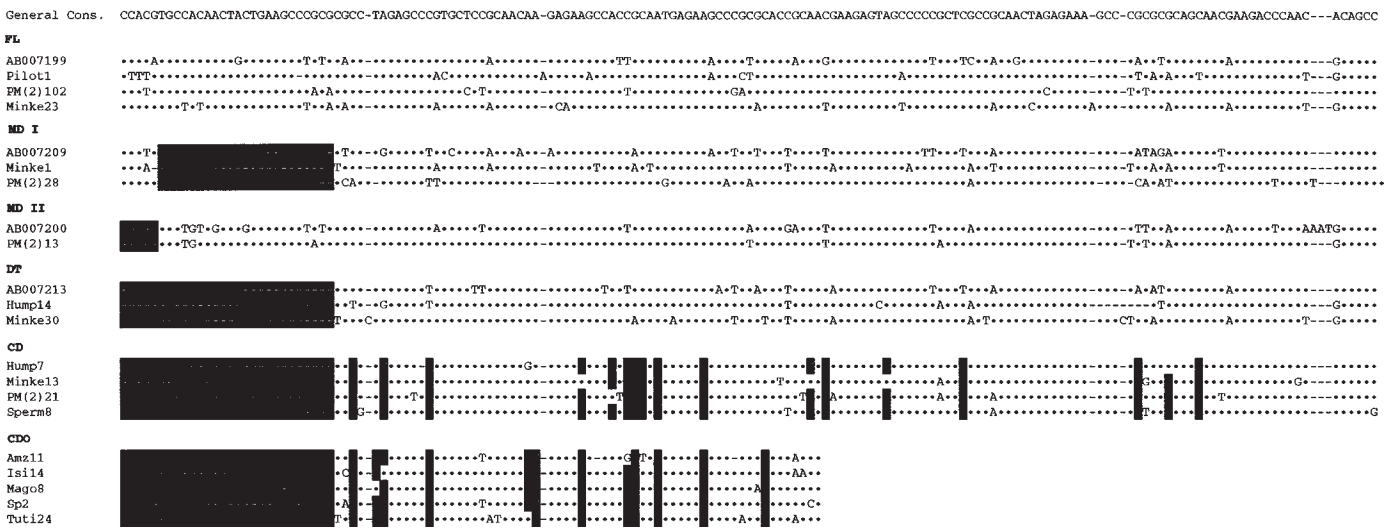


Fig. 7. *(continued)*

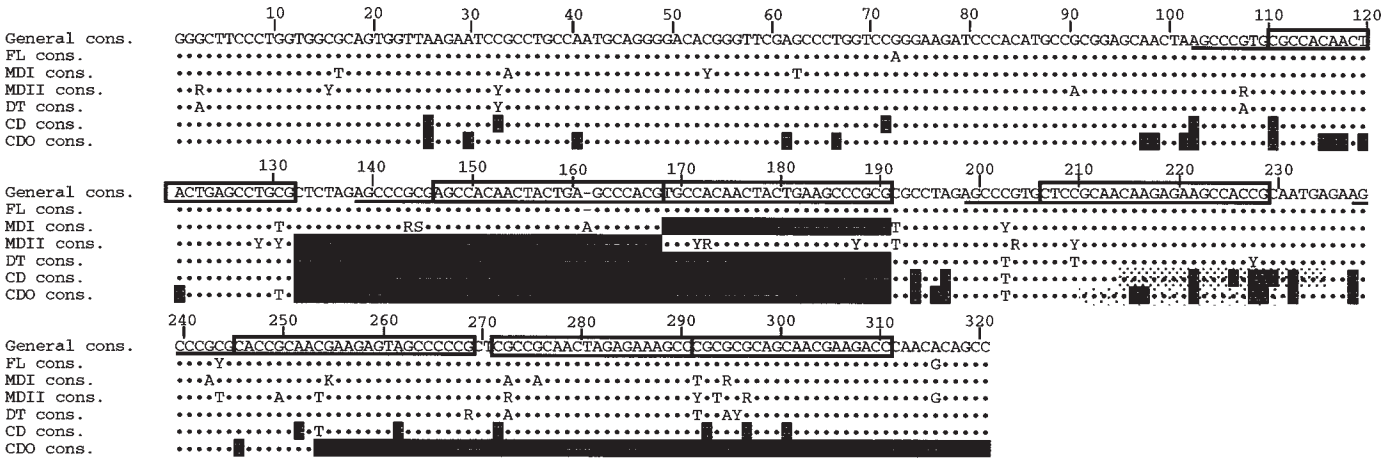


Fig. 8. An alignment of the consensus sequences of six CHR-2 subfamilies. The positions of probes used for the dot-hybridization analysis are shown in shaded boxes. Duplicated regions are boxed in the general consensus sequence, and partially related sequences are underlined.

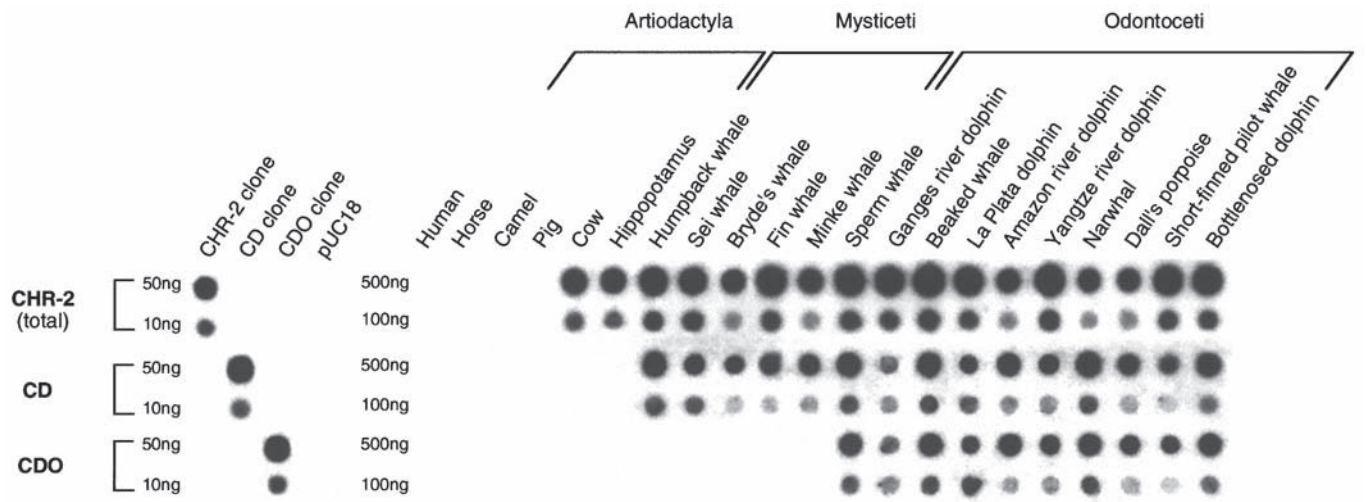


Fig. 9. Distribution of subfamilies of CHR-2 SINEs, as revealed by dot-blot hybridization. (Reprinted from **ref. *50*** with permission.)

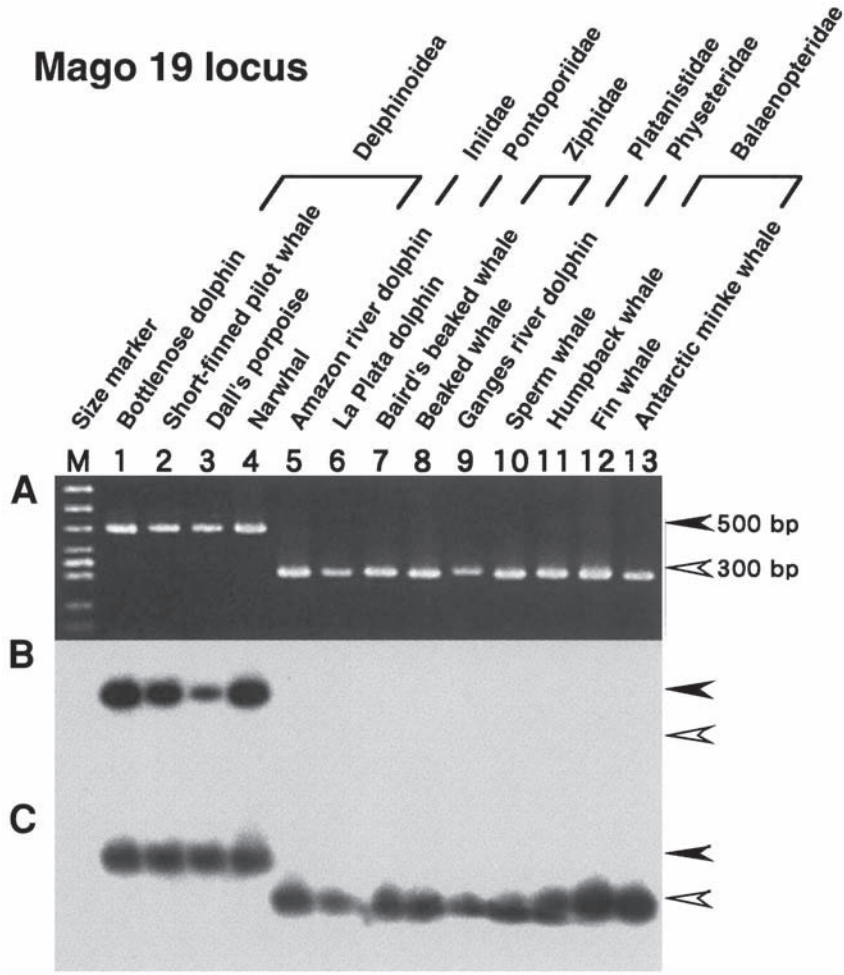


Fig. 10. An example of the SINE method. **(A)** The electrophoresis of PCR product of the locus Mago 19. **(B)** Southern hybridization experiment using SINE as probe. **(C)** Southern hybridization experiment using the SINE flanking DNA of this locus as probe.

3.7. General Considerations and Perspectives

3.7.1. Inferring Phylogeny of a Mammalian Group Using SINEs

Generally, most mammalian genomes are characterized by a large amount of SINEs. These SINEs are apparently specific to orders, suborders, superfamilies, families, genera, or a few species based on the hybridization patterns evident among species examined (*see* **Fig. 9**). Such empirical evidence indicates that a

SINE family was newly generated in many ancestral mammalian lineages, although the mechanism of the generation is not fully understood. The reason why there are so many SINEs, or retroposons in general, in mammalian genomes may be that RTase encoded by mammalian L1 altered its template-recognition specificity in a common ancestor of mammals. This change would have enabled recognition of the poly-A tail required for retroposition, whereas many LINEs strictly recognize the 3' tail that forms a stem-loop structure for retroposition ***(13,15,17)***. Such a scenario could have allowed poly-A containing RNAs to become pseudogenes via L1 RTase in mammalian genomes.

Figure 11 shows a recently proposed mammalian phylogeny ***(52–55)*** with mammalian SINE families characterized to date mapped onto the hypothesis.

Briefly, the oldest SINE family distributed in all mammalian genomes is MIR ***(49,56)***. The *Alu* family is apparently specific to the primate genomes as one of the most famous SINE families. The *Alu*-related SINE families have been characterized in the scandentian (tupai) genome ***(12)***. However, these SINE families have not been characterized in presumably closely related species, e.g., *dermopterans* (flying lemur). The SINE family specific to *dermopterans* have been characterized by Piscurek, O., et al. ***(57)***, and named t-SINE. The rodent B1, B2, and ID families are specific to rodent genomes. The rabbit C family was reported in the rabbit genome, although its distribution in closely related species was not reported. SINE families present in cetartiodactyl genomes, such as CHR-1, CHR-2, CHRS, CHRS-S, PRE-1, and *BovtA*, have been examined in detail ***(19,21,58)***. The Can SINE was first reported from the *Canidae* genome ***(59,60)***, but has been shown to occur later during the evolution of many other carnivore genomes ***(61)***. The horse SINE, designated the ERE family, was reported and its distribution was examined ***(62,63)***. The bat SINE family was isolated by Borodulina and Kramerov ***(64)*** and designated VES, and another bat SINE family was characterized recently (Kawai, K., et al., unpublished). The SINE families from eulipotyphlans (core insectivores) have been isolatd by Borodulina and Kramerov ***(65)***, and were named ER1 and 2 (hedgehogs), SOR (shrew), and TAL (moles). The elephant SINE family was recently isolated and was shown to be distributed among species of Afrotheria ***(44)***.

Although many families of SINEs in mammalian genomes have been isolated to date, they have not yet been fully characterized in terms of their subfamily structures, and it is possible that new mammalian SINE families will be isolated from these investigations.

SINEs mapped onto the proposed phylogeny in **Fig. 11** have been used to independently establish common ancestry for numerous mammalian groups. For instance, the relationships of cetartiodactyls were determined by using

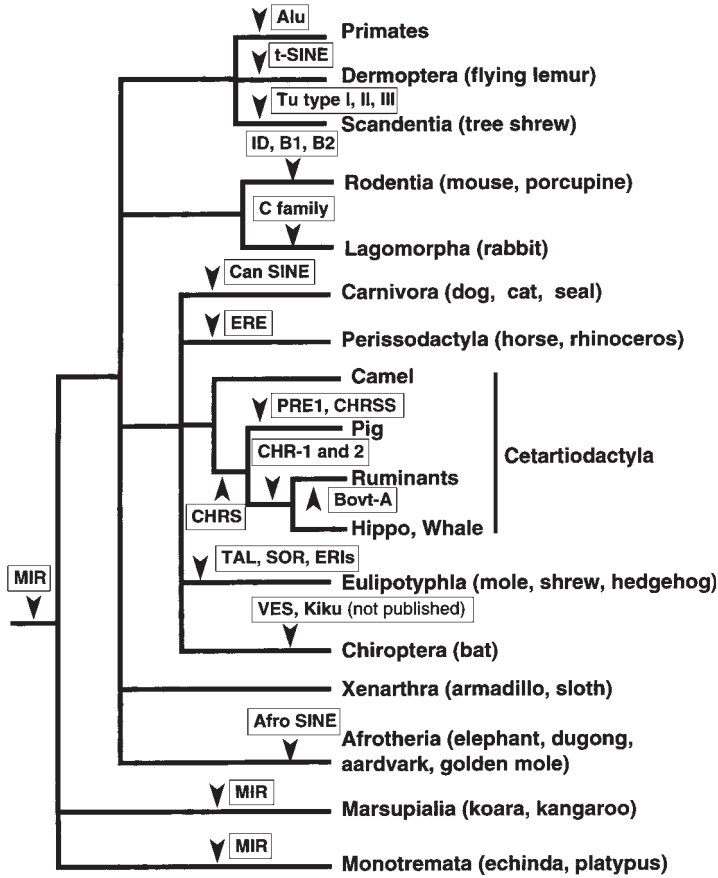


Fig. 11. Distribution of SINEs characterized to date mapped onto the phylogenetic tree of mammals.

CHR-1 SINEs, CHR-2 SINEs, and other retroposons ***(19,21,51)***. By this line of investigation, hippopotamuses were shown to be the most closely related extant relatives of whales based on characterization of several CHR-1 SINE loci ***(21)***. Recently, the major lineages of toothed whales have been resolved by CHR-2 SINEs ***(50)***. The primate *Alu* SINEs were used to elucidate primate phylogeny ***(66)***. Rodent SINEs are now beginning to be applied to infer the complex evolutionary histories of rodent groups ***(67,68)***.

The protocol outlined in this chapter provides a strategy for isolating SINEs pertinent to establishing the phylogeny of species for which no SINE data have been previously reported. The value of SINEs as systematic tools

is founded upon understanding the molecular evolution of the elements themselves, in addition to employing them experimentally to diagnose clades. Even if a given SINE of interest has already been isolated and described in the literature, it should be properly characterized for subfamily structures and taxonomic distribution, depending on the phyla of interest.

3.7.2. Inferring Phylogeny of a Nonmammalian Vertebrate Group Using SINEs

Among nonmammalian vertebrates, the SINEs of salmonid fishes have been examined in greatest detail and used as a model system for investigating SINE evolution. Three different SINE families, namely, *Sma*I, *Fok*I, and *Hpa*I were characterized and shown to be specifically distributed in the genomes of salmon ***(69)***. The *Sma*I family of SINEs is specific to chum and pink salmon in the genus Oncorhynchus. The *Fok*I family of SINEs is specific to genomes of charr, namely, the genus *Salvelinus*. The *Hpa*I family of SINEs is apparently distributed among genomes throughout the family Salmonidae. Accordingly, the phylogenetic relationships of *Oncorhyncus* species were determined by using *Hpa*I SINEs that were differentially inserted in these genomes ***(18,70)***. The body of literature regarding salmonid SINEs follows typical procedures for retroposon analysis, such as isolation of new SINEs using total genomic DNA transcription in vitro ***(43)***, characterization of the taxonomic distribution of families of SINEs ***(69,71)***, subfamily characterization ***(33,72)***, and determination of phylogeny ***(18,70)***.

SINEs for other fish groups have not yet been extensively characterized except for the AFC (African cichlid) SINEs ***(16,20)*** and DANA SINEs in zebrafish ***(73)***. AFC SINEs were used to demonstrate monophyly of major cichlid tribes in Lake Tanganyika ***(20)*** and were also recently applied to demonstrate common ancestry of cichlids from lake Malawi ***(74)***. The presence of SINEs similar to those in zebrafish was reported in the genomes of Medaka and Fugu ***(75)***. SINEs in vertebrates other than mammals and fishes are largely unknown. The tRNA-derived repetitive sequences were characterized in genomes of the newt ***(76)*** and frog ***(77)***, but most of them are shown to be tandemly repeated. Since these repetitive sequences were originally amplified by retroposition ***(76)***, a portion of them should be distributed in a dispersed pattern among these genomes. However, dispersed SINEs have yet to be carefully characterized in the genomes in amphibians and reptiles. Presently, the SINE method has not been applied to infer the phylogeny of amphibians and reptiles, although preliminary evidence from the genome of lizards suggests the presence of a possible new SINE family of potential phylogenetic utility (Austin, C. and Okada, N., unpublished).

3.7.3. Inferring Phylogeny of an Invertebrate Group Using SINEs

Information on invertebrate SINEs is severely limited, and there is no application of SINEs to date for inferring invertebrate phylogeny. Attempts in our laboratory to isolate a new SINE family from several species of mollusk failed to indicate the presence of these elements in all species examined. Octopus SINEs and squid SINEs were isolated and characterized ***(78,79)***. Squid SINEs should be useful for the determination of phylogeny of related squid species. SINEs from some bivalve species could not be isolated, suggesting additional work is required before application of the SINE method could be realized for numerous bivalve taxa. SINE families of invertebrates have been compiled elsewhere ***(3)***.

3.7.4. Inferring Phylogeny of a Plant Group by Using SINEs

At present, the SINE method has yet to provide strong inference for problems in systematic botany. Two trials toward this end, however, were published in the cases of rice ***(80)*** and cruciferous species ***(81)***. In each case, closely related species, including plants cultivated by humans, were analyzed, and the patterns of SINE insertion were not entirely consistent with each other. This result is likely to have occurred in part for two reasons: (1) these species groups were too young for a SINE to become fixed among populations; and (2) the plants under investigation are subject to extensive hybridization between species. This problem could also be exacerbated by the possibility that SINEs in some plants may take longer to become fixed than in animals because such plants are subject to self-fertilization, and thus their SINEs do not become dispersed by genetic drift as is typical for animal SINEs. Therefore, it is not yet clear to what extent the SINE method can be used effectively for plant systematics. In this vein, applying the SINE method to a natural species group that is more distantly related to cultivated plants such as those examined for rice and crucifers may provide valuable insight. Plant SINEs have also been characterized in tobacco ***(82)*** and recently in *Arabidopsis* ***(83)***.

3.7.5. What Kind of Taxonomic Problem Can Be Addressed by the SINE Method?

The SINE method is technically established in the literature, and a detailed protocol has been outlined in this chapter. In order to effectively apply the method, one may ask, "What kind of taxonomic problem can be addressed by the SINE method?" As described above, the method can be effectively applied to resolve mammalian phylogeny if the study is appropriately designed. Its application is greatly facilitated by the abundance of SINEs in mammalian genomes. SINEs also appear well matched for resolving phylogenies of nonmammalian vertebrates, but this statement requires some reservation,

because it seems evident at this early phase of comparative investigation that SINEs are not as abundant in the genomes of selected nonmammalian vertebrates as they are in mammals. The question of applying SINEs to bird studies is still wide open, and it seems promising that some lizard systematics could be addressed by this method (Austin, C. and Okada, N., unpublished). SINEs have been valuable for understanding the complex evolution of certain fishes, such as salmonids and African cichlids; however, their application to resolving tuna speciation has been less straightforward than expected (Akazaki, T. and Okada, N., unpublished). General applicability of the SINE method to invertebrate and plant systematics remains uncertain at present. Phylogenies of some molluscan species groups, such as octopus and squid, in which certain families of SINEs could be amplified sporadically, may be addressed by this method. Application of the SINE method to plant phylogeny has inherent complications not expected for typical animal investigations, such as frequent hybridization among different plant species and self-fertilization.

3.7.6. Problems With Recent Amplification of SINEs: Fixation and Rapid Speciation

If a SINE was amplified very recently in the evolution of a genome, and it is not fixed among populations of a species, then its status as a shared, derived character remains unclear, and it should not be used for cladogram construction. The distribution of such SINEs, however, can be used for analysis of population structure ***(84)***. If speciation has occurred rapidly, namely, before most SINEs have been fixed among populations via genetic drift, ancestral polymorphism followed by incomplete lineage sorting may produce inconsistent patterns of SINE insertion ***(85,86)***. This phenomenon, coupled with the irreversible nature of SINE retroposition, provides a basis for employing SINEs to investigate the historical pattern of lineage sorting in a taxonomic group that has undergone explosive radiation. In cases where ancestral polymorphism and incomplete lineage sorting has produced inconsistent insertion patterns, it is useful to evaluate the SINE character matrix for insertion presence or absence at each locus, using maximum parsimony methods of phylogenetic inference. Although unfixed, polymorphic SINEs are not useful for higher-level phylogenetics, they make excellent systematic tools for population analysis within species, in part because of their ability to distinguish identity by descent vs identity by character state only ***(22,84,87)***. These aspects of SINE analysis are discussed in detail in a forthcoming review (Shedlock, A. M., et al., unpublished).

3.7.7. The Value of Flanking Sequences

Another important expansion of the basic SINE method lies in the nonfunctional nucleotide flanking-sequence information available at the loci examined

for SINE insertion. These sequence data are readily gathered as part of the procedure for isolation and characterization of SINEs as described in detail above (*see* **Subheadings 3.2.** and **3.3.**). Although insertion data alone can only be used to establish tree topologies among taxa, their integration with flanking sequences may provide useful branch-length information between clades defined by independent SINE insertion events ***(3,88)***. Furthermore, because the flanking sequences associated with a given inserted element are by definition linked, the consistency between topologies derived from SINEs vs flanking sequences provides a numerical approach for evaluating the basic assumption of irreversible insertion at each locus ***(88)***. One would expect obvious incongruence between trees if there were any homoplasy, or character conflict, at independent SINE loci. This approach is emerging as a valuable new dimension of SINE analysis and provides a foundation for broadening the statistical evaluation of SINE method in general.

4. Notes

1. Since the presence of SDS inhibits the transcription reaction, SDS must be removed carefully during the DNA preparation.
2. The CPrimer program is free and can be obtained via the Web, although it only works on a Macintosh Computer (the URL is http://iubio.bio.indiana.edu/soft/molbio/mac/).
3. To minimize this sampling problem, one can select species from the more derived, or crown group of taxa, if hypotheses of relationship already exist from other phylogenetic studies, such as those based on morphology or DNA sequence information. Thus, before employing the SINE method, it is clearly valuable to consult comparative studies for the taxonomic groups being considered. It should be noted that, just as with hypothesis formation, a strategy for selecting a species for genomic library construction may be confounded by the artifacts of ancestral polymorphism and differential lineage sorting (*see* **Subheading 5.6.**). Cases of rapid radiation among lineages are classic evolutionary scenarios that warrant extra attention to these sampling issues and to the potential need to create libraries from an expanded set of species under investigation.
4. In our experience, when there are more than ten thousand copies of SINEs present in vertebrate and/or invertebrate genomes, they can be detected by in vitro transcription of total genomic DNA. Curiously, this method has been unsuccessful for investigating plant SINEs, because these SINEs appear not to be transcribed in a HeLa cell extract.
5. Occasionally, accumulation of mutations in the primer binding region for B, C, and D may preclude efficient primer–template annealing during the reaction. In this case, the annealing temperature should be brought down to 45–50°C, accordingly. If no PCR products are amplified by PCR with initial primers, new primers should be designed with additional precaution for potential artifacts that may reduce PCR efficiency.

6. When there are inconsistent patterns of insertion among independent loci examined, it is likely that ancestral polymorphism followed by incomplete lineage sorting has been operating in the system under investigation. The issue of ancestral polymorphism and its relevance to SINE analysis is discussed later in this chapter and is the focus of a detailed forthcoming review (Shedlock, A. M., et al., unpublished).

References

1. Kazazian, H. H., Jr., and Moran, J. V. (1998) The impact of L1 retrotransposons on the human genome. *Nature Genet.* **19,** 19–24.
2. Brosius, J. (1991) Retroposons—seeds of evolution. *Science* **251,** 753.
3. Shedlock, A. M. and Okada, N. (2000) SINE insertions: Powerful tools for molecular systematics. *BioEssays* **22,** 148–160.
4. Rogers, J. (1983) Retroposons defined. *Nature* **301,** 460.
5. Weiner, A., Deininger, P. L., and Efstratiadis, A. (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55,** 631–661.
6. Schmid, C. (1996) Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. *Prog. Nucl. Acid Res. and Mol. Biol.* **53,** 283–319.
7. Clark, J. B. and Kidwell, M. (1997) A phylogenetic perspective on *P* transposable element evolution in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **94,** 11,428–11,433.
8. Hartl, D. L., Lohe, A. R., and Lozovskya, E. R. (1997) Modern thoughts on an ancient *mariner*: function, evolution, regulation. *Annu. Rev. Genet.* **31,** 337–358.
9. Okada, N. (1991) SINEs: short interspersed repeated elements of the eukaryotic genome. *Trends Ecol. Evol.* **6,** 358–361.
10. Okada, N. (1991) SINEs. *Curr. Opin. Genet. Dev.* **1,** 498–504.
11. Eickbush, T. H. Origin and evolutionary relationships of retroelements. In *The Evolutionary Biology of Viruses* (Morse, S. S., ed.), Raven, New York, NY, 1994, pp. 121–157.
12. Nishihara, H., Terai, Y., and Okada, N. (2002) Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. *Mol. Biol. Evol.* **19,** 1964–-1972.
13. Ohshima, K., Hamada, M., Terai, Y., and Okada, N. (1996) The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. *Mol. Cell. Biol.* **16,** 3756–3764.
14. Okada, N. and Ohshima, K. Evolution of tRNA-derived SINEs. In *The Impact of Short Interspersed Elements (SINEs) on the Host Genome* (Maraia, R. J., ed.), RG Landes Co., Austin, TX, 1995, pp. 62–79.
15. Okada, N., Hamada, M., Ogiwara, I., and Ohshima, K. (1997) SINEs and LINEs share common 3' sequences: a review. *Gene* **205,** 229–243.
16. Terai, Y., Takahashi, K., and Okada, N. (1998) SINE Cousins: The 3' end tails of the two oldest and distantly related families of SINEs are descended from the 3' ends of LINEs with the same genealogical origin. *Mol. Biol. Evol.* **15,** 1460–1471.
17. Kajikawa, M. and Okada, N. (2002) LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* **111,** 433–444.

18. Murata, S., Takasaki, N., Saitoh, M., and Okada, N. (1993) Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc. Natl. Acad. Sci. USA* **90,** 6995–6999.
19. Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., et al. (1997) Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* **388,** 666–670.
20. Takahashi, K., Terai, Y., Nishida, M., and Okada, N. (1998) A novel family of short interspersed repetitive elements (SINEs) from cichlids: the pattern of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. *Mol. Biol. Evol.* **15,** 391–407.
21. Nikaido, M., Rooney, A. P., and Okada, N. (1999) Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA* **96,** 10,261–10,266.
22. Deininger, P. L. and Batzer, M. A. (1993) Evolution of retroposons. *Evol. Biol.* **27,** 157–196.
23. Rokas, A. and Holland, P. W. (2000) Rare molecular changes as a tool for phylogenetics. *Trends Ecol. Evol.* **15,** 454–459.
24. Shedlock, A. M., Milinkovitch, M. C., and Okada, N. (2000) SINE evolution, missing data, and the origin of whales. *Syst. Biol.* **49,** 808–817.
25. Hillis, D. M. (1999) SINEs of the perfect character. *Proc. Natl. Acad. Sci. USA* **96,** 9979–9981.
26. Miyamoto, M. M. (2000) Perfect SINEs of evolutionary history? *Current Biology* **9,** 816–819.
27. Schmid, C. W. and Maraia, R. (1992) Transcriptional regulation and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Devel.* **2,** 874–882.
28. Kim, J., Martignetti, J. A., Shen, M. R., Brosius, J., and Deininger, P. (1994) Rodent BC1 RNA gene as a master gene for ID element amplification. *Proc. Natl. Acad. Sci. USA* **91,** 3607–3611.
29. Shen, M. R., Batzer, M. A., and Deininger, P. L. (1991) Evolution of the Master Alu Gene(s). *J. Mol. Evol.* **33,** 311–320.
30. Matera, A. G., Hellman, U., and Schmid, C. W. (1990) A transpositionally and transcriptionally competent Alu subfamily. *Mol. Cell. Biol.* **10,** 5424–5432.
31. Leeflang, E. P., Liu, W. M., Hashimoto, C., Choudary, P. V., and Schmid, C. W. (1992) Phylogenetic evidence for multiple Alu source genes. *J. Mol. Evol.* **35,** 7–16.
32. Kido, Y., Himberg, M., Takasaki, N., and Okada, N. (1994) Amplification of distinct subfamilies of short interspersed elements (SINEs) during evolution of the Salmonidae. *J. Mol. Biol.* **241,** 633–644.
33. Takasaki, N., Murata, S., Saitoh, M., Kobayashi, T., Park, L., and Okada, N. (1994) Species-specific amplification of tRNA-derived short interspersed repetitive elements (SINEs) by retroposition: A process of parasitization of entire

genomes during the evolution of salmonids. *Proc. Natl. Acad. Sci. USA* **91,** 10,153–10,157.
34. Hennig, W. (ed.) *Phylogenetic Systematics*. University of Illinois Press, Urbana-Champagne, IL, 1996.
35. Hendy, M. D. and Penny, D. (1989) A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38,** 297–309.
36. Nei, M. and Kumar, S. (eds.) *Molecular Evolution and Phylogenetics.* Oxford University Press, New York, NY, 2000.
37. Manley, J. L., Fire, A., Cano, A., Sharp, P. A., and Gefter, M. L. (1980) DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. *Proc. Natl. Acad. Sci. USA* **83,** 3156–3160.
38. Talkington, C. A., Nishioka, Y., and Leder, P. (1980) *In vitro* transcription of normal, mutant, and truncated mouse α-globin genes. *Proc. Natl. Acad. Sci. USA* **77,** 7132–7136.
39. Blin, N. and Stafford, D. W. (1976) A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res*. **13,** 537–556.
40. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.
41. Sherry, S. T., Harpending, H. C., Batzer, M. A., and Stoneking, M. (1997) Alu evolution in human populations: Using the coalescent to estimate effective population size. *Genetics* **147,** 1977–1982.
42. Endoh, H. and Okada, N. (1986) Total DNA transcription *in vitro*: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. *Proc. Natl. Acad. Sci. USA* **83,** 251–255.
43. Matsumoto, K., Murakami, K., and Okada, N. (1986) Gene for lysine $tRNA_1$ may be a progenitor of the highly repetitive and transcribable sequences present in the salmon genome. *Proc. Natl. Acad. Sci. USA* **83,** 3156–3160.
44. Nikaido, M., Nishihara, H., Fukumoto, Y., and Okada, N. (2003) Ancient SINEs from African Endemic Mammals. *Mol. Biol. Evol.* **20,** 522–527.
45. Sprinzl, M., Hartmann, T., Meissner, F., Moll, J., and Vorderwülbecke, T. (1987) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **15,** r53–r188.
46. Galli, G., Hofstetter, H., and Birnstiel, M. L. (1981) Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* **294,** 626–631.
47. Smit, A. F. A., Toth, G., Riggs, A. D., and Jurka, J. (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Evol.* **246,** 401–417.
48. Britten, R. J., Baron, W. F., Stout, D. B., and Davidson, E. H. (1988) Sources and evolution of human Alu repeated sequences. *Proc. Natl. Acad. Sci. USA* **85,** 4770–4774.
49. Jurka, J., Zeitkiewicz, E., and Labuda, D. (1995) Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the Mesozonic era. *Nucleic Acids Res*. **23,** 170–175.
50. Nikaido, M., Matsuno, F., Abe, H., Shimamura, M., Matsubayashi, H., and Okada, N. (2001) Evolution of CHR-2 SINEs in cetartiodactyl genomes: possible evidence for the monophyletic origin of toothed whales. *Mamm. Genome* **12,** 909–915.

51. Nikaido, M., Matsuno, F., Hamilton, H., Brownell, R. L., Jr., Cao, Y., Wang, D., et al. (2001) Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci. USA* **98,** 7384–7389.
52. Waddel, P. J., Okada, N., and Hasegawa, M. (1999) Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48,** 1–5.
53. Cao, Y., Fujiwara, M., Nikaido, M., Okada, N., and Hasegawa, M. (2000) Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. *Gene* **259,** 149–158.
54. Nikaido, M., Harada, M., Cao, Y., Hasegawa, M., and Okada, N. (2000) Monophyletic origin of the order Chiroptera and its phylogenetic position among mammalia, as inferred from the complete sequence of the mitochondrial DNA of a Japanese megabat, the Ryukyu flying fox (*Pteropus dasymallus*). *J. Mol. Evol.* **51,** 318–328.
55. Nikaido, M., Kawai, K., Cao, Y., Harada, M., Okada, N., and Hasegawa, M. (2001) Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* **53,** 508–516.
56. Smit, A. F. A. and Riggs, A. D. (1995) MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.* **23,** 98–102.
57. Piscurek, O., Nikaido, M., Boeadi, M., and Okada, N. (2003) Unique mammalian tRNA-derived repetitive elements in Dermopterans: the t-SINE family and its retrotransposon through multiple sources. *Mol. Biol. Evol.* **20,** 1659–1668.
58. Shimamura, M., Abe, H., Nikaido, M., Ohshima, K., and Okada, N. (1999) Genealogy of families of SINEs in cetacean and artiodactyls: The presence of a huge superfamily of tRNAGlu-derived families of SINEs. *Mol. Biol. Evol.* **16,** 1046–1060.
59. Minnick, M. F., Stillwell, L. C., Heineman, J. M., and Stiegler, G. L. (1992) A highly repetitive DNA sequence possibly unique to canids. *Gene* **110,** 235–238.
60. Coltman, D. W. and Wright, J. M. (1994) Can SINEs: a family of tRNA-derived retroposons specific to the superfamily *Canoidea*. *Nucleic Acids Res.* **22,** 2726–2730.
61. van der Vlugt, H. H. J. and Lenstra, J. A. (1995) SINE elements of carnivores. *Mamm. Genome* **6,** 49–51.
62. Sakagami, M., Ohshima, K., Mukoyama, H., Yasue, H., and Okada, N. (1994) A novel tRNA species as an origin of short interspersed repetitive elements (SINEs). Equine SINEs may have originated from tRNA(Ser). *J. Mol. Biol.* **239,** 731–735.
63. Gallagher, P. C., Lear, T. L., Coogle, L. D., and Bailey, E. (1999) Two SINE families associated with equine microsatellite loci. *Mamm. Genome* **10,** 140–144.
64. Borodulina, O. R. and Kramerov, D. A. (1999) Wide distribution of short interspersed elements among eukaryotic genomes. *FEBS Lett.* **457,** 409–413.
65. Borodulina, O. R. and Kramerov, D. A. (2001) Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. *Mamm. Genome* **10,** 779–786.

66. Schmitz, J., Ohme, M., and Zischler, H. (2001) SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other Primates. *Genetics* **157,** 777–784.
67. Kramerov, D., Vassetzky, N., and Serdobova, I. (1999) The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. *Mol. Biol. Evol.* **16,** 715–717.
68. Kass, D. H., Raynor M. E., and Williams, T. M. (2000) Evolutionary history of B1 retroposons in the genus Mus. *J. Mol. Evol.* **51,** 256–264.
69. Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M., et al. (1991) Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposon during evolution. *Proc. Natl. Acad. Sci. USA* **88,** 2326–2330.
70. Murata, S., Takasaki, N., Saitoh, M., Tachida, H., and Okada, N. (1996) Details of retropositional genome dynamics that provide a rationale for a genetic division: the distinct branching of all the Pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). *Genetics* **142,** 915–926.
71. Kido, Y., Saitoh, M., Murata, S., and Okada, N. (1995) Evolution of the active sequences of the HpaI short interspersed elements. *J. Mol. Evol.* **41,** 986–995.
72. Takasaki, N., Park, L., Kaeriyama, M., Gharrett, A. J., and Okada, N. (1996) Characterization of species-specifically amplified SINEs in three salmonid species—chum salmon, pink salmon, and kokanee: the local environment of the genome may be important for the generation of a dominant source gene at a newly retroposed locus. *J. Mol. Evol.* **42,** 103–106.
73. Izsvak, Z., Ivics, Z., Garcia-Estefania, D., Fahrenkrug, S. C., and Hackett, P. B. (1996) DANA elements: a family of composite, tRNA-derived short interspersed DNA elements associated with mutational activities in zebrafish (*Danio rerio*). *Proc. Natl. Acad. Sci. USA* **93,** 1077–1081.
74. Takahashi, K., Nishida, M., Yuma, M., and Okada, N. (2001) Retroposition of the AFC family of SINEs (Short Interspersed Repetitive Elements) before and during the adaptive radiation of cichlid fishes in lake Malawi and related inferences about phylogeny. *J. Mol. Evol.* **53,** 496–507.
75. Ogiwara, I., Miya, M., Ohshima, K., and Okada, N. (2002) V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. *Genome Res.* **12,** 316–324.
76. Nagahashi, S., Endoh, H., Suzuki, Y., and Okada, N. (1991) Characterization of a tandemly repeated DNA sequence family originally derived by retroposition of tRNA(Glu) in the newt. *J. Mol. Biol.* **222,** 391–404.
77. Unsal, K. and Morgan, G. T. (1995) A novel group of families of short interspersed repetitive elements (SINEs) in *Xenopus*: evidence of a specific target site for DNA-mediated transposition of inverted-repeat SINEs. *J. Mol. Biol.* **248,** 812–823.
78. Ohshima, K., Koishi, R., Matsuo, M., and Okada, N. (1993) Several short interspersed repetitive elements (SINEs) in distant species may have originated from a common ancestral retrovirus: characterization of a squid SINE and a possible mechanism for generation of tRNA-derived retroposons. *Proc. Natl. Acad. Sci. USA* **90,** 6260–6264.

79. Ohshima, K. and Okada, N. (1994) Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of three different tRNA-derived retroposons in the octopus. *J. Mol. Biol.* **243,** 25–37.
80. Mochizuki, K., Umeda, M., Ohtsubo, H., and Ohtsubo, E. (1992) Characterization of a plant SINE, p-SINE1, in rice genomes. *Jpn. J. Genet.* **67,** 155–166.
81. Deragon, J. M., Landry, B. S., Pelissier, T., Tutois, S., Tourmente, S., and Picard, G. (1994) An analysis of retroposition in plants based on a family of SINEs from *Brassica napus*. *J. Mol. Evol.* **39,** 378–386.
82. Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N., and Machida, Y. (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proc. Natl. Acad. Sci. USA* **90,** 6562–6566.
83. Surzycki, S. A. and Belknap, W. R. (1999) Characterization of repetitive DNA elements in Arabidopsis. *J. Mol. Biol.* **48,** 684–691.
84. Hamada, M., Takasaki, N., Reist, J. D., DeCicco, A. L., Goto, A., and Okada, N. (1998) Detection of the ongoing sorting of ancestrally polymorphic SINEs toward fixation or loss in populations of two species of charr during speciation. *Genetics* **150,** 301–311.
85. Takahashi, K., Terai, Y., Nishida, M., and Okada, N. (2001) Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. *Mol. Biol. Evol.* **18,** 2057–2066.
86. Terai, Y., Takahashi, K., Nishida, M., Sato, T., and Okada, N. (2003) Using SINEs to probe ancient explosive speciation: "Hidden" radiation of African Cichlids? *Mol. Biol. Evol.* **20,** 924–930.
87. Stoneking, M., Fontius, J. J., Clifford, S. L., Soodyall, H., Arcot, S. S., Saha, N., et al. (1997) Alu insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. *Genome Res.* **7,** 1061–1071.
88. Lum, J. K., Nikaido, M., Shimamura, M., Shimodaira, H., Shedlock, A. M., Okada, N., et al. (2000) Consistency of SINE insertion topology and flanking sequence tree: Quantifying relationships among cetartiodactyls. *Mol. Biol. Evol.* **17,** 1417–1424.

14

Transformation Systems in Insects

David A. O'Brochta and Peter W. Atkinson

Summary

Genetic transformation is an important technology that provides unique opportunities to find, isolate, and analyze genes, as well as to create organisms with unique functional characteristics. Insect biologists have been developing genetic transformation technologies that rely extensively on transposable elements. A number of class II transposable elements isolated originally from insects have been converted into broad host range insect gene vectors. Class II transposable elements are particularly amenable to gene vector development, although they suffer from some limitations such as low rates of recombination. Use of these gene vectors requires the physical introduction of the vectors into developing insect embryos by microinjection. Microinjection methods vary to accommodate the unique physical and developmental characteristics of the target insects. All methods rely on the use of fine glass needles in conjunction with micromanipulators and a microscope. A serious constraint on the use of existing systems can be the inefficiency of successfully delivering the gene vectors to the germ cells of the developing embryo. The general method for vector delivery to insect germ cells is described, as well as variations that are useful under some conditions.

Key Words: *P* elements; *Hermes*; *mariner*; *Minos*; *piggyBac*; transposable elements; *Drosophila melanogaster*; mosquitoes; microinjection; transgenic insects.

1. Introduction

Creating transgenic organisms has become an important step in modern efforts to identify, isolate, and/or analyze genes. Genetic transformation technologies have had profound effects on our understanding of the molecular genetics of those organisms in which they have been developed through providing a means by which the identification and function of cloned genes can be validated. Not only have transformation technologies permitted the biology of organisms to be investigated in new and creative ways, but they have also led to the creation of organisms with novel genotypes that might never be encountered in nature. Organisms possessing these new and unique genotypes offer

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

many exciting opportunities for addressing agricultural and biomedical problems of great practical significance. Insect biologists share this interest in transgenic technologies, and in fact, their interest predates their abilities to isolate insect genes. The history of these efforts has been reviewed by others and will not be discussed here ***(1–4)***.

The first successful insect transformation system was developed in the early 1980s for the creation of transgenic *Drosophila melanogaster*, but only within the last five years has similar technology become available to insect biologists working on "nonmodel" insect systems. The system developed for *D. melanogaster* was based on a class II transposable element, called *P*, whose movement is highly species-specific. Consequently, the *P* element transformation system is of no use in insects outside the family *Drosophilidae*. A second *D. melanogaster* system based on the *hobo* transposable element appears less restricted ***(5,6)***. Nevertheless, the *D. melanogaster* transformation systems have served as useful paradigms for the development of new insect transformation systems, and all currently available "non-drosophilid" transformation systems greatly resemble them in design. Currently there are four distinct insect transformation systems other than the *D. melanogaster* systems, and all are based on exploiting the mobility properties of class II transposable elements. All of these systems should be considered prototypes, and they have been used successfully in only a limited number of species. They all show great promise and are the focus of continuing efforts of research and development. Consequently, the reader should be aware that this is a dynamic area of insect biology, and the methods described are still evolving.

This chapter will focus exclusively on the creation of insects with transgenes integrated into chromosomes such that they are inherited vertically; so called germline transformation. Methods exist for expressing transgenes in insects from nonintegrating virus-based vectors, but they will not be considered here because they do not lead to the production of stable, transgenic insect lines and do not involve the use of transposable elements ***(7)***.

As stated above, current methodologies for creating transgenic insects have evolved from methods developed by *Drosophila* geneticists for the manipulation of early *D. melanogaster* embryos and the creation of transgenic *D. melanogaster*. Familiarity with these methods will be valuable to the novice, and the details of the methods have been well documented ***(8–12)***. Injection methods for mosquitoes have also been described ***(13)***. Although the methods developed for *D. melanogaster* serve as a reasonable starting point for the creation of any transgenic insect they may be impractical for some non-drosophilids. We strongly encourage readers of this chapter to consider their particular transformation problem within the context of the biology of the species being considered and to be creative. Many of the problems encountered in transforming

insects (e.g., DNA delivery) are general problems being faced by researchers in a variety of fields including plant biology and human gene therapy.

Germline transformation requires the delivery of DNA containing the transgene either directly to germ cells or to the cells that will give rise to germ cells. Insect development provides researchers with unique opportunities to access developing germ cells because young embryos of most holo- and hemimetabolous insects are a syncytium. In these insects early embryonic development proceeds through a succession of nuclear divisions that usually occur uniformly and synchronously. Eventually the resulting nuclei migrate to the periplasm (cortex) of the egg, where cell formation takes place resulting in a blastoderm. The exact timing of events of early embryogenesis varies among species, and some understanding of this will be valuable in establishing a transformation protocol. In some species, primordial germ cells become established prior to blastoderm formation as a group of rounded cells separated from the remainder of the embryo and located at the posterior pole of the egg (called pole cells). In many hemimetabolous insects, the germ cells arise from abdominal mesoderm tissue later in development (*see* **refs. *14*** and ***15*** for an introduction to the diversity of insect embryogenesis). Subtle differences in early embryology of Diptera, Lepidoptera, and Coleoptera have not prevented the successful transformation of insects from these orders. Nevertheless, an appreciation of the nature of these differences is likely to result in the development of customized protocols and optimized results. Successful transformation of hemimetabolous insects has not been reported, although the general strategy outlined here should be effective.

Introducing DNA into preblastoderm embryos has been a general strategy for physically delivering DNA to germ cells. This has been accomplished almost exclusively by direct injection using glass microinjection needles. Biolistic- and electroporation-based methods have been reported but not widely employed ***(16–19)***. Because microinjection of DNA is currently the most effective way to deliver gene vectors to insect germ cells, a central problem of all insect transformation efforts is devising methods whereby insect eggs with a variety of morphologies and characteristics can be injected without killing or sterilizing the resulting adult.

DNA introduced into eukaryotic cells will recombine with resident chromosomes via both homologous and illegitimate recombination processes. Our abilities to exploit these basal recombination activities depend greatly on the type of cells being transformed and the methods by which transgenic cells or organisms will be recognized. Introducing plasmid DNA into insect pole cells by injection of preblastoderm embryos rarely leads to the production of transgenic insects in the next generation if only basal recombination activities are present ***(20–22)***. Obtaining routine integration of transgenes into the chro-

mosomes of pole cells requires linking transgenes to highly-recombinogenic transposable elements. If linked appropriately, the transgene will assume the properties of the transposable element and will integrate in chromosomal locations, and at rates that are characteristic of the native transposable element.

All current germline transformation systems for insects rely on the use of class II transposons that transpose via a cut-and-paste mechanism. There are four transposable element systems (excluding the *hobo* and *P* systems from *D. melanogaster*) that can serve as vectors of transgenes in insects; *Hermes* ***(23)***, *mariner* ***(24)***, *Minos* ***(25,26)*** and *piggyBac* ***(27)***. Class II transposable elements are particular useful for creating gene vectors because they are usually simple recombination systems requiring only an element-encoded transposase protein. The terminal inverted repeat (TIR) sequences of class II transposable elements serve as recombination signal sequences for the direct cutting and pasting by transposase. When attached to a transgene, the TIRs convert the transgene into a nonautonomous class II transposable element that can undergo excision and transposition when functional transposase is present. Because the transposases of class II transposable elements can act *in trans* to the TIRs, it is possible to construct a gene vector from the TIRs that is incapable of remobilizing once it becomes integrated and the source of transposase is removed. Transposase is typically provided by the transient expression of the transposase gene from a nonintegrating "helper" plasmid that is co-injected into preblastoderm embryos with the vector.

The criteria for choosing an insect gene vector system are rather ill defined at this point. Based on the known mobility properties of the elements, such as host range and rates of integration, nothing at the present time greatly distinguishes these elements (**Table 1**). All seem to have comparable host ranges within insects, and any differences simply reflect the degree of effort expended in using a given system. Most transformation efforts to date have focused on Diptera, and rates of transformation have been 10% or less. The rate of transformation refers to the number of fertile adults arising from embryos injected with a transgene-containing vector that give rise to at least one transgenic progeny. A reasonable criterion at this point for choosing a gene vector is whether there have been any reported successes with the vector, either in the target species or in closely related species.

A second criterion for choosing an insect gene vector might be whether the host insect contains transposable elements related to any of the current vectors. The transposable elements used as insect gene vectors belong to families of elements that are widely distributed in nature. If the species to be transformed contains functional endogenous transposable elements belonging to the same family as the gene vector, then there may be a risk of vector instability because of a phenomenon of "crossmobilization." Crossmobilization has only been

Table 1
Settings for P97 Micropipet Puller[a]

	Heat[b]	Pull[c]	Veloc.[d]	Time[e]
Drosophila				
1st cycle[f]	567	0	20	20
2nd cycle	567	45	40	200
Mosquitoes				
1st cycle	570	0	1	250
2nd cycle	570	150	13	250

[a]Sutter Instrument Company, Novato, CA
[b]Filament heat setting with a tungsten filament; heat affects the length and size of the pipet. Higher settings produce longer and finer tips.
[c]Low values of pull strength produce larger tips, higher values produce smaller tips.
[d]Velocity reflects the speed at which the two carrier bars are moving during the weak pull.
[e]Time controls the length of time cooling gas bathes the tip. Longer cooling times lead to shorter needle tapers.
[f]Two heating and pulling stages are used. The first stage reduces the diameter of the capillary and determines the length and severity of taper. The second stage completes the process by forming the tip of the needle.

demonstrated experimentally for the *Hermes* gene vector ***(28)***. However, until experimental data are reported that indicate otherwise, all insect gene vectors should be considered vulnerable to this form of instability. The presence of related transposable elements can be detected using PCR strategies employing degenerate primers to highly conserved regions of the transposase open reading frame ***(29,30)***.

Insect transformation is difficult because it requires a great deal of technical skill to introduce vector DNA into the appropriate cell type, and because current gene vectors do not have high rates of recombination. In addition, current protocols have not been optimized and are essentially modified forms of protocols developed for transforming *D. melanogaster*. Despite these limitations, determined researchers have a good chance of successfully creating the transgenic insects they need.

2. Materials

2.1. Reagents

1. Calcium sulfate, anhydrous (desiccant).
2. Compressed air or nitrogen.
3. 30% Dextran sulfate (0.5–2.0% phosphate buffer salts, pH 6.0–8.0; D-6001, Sigma Chemical Co.)
4. 100% Ethanol.

5. 70% Ethanol.
6. Halocarbon oil, Series 700 and/or Series 27 (Halocarbon Products Corporation).
7. 8 *M* Lithium chloride.
8. 0.1 m*M* *p*-nitrophenyl *p*'-gunidinobenzoate (pNpGB).
9. 5 m*M* Potassium chloride, 0.1 m*M* sodium phosphate, pH 6.8 (injection buffer).
10. 2.5% Sodium hypochlorite (this is a 50% solution of household bleach).
11. 0.1% Sodium dodecyl sulfate.
12. 0.02% Triton X-100.

2.2. Supplies

1. Scotch® Double Stick Tape (3M Corporation).
2. 3M™ Double Coated Tape 415 (3M Corporation).
3. Tegaderm™ (3M Corporation).
4. Toupee tape (TopStick™, Vapon Inc.).
5. Microcapillaries (type used varies depending on needs and micropipet puller) (World Precision Instruments and Sutter Instrument Company).
   a. Aluminosilicate microcapillaries (id 0.75 mm, od 1.00 mm).
   b. Borosilicate microcapillaries (id 0.75 mm, od 1.00 mm).
   c. Quartz glass microcapillaries (id 0.75 mm, od 1.00 mm).
6. 100-µL Microcap® (Drummond Scientific Company).
7. Glass microscope slides.
8. 40 × 100 mm no. 1 cover slips.
9. 22 × 22 mm plastic cover slips.

2.3. Equipment

1. Micropipet puller: This piece of equipment is essential if you intend to fabricate your own needles. There are many manufactures and models to choose from. We have used a Flaming/Brown-type pipet puller (Model P97, Sutter Instrument Company) and found it to be excellent for producing aluminosilicate and borosilicate needles. We have also used a Model 730 (David Kopf Instruments) and have found it adequate for many things, but not for producing needles for mosquito injection.
2. Microscope: A dissecting microscope or inverted compound microscope is often used for insect embryo microinjections. Standard compound microscopes have been used, but we do not recommend them for this application. Each investigator will have his or her own preferences. Our preference is a quality dissecting microscope equipped with an equally high-quality mechanical stage (e.g., Olympus SZ series). Dissecting microscopes have certain advantages over inverted and standard compound microscopes. First, dissecting microscopes have more depth of field than do compound microscopes, and because one will be constantly adjusting the position of embryos and the position of the microinjection needle relative to the embryo, enhanced depth perception is very valuable. Second, one often needs to touch or adjust the injection needle during the procedure.

For example, "flicking" the end of the needle with a pair of forceps can sometimes unclog clogged needles. The greater working distances and the comfortable positioning of the hands below one's eyes make these delicate maneuvers much easier. In the end, however, which microscope one uses is determined by what is available and by personal preferences. All can be configured for effective microinjection of insect embryos.

3. Micromanipulator: This instrument should be of high quality and should be mounted either onto the microscope table or onto the microscope such that vibrations and unwanted movements are minimized. We recommend a micromanipulator that has coarse and fine movements in the *x*, *y*, and *z* axes. Fine control in the vertical axis relative to the microscope's mechanical stage is very important, because it will permit easy adjustment of the position of the needle relative to the egg. We find a joystick control to be extremely convenient for this type of fine positioning.
4. Picopump: We have used both the World Precision Instrument PV820 and PV830 picopumps to regulate the pulse of compressed gas used to drive the DNA into the insect embryo. Both provide control over the timing of the pulse and dual pressure control: (1) the pressure required to inject the DNA solution into the embryo, and (2) a second "hold" pressure that prevents the contents of the embryo from entering the needle both before and after the injection pulse.
5. Pipet holder: We use a pipet holder from World Precision Instruments (MPH6S-1.0 mm). Others are available, e.g., Narishige International USA, Inc.
6. Beveler: We use a BV-10 needle beveler available from Sutter Instrument Company. The grinding stones we use are suitable for beveling tips of approx 1 micron in size and can be made of either alumina abrasive or diamond.

3. Methods

3.1 Preparing the Microinjection Needles

The microinjection needle is the most critical tool used in the production of transgenic insects and is a major determinant of the success of the entire process. Time spent making a "good" needle is time saved performing injections that are doomed to failure because of poor needle shape. Almost anyone who has spent any time injecting anything from insect eggs to individual cells will posit that a good needle is critical for success. Furthermore, functional gene vectors are useless if they cannot be delivered to the germ cells without killing or sterilizing the organism through the loss of germ cells resulting from poor injection technique.

Microinjection needles used in introducing vector DNA into developing insect embryos are prepared by melting a small region of a glass capillary and then pulling the two halves of the capillary in opposite directions to draw the glass into a fine point (**Fig. 1**). Fabricating microinjection needles with the appropriate shape and diameter from glass capillaries requires specialized

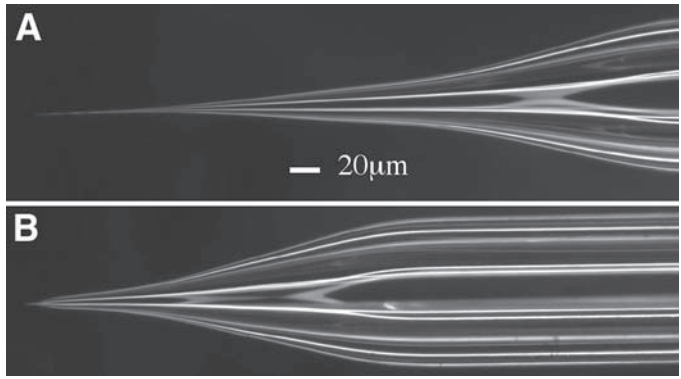


Fig. 1. Aluminosilicate glass microinjection needles fabricated on a Sutter P97 micropipet puller. **(A)** Needle used for *D. melanogaster*, *M. domestica*, *C. capitata*, and *S. calcitrans* injections. The taper is gradual, resulting in a needle with moderate stiffness. **(B)** Needle used for *Aedes aegypti* and *Culex quinquefasciatus* injections. The taper is much more severe resulting in a needle that is very stiff. Both needles have beveled tips with openings of about two microns.

equipment. However, because microinjection techniques are used extensively in a number of fields of biology, there are many options available for all aspects of the needle manufacturing process (*see* **Note 1**).

We use a Flaming/Brown-type micropipet puller (Model P97, Sutter Instrument Company), that allows for control of filament temperature, heating duration, number of heating and pulling cycles, the force of pulling, and the velocity of the jet of air that bathes the pipet at the point of heating. Varying these parameters, as well as varying the ratio of the internal diameter to the external diameter of the pipet, permits the generation of needles of varying lengths and taper. We use aluminosilicate glass microcapilleries that have an internal diameter of 0.75 mm and an external diameter of 1.00 mm with a solid glass filament fused to the inner surface (World Precision Instruments) (*see* **Note 1**). **Table 1** shows typical settings for a Sutter P97 micropipet puller for the production of aluminosilicate needles suitable for injecting *D. melanogaster*, *Aedes aegypti*, and *Culex quinquefasciatus* embryos, and **Fig. 1** illustrates typical needles produced under these conditions.

Needle sharpness is also very important. Needles that are not sharp can tear the chorion and vitelline membrane, causing excessive loss of ooplasm, yolk, and nuclei that leads to death of the embryo. The tip of the microinjection needle immediately following the fabrication process is sealed and must be opened before it is functional. For easily penetrated materials such as the soft vitelline membranes of *D. melanogaster*, *Ceratitis capitata*, *Stomoxys calcitrans*,

and *Musca domestica*, one can crudely open the end of the sealed needle by delicately touching the tip, with help of a micromanipulator and mechanical stage, to the edge of a glass cover slip. With a bit of practice one can create a needle that is sufficiently sharp to penetrate a soft vitelline membrane without excessive tearing (*see* **Note 2**).

3.2. Preparing the Injection Cocktail

DNA used for injections should be highly purified and should be prepared using a cesium chloride gradient or solid-phase anion exchange chromatography. A mixture of plasmid DNAs containing the nonautonomous gene vector with associated transgenes and the transposase gene-containing helper plasmid should be made in a buffered aqueous solution (5 m*M* KCl, 0.1 m*M* sodium phosphate, pH 6.8) (*see* **Note 3**). The concentration of nucleic acids in the injection mixture is usually kept below 1 mg/mL (**Table 2**). However, current insect transformation protocols have not been optimized with respect to this parameter. The use of dilute solutions of vector and helper DNA have the advantage of flowing more easily from the needle, and they are less likely to disrupt the physiology of the developing embryo. In addition, high concentrations of vector- and helper-plasmid DNA may result in reduced recombination activity of the transposable element vector. In the case of *mariner*, it is known that a high concentration of transposase leads to the phenomenon of overexpression inhibition and reduced *mariner* transposition activity ***(31)***. This phenomenon has also been observed for other transposable elements such as the *Ac* element of maize ***(32)***. *Ac* is a member of the *hAT* element, as are the *Hermes* and *hobo* elements from insects. A high concentration of *Ac* transposase in vivo leads to nonspecific association of transposase monomers, forming large inactive complexes of protein in the nucleus ***(33)***. Whether this type of overexpression inhibition also occurs with other *hAT* elements remains to be determined.

1. Coprecipitate plasmid DNA using 1/10 volume of 8 *M* LiCl and 2 volumes of 100% ethanol.
2. Wash the pellet extensively with 70% ethanol to dissolve residual salt.
3. Dissolve the pellet thoroughly in injection buffer (5 m*M* KCl, 0.1 m*M* sodium phosphate, pH 6.8).
4. Filter the DNA-containing injection cocktail to remove insoluble particulate matter that can clog needles using 0.45-μm Ultrafree®-MC Centrifugal Filter Units (Millipore Corporation) or the equivalent (*see* **Note 4**).

3.3. Filling Microinjection Needle

The microinjection needle is partially filled with 0.5–1.0 μL of the filtered DNA injection cocktail by the process of "back-filling."

Table 2
Non-Drosophilid Insect Transformation

		[Vector][a]	[Helper][a]	Frequency	Ref.
Hermes					
Aedes aegypti	a	0.5	0.3	8	***(41)***
	b	0.5	0.3	NR[c]	***(42)***
	c	0.5	0.3	2.3–3.8	***(43)***
	d	0.5	0.3	4.3	***(44)***
	e	0.5	0.3	0.83	***(45)***
Ceratitis capitata		0.125	0.125	0.6	***(46)***
Stomoxys calcitrans		0.150	0.125	4	***(47)***
Culex quinquefasciatus		0.5	0.5	11.8	M. Allen (pers. comm.)
Tribolium castaneum		0.5	0.3	1	***(48)***
piggyBac					
Ceratitis capitata	a	0.5	0.15	5.3	***(49)***
	b	0.5	0.3	3.0	***(49)***
Anastrepha suspensa		0.6	0.4	2.0	***(50)***
Bactrocera dorsalis		0.5	0.3	4.5	***(51)***
Musca domestica		0.4	0.08	17	***(52)***
Anopheles albimanus		0.3	0.15	10	A. M. Handler (pers. comm.)
Tribolium castaneum		0.5	0.3	60	***(48)***
Pectinophora gossypiella		0.5	0.3	3.5	***(53)***
Bombyx mori	a	0.4	0.4	0.7	***(54)***
	b	0.4	0.4	3.9	***(54)***
mariner					
Aedes aegypti	a	0.5	0.5	6.6	***(55)***
	b	0.5	0.5	4	***(44)***
	c	0.5	0.00038[b]	1.0	***(35)***
	d	0.5	0.038[b]	5.9	***(35)***
Minos					
Ceratitis capitata	a	0.4	0.1	1.3	***(26)***
	b	0.6	0.3	1.9	***(56)***
Anopheles stephensi		0.4	0.1	7	***(36)***

[a]Concentrations in mg/mL.
[b]Concentration of purified *mariner* transposase in injection cocktail instead of helper plasmid.
[c]NR, not reported.

1. Fabricate a filling capillary by drawing a 100-µL Microcap® (Drummond Scientific Company) into a fine tube by melting the middle of the capillary using a small Bunsen burner and pulling the two halves apart, following the removal of the capillary from the heat.
2. Fill the capillary with the injection solution by capillary action.
3. Carefully insert the drawn and filled capillary inside the microinjection needle and expel the injection solution. This step will require practice before one will be able to deposit the liquid at the tip of the microinjection needle. A common problem at this step is to have the liquid leave the filling capillary and flow up to the top of the injection needle instead of down toward the tip, resulting in many small droplets of liquid being deposited in the injection needle when the filling capillary is removed. If droplets of solution become widely distributed within the microinjection capillary tube, one can shake the needle vigorously to try to drive the liquid toward the tip. In addition, lying the filled needle horizontally for a few minutes usually will result in the removal of most air bubbles from the tip of the needle. Stubborn bubbles can usually be removed by holding and vibrating the needle by flicking with your finger or by touching the needle gently to a laboratory tube vortexer. Alternatively, any remaining air bubbles can be removed by expelling them from the tip of the needle after attaching it to the needle holder and compressed gas line (*see* **Subheading 3.6.** and **Note 5**). We usually fill three needles, seal the ends of the two spare needles with Parafilm, and keep them in a closed container. Spare needles filled with DNA allow for rapid needle replacement should a needle become clogged or broken during the injection procedure.
4. Attach the needle to the needle holder and when it is not being used it should be positioned so that the tip is in a droplet of oil on a glass slide. We use Series 700 Halocarbon oil (Halocarbon Products Corporation).

3.4. Preparing Embryos for Injection

The ability to collect eggs of the appropriate age is critical for the successful introduction of vector DNA into developing germ cells and the creation of transgenic insects. The methods used to collect appropriately aged eggs vary greatly depending on the species. In all cases however, one must be able to collect eggs and prepare them for injection before they reach the cellular blastoderm stage. In *D. melanogaster*, this stage occurs after the 13th mitotic cycle in the embryo. Exactly how long it will take to reach this stage after egg deposition must be determined empirically for each insect species (*see* **Note 6**). One must consider the possibility that a female insect may retain her eggs for some time following fertilization, and such behavior can limit the time available for injection. Here we will simply assume that one can collect eggs in the syncitial stage of embryonic development, and that one can also distinguish the anterior from the posterior end of the egg.

For cases in which chorions can be removed without killing the embryo, a standard method is to treat the embryos with a 2.5% solution of sodium hypochlorite (*see* **Note 7**).

1. Prepare a 50% solution of household bleach in water.
2. Eggs of the appropriate age are placed in a small watchglass containing the sodium hypochlorite solution for approx 3 min with regular agitation (*see* **Note 8**). Monitor the dechorionation process under a dissecting microscope.
3. Dechorionated embryos are thoroughly washed with either water, 0.1% SDS, or 0.02% Triton X-100. Failure to adequately wash the dechorionated embryos will result in the embryos being covered in a residue from the dissolved chorions, preventing subsequent desiccation and eventually killing the embryo. We find it convenient to collect and wash the dechorionated eggs in a small vacuum-filtering apparatus such as a Buchner funnel lined with filter paper.
4. Fix the dechorionated eggs to a glass slide or a 40 mm × 100 mm no. 1 cover slip by laying them on a thin strip of Scotch Double Stick Tape (3M Corporation) (*see* **Note 9**). The surface of the embryos should be dry at this stage to facilitate sticking. Eggs are carefully transferred to the tape with the aid of a dissecting microscope such that they are aligned "shoulder to shoulder" with all posterior ends facing in one direction (*see* **Note 10**). The transfer procedure can be done using fine forceps. Dechorionated eggs are extremely delicate and are naturally a bit sticky and will readily stick to the tips of the forceps, making transfer simple and quick (*see* **Note 11**). The tips of the forceps can be made sticky by scraping a bit of glue from the double-sided sticky tape and holding it with the forceps (*see* **Note 12**).
5. Desiccate the eggs slightly to allow a small volume of the DNA injection mixture to be introduced. Embryos are desiccated either by placing the cover slip with the aligned embryos in a jar with desiccant (e.g., anhydrous calcium sulfate), or by simply exposing the embryos to air. The amount of desiccation that is necessary or possible will vary with the species being injected and with the ambient humidity of the laboratory (which can fluctuate seasonally). Dechorionated eggs desiccate easily with times ranging from 3 to 15 min (*see* **Note 13**). Some species are very sensitive to desiccation at this stage of development, and care should be taken. Young *Aedes aegypti* embryos (as indicated by the light gray color of the chorion) readily become desiccated if exposed to dry air. Their chorions "dimple" slightly when they have desiccated just enough for injection. One will have to experiment with a particular species to determine if and under what conditions desiccation can be achieved.
6. Desiccated embryos are covered with viscous (700 Centistokes) halocarbon oil (Series 700, Halocarbon Products Corporation) and mounted on the stage of a microscope for injection (*see* **Note 14**). Although viscous halocarbon oils usually will not flow off of the slide, one can aid retention of the oil by encircling the embryos with a grease-pencil line prior to covering them with oil. The grease line provides enough of a barrier to prevent loss of oil. Heated high viscosity oils (for example during a heat shock step) and low viscosity oils are more difficult to contain (*see* **Note 15**).

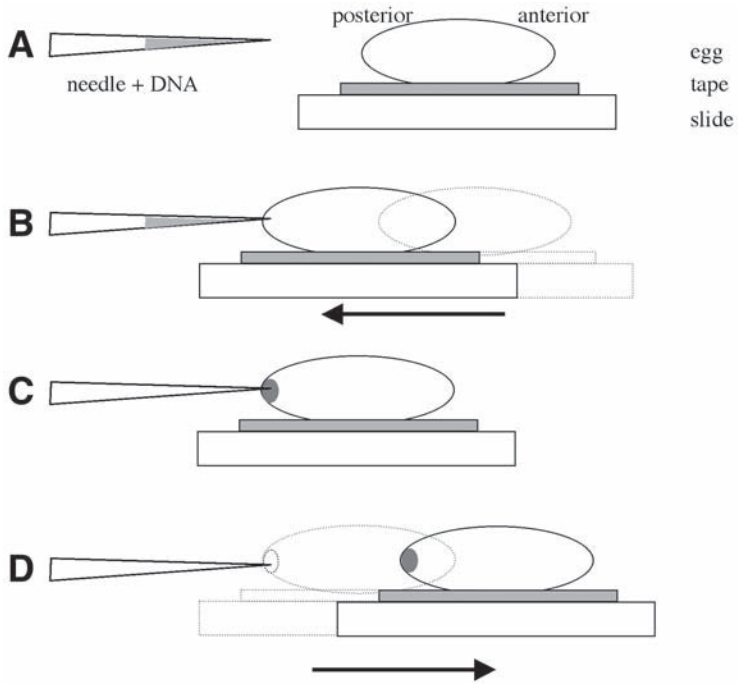


Fig. 2. Insect embryo microinjection by impaling the embryo on a fixed needle. **(A)** Needle is positioned horizontally and is fixed in the field of view of the microscope. **(B)** The egg is impaled on the needle by moving the mechanical stage of the microscope. **(C)** DNA is injected into the posterior pole region. **(D)** The embryo is removed from the needle and the next embryo is moved into position.

3.5. Inserting the Needle Into an Embryo

1. Attach the filled needle to the needle holder (*see* **Note 16**).
2. Position the needle at the angle of injection desired, relative to the immobilized embryos. The most common procedure for injecting insect embryos is to position the needle almost horizontal to the stage of the microscope and to place the tip of the needle within the field of view (**Fig. 2**). We have constructed a slightly raised platform on the mechanical stage of our microscope in order to position the needle in this way (*see* **Note 17**). To mount the prepared embryos to the stage, the needle is raised vertically to provide sufficient clearance during the mounting process. The first of the aligned embryos is positioned within the field of view and the needle is lowered until it comes into focus. The embryo is then positioned such that its posterior end is almost touching the tip of the needle. Fine adjustments are made in the vertical position of the needle such that it will penetrate the chorion and/or vitilline membrane from an angle as close to horizontal as possible (**Fig. 2**).

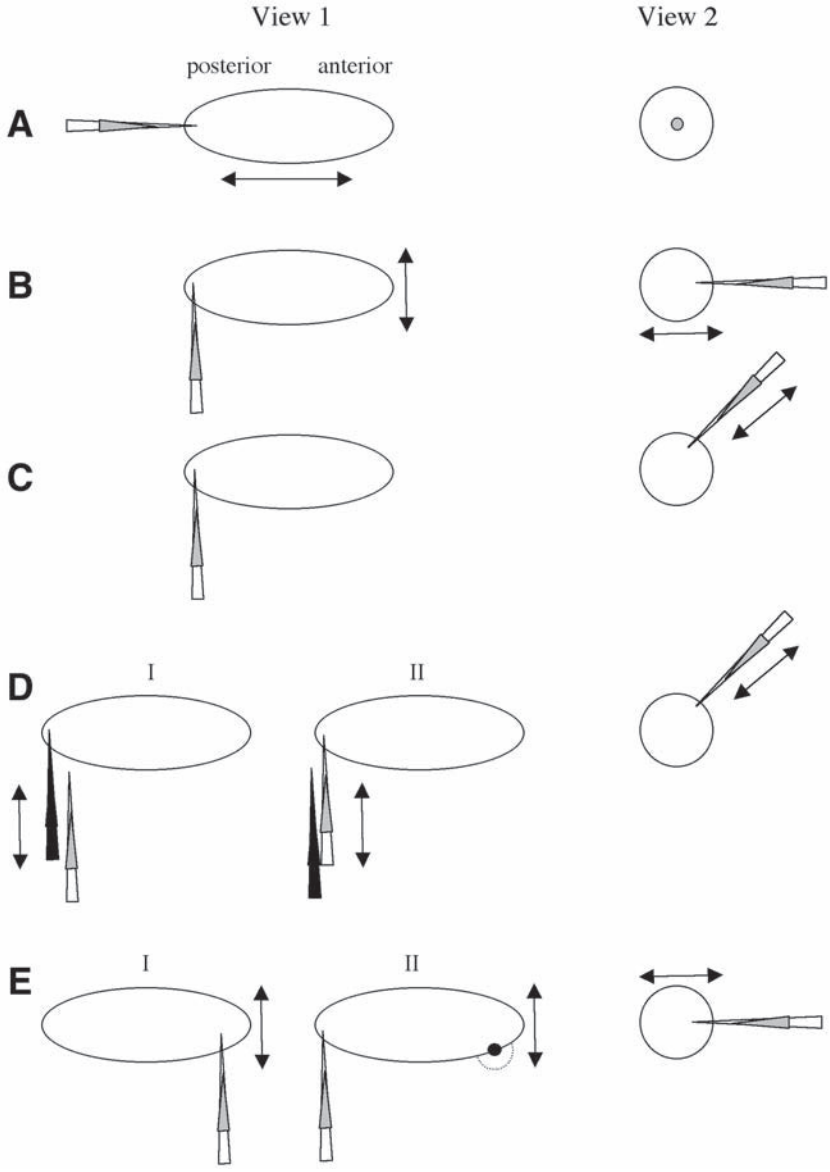


Fig. 3. A variety of strategies that have been employed to inject insect embryos. View 1 is looking straight down on the embryos as they are on the microscope stage. View 2 is an end-on view of the posterior pole. **(A)** The common method of impaling the embryos on a horizontally fixed needle by moving the embryos in the direction indicated. **(B)** A variation of the impaling method that permits *(continued on next page)*

3. To inject the embryo, impale it on the fixed needle by slowly moving the embryo toward the tip of the needle (*see* **Fig. 2** and **Note 18**). Impaling the embryo with the microinjection needle at the posterior pole of the embryo can lead to sterility of the resulting animal because the pole cells fail to develop or because the pole plasm leaks from the embryo after the needle is withdrawn. Great care should be taken, therefore, to minimize disrupting embryonic development (*see* **Fig. 3** and **Notes 19** and **20**).

3.6. Injecting DNA Into an Embryo

We use a microinjection needle holder that allows it to be attached to a pressurized gas (air or nitrogen) line (*see* **Note 16**). Applying air pressure to the needle expels the DNA solution within the microinjection needle. Air pressure-driven microinjection systems provide the operator with a large amount of control of fluid movement. By regulating the pressure and the duration of a pulse of pressure one can accurately and repeatedly deliver a volume of liquid. Simple inexpensive systems can be assembled using low voltage solenoids, a hand held momentary pushbutton switch and a 9-volt battery (*see* **Fig. 4** and **Note 21**).

3.7. Postinjection Procedure

Postinjection handling of the embryos is a critical step, but it varies widely depending on the species being injected. The simplest procedure is to leave the injected embryos attached to the cover slip covered with halocarbon oil and to place them in a humid chamber at the appropriate temperature. Hatched larvae are collected from the oil and transferred to the appropriate larval-rearing media (*see* **Note 21**).

Fig. 3. *(continued)* placement of the needle in the posterior pole region by puncturing the embryo laterally. **(C)** An injection method involving the movement of the injection needle. The needle can be positioned at almost any angle and position relative to the embryo. This method requires a micromanipulator with fine control and minimal vibration. **(D)** This method relies on the use of two needles. First, a metal needle is used to puncture the chorion. The needles are repositioned and the glass microinjection needle is inserted into the embryo through the hole created using the metal needle. This method is useful for insects with extremely hard chorions. **(E)** This method has been used occasionally by the authors in situations where leakage of the ooplasm is unavoidable. The embryo is punctured with the injection needle in an anterior lateral position resulting in the possible loss of some ooplasm. The embryo is repositioned and injected in the posterior pole region. Ooplasm and the injected DNA will not leak from the hole created by the posterior injection. Many insects are tolerant of the loss of small amounts of ooplasm.

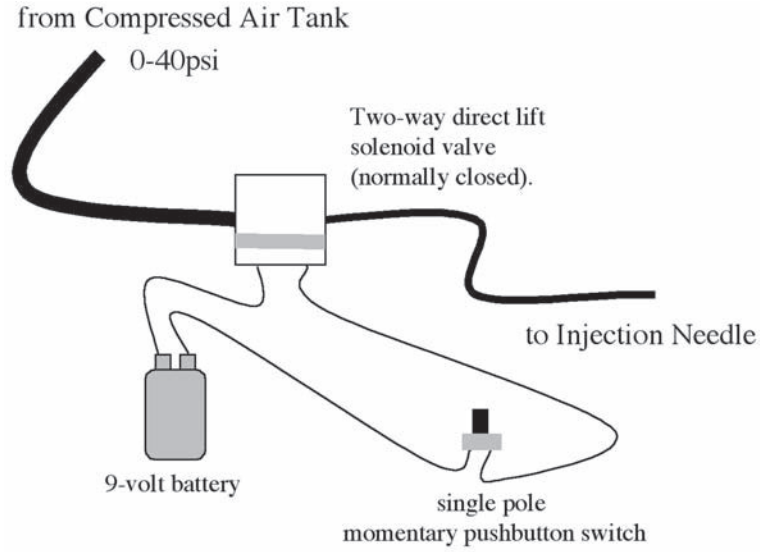


Fig. 4. A simple inexpensive air-pressure regulated injection system. A two-way solenoid that functions on low voltage is placed in line with a regulated compressed air source. The solenoid is controlled with a momentary pushbutton switch (push on, release off) and a small battery. Solenoids for this application cost less than US$100 (e.g., Cole-Parmer model U98300-60).

3.8. Screening for Transgenics

Successful injections result in adults developing from injected embryos (so called G_0 adults) that contain mosaic germlines, in which some of the germ cells have chromosomes with at least one integrated gene vector. To find and recover these chromosomes, G_0 adults are mated to an appropriate test strain and the progeny examined for the presence of the transgene and/or a genetic marker contained on the vector. Progeny arising from transgenic gametes will be transgenic and, by using dominant visible markers such as the green fluorescent protein (GFP) and its many derivatives, these individuals often can be easily recognized. Confirmation of putative transgenic strains typically involves transmission studies demonstrating Mendelian inheritance, Southern blot analysis to estimate copy number and integrity of the vector, and DNA sequence analysis of the junction points between the gene vector and the chromosome, to demonstrate that the vector has integrated into the host's genome. A description of the various filter sets required for the detection of the most commonly used fluorescent proteins for transgenic insect detection can be found in Horn and Wimmer ***(34)***.

4. Notes

1. Pre-made needles can be purchased (μTip™, World Precision Instruments, Inc.), and they may be adequate for certain insect embryos, such as those of *D. melanogaster*, that present few major technical challenges to the microinjecter. For most other insects, commercially manufactured needles are inadequate and custom manufactured needles are essential.

   Insect microinjection needles are used to penetrate resilient chorions and vitelline membranes, and consequently must be strong, stiff, and sharp. The degree to which these characteristics can be controlled during the manufacturing process largely determines the quality of the needles that are produced. The choice of glass capillaries plays a large role in determining the strength of the resulting microinjection needle. Borosilicate glass capillaries, while soft and easy to shape into microinjection needles, are not very strong, break easily, and quickly loose their sharpness. Aluminosilicate glass capillaries are a superior alternative in terms of strength and durability, but require higher filament temperatures in order to melt the glass sufficiently for shaping into a needle. Virtually all conventional micropipet manufacturing devices (micropipet pullers) permit the use of both borosilicate and aluminosilicate glass capillaries. However, for the ultimate in strength and hardness, one should manufacture microinjection needles from quartz glass capillaries. These needles are extremely strong and hard, but they cannot, unfortunately, be manufactured on conventional micropipet pullers that rely on a platinum or tungsten filament to melt the glass. These filaments cannot develop sufficient heat to melt quartz glass. To use quartz glass capillaries, one will need a highly-specialized micropipet puller that can generate the heat required to melt this material. A micropipet puller equipped with a microprocessor-controlled CO_2 laser capable of melting quartz glass and fabricating quartz glass microinjection needles is available (Model P-2000, Sutter Instrument Company).
2. An alternative method for creating sharp needles with excellent penetrating ability is to bevel the tip of the needle with a needle beveler. Needle bevelers consist of a small turntable to which a fine abrasive disc is attached. The needle is attached to a positioning device at the desired angle of the bevel and lowered onto the beveling surface; The process is monitored under a microscope. Beveled needles significantly reduce the amount of mechanical damage caused by needle penetration and increase survival of the injected embryos. They are mandatory for the successful injection of mosquito embryos.

   We use a micropipet beveler equipped with either a fine or extra-fine diamond or alumina abrasive plate (Model BV-10, Sutter Instrument Company). We typically position our needles (aluminosilicate glass) at an angle of approximately 30° to the dry abrasive surface. The needle is viewed through a microscope and is gradually lowered to touch the surface of the abrasive plate. Typically the micropipet is left in contact with the surface for 1–5 min.
3. All current insect gene vectors are designed to function as binary systems consisting of the vector and a plasmid-encoding transposase (helper plasmid).

The vector typically consists of the transgene of interest and a gene that serves as a dominant visible (GFP) that permits transgenic organisms to be recognized. These two components are flanked by the terminal sequences including the TIRs of the transposable element. Together these sequences act as a nonautonomous transposable element, i.e., an element fully capable of undergoing excision and integration, but incapable of producing its own transposase. Transposase is usually provided through the transient expression of the transposase gene contained on a helper plasmid. Direct injection of purified *mariner* transposase protein has been reported instead of helper plasmids ***(35)***. In addition, injecting transposase mRNA, transcribed and capped in vitro, is also an option. Binary systems are used because they enhance the stability of the integrated transgene. Once the nonautonomous gene vector integrates, it cannot undergo further excision or transposition because the plasmid (mRNA or protein) responsible for supplying functional transposase gradually disappears from the developing embryos and larvae, leaving the integrated vector in a transposase-free environment. It is not clear at this time what way is the best for supplying transposase to the injected embryo. Direct injection of transposase protein or mRNA provides a more immediate source of transposase compared to a helper plasmid, but whether this will consistently lead to higher rates of transformation is not known.

4. If an unfiltered DNA injection mixture is to be used it should be centrifuged prior to use to concentrate any particulate matter.
5. Making microinjection needles from capillaries containing a solid filament fused to the internal surface of the needle usually results in needles that are much easier to fill and less prone to having air bubbles.
6. For many insects the fertilized egg at oviposition contains the female and male pronuclei located around or anterior to the equatorial plane of the embryo. Shortly after oviposition the female pronucleus completes meiosis, followed immediately by karyogamy and the onset of nuclear divisions. Differences between species exist in the rates of nuclear division, the presence or absence of pole cells, the timing of pole cell formation, the number of pole cells formed, and the fate of the pole cells. For example, *Nematocera* (Diptera) typically have few pole cells, while *Cyclorrhapha* (Diptera) have as many as 80. The germline of many hemimetabolous insects do not arise from pole cells formed early in embryogenesis, but instead develop from abdominal mesoderm tissue later in embryonic development.

   The total number of mitotic divisions that occur prior to cellularization of the embryo and the time required to reach this stage also vary from species to species. For most insects there are from 9 to 13 nuclear divisions prior to cellularization resulting in a blastoderm embryo with 500–8000 cells.

   Some understanding of the early embryology of the insect being transformed will result in more effective protocols and will be likely to increase the chances for a successful transformation.
7. Chorions can sometimes be removed by placing eggs with dry chorions on a glass microscope slide to which a small piece of double-sided sticky tape has

been attached. By gently pressing and rolling the eggs on this sticky surface, one can sometimes crack the chorion and roll the vitelline membrane-bound embryo from the egg.

8. The chorions of some insects, such as *M. domestica*, are not thoroughly dissolved by sodium hypochlorite. Instead only a specialized region of the chorion running laterally along one of the surfaces is dissolved resulting in a long lateral split in the chorion from which the embryos can be removed by simply agitating the suspension of embryos.

   Mosquitoes such as *Aedes aegypti*, *Culex quinquefasciatus* and *Anopheles gambiae* present a number of challenges because they cannot be dechorionated and must be injected before the melanization of the chorion makes them too resistant to penetration. The melanization process of Anopheline eggs can be delayed by allowing mosquitos to deposit their eggs on filter paper soaked in 0.1 m*M* *p*-nitrophenyl *p*'-gunidinobenzoate (pNpGB). This treatment effectively increases the time during which the eggs can be injected, allowing eggs 90–120 min old to be injected. This treatment may even improve survival of injected eggs by reducing mechanical damage during the injection process ***(36)***.
9. The tape used in this procedure is a potential source of toxins that can kill developing insect embryos. Scotch Double Stick Tape (3M) works reliably for many users; however, it has been known to be the cause of embryo mortality during microinjection. 3M Double Coated Tape 415 is an excellent alternative that appears to be consistently nontoxic. The above tapes function well in dry or non-aqueous environments, e.g., when covered with halocarbon oil (*see* **Note 10**).

   Other tapes can also be used and are more effective in aqueous environments, e.g., when covered with an aqueous solution of dextran sulfate (*see* **Note 14**). Toupee tape is double sided and functions well in moist environments (TopStick™, Vapon Inc.). Embryos adhere to this tape even when covered with water. Tegaderm™ (3M) is a transparent medical dressing that also remains adhesive in moist environments. Tegaderm is not double sided but can be fixed to a slide with the glue side up by placing a piece of toupee tape on the slide or cover slip first. We have used Tegaderm to immobilize mosquito embryos prior to covering them with an aqueous solution of dextran sulfate (*see* **Note 14**).
10. We have found that when embryos are covered with halocarbon oil they can be sensitive to crowding during alignment. For example, we have observed that dechorionated *D. melanogaster* embryos can be spaced so that they are touching, and survival is excellent. For other species, such as tephritids and muscids, we have observed that if the embryos are spaced very close together with more than 25 embryos per 40 × 100-mm coverslip we begin to see decreased survival. For *Culex quinquefasciatus* it is important that the width of the tape be approximately equal to the width of a single egg. In some cases we have found that placing embryos in a humid chamber with a high O_2 tension increases survival (*see also* **Note 21**).
11. Some very young embryos, such as those of *M. domestica*, *S. calcitrans*, and *C. capitata*, are extremely delicate following dechorionation. In fact, the vitelline

membrane is so delicate and the embryo so flaccid that merely touching the embryos can destroy them. Transferring these very young embryos is almost impossible, and we find their survival rate to be very low. Allowing newly laid eggs to "age" for 30 min prior to dechorionation results in plump, turgid embryos covered by a relatively tough vitelline membrane. These embryos can be readily transferred and aligned. *Culex quinquefasciatus* egg rafts are collected immediately following oviposition and allowed to mature for 30 to 45 min or until their color changes from creamy yellow to gray. The rafts are then gently broken up, and eggs collected from the central portion of the raft ***(37)***.

12. Other methods can be employed to align and transfer embryos on a glass slide or coverslip prior to injection. We align *Aedes aegypti* embryos on a piece of moist filter paper prior to transferring them to a cover slip. The filter paper is kept wet, permitting the embryos to be pushed easily into place while preventing desiccation. When all of the desired embryos are aligned, the excess moisture is removed from the filter paper by blotting; excess moisture is removed from the surface of the *Aedes aegypti* embryos by lightly toughing the surface of the chorions with the edge of a dry piece of filter paper. The aligned embryos are then transferred *en masse*, to a coverslip containing a strip of double sided sticky tape by inverting the flexible plastic coverslip and lowering it slowly onto the embryos until the embryos contact the surface of the tape. The coverslip is pressed gently to insure adequate adhesion of the embryos and the coverslip is then inverted ***(13)***. An alternative to moist filter paper is to use a block of 2% agar. A 1/4-inch-thick slab of 2% agar is cast in the casting tray of an electrophoresis minigel apparatus. Blue food coloring can be added to the agar to enhance its contrast with the embryos. Once the slab is solidified a piece approx 1 in^2 is removed and placed on the stage of a dissecting microscope. Embryos are transferred to the surface of the agar with a moist, fine-bristled brush. The embryos are kept moist and can be aligned by sliding them across the surface of the agar. Once aligned, the excess moisture is removed by blotting with filter paper, and the embryos transferred to a coverslip as described above.
13. Desiccation times are influenced by the method of dechorionation. Where dechorionation is done manually, desiccation times are minimized and are usually less than 5 min. When dechorionation is done chemically, desiccation times are longer and are usually more than 10 min.
14. Halocarbon oils have been commonly used because they are as clear as water, are nontoxic to a number of species, and permit adequate gas exchange with rapidly respiring embryos. Series 700 oil is very viscous (700 centistokes) and is popular, but may need to be avoided when working with some insects. *M. domestica* embryos that develop and hatch in Series 700 oil have poor survival rates. If allowed to develop under a mixture of 65% Series 700 and 35% Series 27 oil (with low viscosity) survival improves. Mosquito embryos fail to develop if covered with halocarbon oil for more than 1 or 2 h.

    While the continued use halocarbon oils has been largely the result of tradition, their use with some species is contraindicated (e.g., mosquitoes). We have

explored the use of other materials and have found that a buffered neutral solution of 30% dextran sulfate is sufficiently viscous and clear to serve as a substitute for halocarbon oil in some cases (e.g., *Aedes aegypti*).

We have used 30% dextran sulfate to temporarily cover *Aedes aegypti* and *Anopheles gambiae* embryos during the injection process. While comparative toxicity studies have not been done, mosquito embryos appear to tolerate this treatment. Dextran sulfate has an advantage over halocarbon oils in that it can be completely and easily removed from the embryos post injection (*see* **Note 21**).

Unfortunately, Scotch Double Stick Tape and 3M Double Coated Tape 415 do not function well in aqueous environments and consequently other tapes must be used (*see* **Note 6**).

Culex quinquefasciatus eggs are not covered with halocarbon oil or dextran sulfate. Instead, the eggs are injected dry with only a piece of moistened filter paper placed immediately adjacent to the anterior end of the aligned eggs in order to prevent excessive desiccation ***(37)***.

Lepidoptera eggs are not covered with oil at anytime during the injection process.

15. If halocarbon oil is to be left on the embryos post injection, it is important to prevent the oil from running off them and resulting in desiccation. Either the slides with the embryos must be kept perfectly horizontal or a ring of silicon vacuum grease can be formed around the pool of oil. Fill a 10-mL syringe with vacuum grease and use this to dispense a narrow bead of grease. Consider the grease a potential source of toxins and discontinue use if necessary.
16. We routinely use a needle holder from World Precision Instruments (Model MPH6S) but have also used a Narashige (Model HI-4A) needle holder with good success.
17. The angle of attack of the needle can vary depending on the species being injected. Some investigators position the needle at approximately a 45° angle in the vertical dimension to the horizontally aligned embryos. For example, this is done for injecting *Culex quinquefasciatus* embryos. This alignment requires that injections be done by moving the needle into the egg, as opposed to impaling the embryo on a fixed horizontal needle. Either method works; however, an important consideration is the degree of control the operator has over the movement of the injection needle. One will want to employ a method that provides maximal penetration control and minimal vibration (**Fig. 3**).

    Peloquin et al. ***(38)*** reported the use of a programmable electromechanical positioning device (Model 5171, Eppendorf) to move the needle into immobilized pink bollworm (*Pectinophora gossypiella*) eggs. Because the system used by these investigators was programmable and mechanized, they could minimize vibrations and obtain repeatable movements of the needle in any dimension. We have injected *Culex quinquefasciatus* eggs using a similar method, only using a manually operated micromanipulator ***(37)***.
18. It is possible to deposit a small amount of DNA solution precisely at the posterior pole by inserting the needle into the posterior lateral position of the embryo (**Fig. 3**).

The tip of the needle can then be precisely placed at the posterior pole, and disruption of the posterior region of the embryo is minimized. We use this method routinely to inject *Culex quinquefasciatus* eggs.

Leakage of polar plasm following the injection process results in loss of injected DNA and development of a sterile adult because of the failure of pole cells to develop. We have noticed that some insect embryos (e.g., *C. capitata*) are remarkably tolerant of the loss of ooplasm. When leakage of ooplasm has been a persistent problem even following desiccation, we have sometimes resorted to injecting the embryo twice. The first injection consists of simply puncturing the embryo in an anterior dorsal lateral position, resulting in the loss of ooplasm and relieving internal pressure within the embryo (**Fig. 3**). Often loss of this ooplasm will be inconsequential, and the embryo will develop completely. The second injection involves penetration of the needle at the posterior pole and deposition of the vector DNA. Because the internal pressure within the embryo was relieved by the initial puncturing of the embryo, the second injection does not result in any loss of DNA and ooplasm (**Fig. 3**).

Lepidoptera eggs often have elaborate chorions that are thick and quite difficult to penetrate. Some workers have resorted to using a surrogate metal needle to puncture a hole in the chorion and then inserting the glass microinjection needle through it and into the embryo (**Fig. 3**). Following injection, the hole in the chorion is then sealed with paraffin or a small drop of glue ***(39)***.

19. A limitation of the methods most commonly employed for penetrating eggs is that movement of the needle relative to the embryo is slow. The slow movement of the needle results in a gradual pushing action of the needle. If chorions are resilient, the amount of force required to penetrate the egg may exceed the amount of force needed to push the egg off the sticky mounting surface. In addition, this slow pushing motion can deform the embryo and can result in tearing the chorion instead of puncturing it. The low velocity movement of the needle has made the use of very stiff and sharp needles essential because penetration may otherwise be impossible. This problem has been commonly encountered by microinjecters. A solution to this problem is to move the needle at high velocities to achieve a stabbing effect. Many reports exist of microinjection needles being coupled to high velocity actuators ***(40)***. We have been experimenting with a system that couples the injection needle to a piezoelectric actuator that is capable of producing a rapid stabbing motion over a linear range of up to 100 μm. A piezoelectric actuator is essentially a stack of piezoceramic discs made of Lead-Zirconium-Titanate oxide. When a voltage is applied, this stack of discs expands in height. A 10-cm stack will expand by approx 100 μm when the appropriate voltage is applied. This expansion can occur in milliseconds if the voltage is applied in an appropriate way. Consequently one can drive a needle forward in a rapid stabbing motion. Piezoelectric actuators have been used to drive microinjection needles through cell membranes ***(40)***. We envision that the high-speed stabbing motion will facilitate needle penetration, minimize embryo deformation, minimize mechanical damage, and minimize the need to mount the embryos on

extremely sticky materials. Furthermore, with a needle attached to a stabbing actuator, the needle can be positioned at any angle relative to the embryo, eliminating the need to align eggs as is done using a horizontal needle to impale embryos.

20. A number of manufactures produce a pressure-regulatory system for microinjection (e.g., World Precision Instruments, Eppendorf, Narashige). There are a number of advantages to using these more sophisticated regulatory systems. First, many of the systems allow a holding pressure to be maintained in the injection needle in the standby mode. Pressure in the needle in the standby mode helps to prevent any backflow of ooplasm into the needle during the injection process. Another advantage is that injection pressure can be applied in either a gated or timed mode. In timed mode, pressure is applied for an operator-determined interval of time. This allows very reproducible volumes of DNA to be delivered from injection to injection. Overinjecting or underinjecting become less of a concern. We inject very small volumes (10–100 picoliters) of DNA in the precise location of the embryo where pole cells will form. Large volumes are not needed and only increase the risk of developmental defects leading to sterility or embryonic lethality. Survival is highest when very small amounts of DNA are injected.

    An alternative approach to expelling the DNA solution from the microinjection needle is to maintain a constant, slow flow of DNA even from the resting micropipet. The constant flow of material out of the needle helps to keep the needle free of clogs that can occur because of backflow of cytoplasm into the needle or the adhesion of material to the exterior of the needle.

21. We often place the injected embryos in a sealed chamber in which the humidity is very high (approaching 100%) and with an elevated oxygen tension. We have found that elevated oxygen tension is useful for large embryos maintained under halocarbon oil, such as *M. domestica*, *S. calcitrans*, and *C. capitata*. For small embryos such as *D. melanogaster* this is not necessary.

    Some insect embryos, such as those of *Aedes aegypti*, must be removed from the oil soon after injection. Removal of oil requires handling of the embryos as they must be removed from the double-sided tape and place on filter paper ***(13)***. This handling process is slow and can result in damaged embryos. Furthermore, all of the oil is never removed from these embryos. For this reason we have begun to use 30% dextran sulfate to cover mosquito embryos during injection because it can be removed by simply running a gentle stream of water over the embryos (*see* **Note 14**). Furthermore, *Aedes aegypti* and *Anopheles gambiae* embryos do not have to be removed from the tape after this step. *Aedes aegypti* embryos are allowed to continue development in a nonaqueous environment for 5–7 d after which hatching is induced by submersion in deoxygenated water. *Anopheles gambiae* embryos will develop while attached to Tegaderm, but must be covered by a thin film of water (not totally immersed). Post hatching, the larvae should be able to swim from the thin film of water covering the cover slip to a larger volume in a rearing tray. For *Culex quinquefasciatus*, the entire slide, still with the embryos and moistened filter paper attached, is set in a small box allowing the

embryos to be incubated in a vertical orientation with their anteriors facing down (as they would in a natural environment). A level of water sufficient to immerse the filter paper, but not the embryos, is maintained. Following an incubation time of 24–30 h, the filter paper is removed from the coverslip and additional water added to the box, so that the water surface is at the level of the anterior end of the embryos. Larvae then hatch into the water where they are supplied with food.

Newly hatched *M. domestica* larvae should be removed from the halocarbon oil as soon as possible and place on a larvae diet.

Rearing insects in the laboratory at low densities can be difficult. Not unlike cell culture, high densities of larvae "condition" the media in ways that make it conducive to development. At low densities, these group effects are lost. One must be very familiar with the insect if this problem is to be overcome. The importance of excellent animal husbandry practices cannot be overemphasized.

Acknowledgments

The development of effective transformation protocols depends largely on the free exchange of ideas. Current methods have evolved through the efforts of many, and we acknowledge those efforts. We particularly thank Dr. A. A. James and members of his laboratory for information and advice on injecting *Aedes aegypti*. Margaret Allen provided us with information about *Culex quinquefasciatus* injection methods.

References

1. Atkinson, P. W., Pinkerton, A. C., and O'Brochta, D. A. (2001) Genetic transformation systems in insects. *Annu. Rev. Entomol.* **46,** 317–346.
2. Handler, A. M. An introduction to the history and methodology of insect gene transformation. In *Transgenic Insects: Methods and Applications* (Handler, A. M. and James, A. A., eds.), CRC, Boca Raton, FL, 2000, pp. 3–26.
3. Handler, A. M. and O'Brochta, D. A. (1991) Prospects for gene transformation in insects. *Annu. Rev. Entomol*. **36,** 159–183.
4. Walker, V. K. (1989) Gene transfer in insects. *Adv. Cell Culture* **7,** 87–124.
5. Blackman, R. K., Macy, M., Koehler, D., Grimaila, R., and Gelbart, W. M. (1989) Identification of a fully-functional *hobo* transposable element and its use for germline transformation of *Drosophila. EMBO J.* **8,** 211–217.
6. O'Brochta, D. A., Warren, W. D., Saville, K. J., and Atkinson, P. W. (1994) Interplasmid transposition of *Drosophila hobo* elements in non-drosophilid insects. *Mol. Gen. Genet.* **244,** 9–14.
7. Handler, A. M. and James, A. A. (eds.) *Insect Transgenesis*. CRC, Boca Raton, FL, 2000.
8. Ashburner, M. (ed.) *Drosophila: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.
9. Karess, R. E. *P* element mediated germ line transformation of *Drosophila*. In *DNA Cloning: A Practical Approach* (Glover, D., ed.), IRL (OUP USA), New York, NY, 1985, pp. 121–141.

10. Pirrotta, V. Cloning *Drosophila* genes. In *Drosophila: A Practical Approach* (Roberts, D. B., ed.), IRL (OUP USA), New York, NY, 1986, pp. 83–110.
11. Pirrotta, V. Vectors for *P*-element transformation in *Drosophila*. In *Vectors: A Survey of Molecular Cloning Vectors and Their Uses* (Rodriguez, R. L. and Denhardt, D. T., eds.), Butterworth, Boston, MA, 1988, pp. 437–456.
12. Spradling, A. C. *P* element-mediated transformation. In *Drosophila. A Practical Approach* (Roberts, D. B., ed.), IRL (OUP USA), New York, NY, 1986, pp. 175–198.
13. Morris, A. C. Microinjection of mosquito embryos. In *The Molecular Biology of Insect Disease Vectors* (Crampton, J. M., Beard, C. B., and Louis, C., eds.), Chapman and Hall, London, 1997, pp. 423–429.
14. Anderson, D. T. Development of hemimetabolous insects. In *Developmental Systems: Insects* (Counce, S. J. and Waddington, C. H., eds.), Academic, London, 1972, pp. 96–162.
15. Anderson, D. T. Development of holometabolous insects. In *Developmental Systems: Insects* (Counce, S. J. and Waddington, C. H., eds.), Academic, London, 1972, pp. 166–241.
16. Baldarelli, R. M. and Lengyel, J. A. (1990) Transient expression of DNA after ballistic introduction into *Drosophila* embryos. *Nucleic Acids Res.* **11,** 5903–5904.
17. Leopold, R. A., Hughes, K. J., and DeVault, J. D. (1996) Using electroporation and a slot cuvette to deliver plasmid DNA to insect embryos. *Genet. Anal.* **12,** 197–200.
18. Kamdar, K. P., Wagner, T. N., and Finnerty, V. Electroporation of Drosophila embryos. In *Animal Cell Electroporation and Electrofusion Protocols* (Nickoloff, J. A., ed.), Humana, Totowa, NJ, 1995, pp. 239–243.
19. Mialhe, E. and Miller, L. H. (1994) Biolistic techniques for transfection of mosquito embryos (*Anopheles gambiae*). *Biotechniques* **16,** 924–931.
20. McCrane, V., Carlson, J. O., Miller, B. R., and Beaty, B. J. (1988) Microinjection of DNA into *Aedes triseriatus* ova and detection of integration. *Amer. J. Trop. Med. Hyg.* **39,** 502–510.
21. Miller, L. H., Sakai, R. K., Romans, P., Gwadz, R. W., Kantoff, P., and Coon, H. G. (1987) Stable integration and expression of a bacterial gene in the mosquito *Anopheles gambiae*. *Science* **237,** 779–781.
22. Morris, A. C., Eggelston, P., and Crampton, J. M. (1989) Genetic transformation of the mosquito *Aedes aegypti* by micro-injection of DNA. *Med. Vet. Entomology* **3,** 1–7.
23. Atkinson, P. W. and O'Brochta, D. A. *Hermes* and other *hAT* elements as gene vectors in insects. In *Transgenic Insects: Methods and Applications* (Handler, A. M. and James, A. A., eds.), CRC, Boca Raton, FL, 2000, pp. 219–235.
24. Lampe, D. J., Walden, K. K. O., Sherwood, J. M., and Robertson, H. M. Genetic engineering of insects with *mariner* transposons. In *Transgenic Insects: Methods and Applications* (Handler, A. M. and James, A. A., eds.), CRC, Boca Raton, FL, 2000, pp. 237–248.

25. Loukeris, T. G., Arca, B., Livadaras, I., Dialektaki, G., and Savakis, C. (1995) Introduction of the transposable element *Minos* into the germ line of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **92,** 9485–9489.
26. Loukeris, T. G., Livadaras, I., Arca, B., Zabalou, S., and Savakis, C. (1995) Gene transfer into the Medfly, *Ceratitis capitata*, using a *Drosophila hydei* transposable element. *Science* **270,** 2002–2005.
27. Fraser, M. J. The TTAA-specific family of transposable element: identification, functional characterization, and utility for transformation of insects. In *Transgenic Insects: Methods and Applications* (Handler, A. M. and James, A. A., eds.), CRC, Boca Raton, FL, 2000, pp. 249–268.
28. Sundararajan, P., Atkinson, P. W., and O'Brochta, D. A. (1999) Transposable element interactions in insects: Crossmobilization of *hobo* and *Hermes*. *Insect Mol. Biol.* **8,** 359–368.
29. Atkinson, P. W., Warren, W. D., and O'Brochta, D. A. (1993) The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc. Natl. Acad. Sci. USA* **90,** 9693–9697.
30. Robertson, H. M. (1993) The *mariner* transposable element is widespread in insects. *Nature* **362,** 241–245.
31. Lohe, A. R. and Hartl, D. L. (1996) Autoregulation of *mariner* transposase activity by overproduction and dominant-negative complementation. *Mol. Biol. Evol.* **13,** 549–555.
32. Scofield, S. R., English, J. J., and Jones, J. D. G. (1993) High level expression of the *Activator* transposase gene inhibits the excision of *Dissociation* in tobacco cotyledons. *Cell* **75,** 507–517.
33. Heinlein, M., Brattig, T., and Kunze, R. (1994) *In vivo* aggregation of maize *Activator* (Ac) transposase in nuclei of maize endosperm and *Petunia* protoplasts. *Plant J.* **5,** 705–714.
34. Horn, C. and Wimmer, E. A. (2000) A versatile vector set for animal transgenesis. *Dev. Genes Evol.* **210,** 630–637.
35. Coates, C. J., Jasinskiene, N., Morgan, D., Tosi, L. R. O., Beverly, S. M., and James, A. A. (2000) Purified *mariner* (*Mos1*) transposase catalyzes the integration of marked elements into the germ-line of the yellow fever mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **30,** 1003–1008.
36. Catteruccia, F., Nolan, T., Loukeris, T. G., Blass, C., Savakis, C., Kafatos, F. C., et al. (2000) Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* **405,** 959–962.
37. Allen, M., Levesque, C. S., O'Brochta, D. A., and Atkinson, P. W. (2001) Stable germline transformation of *Culex quinquefasciatus* (Diptera: Culicidae) *J. Med. Entomol*. **38,** 701–710.
38. Peloquin, J. J., Thibault, S. T., Schouest, L. P., and Miller, T. A. (1997) Electromechanical microinjection of pink bollworm *Pectinophora gossypiella* embryos increases survival. *Biotechniques* **22,** 496–499.
39. Kanda, T. and Tamura, T. (1991) Microinjection method for DNA in early embryos of the silkworm. *Bull. Natl. Inst. Seri. Ento. Sci.* **2,** 31–46.

40. Brown, K. T. and Flaming, D. G. *Advanced Micropipette Techniques in Cell Physiology*. IBRO Handbook Series: Methods in the Neurosciences, Vol. 9, Wiley and Sons, New York, NY, 1986.
41. Jasinskiene, N., Coates, C. J., Benedict, M. Q., Cornel, A. J., Rafferty, C., Salazar-Rafferty, C., et al. (1998) Stable, transposon mediated transformation of the yellow fever mosquito, *Aedes aegypti*, using the *Hermes* element from the housefly. *Proc. Natl. Acad. Sci. USA* **95,** 3743–3747.
42. Coates, C. J., Jasinskiene, N., Pott, G. B., and James, A. A. (1999) Promoter-directed expression of recombinant fire-fly luciferase in the salivary glands of *Hermes*-transformed *Aedes aegypti*. *Gene* **226,** 317–325.
43. Pinkerton, A. C., Michel, K., O'Brochta, D. A., and Atkinson, P. W. (2000) Green fluorescent protein as a genetic marker in transgenic *Aedes aegypti. Insect Mol. Biol.* **9,** 1–10.
44. Moreira, L. A., Edwards, M. J., Adhami, F., Jasinskiene, N., James, A. A., and Jacobs-Lorena, M. (2000) Robust gut-specific gene expression in transgenic *Aedes aegypti* mosquitoes. *Proc. Natl. Acad. Sci. USA* **97,** 10,895–10,898.
45. Kokoza, V., Ahmed, A., Cho, W. L., Jasinskiene, N., James, A. A., and Raikhel, A. (2000) Engineering blood meal-activated systemic immunity in the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **97,** 9144–9149.
46. Michel, K., Staminova, A., Pinkerton, A. C., Franz, G., Robinson, A. S., Gariou-Papalexiou, A., et al. (2001) *Hermes*-mediated germline transformation of the Mediterranean fruit fly *Ceratitis capitata*. *Insect Mol. Biol.* **10,** 155–162.
47. O'Brochta, D. A., Atkinson, P. W., and Lehane, M. J. (2000) Transformation of *Stomoxys calcitrans* with a *Hermes* gene vector. *Insect Mol. Biol.* **9,** 531–538.
48. Berghammer, A. J., Klingler, M., and Wimmer, E. A. (1999) A universal marker for transgenic insects. *Nature* **402,** 370.
49. Handler, A. M., McCombs, S. D., Fraser, M. J., and Saul, S. H. (1998) The lepidopteran transposon vector, *piggyBac*, mediates germ-line transformation in the Mediterranean fruit fly. *Proc. Natl. Acad. Sci. USA* **95,** 7520–7525.
50. Handler, A. M. and Harrell, R. A. (2001) Transformation of the Caribbean fruit fly with a *piggyBac* transposon vector marked with polyubiquitin-regulated GFP. *Insect Biochem. Mol. Biol.* **31,** 199–205.
51. Handler, A. M. and McCombs, S. D. (2000) The *piggyBac* transposon mediates germ-line transformation of the Oriental fruit fly and closely related elements exist in its genome. *Insect Mol. Biol.* **9,** 605–612.
52. Hediger, M., Niessen, M., Wimmer, E. A., Dubendorfer, A., and Bopp, D. (2001) Genetic transformation of the housefly *Musca domestica* with the lepidopteran derived transposon *piggyBac*. *Insect Mol. Biol* **10,** 113–119.
53. Peloquin, J. J., Thibault, S. T., Staten, R., and Miller, T. A. (2000) Germ-line transformation of pink bollworm (Lepidoptera: Gelechiidae) mediated by the *piggyBac* transposable element. *Insect Mol. Biol.* **9,** 323–333.
54. Tamura, T., Thibert, C., Royer, C., Kanda, T., Abraham, E., Kamba, M., et al. (2000) Germline transformation of the silkworm *Bombyx mori* L. using a *piggyBac* transposon-derived vector. *Nat. Biotechnol.* **18,** 81–84.

55. Coates, C. J., Jasinskiene, N., Miyashiro, L., and James, A. A. (1998) *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **95,** 3748–3751.
56. Christophiles, G. K., Livadaras, I., Savakis, C., and Komitopoulou, K. (2000) Two medfly promoters that have originated by recent gene duplication drive distinct sex, tissue and temporal expression patterns. *Genetics* **156,** 173–182.

15

Transposable Elements for Transgenesis and Insertional Mutagenesis in Vertebrates

A Contemporary Review of Experimental Strategies

Zoltán Ivics and Zsuzsanna Izsvák

Summary

Functional genomic analyses in vertebrate model systems, including fish, frogs, and mice, have greatly contributed to our understanding of embryonic development and human disease. However, new molecular tools and strategies are needed to meet the increasing demands of linking sequence information to gene function. Transposable elements (TEs) are very efficient at integrating into DNA, and are therefore useful vectors for transferring new genetic material into genomes. In particular, members of the Tc*1*/*mariner* superfamily of elements are able to transpose in species other than their hosts, and are therefore emerging tools for functional genomics in several organisms. This chapter describes strategies of using retrovirus vectors and DNA-based TEs for transgenesis and insertional mutagenesis in vertebrates, with special emphasis on the *Sleeping Beauty* (SB) element, a reconstructed Tc*1*/*mariner*-like transposon from fish. SB jumps efficiently in cells of diverse vertebrate species in culture, as well as in somatic and germline tissues of the mouse in vivo. Simple structure and easy laboratory handling of transposon vectors are coupled with efficient and stable transgene integration and persistent, long-term transgene expression by transposon-mediated gene transfer. These features all contribute to the usefulness of TEs as tools for vertebrate functional genomics, as well as for animal biotechnology and human gene therapy.

Key Words: Transgenesis; insertional mutagenesis; transposon; transposition; functional genomics; gene expression; gene tagging; gene therapy.

1. Introduction

The international efforts of the Human Genome Project forecast the presence of approx 30,000–40,000 protein-coding genes in the human genome, only about twice as many as in worms and flies ***(1)***. Determination of the human genome sequence is undoubtedly one of the greatest achievements of biology. However, a major bottleneck in obtaining meaningful nucleic acid sequence

From: *Methods in Molecular Biology, vol. 260: Mobile Genetic Elements*
Edited by: W. J. Miller and P. Capy © Humana Press Inc., Totowa, NJ

information for biomedical research is the difficulty of obtaining, or the complete lack of, knowledge of the function of genes and their roles in diverse cellular functions.

One powerful approach is to study the function of a particular gene using model systems. The use of vertebrate model organisms in genomics allows human genes and their homologs to be studied in simpler organisms, while maintaining similar physiological conditions. This approach is possible because of the remarkable conservation of genes and expressed proteins between organisms. However, a serious problem associated with large-scale gene identification in vertebrates has been the lack of technology required for efficient recovery of mutant genes. Two large-scale mutagenesis screens have been carried out in the zebrafish (*Danio rerio*) using the chemical mutagen ethyl-nitroso-urea (ENU) ***(2,3)***. ENU is efficient at mutating genes. However, ENU introduces base-pair changes into DNA, and thus identification of the affected genes by positional cloning is extremely difficult. As a result, the number of cloned zebrafish genes is modest. Similarly, a number of mutant mice have been developed using ENU followed by identification of associated phenotypic changes for some genes. However, this approach is still daunting for all of the approximately 35,000 genes. Clearly, new tools and strategies are needed to meet the increasing demands of linking sequence information to gene function.

In several experimental organisms the problem of identification of mutant genes has been approached by transposon tagging in the germline. Indeed, many of the genes isolated from *Drosophila* and *C. elegans* have been identified by isolating a transposon insertion allele. Essentially, transposons can be mobilized to "hop" into genes, thereby often inactivating them by insertional mutagenesis. In the process, the inactivated genes are "tagged" by the transposable element (TE), and they can be used for subsequent recovery of the mutated allele. Transposon-tagging can also be used for finding and identifying tissue-specific and development-specific transcriptional regulatory signals in gene-trap and enhancer-trap screens. With quick PCR techniques, gene identification can be reduced from several years to months or even weeks.

The function of novel genes identified from an insertional mutagenesis screen can be validated by reinserting a functional copy into mutant animals and looking for restoration or "rescue" of a normal phenotype. There are several methods and vectors in use for gene delivery in vertebrates for genetic analyses in model species, and for the purposes of the production of transgenic animals for medical and agricultural biotechnology and human gene therapy. These methods can be broadly classified as viral and nonviral technologies, and all have advantages and limitations ***(4)***. Viral vectors, where available, are efficient at introducing and expressing genes in cells. However, adapting

viruses for gene transfer restricts genetic design to the constraints of the virus in terms of size, structure, and regulation of expression. Nonviral methods, including DNA condensing agents, liposomes, microinjection, and "gene guns," might be easier and safer to use than viruses, but they do not promote integration into chromosomes. As a result, stable gene-transfer frequencies using nonviral systems have been very low. Moreover, most nonviral methods often result in concatamerization as well as random breaks in input DNA that might lead to gene silencing.

Transposons can be harnessed not only as vehicles for destroying endogenous genes by insertional mutagenesis (loss-of-function mutations), but also for bringing new phenotypes into genomes by transgenesis (gain-of-function mutations). Although the underlying technology and the vectors used for both transgenesis and insertional mutagenesis can be similar, the requirements that must be fulfilled can be very different. For example, it is desirable that a transposon vector have the capacity to mobilize long genes for transgenesis, whereas this mobilization is not an issue in insertional mutagenesis, in which the sole purpose of the integrating transposon is to inactivate a gene and provide a short sequence tag for subsequent recovery of the allele. Furthermore, insertion of a single copy of the transposon, preferably away from other, endogenous genes, is desired in transgenesis. Clearly, just the opposite is the goal in insertional mutagenesis, where one aims at multiple integrations causing as much damage as possible. Thus, although transposon vectors can be adapted for diverse purposes, their use needs to be adjusted according to the particular experimental goals.

Transposable elements have not been used for the investigation of vertebrate genomes for two reasons. First, until very recently, there have not been any well-defined, DNA-based mobile elements in these species. Second, most transposon systems that are active in certain species cannot easily be transferred to vertebrates, because of species-specificity of transposition due to the requirement of factors produced by the natural host. The last five or so years have brought major discoveries and technological advances that allow TEs to be used as experimental tools for the exploration of the vertebrate genome.

The following sections are meant to provide experimental strategies and important considerations for the use of transposons in vertebrate genetics, rather than detailed protocols. The reason for this is twofold. First, as said above, the application of TEs in vertebrate species is still in its infancy. Much work is needed to build up experience, preferably from different laboratories, that could be synthesized into unified experimental protocols. Second, the methodology required for generating transposon insertions in vertebrate animals will be specific to the species in question. This situation is illustrated in **Table 1**, which summarizes the different transgenic experiments in vertebrates

Table 1
Horizontal Gene Transfer in the Laboratory[a]

Transposon	Host		Method of transfer
	Origin	Target	
Tc*1*	Nematode (*C. elegans*)	Human	Transfection/DNA
Tc*3*	Nematode (*C. elegans*)	Zebrafish	Microinjection/RNA
Minos	Insect (*D. hydei*)	Human Mouse	Transfection/DNA chromosomal mobilization
Sleeping Beauty		Zebrafish μ, # Carp # Sea bream # Medaka # Trout # Sword tail #	
	Fish (reconstructed)	Fathead minnow # *Xenopus* # Sheep # Cow # Dog # Rabbit # Hamster # Mouse #, *, Å Mouse ES # Monkey # Human #	(μ) Microinjection/RNA (#) Transfection/DNA (*) Tail vein injection/DNA (Å) Adenoviral infection
*Himar*1	Insect (*H. irritans*)	Human	Adenoviral infection
*Mos*1	Insect (*D. mauritiana*)	Chicken Zebrafish	Microinjection
*Tol*2	Fish (*O. latipes*)	Zebrafish	Microinjection/RNA
MoMLV/VSV	Murine retrovirus	Zebrafish	Microinjection/virus

[a]Summary of genetic transformation experiments using Tc*1*/*mariner* transposons, *Tol*2 elements of the medakafish, and pseudotyped murine retrovirus vectors in heterologous vertebrate species.

done to date, and the diverse methods that have been used for gene transfer. For example, the technology required for gene trapping with transposon vectors in mouse embryonic stem (ES) cells will be different from the technology required for generating transgenic zebrafish with transposon vectors. This chapter describes the types of transposons currently tested for transgenesis and insertional mutagenesis in vertebrates.

2. Methods

2.1. Insertional Mutagenesis of Zebrafish Genes by Retroviruses

There is a remarkable similarity between the biochemical steps of retrovirus integration, and DNA transposition ***(5)***. Retroviruses make a cDNA copy of themselves that is inserted into chromosomal DNA by the virus-encoded recombinase, the integrase. The inserted provirus is stably inherited by the descendants of the infected cell. Because of stable genomic integration and maintenance in the genome, retroviruses are excellent vectors for introducing genetic material into cells.

A large-scale insertional mutagenesis screen using retrovirus vectors has been initiated with the aim of generating about 1000 embryonic mutations in the zebrafish ***(6)***. Because there is no known natural retrovirus that infects zebrafish cells, viral vectors based on the Moloney Murine Leukemia Virus (MoMLV) genome and pseudotyped with the envelope glycoprotein of vesicular stomatitis virus (VSV) had to be developed. VSV can infect cells of diverse species, including zebrafish. Indeed, the key for success in generating mutations in useful numbers in a genome as complex as that of the zebrafish (about 1.7×10^9 bp) was the generation of recombinant viral stocks that had the ability to infect cells of a developing zebrafish embryo, and that were stable enough to be concentrated 1000-fold to derive high titers ***(7)***.

The protocol is based on microinjection of approx $1–2 \times 10^4$ virus particles into blastula-stage embryos (on average 1000 cells) ***(8)***. The need for injecting virus into developing embryos one at a time is unfortunate, and it arises from the observation that the virus cannot penetrate embryonic membranes. It took about one year for two researchers working five days per week to establish 36,000 founder fish, by injecting a total of about 250,000 embryos. Proviruses are able to integrate into the zebrafish germline, but because of the delay in vector delivery, the transgenic fish obtained by this protocol will be mosaic. Mosaic founders are outcrossed to establish F_1 fish with retrovirus insertions; founders transmit insertions to between 1% and 40% of their F_1 progeny, and can transmit on average 20 insertions to individual F_1 fish ***(9)***. A critical parameter of the whole screen is the average number of unique inserts in F_1 animals used to generate the F_2 families. Thus, founders are first crossed in

order to enrich for fish with multiple insertions, and these F_1 fish are used to generate F_2 families with six or more segregating insertions ***(6)***. Sibling crosses of F_2 yield F_3 embryos that are subjected to visual examination using a dissecting microscope to identify phenotypic mutations. Linkage between a particular proviral insertion and a mutation is established using Southern blot hybridization. A specific Southern band must be shared by both parents of every cross that shows the phenotype, and it must be found in only one or in neither of the parents of crosses that do not show the phenotype. Junction fragments of candidate inserts are recovered by inverse PCR, and these fragments are then used to clone out the affected genes. The essential numbers and parameters of a large-scale insertional mutagenesis screen carried out in the Hopkins laboratory ***(6)*** are as follows: (1) Mutagenic frequency was about 1 embryonic lethal mutation per 85 insertions; insertional mutagenesis with retroviruses is about 100 times less efficient in zebrafish than ENU mutagenesis; (2) Frequency of recovering mutations was about 1 embryonic lethal mutation per F_2 family; thus the frequency of recovering insertional mutations is about one-ninth of that observed with ENU; and (3) 18 insertional mutants were identified, and 10 of these genes have been cloned.

2.2. DNA-Based TEs as Genetic Tools

Especially useful for genetic analyses might be members of a class of TEs that move via a "cut-and-paste" mechanism: The transposase catalyzes the excision of the transposon from its original location and promotes its reintegration elsewhere in the genome. Transposase-deficient elements can be mobilized if the transposase is provided *in trans*. It is this *trans*-complementary nature of DNA transposition that forms the basis of powerful experimental manipulation: The experimentalist can insert a gene of interest into a nonautonomous element that is stable in cells in the absence of a transposase. The transposase can then be added to induce mobilization of the element. Thus, transposition can be controlled by simply adding or taking away the transposase. The next sections describe the current status and methodology of genetic applications of these elements.

2.2.1. The Tol*2 TE in the Medakafish: A Promising Candidate for a Gene Vector in Vertebrate Species*

The *Tol*2 TE was discovered as an insertion in the tyrosinase gene of the medakafish (*Oryzias latipes*) ***(10)***. The insertion caused a loss-of-function mutation in the gene by disrupting the reading frame, and resulted in an albino phenotype. The cloned insertion turned out to be a 4.7-kb element that showed similarity in its sequence to the hAT superfamily of transposons, which includes *hobo* elements in *Drosophila*, *Ac* elements in maize, and *Tam*3 in

snapdragon. Medaka contains about ten copies of *Tol*2 per haploid genome. First evidence for mobility of *Tol*2 elements was provided by finding the element in the tyrosinase gene to undergo excision during embryogenesis ***(10)***, arguing for at least a single transposase-producing element somewhere in the medaka genome. In contrast to many transposable elements in which substantial numbers of mutations build up over evolutionary time as a result of the lack of selective pressure that would maintain their DNA sequences, *Tol*2 elements show a remarkable homogeneity in medaka. These sequences are either identical or very similar to each other, and it thus appears reasonable to speculate that this element is a relative newcomer in the medaka genome ***(11)***. Based on these observations, looking at one particular copy in a search for a transposase is as good as looking at any other. Indeed, there was no need to go further than the original *Tol*2 copy identified in the tyrosinase locus, which turned out to encode a fully functional transposase.

When *Tol*2 transposase mRNA and a plasmid containing a nonautonomous *Tol*2 element are coinjected into zebrafish embryos, the element undergoes excision and integrates into chromosomes ***(12)***. Transposon insertion apparently can occur in cells of the germ lineage, as insertions can be detected in the F_1 generation. Four transposon insertions in one fish out of eight injected have been found and analyzed. Clearly, the number of transgenic fish obtained by this method and the number of transposon insertions obtained so far are far too low to allow conclusions as to the usefulness of the element for transgenesis and/or insertional mutagenesis. *Cis* requirements for element mobility are practically unknown, as are size constraints for transposition or the ability of a marker gene to be expressed from within the integrated transposon. However, the *Tol*2 element undoubtedly holds promise as a potential gene vector in fish, and possibly in other vertebrates.

*2.2.2. Tc1/*mariner*-Like TEs in Transgenic Experiments*

Homologs of the Tc*1* element in *Caenorhabditis elegans* and those of the related *mariner* transposon found in *Drosophila mauritiana*, are probably the most widespread DNA transposons in nature ***(13)***. Tc*1*/*mariner* elements are approx 1300–2400 bp in length and contain a single gene encoding a transposase enzyme flanked by terminal inverted repeats. Although quite divergent in primary sequence, members of the Tc*1*/*mariner* superfamily are probably monophyletic in origin ***(14)*** and have similar structures and molecular mechanisms of transposition ***(13,15)***.

Many prokaryotic elements require specific host proteins for transposition, which limits their mobility outside their natural hosts. Similarly, *P* elements require a *Drosophila*-encoded factor for transposition, and thus do not seem to jump in vertebrate cells ***(16)***. Two observations indicated that Tc*1*/*mariner*

elements are less restricted in their transposition in different hosts. First, on an evolutionary time scale there is an indication that these elements are promiscuous, because sequence comparisons of certain transposons in species that are thought to have diverged more than 100 million years ago showed elements that were virtually identical. A likely interpretation of this observation was that the elements spread fairly recently from one species to the other by horizontal transfer ***(17,18)***. Second, it was found, using in vitro transposition assays, that no species-specific protein other than the transposase was required for transposition of Tc*1*/*mariner* transposons ***(19,20)***.

Thus, the prediction was that expression of transposase in any host should be sufficient to trigger transposition of the corresponding transposon. Indeed, the natural process of horizontal transfer can be mimicked in the laboratory. **Table 1** summarizes experiments in which transposition of various Tc*1*/*mariner* elements was shown in different vertebrate species. These elements have been shown to undergo transposition in cultured human cells ***(21–24)***, and have been used to generate transgenic zebrafish ***(25,26)*** and chickens ***(27)***.

2.2.3. The Sleeping Beauty *Transposon*

2.2.3.1. *Sleeping Beauty* Awakened

Tc*1*/*mariner*-like elements have been found in several vertebrate genomes ***(18,28–31)*** including the human genome ***(32–34)***. However, in sharp contrast to the *Tol*2 elements in the medakafish, all of the transposon copies isolated to date from vertebrates are clearly dead remnants of once active transposons that, after successfully colonizing genomes, have become inactivated by mutations. This is most likely because of a fairly long history of these elements in vertebrate genomes.

In an attempt to derive an active Tc*1*-like transposon from vertebrates, we have analyzed a particular subfamily of elements that was presumed to have been active more than 10–15 million years ago, and that appeared to have been able to invade different fish genomes through horizontal transmission ***(18)***. We reasoned that a consensus sequence generated from a sequence alignment of defective copies isolated from different fish genomes would likely represent an active archetypal sequence. We engineered this sequence to reconstruct an active ancestral element, which was named *Sleeping Beauty* (SB) ***(35)***.

The SB transposon system consists of two main functional components: The transposase protein and the terminal inverted repeats of the transposable element (**Fig. 1A**). The transposase has an N-terminal DNA-binding domain overlapping with a nuclear localization signal ***(18)***, that is followed by a second major domain that is responsible for the catalytic steps of transposition. This domain is characterized by a conserved amino acid triad, the DDE signa-

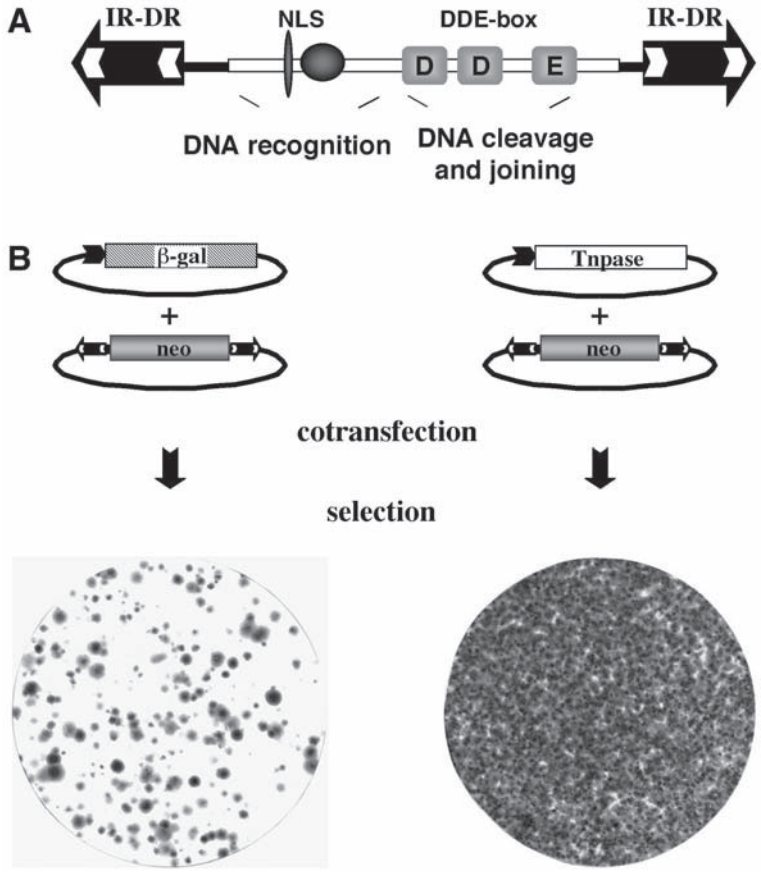


Fig. 1. *Sleeping Beauty* transposition in cultured vertebrate cells. **(A)** Schematic representation of the two major components of the *Sleeping Beauty* TE system. Under experimental conditions the two components are separated. The terminal inverted repeats of SB elements have a characteristic structure (IR/DR) and contain multiple binding sites for the transposase enzyme. The transposase has an N-terminal DNA-binding domain overlapping with a nuclear localization signal, which is followed by the catalytic domain responsible for the DNA cleavage and joining reactions and characterized by the conserved DDE signature. **(B)** Assay for transposition in cultured cells. A selectable marker such as an antibiotic-resistance gene is cloned between the inverted repeats of the SB transposon. Transposon donor plasmids are introduced into cells together with transposase-expressing helper plasmids by transfection. Cells are placed under antibiotic selection; only cells that express the antibiotic-resistance gene as a result of chromosomal integration survive. Resistant cells give rise to colonies that can be harvested for DNA analysis, picked and expanded into larger cultures or stained for documentation. Shown are two Petri dishes with stained human HeLa cell colonies obtained in the absence (*left dish*) or in the presence (*right dish*) of transposase. The marked difference in the numbers of resistant clones is a result of transposition of the marked transposable elements into chromosomes.

ture (designating two conserved aspartic acid and a glutamic acid residue ***(36)***, and contains a glycine-rich block of amino acids of yet unknown function ***(35)***. The transposase gene is flanked by terminal inverted repeats (IRs), that contain binding sites for the transposase. The transposase binding sites of SB elements are repeated twice per IR in a direct orientation (DRs) ***(35)***. This special organization of inverted repeat, termed IR/DR ***(37)***, appears to be an evolutionarily conserved feature of a group of elements within the Tc*1* family ***(13)***.

The two components of the SB system have to be put together in a cell for transposition to occur. We established an in vivo transposition assay to detect SB transposition events from plasmids to the chromosomes of vertebrate cells ***(35)***. The assay, shown in **Fig. 1B**, is based on cotransfection of a transposon donor plasmid and a transposase-expressing helper plasmid into cultured cells. The TE carries an antibiotic resistance gene such as *neo*, so that inserted transposons that express their genes confer an antibiotic-resistant phenotype to cells. Cells are then placed under G-418 selection, and resistant colonies counted. The ratio between numbers obtained in the presence and absence of transposase is the readout of the assay, and it is a measure of the efficiency of transposition. The result of a typical experiment on human HeLa cells is shown in **Fig. 1B**.

Two useful variations of the basic assay have recently been developed. One allows positive/negative selection to be applied in order to enrich populations of cells harboring inserted transposons in their chromosomes. This assay positively selects for transposition events using the antibiotic-resistance gene within the transposon, and negatively selects against integration of plasmid vector sequences, using the *thymidine kinase* (*TK*) gene of herpes simplex virus type 1 within the vector backbone of transposon donor plasmids ***(38)***. Upon cotransfection of this construct into cells together with a helper plasmid, G-418-resistant cell clones are selected in the presence of gancyclovir, which is toxic to cells expressing the *TK* gene. The other useful feature added to the system is the incorporation of a bacterial promoter upstream of the antibiotic-resistance gene and a plasmid origin of replication within the transposable element ***(38,39)***. This modified vector allows for rapid isolation of the chromosomal insertion site of the element by plasmid rescue.

The basic transposition assay has been employed to compare the activities of several Tc*1*/*mariner* transposable elements under identical conditions ***(23)***. Human HeLa cells were cotransfected with helper plasmids expressing the Tc*1* and Tc*3* transposases from *C. elegans*, the Mos1 and the Himar1 transposases (and hyperactive versions thereof) from insects and the SB transposase together with the corresponding TEs. *Sleeping Beauty* was found an order of magnitude more efficient in transposition than the other elements ***(23)***.

2.2.3.2. *Sleeping Beauty* Is Active in Diverse Vertebrate Species

To assess the limitations of host specificity of SB among vertebrates, we applied the above transposition assay to cultured cells of representatives of different vertebrate classes ***(38)***. Cell lines from different fish species, from mouse, human, frog, quail, sheep, cow, dog, rabbit, hamster, and monkey were tested. As summarized in **Table 1**, SB was able to increase the frequency of transgene integration in all of these cell lines. These results indicate that SB is active in most vertebrate species. We found extensive variation in the extent to which transposase stimulates integration between different species and even between different cell lines of the same species. Such variability can be explained by (1) different transfectability of the different cell lines, that can affect both the amount of transposase and the number of available transposon substrate molecules per cell; (2) different activities of the CMV promoter in the different cell lines; and (3) interaction of certain host factors with the transpositional machinery, that might lead to different efficiencies of transposition in different species and/or cells. Thus, use of the TE vector system must be optimized for each species in terms of nucleic acid delivery and transcriptional regulation of transgenes.

2.2.3.3. *Sleeping Beauty* Transposition in Somatic Tissues of the Mouse

In vitro induction of transposition and selection of cultured cells that harbor integrated transposons in their chromosomes can be useful when combined with embryonic stem cell technology ***(40)***. However, there is considerable interest in technologies that allow the delivery and expression of genes in certain tissues or organs in vivo, for the correction of genetic diseases.

Evidence that the SB system can potentially be developed as a useful vector for gene therapy came from experiments in which the two components of the transposon system were administered into living mice by tail vein injection ***(39)***. Using this simple technology, about 5% of hepatocytes of the experimental animals expressed a foreign marker protein, β-galactosidase, from the *lac*Z gene within the transposon vector. Taking into account that the transposon vector cannot infect cells (thus active cellular uptake is not promoted), a 5% transformation efficiency is a significant result, because other integrating and infectious vectors, such as retroviruses and adeno-associated virus vectors, also transform hepatocytes in vivo with similar efficiencies. Thus, SB can mediate efficient chromosomal integration of transgene constructs in vivo in a mammalian model system. For gene therapy, chromosomal integration of transgene constructs itself does not solve the problem, because in many cases the transferred gene has to be expressed for a prolonged period of time, and the gene product must be produced at a level that will have a therapeutic effect.

The following three experiments clearly demonstrated that the SB transposon can fulfill both requirements. First, transgenic mice generated with an SB vector containing the human α-1-antitrypsin (hAAT) cDNA expressed hAAT in their blood for more than six months ***(39)***. Second, transfer of an SB vector containing a human Factor IX (FIX) expression cassette resulted in partial correction of the bleeding disorder in hemophilic mice ***(39)***, and the production of biologically active FIX was sustained at levels that would convert a severely affected patient with hemophilia B to one with a much milder phenotype. Finally, SB-mediated gene therapy in fumarylacetoacetate hydrolase (FAH)-deficient mice has been shown to correct hereditary tyrosinemia type 1 in 62 % of the animals receiving a FAH-expressing transposon construct and the transposase ***(41)***. This last study demonstrated an average transposon copy number of 1 per diploid cellular genome in the liver, and a long-lasting transgene expression even after serial transplantation of hepatocytes.

Together, the above studies have demonstrated the potential usefulness of *Sleeping Beauty* for the correction of human genetic diseases. One problem with respect to further applications is that the efficiency of in vivo gene transfer into many types of tissue with naked DNA constructs is rather low; therefore, the overall transformation rates with plasmid-borne SB vectors can be insufficient in clinical applications. A second problem is that the hydrodynamic injection method used in the above studies (injection of a large volume of DNA solution into the bloodstream in a couple of seconds) is hardly applicable in humans. A potential solution to both problems was offered by engineering an adenovirus/SB hybrid vector ***(42)***. Adenovirus vectors are very efficient at infecting cells, but transgene expression from these vectors is transient because of the lack of stable genomic integration. Repeated administration of adenovirus vectors can induce an immune response against viral proteins and the elimination of transduced cells. The adenovirus/SB hybrid combines the advantages of the two systems: high efficiency gene transfer, and stable transgene integration and expression.

Recently, another Tc*1*/*mariner*-type element from *Drosophila hydei*, *Minos*, was successfully mobilized in somatic tissues of the mouse ***(43)***. Transgenic mice expressing the transposase in a tissue-specific manner in the thymus and spleen were generated. These animals also contained GFP-marked *Minos* elements as substrates for transposition. Excision of the element was tissue-specific, but reintegration took place at a relatively low efficiency. About 0.6% of thymus and spleen cells were found to contain new transposon insertions. If rates of somatic transposition were high enough, similar schemes could prove useful for the discovery of novel cancer genes.

2.2.4. Experimental Strategies to Induce Transposition in the Vertebrate Germline

Whether the goal is the expression of an exogenously supplied transgene in certain vertebrate species for agricultural and medical biotechnology, or the insertional inactivation of an endogenous gene in an experimental model, it is important that genetic changes can be directed to the germline so that mutations can be passed on to the next generation, and lines of transgenic animals can be established. Classical methods to express foreign genes in vertebrate animals rely on injection of nucleic acids into oocytes or fertilized eggs. These techniques are relatively simple, but because plasmid-borne genes are not equipped to promote chromosomal integration, their presence and expression is usually transient and mosaic, and it very rarely results in genomic integration. The different methods that can be considered for the introduction of transposon vectors into the germline of vertebrate animals are shown in **Fig. 2**, using zebrafish as an example, and are described next.

2.2.4.1. Microinjection of Plasmids Containing Marker Transposons Together With Transposase mRNA or Purified Transposase Protein Into Oocytes or Early Embryos

This method is the most straightforward one to use (**Fig. 2A**). It has been employed to generate transgenic zebrafish with the Tc*3* ***(26)***, *mariner* ***(25)***, *Tol*2 ***(12)*** and SB elements ***(44)***, and transgenic chickens with the Mos1 *mariner* element ***(27)*** (**Table 1**). If transposase mediates the insertion of marked transposons into the genome early enough (i.e., before the first cell division), then the integrated transposon should be passed on to all of the daughter cells. Ultimately, all of the cells of the developing animal, including the germline, would be expected to contain the transgene. However, all the transgenic founder animals that have been generated in the above experiments were without exception mosaic with respect to the presence of integrated transposons in their germline, suggesting a delayed transposition of the elements during embryogenesis.

2.2.4.2. Microinjection of Plasmids Containing Marker Transposons Into Oocytes of Transgenic Females That Express Transposase in the Germline

In this approach, which might help to circumvent the problem of mosaicism due to delayed transposition, a transgenic line of the experimental animal species needs to be generated first that will express the transposase in the female germline (**Fig. 2B**). Specificity of transposase expression can be ensured by transcriptional regulatory elements that direct the expression of the transposase gene in developing oocytes, so that transposase is deposited in the egg before microinjection of the DNA transposon substrates takes place. Thus, a presum-

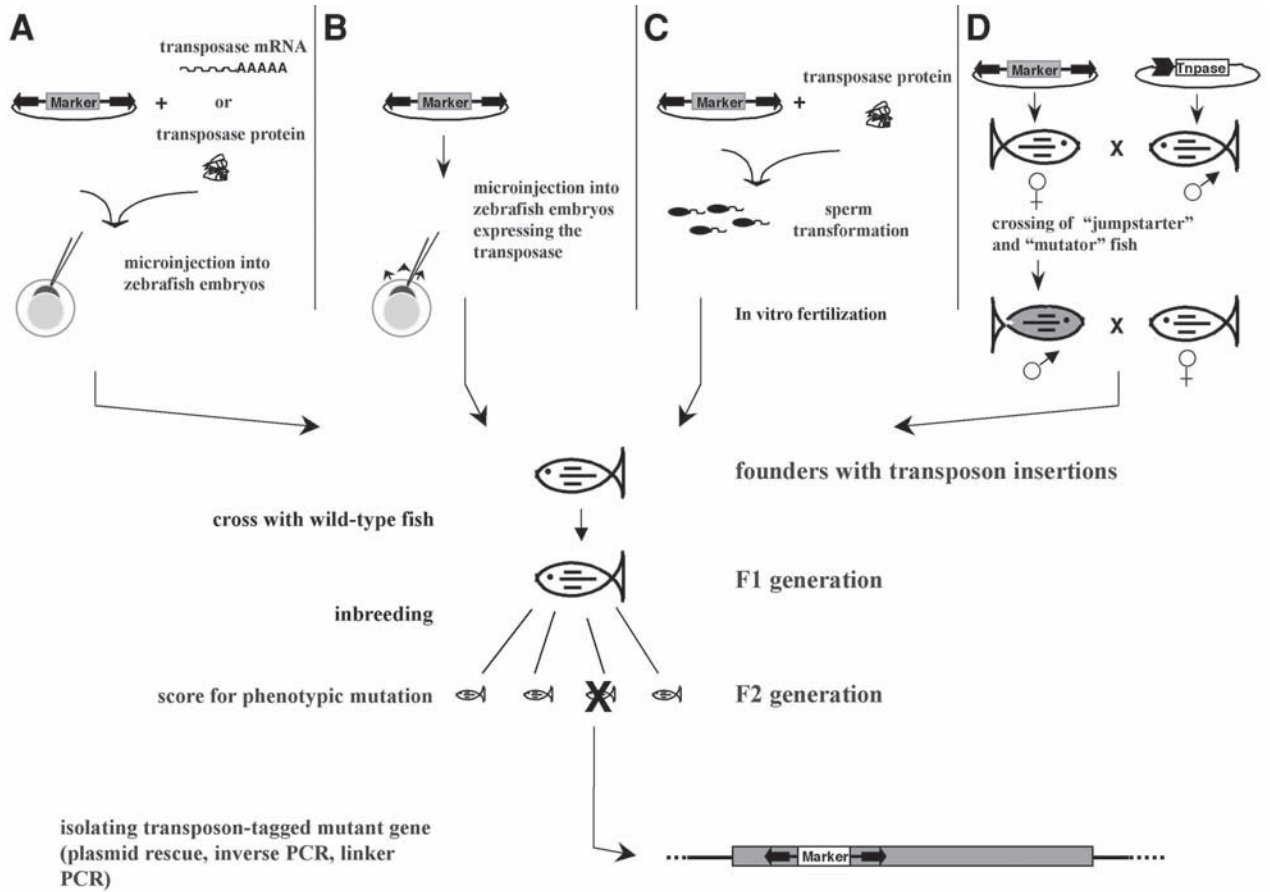

A
transposase mRNA
AAAAA
Marker
+
or
transposase protein
microinjection into zebrafish embryos
B
Marker
microinjection into zebrafish embryos expressing the transposase
C
transposase protein
Marker
+
sperm transformation
In vitro fertilization
D
Marker
Tnpase
x
crossing of "jumpstarter" and "mutator" fish
x
founders with transposon insertions
cross with wild-type fish
F1 generation
inbreeding
score for phenotypic mutation
F2 generation
isolating transposon-tagged mutant gene (plasmid rescue, inverse PCR, linker PCR)
Marker

Fig. 2.

ably active transposase is available for transposition prior to injection of transposon donor plasmids into in vitro fertilized eggs of the transposase producer line. One consideration that might be important in order to circumvent a potential problem associated with continuing expression of the transposase gene in transgenic lines is the use of inducible expression systems, such as tetracycline-inducible gene expression.

2.2.4.3. Sperm Transformation In Vitro With Transposon Plasmids and Purified Transposase

Sperm cells can take up genetic material in vitro, and these transgenic sperm contribute to the genome of a developing animal upon fertilization of oocytes (**Fig. 2C**). This approach has been shown to work by simply incubating sperm cells of mammalian species in DNA solution ***(45)***, by electroporation of fish sperm ***(46)***, and by treatment of *Xenopus* sperm with restriction enzymes (restriction-enzyme mediated integration, REMI) prior to exposure to exogenous DNA ***(47)***. In essence, any of these methods could be adapted for transposition by applying a treatment to sperm cells that would allow access of the transpositional machinery to the chromosomal DNA.

2.2.4.4. Classical Breeding of "Jumpstarter" and "Mutator" Stocks to Induce Transposition in the Germline of the Hybrid

This method is most likely the preferred one for generating large numbers of transposon insertions for insertional mutagenesis, a method that cannot be applied for retrovirus vectors (**Fig. 2D**). In this experimental setup, two transgenic lines need to be generated first; a "jumpstarter" stock expressing the transposase in the male germline, and a "mutator" stock containing the transposon to be mobilized. These two stocks are crossed to bring the two components of the transposon system together, and transposition is expected to

---

Fig. 2. *(previous page)* Experimental strategies to induce transposition in the vertebrate germline. The figure shows possible strategies for the zebrafish, but some of these methods can be adapted to other vertebrate species. **(A)** Microinjection of plasmids containing marker transposons together with transposase mRNA or purified transposase protein into oocytes or early embryos. **(B)** Microinjection of plasmids containing marker transposons into oocytes of transgenic females that express transposase in the germline. **(C)** Sperm transformation in vitro with transposon plasmids and purified transposase. **(D)** Breeding of "jumpstarter" and "mutator" stocks to induce transposition in the germline of the hybrid. All of these strategies are expected to yield transgenic founders containing integrated TEs in their germline cells. These founders need to be bred to homozygosity in order to visualize the phenotypic effects of recessive mutations. Mutant genes can easily be cloned by different PCR methods making use of the inserted transposon as a unique sequence tag.

occur in the sperm cells of males of the heterozygous hybrids. Such males would be crossed to wild-type females to segregate the different insertion events in the genomes of their sperm cells into separate animals.

Indeed, such a scheme might be useful for generating insertional mutations in vertebrate model systems ***(23)***. Two separate transgenic mouse lines have been established: one expressing the SB transposase from the protamine 1 promoter, which is active during spermiogenesis, and the other containing an integrated *neo*-marked SB element. In 20% of the offspring of double-transgenic males, the transposon jumped to different genomic locations, and transposon insertions are stably transmitted in the absence of the transposase ***(23)***. Subsequent to this study, two other papers describing a similar experimental design and even more encouraging results were published ***(48,49)***. Both studies employed a ubiquitous promoter to drive the expression of the Sleeping Beauty transposase in transgenic mice, and a multicopy array of transposons as donors for transposition. Horie et al. ***(49)*** have found up to 80% of the progeny of a double-transgenic male to contain transposition events, and they estimated the frequency of germline transposition to be about one event per gamete; whereas Dupuy et al. ***(48)*** estimate that, on average, their double-transgenic males carry about two new transposon insertions per sperm cell. Although further studies are required to establish a protocol for inducing large numbers of transposon insertions in the mouse germline, it appears that promoter choice and a chromosomal pool of transposons with a sufficiently large number of elements available for mobilization will be among the important parameters.

Preferably, the transposon would contain a gene trap construct to allow for selection of transposon insertions into genes. In the mutator stock, the gene trap must not be expressed. Offspring of double-transgenic males can be examined for expression of the gene trap marker, such as the green fluorescent protein (GFP), which is indicative of transposition of the marker elements into expressed genes. Usually, mutations can only be observed when products of both copies of a gene are inactivated. By bringing the insertions identified in the founder animals to homozygosity, animals that contain two mutant copies of the affected gene can be generated. The spatial/temporal expression of the transposon marker and possible phenotypic effects (mutations) can be examined over the course of embryonic development, and the affected tissues/organs/developmental pathways can be colocalized with the marker.

2.3. Considerations for Using TEs for Transgenesis and Insertional Mutagenesis in Vertebrate Species

2.3.1. General Considerations for Gene Transfer

Transposable elements in general and the *Sleeping Beauty* system in particular has several advantages for gene transfer in vertebrates:

1. SB can transform a wide range of vertebrate cells ***(38)***.
2. SB does not appear to be restricted in its ability to transpose DNA of any sequence.
3. SB vectors do not have strict size limitations ***(38)***.
4. SB requires only about 230 bp of transposon inverted repeat DNA flanking a transgene on each side for mobilization ***(38)***.
5. SB vectors are transcriptionally neutral, and thus do not alter endogenous gene function at the site of insertion.
6. Transposition is inducible, and requires only the transposase protein; thus one can simply control the site and moment of jumping by control of transposase expression ***(38)***.
7. SB mediates stable, single-copy integration of genes into chromosomes, that forms the basis of long-term expression throughout multiple generations of transgenic cells and organisms ***(35,38,39,41,42)***.

2.3.1.1. Stability of Integrated Transposon Vectors in Genomes

Stable inheritance and expression of transgenes introduced into genomes by TE vectors is of key importance for maintaining particular genotypes. As indicated in **Fig. 2**, for most experimental species the transposase source can be provided in the form of (1) DNA, in which case expression of the transposase gene should either be tightly controlled, or the transposase gene should be segregated away by outcrossing transgenic animals in order to ensure stability of the inserted transposon; (2) transposase mRNA; or (3) transposase protein. In general, it is expected that the presence of SB transposase is only transitory in cells and is limited to a time window when transposition is catalyzed. However, in case of somatic gene transfer for human gene therapy, the potential danger of integration of transposase-expressing nucleic acid is unacceptable, and thus use of purified transposase protein would be favored.

In the absence of exogenously supplied transposase, integrated SB elements are expected to behave as stable, dominant genetic determinants in the genomes of transgenic cells, because (1) there is no evidence of an endogenous transposase source in vertebrate cells that could activate and mobilize integrated SB elements; and (2) with the exception of some fish species, there are no endogenous sequences in vertebrate genomes with sufficient homology to SB that would allow recombination and release of transpositionally competent (autonomous) elements.

2.3.2. Considerations for Insertional Mutagenesis

2.3.2.1. Target Site Specificity of Transposition: Can TEs Hit All Genes?

One important consideration when using insertional mutagens such as retroviruses or DNA-based transposons is whether it is at all possible to hit all genes in the vertebrate genome. Retroviruses by no means integrate randomly

into chromosomal DNA. On the contrary, integration into the 5' regions of actively transcribed regions is preferred ***(50)***, which is understandable if we assume that the retroviral integration machinery needs access to DNA for integration, and thus chromosomal regions of open chromatin structure will be hit more frequently than others. A similar observation was made for *P* elements in *Drosophila* ***(51)*** and, although preferred integration into or around promoter regions is useful when knocking out genes is the goal, it is estimated that only about one-third to one-half of the fly genes can be mutated with *P* elements ***(52)***. Although the number of retrovirus insertions in zebrafish analyzed to date is too limited to draw firm conclusions on the randomness of the insertions, based on observations in other experimental systems it seems safe to expect that a substantial fraction of genes can be mutated with retroviruses.

Tc*1*/*mariner* elements, including SB, integrate at TA dinucleotides ***(13)***, which occur approximately once every 20 bp, on average, in vertebrate genomes. SB shows no apparent preference for certain chromosomes or chromosomal regions in human cells, and it appears to integrate in a random fashion ***(53)***; thus it may have a different but overlapping set of targets than that of retroviral vectors.

2.3.2.2. Efficiency of Transposition in the Germline

Results with ENU obtained for zebrafish and for mice show about 1 specific locus mutation per 1000 mutagenized gametes. What frequency of transposition is required for optimal mutagenesis? In theory, one needs a mutagenic dose that has the same effect as the optimal mutagenic dose of ENU. Even at the relatively high rates of transposition found in the mouse germline, it is estimated that the mutational frequency is less than 10^{-6} per locus per sperm of a mouse doubly transgenic for both components of the transposon system ***(23)***. This is 1000 times lower than ENU mutagenesis. However, on average, the phenotypic effect of a transposon insertion is more dramatic than that of a single nucleotide substitution, and therefore the total number of insertions required to reach the same mutagenic effect as treatment with ENU should be much lower. Results obtained with retrovirus vectors in zebrafish show that about 1% of insertions will cause an embryonic lethal mutation ***(6)***. Similarly, approx 1% of insertions results in activation of a gene trap marker in *Xenopus*, using REMI ***(54)***. Thus, similar numbers can probably be expected from transposon vectors. Although relatively random insertion of transposon vectors can be a clear advantage for gene identification through insertional mutagenesis, a limiting factor can be the overall frequency of transposition. A critical parameter contributing to the success of insertional mutagenesis with TEs would be whether multiple transposon insertions per gamete can be generated. The relative inefficiency of gene inactivation with insertional mutagens can be

counterbalanced by the fecundity of model species, such as fish or frogs, where several thousand embryos can be generated in a single mating.

2.3.3. Considerations for Human Gene Therapy

Considerable effort has been devoted to the development of in vivo gene delivery strategies for the treatment of inherited and acquired disorders in humans. For gene therapy, a desirable vector must ensure that:

1. Therapeutic genes can be delivered at high efficiency specifically to relevant cells.
2. Expression of the gene occurs for a prolonged period of time.
3. The introduction of the therapeutic gene is not deleterious to endogenous gene function and does not cause any other unwanted side effect.
4. Cost-effective, large-scale manufacture is possible.

As discussed above, TEs can fulfill the requirements of efficient gene transfer and persistent gene expression in mammalian systems. However, a concern with any vector that integrates into chromosomes in a random fashion, including all current retrovirus vectors and most TEs, is the potential disruption (or activation) of endogenous gene function at and near the insertion site. One way to minimize the chance of insertional inactivation (or activation) of an endogenous gene might be to titrate transposition to a single insertion per average genome. Alternatively, it might be feasible to direct transposon integration into desired locations by engineered transposases with designed DNA-binding specificities. Because SB is a DNA-based transposon, its production is easy, inexpensive, and able to be scaled up.

Acknowledgments

We thank D. Fiedler and E. Stüwe for their technical assistance, and O. Walisko and C. Kaufman for their helpful comments on the manuscript.

References

1. International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921.
2. Driever, W., Solnica-Krezel, L., Schier, A., Neuhauss, S., Maliki, J., Stemple, D., et al. (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123,** 37–46.
3. Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., et al. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123,** 1–36.
4. Verma, I. M. and Somia, N. (1997) Gene therapy—Promises, problems and prospects. *Nature* **389,** 239–242.
5. Craig, N. L. (1995) Unity in transposition reactions. *Science* **270,** 253–254.

6. Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., et al. (1999) A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev.* **13,** 2713–2724.
7. Burns, J. C., Friedmann, T., Driever, W., Burrascano, M., and Yee, J. K. (1993) Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc. Natl. Acad. Sci. USA* **90,** 8033–8037.
8. Lin, S., Gaiano, N., Culp, P., Burns, J. C., Friedmann, T., Yee, J. K., et al. (1994) Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. *Science* **265,** 666–669.
9. Gaiano, N., Amsterdam, A., Kawakami, K., Allende, M., Becker, T., and Hopkins, N. (1996) Insertional mutagenesis and rapid cloning of essential genes in zebrafish. *Nature* **383,** 829–832.
10. Koga, A., Suzuki, M., Inagaki, H., Bessho, Y., and Hori, H. (1996) Transposable element in fish. *Nature* **383,** 30.
11. Koga, A., Shimada, A., Shima, A., Sakaizumi, M., Tachida, H., and Hori, H. (2000) Evidence for recent invasion of the medaka fish genome by the *Tol*2 transposable element. *Genetics* **155,** 273–281.
12. Kawakami, K., Shima, A., and Kawakami, N. (2000) Identification of a functional transposase of the *Tol*2 element, an *Ac*-like element from the japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc. Natl. Acad. Sci. USA* **97,** 11,403–11,408.
13. Plasterk, R. H., Izsvák, Z., and Ivics, Z. (1999) Resident aliens: the Tc*1*/*mariner* superfamily of transposable elements. *Trends Genet.* **15,** 326–332.
14. Capy, P., Vitalis, R., Langin, T., Higuet, D., and Bazin, C. (1996) Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? *J. Mol. Evol.* **42,** 359–368.
15. Robertson, H. M. (1995) The Tc*1-mariner* superfamily of transposons in animals. *J. Insect Physiol.* **41,** 99–105.
16. Rio, D. C., Barnes, G., Laski, F. A., Rine, J., and Rubin, G. M. (1988) Evidence for *Drosophila P* element transposase activity in mammalian cells and yeast. *J. Mol. Biol.* **200,** 411–415.
17. Kidwell, M. G. (1992) Horizontal transfer of *P* elements and other short inverted repeat transposons. *Genetica* **86,** 275–286.
18. Ivics, Z., Izsvák, Z., Minter, A., and Hackett, P. B. (1996) Identification of functional domains and evolution of Tc*1*-like transposable elements. *Proc. Natl. Acad. Sci. USA* **93,** 5008–5013.
19. Vos, J. C., De Baere, I., and Plasterk, R. H. (1996) Transposase is the only nematode protein required for *in vitro* transposition of Tc*1*. *Genes Dev.* **10,** 755–761.
20. Lampe, D. J., Churchill, M. E., and Robertson, H. M. (1996) A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. *EMBO J.* **15,** 5470–5479.
21. Schouten, G. J., van Luenen, H. G., Verra, N. C., Valerio, D., and Plasterk, R. H. (1998) Transposon Tc*1* of the nematode *Caenorhabditis elegans* jumps in human cells. *Nucleic Acids Res.* **26,** 3013–3017.

22. Zhang, L., Sankar, U., Lampe, D. J., Robertson, H. M., and Graham, F. L. (1998) The *Himar*1 *mariner* transposase cloned in a recombinant adenovirus vector is functional in mammalian cells. *Nucleic Acids Res.* **26,** 3687–3693.
23. Fischer, S. E. J., Wienholds, E., and Plasterk, R. H. A. (2001) Regulated transposition of a fish transposon in the mouse germ line. *Proc. Natl. Acad. Sci. USA* **98,** 6759–6764.
24. Klinakis, A. G., Zagoraiou, L., Vassilatis, D. K., and Savakis, C. (2000) Genome-wide insertional mutagenesis in human cells by the *Drosophila* mobile element *Minos*. *EMBO Rep.* **1,** 416–421.
25. Fadool, J. M., Hartl, D. L., and Dowling, J. E. (1998) Transposition of the *mariner* element from *Drosophila mauritiana* in zebrafish. *Proc. Natl. Acad. Sci. USA* **95,** 5182–5186.
26. Raz, E., van Luenen, H. G., Schaerringer, B., Plasterk, R. H. A., and Driever, W. (1998) Transposition of the nematode *Caenorhabditis elegans* Tc*3* element in the zebrafish *Danio rerio*. *Curr. Biol.* **8,** 82–88.
27. Sherman, A., Dawson, A., Mather, C., Gilhooley, H., Li, Y., Mitchell, R., et al. (1998) Transposition of the *Drosophila* element *mariner* into the chicken germ line. *Nat. Biotechnol.* **16,** 1050–1053.
28. Radice, A. D., Bugaj, B., Fitch, D. H., and Emmons, S. W. (1994) Widespread occurrence of the Tc*1* transposon family: Tc*1*-like transposons from teleost fish. *Mol. Gen. Genet.* **244,** 606–612.
29. Goodier, J. L. and Davidson, W. S. (1994) Tc*1* transposon-like sequences are widely distributed in salmonids. *J. Mol. Biol.* **241,** 26–34.
30. Lam, W. L., Seo, P., Robison, K., Virk, S., and Gilbert, W. (1996) Discovery of amphibian Tc*1*-like transposon families. *J. Mol. Biol.* **257,** 359–366.
31. Lam, W. L., Lee, T. S., and Gilbert, W. (1996) Active transposition in zebrafish. *Proc. Natl. Acad. Sci. USA* **93,** 10,870–10,875.
32. Auge-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M. H., Meunier-Rotival, M., and Periquet, G. (1995) Human and other mammalian genomes contain transposons of the *mariner* family. *FEBS Lett.* **368,** 541–546.
33. Morgan, G. T. (1995) Identification in the human genome of mobile elements spread by DNA-mediated transposition. *J. Mol. Biol.* **254,** 1–5.
34. Oosumi, T., Belknap, W. R., and Garlick, B. (1995) *Mariner* transposons in humans. *Nature* **378,** 672.
35. Ivics, Z., Hackett, P. B., Plasterk, R. H., and Izsvák, Z. (1997) Molecular reconstruction of *Sleeping Beauty*, a Tc*1*-like transposon from fish, and its transposition in human cells. *Cell* **91,** 501–510.
36. Doak, T. G., Doerder, F. P., Jahn, C. L., and Herrick, G. (1994) A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. *Proc. Natl. Acad. Sci. USA* **91,** 942–946.
37. Izsvák, Z., Ivics, Z., and Hackett, P. B. (1995) Characterization of a Tc*1*-like transposable element in zebrafish (*Danio rerio*). *Mol. Gen. Genet.* **247,** 312–322.
38. Izsvák, Z., Ivics, Z., and Plasterk, R. H. (2000) *Sleeping Beauty*, a wide host-range transposon vector for genetic transformation in vertebrates. *J. Mol. Biol.* **302,** 93–102.

39. Yant, S. R., Meuse, L., Chiu, W., Ivics, Z., Izsvak, Z., and Kay, M. A. (2000) Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nat. Genet.* **25,** 35–41.
40. Luo, G., Ivics, Z., Izsvák, Z., and Bradley, A. (1998) Chromosomal transposition of a Tc*1/mariner*-like element in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **95,** 10,769–10,773.
41. Montini, E., Held, P. K., Noll, M., Morcinek, N., Al-Dhalimy, M., Finegold, M., et al. (2002) *In Vivo* Correction of Murine Tyrosinemia Type I by DNA-Mediated Transposition. *Mol. Ther.* **6,** 759–769.
42. Yant, S. R., Ehrhardt, A., Mikkelsen, J. G., Meuse, L., Pham, T., and Kay, M. A. (2002) Transposition from a gutless adeno-transposon vector stabilizes transgene expression *in vivo*. *Nat. Biotechnol.* **20,** 999–1005.
43. Zagoraiou, L., Drabek, D., Alexaki, S., Guy, J. A., Klinakis, A. G., Langeveld, A., et al. (2001) *In vivo* transposition of *Minos*, a *Drosophila* mobile element, in mammalian tissues. *Proc. Natl. Acad. Sci. USA* **98,** 11,474–11,478.
44. Nasevicius, A. and Ekker, S. C. (2000) Effective targeted gene "knockdown" in zebrafish. *Nat. Genet.* **26,** 216–220.
45. Lavitrano, M., Camaioni, A., Fazio, V. M., Dolci, S., Farace, M. G., and Spadafora, C. (1989) Sperm cells as vectors for introducing foreign DNA into eggs: Genetic transformation of mice. *Cell* **57,** 717–723.
46. Müller, F., Ivics, Z., Erdélyi, F., Papp, T., Váradi, L., Horváth, L., et al. (1992) Introducing foreign genes into fish eggs by electroporated sperm as a carrier. *Mol. Marine Biol. Biotech.* **1,** 276–281.
47. Kroll, K. L. and Amaya, E. (1996) Transgenic *Xenopus* embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation. *Development* **122,** 3173–3183.
48. Dupuy, A. J., Fritz, S., and Largaespada, D. A. (2001) Transposition and gene disruption in the male germline of the mouse. *Genesis* **30,** 82–88.
49. Horie, K., Kuroiwa, A., Ikawa, M., Okabe, M., Kondoh, G., Matsuda, Y., et al. (2001) Efficient chromosomal transposition of a Tc*1/mariner*-like transposon *Sleeping Beauty* in mice. *Proc. Natl. Acad. Sci. USA* **98,** 9191–9196.
50. Scherdin, U., Rhodes, K., and Breindl, M. (1990) Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* **64,** 907–912.
51. Bellen, H. J., O'Kane, C. J., Wilson, C., Grossniklaus, U., Pearson, R. K., and Gehring, W. J. (1989) *P*-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Genes Dev.* **3,** 1288–1300.
52. Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T., and Rubin, G. M. (1995) Gene disruptions using *P* transposable elements: an integral component of the *Drosophila* genome project. *Proc. Natl. Acad. Sci. USA* **92,** 10,824–10,830.
53. Vigdal, T. J., Kaufman, C. D., Izsvák, Z., Voytas, D. F., and Ivics, Z. (2002) Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc*1/mariner* transposable elements. *J. Mol. Biol.* **323,** 441–452.
54. Bronchain, O. J., Hartley, K. O., and Amaya, E. (1999) A gene trap approach in *Xenopus*. *Curr. Biol.* **9,** 1195–1198.