Horizontal Transmission, Vertical Inactivation, and Stochastic Loss of *Mariner*-like Transposable Elements

Allan R. Lohe, Etsuko N. Moriyama,[1] *Dan-Anders Lidholm, and Daniel L. Hartl*

Department of Organismic and Evolutionary Biology, Harvard University

Horizontal transmission has been well documented as a major mechanism for the dissemination of *mariner*-like elements (MLEs) among species. Less well understood are mechanisms that limit vertical transmission of MLEs resulting in the "spotty" or discontinuous distribution observed in closely related species. In this article we present evidence that the genome of the common ancestor of the melanogaster species subgroup of *Drosophila* contained an MLE related to the mellifera (honey bee) subfamily. Horizontal transmission, approximately 3–10 MYA, is strongly suggested by the observation that the sequence of the MLE in *Drosophila erecta* is 97% identical in nucleotide sequence with that of an MLE in the cat flea, *Ctenocephalides felis*. The *D. erecta* MLE has a spotty distribution among species in the melanogaster subgroup. The element has a high copy number in *D. erecta* and *D. orena*, a moderate copy number in *D. teissieri* and *D. yakuba*, and was apparently lost ("stochastic loss") in the lineage leading to *D. melanogaster, D. simulans, D. mauritiana,* and *D. sechellia*. In *D. erecta*, most copies are concentrated in the heterochromatin. Two copies from *D. erecta*, denoted De12 and De19, were cloned and sequenced, and they appear to be nonfunctional ("vertical inactivation"). It therefore appears that the predominant mode of MLE evolution is vertical inactivation and stochastic loss balanced against occasional reinvasion of lineages by horizontal transmission.

Introduction

Mariner-like elements (MLEs) are widely distributed among insects and other invertebrates (Lidholm et al. 1991; Robertson 1993; Robertson and MacLeod 1993; Capy et al. 1994). Functional elements can integrate spontaneously into the germline (Garza et al. 1991) and can support germline transformation with exogenous DNA (Lidholm et al. 1993). MLEs have therefore attracted considerable interest because of their potential for genetic manipulation of insect vectors of human disease and agricultural pests (Kidwell 1993).

MLEs also present important issues from an evolutionary point of view. Although MLEs are widespread, their distribution is "spotty," which means that, among closely related species, a particular type of MLE may be found in some species but not in others; in contrast, distantly related species sometimes share closely related MLEs. Furthermore, a particular type of MLE may be present in thousands of copies in one species but in only a small number of copies in another species. Although several processes contribute to determining the patterns of distribution and abundance of MLEs (as well as other transposable elements) among species, understanding the relative importance of these processes is a major challenge. Three processes that clearly play a role are (1) horizontal transmission, the ability of MLEs to be transferred from a host species into the germline of a different, reproductively isolated species (the vectors and mechanisms of horizontal transmission are as yet unknown); (2) mutation, which may destroy the protein-coding function of an MLE ("mutational inactivation") or impair its ability to transpose, and (3) stochastic loss, the elimination of an active or inactive type of MLE from the genome as a result of a random genetic drift (possibly also with a contribution from natural selection depending on the extent to which presence of the MLE is deleterious).

Horizontal transmission of MLEs in nature was first suggested by the finding that elements in *Zaprionus* were very closely related to those in *Drosophila* (Maruyama and Hartl 1991*a;* Lawrence and Hartl 1992). Several additional cases of probable horizontal transmission were discovered by Robertson (1993) and Robertson and MacLeod (1993) in a study of a large number of insect genomes using primers for the polymerase chain reaction

1. Present address: Department of Biology, Yale University.

Key words: *Drosophila erecta*, horizontal transmission, *mariner*-like element (MLE), stochastic loss, transposable element.

Address for correspondence and reprints: Daniel L. Hartl, Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138.

Mol. Biol. Evol. 12(1):62–72. 1995.
© 1995 by The University of Chicago. All rights reserved.
0737-4038/95/1201-0006$02.00

Downloaded from http://mbe.oxfordjournals.org/ at OUP site access on October 29, 2012
Downloaded from https://academic.oup.com/mbe/article/12/1/62/995568 by guest on 23 April 2024

(PCR) complementary to conserved sequences in MLEs in *Drosophila mauritiana* (Jacobson et al. 1986) and *Hyalophora cecropia* (Lidholm et al. 1991). MLEs can be classified into several distinct subfamilies according to similarities in nucleotide sequence. MLEs in different subfamilies are typically 40%–56% identical at the nucleotide level (Robertson and MacLeod 1993). The major subfamilies all appear to be widely distributed among species, and any particular species may contain MLEs from two or more different subfamilies (Robertson 1993).

Once introduced into a genome, MLEs can persist through multiple speciation events (Maruyama and Hartl 1991*b*). Why, then, are MLEs not present in all insects? Granted that some elements may not amplify with any particular pair of PCR primers, nevertheless only 15.6% (63 of 404) species of insects and related arthropods supported PCR amplification (Robertson and MacLeod 1993). One hypothesis is that MLEs actually have a restricted host range owing to host factors required for transposition; however, this hypothesis is not supported by the wide range of insect orders in which MLEs are found (Robertson and MacLeod 1993). A second hypothesis is that vertical transmission of MLEs is limited by processes resulting in loss of the element ("stochastic loss") as well as by mutational inactivation ("vertical inactivation"). From the standpoint of the host species, the initial inactivating mutation in an MLE may be selectively neutral or perhaps even favorable inasmuch as natural selection may act to minimize the harmful mutagenic effects of transposition. Subsequent mutations in an already-inactivated MLE are presumably selectively neutral. Stochastic loss and vertical inactivation are distinct processes. Stochastic loss results from random genetic drift, whereas vertical inactivation results from mutation.

The role of stochastic loss in the evolutionary dynamics of MLEs is supported by the spotty distribution of the mauritiana subfamily of MLEs among eight closely related species comprising the melanogaster species subgroup of *Drosophila:* it is found in *D. mauritiana, D. simulans, D. sechellia, D. yakuba,* and *D. teissieri* but not found in *D. melanogaster, D. erecta,* and *D. orena* (Capy et al. 1990, 1991; Maruyama and Hartl 1991*b*). The role of vertical inactivation in the evolutionary dynamics of MLEs is supported by the apparent tendency of MLEs that are mutationally inactivated to become prevalent in the genome. Any particular copy of *mariner* may be autonomous (capable of transposing on its own) or nonautonomous (capable of transposing only in the presence of autonomous copies). A nonautonomous element contains one or more inactivating mutations in its open reading frame but may nevertheless be capable of transposition owing to *trans* complementation by active transposase from autonomous copies present in the same genome. The vast majority of MLEs so far discovered contain multiple inactivating mutations (Capy et al. 1991; Maruyama et al. 1991*a*, 1991*b;* Robertson 1993; Robertson and MacLeod 1993): they are essentially pseudogene versions of the autonomous element. At this time, the only MLEs demonstrated to be autonomous are a restricted subset of mauritiana subfamily elements, all closely related to the functional element *Mos1* (Medhora et al. 1991), which are present in *D. mauritiana* and *D. simulans* (Capy et al. 1990, 1992; Maruyama et al. 1991*a,* 1991*b*).

MLEs from different subfamilies can be found in closely related species and often in the same genome (Robertson 1993). In order to evaluate the relative roles of horizontal transmission, vertical inactivation, and stochastic loss in determining the distribution and abundance of MLEs, we have focused on the melanogaster species subgroup because of the rich background of genetic and evolutionary studies in these species. Using a PCR strategy to search for MLE subfamilies in the melanogaster subgroup, we have found that *D. erecta* contains numerous (>100 copies) of an MLE in the mellifera (honey bee) subfamily that is 97% identical in nucleotide sequence to an element from the cat flea *Ctenocephalides felis* (Pulicidae:Siphonaptera) (Robertson and MacLeod 1993). The close relationship strongly suggests horizontal transmission between the ancestors of *C. felis* and *D. erecta.* Furthermore, the species distribution of the mellifera-related element appears to be determined by stochastic loss and vertical inactivation. Stochastic loss is evidenced by the absence of mellifera-related elements in *D. melanogaster* and species of the *D. simulans* complex (*D. simulans, D. mauritiana,* and *D. sechellia*). Vertical inactivation is evidenced by the findings that, in *D. erecta,* most of the mellifera-related elements are located in the pericentromeric heterochromatin and/or the Y chromosome and that two sequenced copies both contain inactivating mutations. Hence, the predominant mode of evolution of MLEs appears to be stochastic loss and vertical inactivation balanced against occasional reinvasion of lineages by horizontal transmission.

Material and Methods

Identification of the *Drosophila erecta* MLE

Oligonucleotide primers corresponding to conserved regions between the mauritiana element defined by the *white-peach* (w^{pch}) allele (Jacobson et al. 1986) and the *Caenorhabditis elegans* element (GenBank accession number M98552, cited in Garcia-Fernàndez et al. [1993]) were synthesized and used to amplify DNA from various sources. Primer "628F" has the sequence

Downloaded from https://academic.oup.com/mbe/article/12/1/62/999568 by guest on 23 April 2024

5′-GTTACTGGCGATGAAAAATGG-3′ and anneals with the *mariner* sequences between nucleotides 628 and 648 (numbered as in Jacobson et al. [1986]); primer "930R" has the sequence 5′-TGNTGNAGCGTTG TCNTGNAA-3′ and anneals with the *mariner* sequences between nucleotides 930 and 910. Conditions for the PCR were 40 cycles of denaturation at 94°C for 1 min, annealing at 60°C for 3 min, and extension at 72°C for 2 min. A 10-ng sample of DNA from *D. erecta* yielded a prominent PCR product of the predicted size of 300 base pairs (bp). This product was purified from the agarose gel with NA45 (Schleicher & Schuell) and cloned into the *Sma*I cloning site of the pBluescript SK+ vector (Stratagene). The resulting clone was designated pDe1210. Although the expected size of the PCR product is 303 bp, the size of the insert in pDe1210 was only 142 bp, owing to a deletion of 161 bp extending from the 930R primer.

Isolation of Full-Length Elements

Plasmid pDe1210 was used to probe a lambda library from *D. erecta* kindly provided by D. Hickey and P. Foster. Hybridization and plaque purification procedures were essentially as described in Sambrook et al. (1989). Restriction digests and DNA hybridizations of independent clones indicated that the putative MLEs were each contained in an *Eco*RI restriction fragment of various size. These fragments were subcloned into pBluescript to yield the plasmids pDe12 (containing a 3-kilobase-pair (kb) *Eco*RI fragment) and pDe19 (containing a 4-kb *Eco*RI fragment). The MLEs in pDe12 and pDe19 were sequenced on both strands by primer walking using as an entry point a conserved oligonucleotide derived from the w^{pch} *mariner* sequence. DNA sequencing was carried out with the dideoxy chain termination method (Sanger et al. 1977) using either Sequenase (United States Biochemical) or *Taq* polymerase (Perkin Elmer) cycle sequencing.

Southern Blots and In Situ Hybridization

Southern-transfer hybridization of genomic DNA (Southern 1975) was carried out essentially as described in Garza et al. (1991). Hybridization in situ to polytene chromosomes was carried out with probe DNA labeled with biotinylated 11-dUTP and detected with the horseradish peroxidase system (ENZO), essentially as described in Lozovskaya et al. (1993).

Phylogenetic Analyses

PCR-amplified coding regions from each of 25 MLEs (Robertson 1993) were taken from GenBank 79.0 to enable comparison of the *D. erecta* elements with the major subfamilies of MLEs. Sequences were chosen to have few or no frame shifts and/or deletions relative to *Mos1*. The MLEs were aligned initially using CLUSTAL V (Higgins et al. 1992) and then adjusted manually to maximize amino acid identity, introducing a minimal number of frame shifts as necessary. Nucleotide sequence alignments in the coding regions were inferred from those of the amino acid alignments. The amino acid alignment is virtually the same as that in Robertson (1993), with a few minor differences in the positions of gaps. The length of the aligned sequences including gaps is 477 bp (159 amino acids). Phylogenetic trees were inferred with maximum parsimony (PAUP, version 3.0s Swofford and Olsen 1990; Swofford 1991) and neighbor joining (Saitou and Nei 1987). For the amino acid alignment, 100 bootstrap replications of maximum parsimony were carried out with heuristic search, 10 random additions of sequences, the tree bisection-reconnection method for branch swapping, and the PROTPARS step matrix for weighing amino acid differences. For neighbor joining, 1,000 bootstrap replications were carried out with the amino acid alignment as well as with the nucleotide alignment using PHYLIP 3.5c (Felsenstein 1993). The numbers of amino acid substitutions were estimated by Kimura's $\ln(1-p-0.2p^2)$ method (Kimura 1983). The numbers of nucleotide substitutions were estimated by the two-parameter method (Kimura 1980). As a check on the accuracy of the two-parameter method for these data, we also used a six-parameter method (Gojobori et al. 1982) and an equal-input model of nucleotide substitution (Tajima and Nei 1984). Phylogenetic trees constructed with any of the three methods showed the same topology; hence we applied the two-parameter method for the bootstrapping.

Results

The 142-bp insert in pDe1210 was identified as an MLE because it showed 51.4% nucleotide identity to a region of the *mariner* element present in the w^{pch} allele of *Drosophila mauritiana* (Jacobson et al. 1986). The similarity included one exact match extending across 1… bp. The pDe1210 plasmid was therefore used to screen a lambda library of *D. erecta* in an attempt to isolate full-length clones.

Full-Length Elements from *D. erecta*

Use of pDe1210 for plaque hybridization of a library from *D. erecta* yielded a large number of positive clones. Restriction mapping identified two clones in which the hybridizing sequence was contained in an *Eco*RI restriction fragment differing in size. Plasmids pDe12 and pDe19 were created by subcloning the *Eco*RI fragments, and the MLEs in the subclones, designated

Downloaded from https://academic.oup.com/mbe/article/12/1/62/988598 by guest on 23 April 2024

```
      1        10         20         30         40         50         60         70         80         90         100        110        120
Mos1  -SS-VPNKEQT-TV-I-C-HLK-T-A-SH-MLVEAF-EQVPTVK-C-R--Q--KS-DF-VD--EH-KP-KRYEDAELQALLDE-DAQTQKQL-EQLE--Q-AVS-R-REM-KIQ-VGR--
De19  ME*F**TNAEIRAILKFSFVKGKSARETFREINGVLGDGTLSLRTAEEWFRRFRAGENDTMDKPAGGRPVTTNTDQIMEYIELDRHVASRDIAQEMGVSHQTILNHLQKAGYKKKLDVWV

               130        140        150        160        170        180        190        200        210        220        230        240
Mos1  --E-NERQMER-K-T-EI--S-YKRKS--H-I--------FFV-P---K-YVDP-QPATST-R-NRFGK-TM------QS-V-YY---KP-E-V-TAR-Q---IN-NR-LQR----YQK-
De19  PHDLTQKNLLDRINACDMLLKRNELDPFLKRMVTGDEKWITYDNIKRKRSWSKAGESSQTVAKPGLTARKVLLCVWWDWKGIIHNELLPYGQTLNSTTYCQQLD#LKQAIDQKRPELANR
10.6      ------------------------------------S-------------------------M--E---------------Y-------M----I--E---R-------------G

               250        260        270        280        290        300        310        320        330        340
Mos1  QHR-I-LH---PS--ARAV-DT-ET-N----P-AA----L--------A--GH--AEQRFD-Y-SVK---D-W--A-DDE--WR--H---E--EKCVASD-K-FE-
De19  KG*VVFHQDNARPHTSLMTRQKLRELGWEVLSHPPYSPDIAPSDYHLFLSMANALGGVKLNSKEACEKWLSEFFANKEGGFYEGGIMKLPSRWKQIIEQNGAYLN.
10.6  ----------------------------------L-
```

FIG. 1.—Inferred amino acid sequences of the putative transposase of the MLEs De19, *Mos1*, and cat flea 10.6 (GenBank accession number L10470), aligned with respect to the open reading frame of *Mos1*. Dashes indicate identity with De19. The symbols # and * indicate one- and three-nucleotide indels, respectively. The period at the end of the De19 sequence indicates the stop codon.

De12 and De19, respectively, were sequenced. Both De12 and De19 are flanked by imperfect inverted terminal repeats (ITRs) of 34 bp (with three mismatches) and an apparent duplication of TA at the site of insertion. The total length of De12 is 1,262 bp, and that of De19 is 1,282 bp, with most of the difference due to three small deletions in De12. When aligned to maximize identity, De12 and De19 are 98% identical in nucleotide sequence. Both copies are flanked at the 5′ end by the sequence 5′-ATATA-3′ and at the 3′ end by the sequence 5′-TATT-3′; the underlined TA is the duplication characteristic of *mariner* insertions. Outside these flanking regions the sequences of genomic DNA in the clones have no resemblance.

De12 and De19 are quite different from *Mos1* from *D. mauritiana:* across the entire element, the nucleotide identity is only 50%; and in the coding region (fig. 1) the amino acid identity is 40%. In addition, both De12 and De19 contain insertions or deletions (indels) that shift the reading frame relative to *Mos1* and are therefore presumably nonfunctional. Figure 1 also includes the amino acid sequence of a 447-bp open reading frame (ORF) of an MLE designated "10.6" amplified by PCR from the cat flea *Ctenocephalides felis* (Robertson and MacLeod 1993). Across the sequenced region, 10.6 exhibits 94% amino acid identity with De19.

Copy Number and Genomic Location in *D. erecta*

The large number of positive plaques observed in the library screening suggested a high copy number. This expectation was borne out by Southern hybridization of *D. erecta* genomic DNA probed with a PCR product from pDe19. Each of 20 different restriction enzymes cleaved genomic DNA into a large number of hybridizing bands, conservatively estimated at 100 or more (data not shown). Furthermore, most copies of the element are located in the pericentromeric heterochromatin and/or the Y chromosome, as indicated by the heavy labeling of the chromocenter of polytene salivary gland chromosomes (fig. 2). The labeling pattern also consistently includes a small region of the nucleolus, two heavily labeled sites in chromosome 4 (plus several weaker sites), and approximately 15 weak euchromatic sites (not readily apparent in the black-and-white micrograph in fig. 2).

Distribution among Selected *Drosophila* Species

Southern hybridization experiments were carried out with genomic DNA from a sample of *Drosophila*

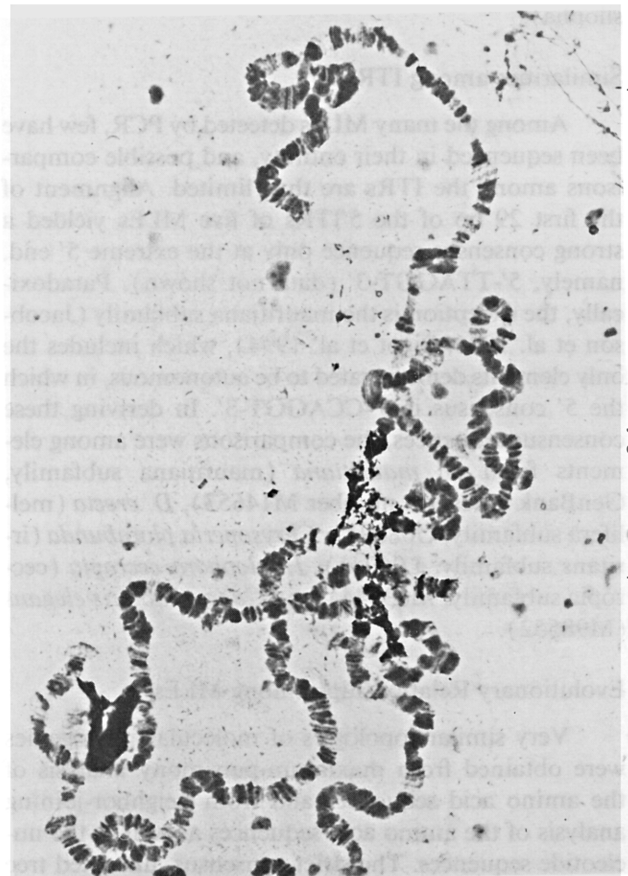

FIG. 2.—In situ hybridization of pDe19 with the polytene salivary gland chromosomes of *D. erecta*. The majority of copies are concentrated in the aggregate of heterochromatin forming the chromocenter. There are also approximately 15 discrete sites of hybridization in the euchromatin, only a few of which are apparent without color.

Downloaded from https://academic.oup.com/mbe/article/12/1/62/999568 by guest on 23 April 2024

species and probed with a full-length PCR product from pDe19. The genomic DNA was digested with *Rsa*I, which cleaves within the De19 sequence at positions 57 and 289. Extensive hybridization, virtually unresolvable into discrete bands, was observed with *D. erecta* (fig. 3). The closely related species *D. orena* also exhibited a large number of hybridizing fragments, but the somewhat more distantly related *D. yakuba* and *D. teissieri* exhibited approximately 10 fragments each. These results are consistent with the phylogenetic relationships because all four species are members of the melanogaster species subgroup. It is interesting that the De19 element was not detected in the phylad of the subgroup consisting of *D. melanogaster, D. simulans, D. mauritiana,* and *D. sechellia.* However, strong hybridization was also observed with *D. ananassae* and somewhat less strong with *D. varians,* both members of the ananassae species subgroup (subgenus Sophophora). No hybridization was detected with DNA from *D. pseudoobscura* (obscura group, subgenus Sophophora), *D. hydei* or *D. virilis* (subgenus Drosophila), or *D. busckii* (subgenus Dorsilopha).

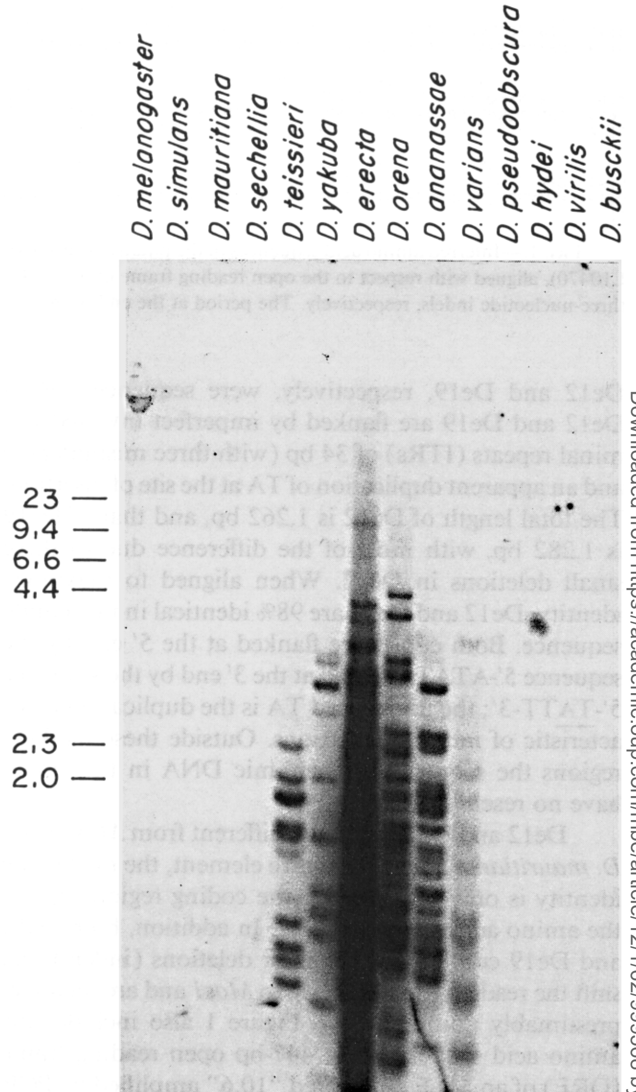

FIG. 3.—Hybridization patterns of genomic DNA from various *Drosophila* species digested with *Rsa*I and probed with a full-length PCR product from pDe19. Numbers are molecular weights of size markers in kb.

Similarities among ITRs

Among the many MLEs detected by PCR, few have been sequenced in their entirety, and possible comparisons among the ITRs are thus limited. Alignment of the first 29 bp of the 5′ITRs of five MLEs yielded a strong consensus sequence only at the extreme 5′ end, namely, 5′-TTAGGT-3′ (data not shown). Paradoxically, the exception is the mauritiana subfamily (Jacobson et al. 1986; Bigot et al. 1994), which includes the only elements demonstrated to be autonomous, in which the 5′ consensus is 5′-CCAGGT-3′. In deriving these consensus sequences, the comparisons were among elements from *D. mauritiana* (mauritiana subfamily, GenBank accession number M14653), *D. erecta* (mellifera subfamily, U08094), *Chrysoperia plorabunda* (irritans subfamily, L06041), *Hyalophora cecropia* (cecropia subfamily, M63844), and *Caenorhabditis elegans* (M98552).

Evolutionary Relationships among MLEs

Very similar topologies of molecular phylogenies were obtained from maximum-parsimony analysis of the amino acid sequences and from neighbor-joining analysis of the amino acid sequences as well as the nucleotide sequences. The strict-consensus, unrooted tree is shown in figure 4. Each of the groupings is strongly supported, although there are a few sequences listed at the bottom whose affinity remains uncertain from this analysis. The brackets indicating the major groupings in the tree correspond to four of the principal subfamilies of MLEs: mauritiana, mellifera, cecropia, and irritans (Robertson and MacLeod 1993). (The capitata subfamily is represented in fig. 4 only by the sequence Earwig 5.10.) In the consensus tree, the De19 element is clearly grouped with the mellifera subfamily and not with the mauritiana subfamily.

Results of a more detailed phylogenetic analysis of 10 MLEs in the mellifera subfamily are shown in figure 5, which is a strict-consensus rooted tree using *Mos1* and *H. cecropia* 1.3 as outgroups. The analysis was carried out as in figure 4. Both De12 and De19 group decisively with an element from *Ctenocephalides felis* (cat flea 10.6), reflecting the close nucleotide-sequence iden-

Downloaded from https://academic.oup.com/mbe/article/12/1/62/999568 by guest on 23 April 2024

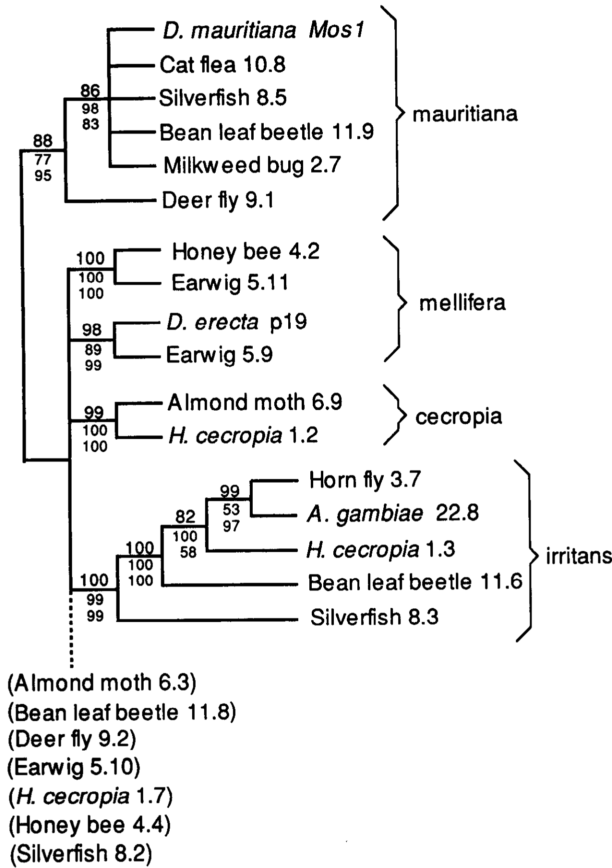

FIG. 4.—Strict-consensus, unrooted tree of selected MLEs. The brackets indicate major subfamilies of MLEs (Robertson and MacLeod 1993). The set of numbers labeling each node are the bootstrap support values (%), from top to bottom, maximum parsimony analysis of amino acid sequences (100 replications), neighbor-joining analysis of amino acid sequences (1,000 replications), and neighbor-joining analysis of nucleotide sequences (1,000 replications).

tity (96% identity in nucleotide sequence between 10.6 and De12, 98% identity in nucleotide sequence between 10.6 and De19). Closer comparison of the sequences also indicates no preference for nucleotide differences to be concentrated in the third position of codons. Indeed, the amino acid identity between the sequences is less than the nucleotide identity (93% amino acid identity between 10.6 and De12, 94% amino acid identity between 10.6 and De19). These comparisons suggest that there are few or no selective constraints on the first and second nucleotides in codons, supporting the inference that De12 and De19 (and possibly also 10.6) are mutationally defective, nonfunctional elements.

Evidence for Horizontal Transmission

The remarkably close similarity between De12, De19, and cat flea 10.6 is suggestive of horizontal transmission, but it is desirable to compare the MLEs with nontransposable sequences from the same or closely related species. The gene for the alpha subunit of the $Na^+ + K^+$ ATPase has been sequenced in both *C. felis* and *D. melanogaster* (Lebovitz et al. 1989; Reeves and Yamanaka 1993). Because the amino acid sequence of this subunit is highly conserved in evolution, the appropriate comparison with the MLEs is the divergence between the third positions of codons, which are expected to be much less constrained. For the sodium-pump gene, 39% of the nucleotides at the third position are identical (N = 1,034); the corresponding identities at the third positions of codons in the sequenced part of the coding region are 96% for 10.6 versus De12 (N = 146) and 99% for 10.6 versus De19 (N = 149). In comparison with the sodium-pump gene, the MLEs are either strongly constrained or have been diverging for a much shorter time.

Discussion

In order to evaluate the evolutionary forces that determine the species distribution of MLEs, it is necessary to study closely related species that have diverged from a common ancestor containing the MLE. We have been particularly interested in MLEs present in the melanogaster species subgroup in view of the extensive background of genetic and evolutionary studies on these species. The data presented in this article support a model in which the species distribution of MLEs is de-

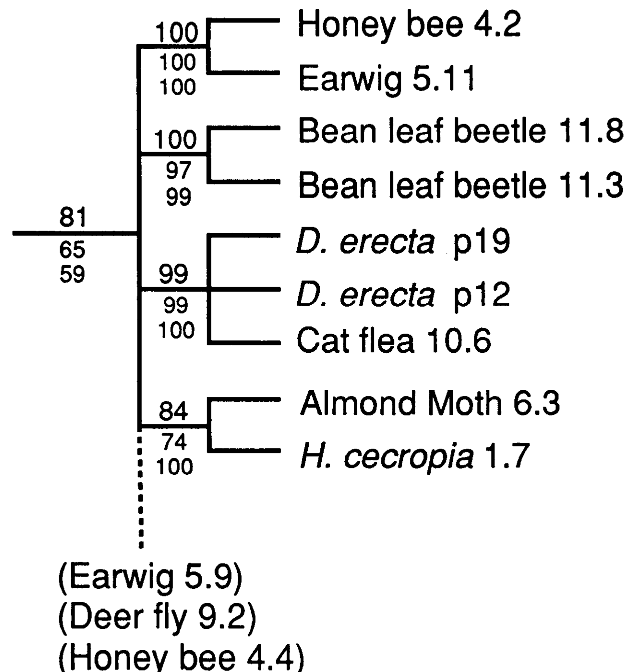

FIG. 5.—Strict-consensus, rooted tree of MLEs in the mellifera subfamily as determined when the *Mos1* and *H. cecropia* 1.3 elements serve as outgroups. The numbers labeling each node are as in fig. 4.

Downloaded from https://academic.oup.com/mbe/article/12/1/62/999568 by guest on 23 April 2024

termined by horizontal transmission balanced against vertical inactivation and stochastic loss.

MLEs from *Drosophila erecta*

The De19 element is 1,282 bp long, very similar to the 1,286-bp length of the *Mos1* element (Medhora et al. 1991). The MLEs in *D. erecta* have ITRs of 34 bp with three mismatches; in comparison, the ITRs in the *Mos1* element have 28 bp with four mismatches. Despite the overall similarity in size, the De12 and De19 elements are only approximately 50% identical in nucleotide sequence with *Mos1*. The 142-bp fragment of another *D. erecta* MLE present in the plasmid pDe1210, which was used to probe the *D. erecta* library, is 84% identical in nucleotide sequence with De19. Thus, the MLE represented in pDe1210 is a member of the same subfamily as De12 and De19 but significantly more divergent from De19 than is De12.

Diversity among Subfamilies

The approximate 50% nucleotide sequence divergence between the *D. erecta* elements (mellifera subfamily) and *Mos1* (mauritiana subfamily) is typical of that between MLEs from different subfamilies (Robertson and MacLeod 1993). Thus far, *Mos1* and closely related elements are the only MLEs demonstrated to function autonomously; all the other sequenced elements, including De12 and De19, are probably nonfunctional, owing to frame shifts or translational-termination mutations. The sequence divergence therefore results in part from relaxed selective constraints and may obscure sequence motifs that may be conserved among autonomous elements in different subfamilies. Nevertheless, a few conserved motifs are notable, such as the conserved acidic sequence motif in the carboxyl part of the putative transposase consisting of Asp (D) and Glu (E) residues fitting the pattern $DEX_{85-95}DX_{35-45}E$, in which the subscripted X's indicate the number of residues between the conserved amino acids (Doak et al. 1994). The "D35E" motif has been implicated as part of the polynucleotidyl transfer center capable of catalyzing part of the transposition reaction (Bushman et al. 1993), which suggests that MLEs may share a common mechanism of transposition with such elements as IS*3* from *Escherichia coli,* IS*630* from *Shigella sonnei,* Tc*1* from *Caenorhabditis elegans, Bari* from *D. melanogaster,* and *Minos* from *D. hydei* (Doak et al. 1994). There are also apparently conserved motifs in the ITRs, particularly at the extreme 5′ and 3′ ends, which may be targets for DNA binding by the transposase.

Species Distribution

In the melanogaster species subgroup, elements from the mellifera subfamily are present in *D. erecta* and *D. orena* (the "erecta complex") and *D. yakuba* and *D. teissieri* (the "yakuba complex") but not in *D. melanogaster, D. simulans, D. mauritiana,* and *D. sechellia* (the "melanogaster complex") (fig. 3). Evidence from various sources, including biogeography and molecular biology (Jeffs et al. 1994), are consistent in indicating that the erecta complex split off the ancestral lineage early (estimated at 2–6 MYA in Lemeunier et al. 1986) and that the melanogaster complex split off late (estimated at 0.8–3 MYA in Lemeunier et al. 1986). The simplest hypothesis to explain the observed species distribution of the MLE is to assume that the element was present in the common ancestor of the species subgroup and retained in the lineages leading to the erecta complex and the yakuba complex but eliminated by stochastic loss in the lineage leading to the melanogaster complex. This hypothesis would imply that the mellifera-related elements in the subgroup are identical by descent. Any alternative model for explaining the observed species distribution must invoke one event of horizontal transmission into the common ancestor of the erecta complex and a separate event of horizontal transmission into the common ancestor of the yakuba complex. Such a scenario may be true but would seem to imply an unrealistically high rate of horizontal transmission to accomplish the introductions in the relatively narrow time intervals available. Assuming that the mellifera-related elements in the subgroup are identical by descent, the observed species distribution implies that the elements were present in the ancestors of the melanogaster complex but subsequently lost.

It should also be noted that the distribution of the mellifera subfamily in the species subgroup is quite different from that of the mauritiana subfamily, as the mauritiana subfamily is present in all species of the melanogaster complex (with the exception of *D. melanogaster* itself) as well as in both species of the yakuba complex. However, DNA-sequence comparisons of MLEs with the *alcohol dehydrogenase* genes in the same species strongly suggest that the mauritiana elements were also present in the common ancestor of the subgroup (Maruyama and Hartl 1991*b*). It should be noted that the elimination of an MLE subfamily from a species qualifies for the term "stochastic loss" inasmuch as the process appears to be unpredictable and random genetic drift must play a role; however, stochastic loss need not be a purely random process because the probability of loss may be determined in part by subtle biological differences among species.

Interaction between Subfamilies

Although distinct subfamilies of MLEs can coexist in the same genome (Robertson 1993; Robertson and MacLeod 1993), as demonstrated in the present study

Downloaded from https://academic.oup.com/mbe/article/12/1/62/1036524 by guest on 23 April 2024

by mellifera and mauritiana elements in the yakuba complex, nothing is known of the extent to which coexisting subfamilies may interact. Some of the questions can be well defined: Can an autonomous element in one subfamily mobilize elements in another subfamily? Does the transposase of one subfamily recognize and bind with elements in another subfamily? Do the processes that regulate transposition in one subfamily also regulate elements in other subfamilies? The extensive sequence divergence among *mariner* subfamilies suggests that sequence divergence may be promoted by diversifying selection, for example, for a new type of MLE to escape from mechanisms that inhibit transposition or that prevent transposase titration by other elements. In addition to functional interactions, structural interactions, such as recombination or gene conversion, may also be important within and between subfamilies. These and related issues can be approached experimentally by means of genetic studies of strains of *D. melanogaster* containing various MLEs introduced by *P*-element germline transformation, as we have demonstrated in the analysis of the mauritiana subfamily of MLEs (Garza et al. 1991; Maruyama et al. 1991).

Concentration in Heterochromatin

More than 85% of the mellifera elements present in *D. erecta* are located in heterochromatin (fig. 2). At least three processes (not mutually exclusive) can account for this concentration. First, *mariner* inserts into the sequence TA and may generally prefer AT-rich regions including long tracts of AT-rich satellites such as are found in the heterochromatin of *D. melanogaster* and, by inference, *D. erecta* (Lohe et al. 1993); recall that both De12 and De19 are inserted into the heptamer 5′-ATATATT-3′. Second, once MLEs have been incorporated into the heterochromatin, the number of copies may increase independently of transposition by means of unequal sister-chromatid exchange and/or other mechanisms of concerted evolution that affect highly repetitive DNA sequences. Third, MLEs that transpose into heterochromatin are expected to have smaller detrimental effects on fitness than those that insert into euchromatin (Golding et al. 1986), so copies may accumulate in heterochromatin through time, owing to relaxed selection.

Vertical Inactivation and Stochastic Loss

Virtually all MLEs so far examined are nonautonomous: they contain one or more mutations that would be sufficient to inactivate or to prevent the synthesis of the putative transposase. At least some nonautonomous elements are probably capable of transposition by means of *trans* complementation by autonomous elements in the same genome; however, this supposition has not been demonstrated directly.

Why is vertical inactivation so prevalent? In seeking the reasons, it is necessary to distinguish between the effect that a mutation in an MLE has on the fate of the MLE itself and the effect that a mutation in an MLE has on the genome of the host. From the standpoint of the MLE, in a genome containing multiple autonomous elements, there may be little or no selection against a new mutation rendering a particular element nonautonomous, since the nonautonomous element may still be able to transpose via *trans* complementation. Once an initial inactivating mutation has occurred, subsequent mutations in the same element are expected to be essentially neutral unless they affect the ability of the element to transpose. On the other hand, an element that is nonautonomous is doomed in a long-term evolutionary sense because its ability to transpose is absolutely dependent on the presence of autonomous elements. In the absence of autonomous elements, each nonautonomous element becomes immobilized. Any particular insertion will be either (1) eliminated by natural selection, owing to deleterious effects that it has on nearby genes, (2) lost by random genetic drift since its initial frequency in the population is small, or (3) divergent in sequence until it is no longer identifiable. The last possibility is by far the least likely. Furthermore, nonautonomous elements are incapable of productive horizontal transmission unless they are transmitted along with autonomous elements or, by chance, are transmitted into a genome already containing autonomous elements of sufficient functional similarity to support *trans* complementation.

Considering the bleak evolutionary prospects of nonautonomous elements, their prevalence among MLEs is all the more remarkable. However, it may be argued that the prevalence of nonautonomous elements results from natural selection at the level of the individual host organisms favoring vertical inactivation. In natural populations of *Drosophila,* most insertions of transposable elements are found at very low frequency, suggesting that each individual insertion is sufficiently deleterious to the host organism that it is prevented from drifting to appreciable frequencies in the population (Golding et al. 1986). If most new insertions are detrimental, then natural selection will tend to eliminate chromosomes containing insertions unless the insertions are present at sites in the genome with a low density of functional genes (e.g., in heterochromatin). Furthermore, owing to the detrimental mutagenic effects of transposition, natural selection will also act to minimize the rate of transposition and thus favor vertical inactivation.

If there is positive selection for vertical inactivation, then the ultimate fate of an MLE in a genome is sto-

Downloaded from https://academic.oup.com/mbe/article/12/1/62/999568 by guest on 23 April 2024

chastic loss. The melanogaster species complex exemplifies at least two episodes of stochastic loss affecting the mellifera subfamily and the mauritiana subfamily of MLEs differently. In the diversification of the melanogaster complex, the mellifera subfamily was lost early, so all present-day members of the complex lack the element. The mauritiana subfamily was lost later—and lost only in the lineage leading to *D. melanogaster* (Maruyama and Hartl 1991*b*).

Horizontal Transmission

If the ultimate fate of an MLE in a genome is vertical inactivation and stochastic loss, then MLEs can only maintain themselves by horizontal transmission into naive or previously infected host genomes. The horizontal transmission of MLEs is well established in various groups of insects (Maruyama and Hartl 1991*a;* Robertson 1993; Robertson and MacLeod 1993). Rigorous statistical proof of horizontal transmission requires the demonstration of a statistically significant inconsistency in the molecular phylogeny of the transposable element when compared with those of single-copy, nontransposable sequences from the same genomes (Lawrence and Hartl 1992; Clark et al. 1994). Formal logical proof of horizontal transmission is even more difficult because an alternative scenario can always be constructed by invoking orders-of-magnitude differences in rates of molecular evolution. For example, the >95% identity in nucleotide sequence between the *D. erecta* elements and the cat flea 10.6 element could be explained by assuming that evolutionary change in the sequences virtually stopped since the species last shared a common ancestor. This argument is difficult to accept, especially in view of the fact that the *D. erecta* elements, and perhaps also the cat flea 10.6 element, are probably nonfunctional and would be expected to evolve at the same rate as pseudogenes. Their divergence should therefore be even greater than observed at third positions in coding sequences, which is 61% for the alpha subunit of the $Na^+ + K^+$ ATPase, and so, in this case, the alternative to horizontal transmission is not so much rigorously refuted as reduced to absurdity.

Assuming an event of horizontal transmission between an ancestor of the melanogaster species subgroup and an ancestor of the cat flea, it is possible to infer the approximate time of the event based on the rate of evolution of MLE sequences. Between copies of the mauritiana subfamily present in the yakuba complex and those present in the melanogaster complex, there is an average 2% difference in nucleotide sequence (Maruyama and Hartl 1991*b*); these complexes are estimated to have diverged approximately 3–5 MYA (based on data in Caccone et al. 1988; Lemeunier et al. 1986). The De12/De19 elements are 2%–4% divergent from the cat flea 10.6. Assuming the same rate of evolution in both the mauritiana and the mellifera subfamilies, the horizontal transmission must have taken place approximately 3–10 MYA. If the donor in the horizontal transmission was the ancestor of the cat flea, then the more ancient times are the more likely because the event probably predated the divergence of the melanogaster complex.

In the context of horizontal transmission, it should also be noted that the hybridization signals observed with genomic DNA from *D. ananassae* (fig. 3) are stronger than expected, given the relatively distant evolutionary relationship between *D. erecta* and *D. ananassae.* (The species are in different species groups.) If the strong hybridization signals are truly indicative of close sequence similarity, then the result suggests yet another case of horizontal transmission.

Sequence Availability

The DNA sequences reported in this article have been deposited in the GenBank database under accession numbers U08093 and U08094.

Acknowledgments

We thank Nathan Blow for technical support, Donal Hickey and Peter Foster for supplying the *D. erecta* library, and Belinda Chang for helpful comments on the maximum parsimony method. We are also grateful to Hugh Robertson for his suggestions for improving the manuscript. This work was supported by grant number GM33741 from the National Institute of General Medical Sciences. D.-A.L. was supported in part by fellowships from the Swedish Cancer Society (3126-B91-01R and 3126-B92-02RAA) and from the Stiftelsen Wenner-Gren Center for Scientific Research.

LITERATURE CITED

BIGOT, Y., M.-H. HAMELIN, P. CAPY, and G. PERIQUET. 1994. *Mariner*-like elements in hymenopteran species: insertion site and distribution. Proc. Natl. Acad. Sci. USA **91:**3408–3412.

BUSHMAN, F. D., A. ENGELMAN, I. PALMER, P. WINGFIELD, and R. CRAIGIE. 1993. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. Proc. Natl. Acad. Sci. USA **90:**3428–3432.

CACCONE, A., G. D. AMATO, and J. R. POWELL. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics **118:**671–683.

CAPY, P., F. CHAKRANI, F. LEMEUNIER, D. L. HARTL, and J. R. DAVID. 1990. Active *mariner* transposable elements are widespread in natural populations of *Drosophila simulans.* Proc. Royal Soc., Lond. [Biol.] **242:**57–60.

Downloaded from https://academic.oup.com/mbe/article/12/1/62/995… by guest on 23 April 2024

CAPY, P., A. KOGA, J. R. DAVID, and D. L. HARTL. 1992. Sequence analysis of active *mariner* elements in natural populations of *Drosophila simulans.* Genetics **130**:499–506.

CAPY, P., T. LANGIN, Y. BIGOT, F. BRUNET, C. BAZIN, M. J. DABOUSSI, G. PERIQUET, J. R. DAVID, and D. L. HARTL. Horizontal transmission versus ancient origin: *mariner* in the witness box. Genetica (in press).

CAPY, P., K. MARUYAMA, J. R. DAVID, and D. L. HARTL. 1991. Insertion sites of the transposable element *mariner* are fixed in the genome of *Drosophila sechellia.* J. Mol. Evol. **33**:450–456.

CLARK, J. B., W. P. MADDISON, and M. G. KIDWELL. 1994. Phylogenetic analysis supports horizontal transfer of *P* transposable elements. Mol. Biol. Evol. **11**:40–50.

DOAK, T. G., F. P. DOERDER, C. L. JAHN, and G. HERRICK. 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.

FELSENSTEIN, J. 1993. PHYLIP (phylogeny inference package) version 3.5c. Department of Genetics, University of Washington, Seattle, Wash.

GARCIA-FERNÀNDEZ, J., G. MARFANY, J. BAGUÑÀ, and E. SALÓ. 1993. Infiltration of *mariner* elements. Nature **364**: 109.

GARZA, D., M. MEDHORA, A. KOGA, and D. L. HARTL. 1991. Introduction of the transposable element *mariner* into the germline of *Drosophila melanogaster.* Genetics **128**:303–310.

GOJOBORI, T., K. ISHII, and M. NEI. 1982. Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. J. Mol. Evol. **18**:414–423.

GOLDING, G. B., C. F. AQUADRO, and C. H. LANGLEY. 1986. Sequence evolution within populations under multiple types of mutation. Proc. Natl. Acad. Sci. USA **83**:427–431.

HIGGINS, D. G., A. J. BLEASBY, and R. FUCHS. 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **7**:189–191.

JACOBSON, J. W., M. M. MEDHORA, and D. L. HARTL. 1986. Molecular structure of a somatically unstable transposable element in Drosophila. Proc. Natl. Acad. Sci. USA **83**:8684–8688.

JEFFS, P. S., E. C. HOLMES, and M. ASHBURNER. 1994. The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. Mol. Biol. Evol. **11**:287–304.

KIDWELL, M. G. 1993. Voyage of an ancient *mariner.* Nature **362**:202.

KIMURA, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

———. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

LAWRENCE, J. G., and D. L. HARTL. 1992. Inference of horizontal genetic transfer from molecular data: an approach using the bootstrap. Genetics **131**:753–760.

LEBOVITZ, R. M., K. TAKEYASU, and D. M. FAMBROUGH. 1989. Molecular characterization and expression of the ($Na^+ + K^+$)-ATPase alpha subunit in *Drosophila melanogaster.* EMBO J. **8**:193–202.

LEMEUNIER, F., J. R. DAVID, L. TSACAS, and M. ASHBURNER. 1986. The melanogaster species group. Pp. 147–256 *in* M. ASHBURNER and H. L. CARSON, ed. The genetics and biology of Drosophila. Academic Press, New York.

LIDHOLM, D.-A., G. H. GUDMUNDSSON, and H. G. BOMAN. 1991. A highly repetitive, *mariner*-like element in the genome of *Hyalophora cecropia.* J. Biol. Chem. **266**:11518–11521.

LIDHOLM, D.-A., A. R. LOHE, and D. L. HARTL. 1993. The transposable element *mariner* mediates germline transformation in *Drosophila melanogaster.* Genetics **134**:859–868.

LOHE, A. R., A. J. HILLIKER, and P. A. ROBERTS. 1993. Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster.* Genetics **134**:1149–1174.

LOZOVSKAYA, E. R., D. A. PETROV, and D. L. HARTL. 1993. A combined molecular and cytogenetic approach to genome evolution in *Drosophila* using large-fragment DNA cloning. Chromosoma **102**:253–266.

MARUYAMA, K., and D. L. HARTL. 1991*a*. Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus.* J. Mol. Evol. **33**:514–524.

———. 1991*b*. Evolution of the transposable element *mariner* in Drosophila species. Genetics **128**:319–329.

MARUYAMA, K., K. D. SCHOOR, and D. L. HARTL. 1991. Identification of nucleotide substitutions necessary for trans-activation of *mariner* transposable elements in *Drosophila:* analysis of naturally occurring elements. Genetics **128**:777–784.

MEDHORA, M., K. MARUYAMA, and D. L. HARTL. 1991. Molecular and functional analysis of the *mariner* mutator element *Mos1* in Drosophila. Genetics **128**:311–318.

REEVES, S. A., and M. K. YAMANAKA. 1993. Cloning and sequence analysis of the alpha subunit of the cat flea sodium pump. Insect Biochem. Mol. Biol. **23**:809–814.

ROBERTSON, H. M. 1993. The *mariner* transposable element is widespread in insects. Nature **362**:241–245.

ROBERTSON, H. M., and E. G. MACLEOD. 1993. Five major subfamilies of mariner transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol. Biol. **2**:125–139.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SANGER, F., S. NICKLEN, and A. R. COULSON. 1977. DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

SOUTHERN, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony, version 3.0s. Illinois Natural History Survey, Champaign.

Downloaded from https://academic.oup.com/mbe/article/12/1/62/999568 by guest on 23 April 2024

SWOFFORD, D. L., and G. J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411–501 *in* D. M. HILLIS, and C. MORITZ, eds. Molecular systematics. Sinauer, Sunderland, Mass.
TAJIMA, F., and M. NEI. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. **1**: 269–285.

JULIAN ADAMS, reviewing editor

Received May 31, 1994

Accepted August 25, 1994

Downloaded from https://academic.oup.com/mbe/article/12/1/62/999568 by guest on 23 April 2024