Molecular Evolution of *P* Transposable Elements in the Genus *Drosophila*. I. The *saltans* and *willistoni* Species Groups

Jonathan B. Clark, Tasha K. Altheide, Michael J. Schlosser,[1] *and Margaret G. Kidwell*
Department of Ecology and Evolutionary Biology, University of Arizona

A phylogenetic survey using the polymerase chain reaction (PCR) has identified four major *P* element subfamilies in the *saltans* and *willistoni* species groups of *Drosophila*. One subfamily, containing about half of the sequences studied, consists of elements that are very similar to the canonical (and active) *P* element from *D. melanogaster*. Within this subfamily, nucleotide sequence differentiation among different copies from the same species and among elements from different species is relatively low. This observation suggests that the canonical elements are relatively recent additions to the genome or, less likely, are evolving slowly relative to the other subfamilies. Elements belonging to the three noncanonical lineages are distinct from the canonical elements and from one another. Furthermore, there is considerably more sequence variation, on the average, within the noncanonical subfamilies compared to the canonical elements. Horizontal transfer and the coexistence of multiple, independently evolving element subfamilies in the same genome may explain the distribution of *P* elements in the *saltans* and *willistoni* species groups. Such explanations are not mutually exclusive, and each may be involved to varying degrees in the maintenance of *P* elements in natural populations of *Drosophila*.

Introduction

P elements form a family of mobile genetic sequences that were first described in *Drosophila melanogaster*, where their mobility is responsible for the phenomenon of hybrid dysgenesis (Kidwell et al. 1977). Since their molecular characterization (Bingham et al. 1982; Rubin et al. 1982), *P* elements have become invaluable as vectors for germline transformation in *D. melanogaster* (Spradling and Rubin 1982). The manifestation of hybrid dysgenesis and the efficiency of transformation are both a result of the high frequency of transposition exhibited by *P* elements in susceptible genetic backgrounds.

In addition to the genus *Drosophila*, sequences with homology to the *P* element have been characterized in the genus *Scaptomyza* (Drosophilidae) (Simonelig and Anxolabéhère 1991) and in the blowfly, *Lucilia cuprina* (Calliphoridae) (Perkins and Howells 1992). These results suggest that *P* elements may be ancient components of higher dipteran genomes. The distribution of sequences with homology to the *P* element of *D. melanogaster* is spotty within the genus *Drosophila*. These sequences are quite common, however, in the subgenus *Sophophora*, where they have been identified by Southern hybridization in 61 of 79 species examined (Daniels et al. 1990). Species in which *P* elements are apparently absent include those most closely related to *D. melanogaster* (Brookfield et al. 1984). There are three explanations for this absence: the elements were never present in these lineages; they have diverged in sequence over time so that they are no longer identifiable, or they have been lost in certain lineages.

In some species (Lansman et al. 1985; Simonelig and Anxolabéhère 1991), *P* elements from the same genome can differ by 4%–10% at the nucleotide level, indicating an extended residence in these genomes accompanied by independent evolution and sequence differentiation of multiple copies. In contrast, the *P* elements of *D. melanogaster* are homogeneous in sequence (O'Hare and Rubin 1983; Sakoyama et al. 1985). Furthermore, the nucleotide sequence of a *P* element from *D. willistoni* is virtually identical to the canonical *P* element from *D. melanogaster*, differing by a single nucleotide substitution in 2907 bp (Daniels et al. 1990). Within *D. melanogaster*, *P* elements are completely absent in old laboratory strains (Bingham et al. 1982; Anxolabéhère et al. 1988) and yet are increasingly abundant in strains collected from the wild during the last 50 yr

Key words: *Drosophila, Sophophora*, molecular evolution, transposable elements, mobile DNA, horizontal transfer.

Address for correspondence and reprints: Margaret Kidwell, University of Arizona, Department of Ecology and Evolutionary Biology, 310 BioSciences West, Tucson, Arizona 85721. E-mail: mkidwell@ccit.arizona.edu.

[1] Present address: Yale University, School of Medicine.

Mol. Biol. Evol. 12(5):902–913. 1995.

0737-4038/95/1205-0019$02.00

(Kidwell 1983). Together these observations provide strong evidence that the canonical *P* element entered the genome of *D. melanogaster* relatively recently, by horizontal transfer from *D. willistoni*. This idea of horizontal transfer is supported by a phylogenetic analysis of *P* element sequences from nine dipteran species (Clark et al. 1994). However, because of the limited scope of that analysis, alternative explanations, such as the coexistence of multiple element types in the same genome, could not be discounted.

In an attempt to understand the evolution of *P* elements in more detail, we have determined partial nucleotide sequences of multiple element copies from over 40 species in the subgenus *Sophophora*. Here, we report the results of a phylogenetic analysis of *P* elements from the *saltans* and *willistoni* species groups, which represent the New World radiation of *Sophophora*. Reported elsewhere are the phylogenetic analyses of the *obscura* (J. García-Planells, N. Paricio, J. B. Clark, R. de Frutos, and M. G. Kidwell, unpublished manuscript) and *melanogaster* (unpublished data) groups. The results confirm the horizontal transfer of *P* to *D. melanogaster* from *D. willistoni* and provide further evidence for additional horizontal transfers in the history of the *Sophophora*. In addition, they support the idea that the evolution of *P* elements is complex, possibly involving the coexistence and independent evolution of multiple subfamilies within certain lineages of flies, and nucleotide sequence divergence and element extinction in others. The findings are discussed in the context of the evolution of the species themselves to provide a comprehensive view of the evolution of one of the best-studied eukaryotic transposable elements.

Material and Methods

DNA Amplification and Sequencing

Samples of flies were obtained from the National Drosophila Species Resource Center (Bowling Green State University, Bowling Green, Ohio 43403), Dr. Linda Strausbaugh (University of Connecticut, Storrs, Conn.), and Dr. Lee Ehrman (SUNY, Purchase, N.Y.). The identities of the species used in this study are given in table 1. The following *P* element sequences were obtained from the literature: *Drosophila guanche* G1 (Miller et al. 1992); *D. nebulosa* N10 (Lansman et al. 1985); *D. subobscura* A1 and G2 (Paricio et al. 1991); *D. willistoni* 13 (Daniels et al. 1990); *Scaptomyza pallida* 2 and 18 (Simonelig and Anxolabéhère 1991); *Lucilia cuprina* P1 (Perkins and Howells 1992).

Total genomic DNA was isolated from pools of about 100 individuals of each species as described elsewhere (Daniels and Strausbaugh 1986). Three degenerate oligonucleotide primers were designed to conserved regions within exon 2 of the *P* element (see fig. 1). Primers 2015, 5′ TGGTTTASC-CATCCWRCRGAYG 3′ (32-fold degeneracy) and 2017, 5′ CCWTCMAGG-GAWGCATTRTTSAC 3′ (32-fold degeneracy) amplify a region of 550 bp between *D. melanogaster P* element positions 1230 and 1780 (O'Hare and Rubin 1983). This primer pair was used to amplify 79 of 92 sequences described here. For the remaining 13 sequences, primer 2015 was replaced by a third primer, 2016, 5′ CGWRACCANTAYGTKGANTCCGG 3′ (256-fold degeneracy), yielding a fragment of 475 bp.

Amplification reactions were generally carried out in 50-μl volumes, with 100-ng template DNA and 0.25 units of *Taq* polymerase (Cetus, Emeryville, Calif.; GIBCO-BRL, Gaithersburg, Md.). The reaction conditions were template denaturation for 1 min., 94°C; primer annealing for 1 min., 50°C; and primer extension for 1 min., 72°C (with 2 s added for each cycle), for a total of 30 cycles. PCR products were purified using Microcon-100 filtration units (Amicon, Beverley, Mass.) and were then ligated into the vector pCR-script (Stratagene, La Jolla, Calif.). Following transformation into *Escherichia coli*, plasmid DNA from individual clones was isolated and sequenced using the Sequenase 2.0 kit (United States Biochemical, Cleveland, Ohio). For each species, between four and nine individual clones were chosen at random as representative of the *P* elements

Table 1
List of *Drosophila* Species Used in This Study and Their Sources

Subgroup	Species	Source
saltans Species group:		
cordata	*D. neocordata*	BG 14041-0831.0
elliptica	*D. emarginata*	BG 14042-0841.0
parasaltans	*D. subsaltans*	BG 14044-0872.0
saltans	*D. austrosaltans*	BG 14045-0881.0
	D. lusaltans	BG 14045-0891.0
	D. prosaltans	BG 14045-0901.0
	D. saltans	BG 14045-0911.0
sturtevanti	*D. sturtevanti*	BG 14043-0871.0
willistoni Species group:		
bocainensis	*D. capricorni*	BG 14030-0721.0
	D. fumipennis	BG 14030-0751.0
	D. nebulosa	BG 14030-0761.0
	D. sucinea	BG 14030-0791.0
willistoni	*D. equinoxalis*	BG 14030-0741.0
	D. insularis	LE
	D. paulistorum	BG 14030-0771.0
	D. pavlovskiana	LE
	D. tropicalis	BG 14030-0801.0
	D. willistoni E	LE
	D. willistoni S	LS

NOTE.—BG, obtained from the Bowling Green stock center; LE, obtained from Dr. Lee Ehrman, SUNY, Purchase, N.Y.; LS, obtained from Dr. Linda Strausbaugh, University of Connecticut, Storrs.

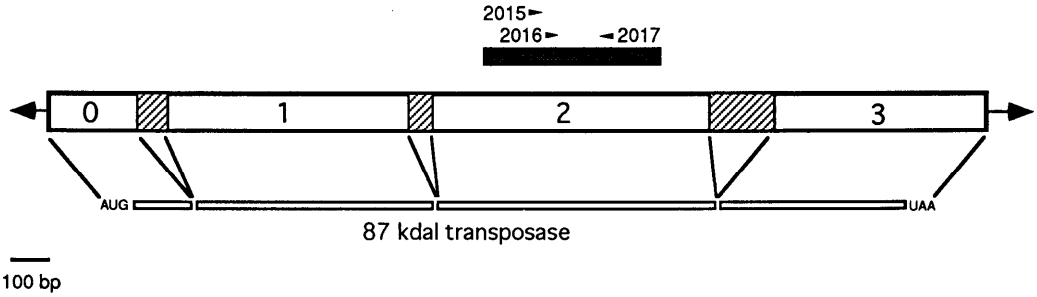

FIG. 1.—Schematic representation of the canonical *P* element from *Drosophila melanogaster*. The complete element is 2,907 bp in length and comprises four exons (denoted 0–3) separated by three introns (*shaded boxes*). The exons encode an 87-kdal transposase necessary for element mobility. The complete element is flanked by 31-bp perfect inverted repeats (*arrowheads*). The relative locations of the primers (2015, 2016, 2017) used to amplify the DNA fragment used for phylogenetic analysis are indicated above exon 2.

amplified in a particular reaction. Sequences are available upon request from the authors.

Phylogenetic Analysis

Regardless of which primer pair was used for amplification, the phylogenetic analysis was confined to the region flanked by primers 2016 and 2017. These nucleotide sequences were aligned by hand after consideration of the codon assignments. The alignment was relatively straightforward, with only occasional gaps needed in certain sequences to preserve the reading frame. The alignment was also examined using CLUSTAL V (Higgins et al. 1992) and differed in only minor respects from the alignment done by hand. In general, the latter was judged to be better on the basis of conservation of codon assignment. Parsimony analysis of the aligned data matrix was performed using the heuristic option of PAUP 3.1.1 (Swofford 1993) with random stepwise addition of sequences and TBR branch swapping. Ten separate searches were performed, each terminating after the accumulation of 1,000 equally parsimonious trees. One of the most parsimonious trees was arbitrarily chosen for presentation. A strict consensus tree was computed allowing the identification of those nodes that were present in all of the 1,000 trees. Bootstrapping of 100 replicates was done on a reduced data set of 35 sequences that included representative taxa from each of the major clades shown in figure 2. Neighbor-joining analysis (Saitou and Nei 1987) was performed on a distance matrix constructed with Kimura's (1980) two-parameter model. Neighbor joining was implemented with the program NJBOOT2 (K. Tamura, Penn State University). Data manipulations, character state analyses and statistics were performed with MacClade 3.01 (Maddison and Maddison 1992).

Results

Amplification of *P* Element DNA Sequences

Oligonucleotide primers were based on the published *P* element sequences available. After identifying suitable conserved regions in the alignment of the *Drosophila* and *Scaptomyza* sequences, the transposase codons were translated into amino acids. The primers were designed to preserve the amino acid identities, but codon usage preferences of these sequences were also considered so that the degeneracy of the primers could be minimized. Because no attempt was made to account for each possible codon for a given amino acid, it is possible that certain (more divergent) subfamilies of *P* elements were missed with these primer combinations. Thus, these results should be viewed not as an exhaustive survey but rather as a sample of the *P* element diversity that may exist in a given genome.

The DNA sequence which was used for this study was confined to exon 2 of the *P* element (see fig. 1). This exon encodes part of the transposase necessary for element mobility (O'Hare and Rubin 1983) as well as a portion of a protein that represses mobility, which is produced from an incompletely spliced *P* element mRNA (Misra and Rio 1990). *P* element-specific PCR products of the expected size were obtained with template DNA from each species examined except for *D. neocordata* and *D. emarginata* of the *saltans* species group. Negative results for the latter two species are consistent with those reported from Southern blots using the canonical *P* element as a probe in low-stringency hybridizations (Daniels et al. 1990). It can be concluded, therefore, that these two species do not possess *P* elements that can be detected by the means employed in these two independent studies.

The apparent absence of *P* elements in *D. emarginata* and *D. neocordata* is a significant finding, as *P* elements are found in all of the other members of the *saltans* species group that were examined and in the closely related *willistoni* species group. These two groups together constitute the New World lineage of *Sophophora*. Furthermore, *P* elements are also common in the two Old World *Sophophora* species groups, *obscura* and *melanogaster*.

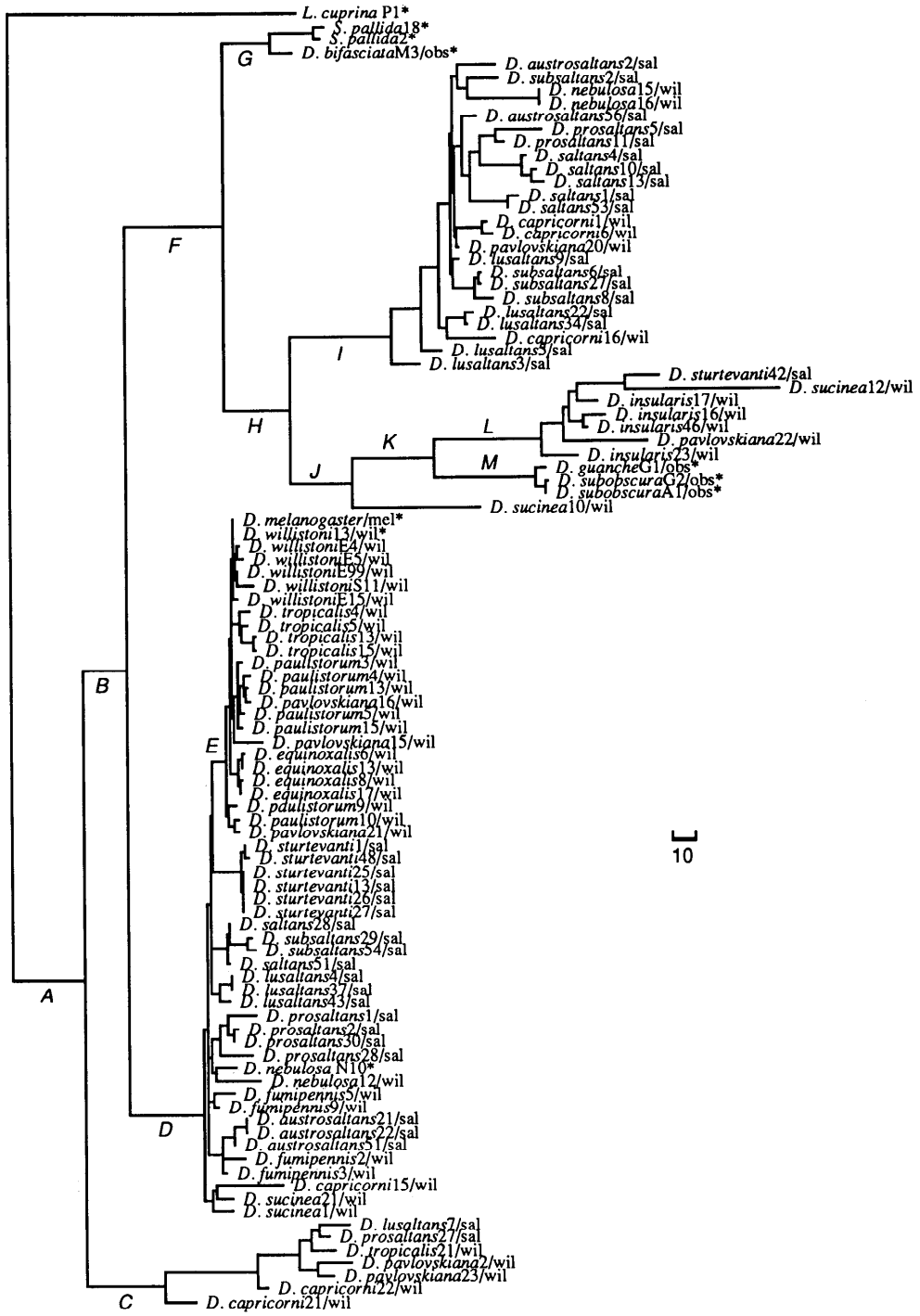

FIG. 2.—Phylogenetic analysis of *P* element nucleotide sequences. This cladogram was generated by parsimony analysis using the heuristic search algorithm of PAUP (Swofford 1993). This tree is an arbitrarily chosen representative of 1,000 equally parsimonious trees, each requiring 1,689 steps. The consistency index is 0.46, and the retention index 0.85. The letters refer to clades discussed in the text. The scale bar gives branch lengths in terms of the number of reconstructed changes. Species names are given in italics followed by the element number. Species group designations for *Sophophora* are given after the slash: *mel, melanogaster; obs, obscura; sal, saltans; wil, willistoni.* Elements identified by an asterisk have been previously published.

Sequence Variation and Phylogenetic Analysis

Ninety-two sequences from 16 species were determined for this study of the *saltans* and *willistoni* lineages. For the analysis, the data matrix included an additional eight sequences previously reported. In all, the data matrix comprises a total of 44,800 bp. A total of 448 positions were aligned in the data matrix, including occasional gaps deemed necessary for correct alignment and excluding the primer sequences. Of these, 418 characters were variable and 260 were phylogenetically informative. Ten separate heuristic searches were performed, each terminating at 1,000 trees. The most parsimonious tree was one of 1,689 steps; trees of an identical number of steps were obtained for 6 of the 10 searches. Numerous additional searches, saving fewer trees, were tried, but none yielded a tree more parsimonious than 1,689 steps. One of these trees, chosen at random, is presented in figure 2.

The branching patterns of the major subfamilies or clades, designated *A-M,* are present in all of the most parsimonious trees. Variation in the branching among the equally parsimonious trees is observed only among the closely related sequences of clade *D* and, occasionally, clade *I*. Support for the placement of the *D. sucinea* 10 element (clade *J* in fig. 2) is not strong. Other nodes that are not well supported are obvious in the strict consensus tree shown in figure 3. The nucleotide sequence differentiation within a clade is summarized in table 2. Because of the extreme divergence of some of these nucleotide sequences, the phylogeny was also examined using neighbor joining (Saitou and Nei 1987). A distance matrix based on Kimura's (1980) two-parameter model was used to construct a neighbor-joining tree (not shown). The composition of the major clades and the tree topology are identical to that of the parsimony tree in figure 2.

Clade *A* includes all of the *Drosophila P* element sequences as well as those from the sister genus, *Scaptomyza.* These sequences are differentiated from that of the blowfly, *Lucilia cuprina,* in both structure and sequence composition (Perkins and Howells 1992). Clade *D* includes those sequences that are very similar to the canonical *P* element sequence described from *D. melanogaster* (O'Hare and Rubin 1983). Within this canonical clade, nucleotide sequence variation is relatively small, ranging from 0% to 10%. Clade *E* includes the canonical sequences of *D. melanogaster* and the *willistoni* subgroup of the *willistoni* species group. This clade is well supported by bootstrapping and was found in all of the most parsimonious trees (see fig. 3). Clade *E* includes sequences from five of the six sibling species of the *willistoni* subgroup, while sequences from the *bocainensis* subgroup are interspersed throughout clade *D* among those of the *saltans* species group.

Among the canonical elements, there is a tendency for *P* elements from one species to form a clade, with only minor sequence differentiation among different elements. For example, the *P* elements from *D. equinoxalis, D. lusaltans,* and *D. sturtevanti* form well-defined clades that exhibit only minor sequence variation within a species. It is assumed that these minor variations are a result of nucleotide substitutions which differentiate the multiple copies within a species. In general, this within-species divergence ranges from 0% to 1.8%, suggesting either extremely conservative evolution or a short time of residence in the genome (see Discussion). However, the elements of some species in clade *D* show considerable variation. For example, *D. nebulosa* N10 and 12 differ by 11% in this region of exon 2, a figure consistent with that reported for differences between complete element sequences from the same species (Lansman et al. 1985).

Divergence within the other three *saltans-willistoni* clades ranges from 17% (*I*) to 30% (*L*) to 46% (*J*). There appears to be more intraspecific nucleotide sequence variation among the *P* elements within these divergent clades than was seen for the elements in clade *D*. For example, the nucleotide sequences of *D. austrosaltans* 2 and 56 (clade *I*) differ by 8%, while those of *D. austrosaltans* 51 and 21 (clade *D*) differ by only 1.4%. For *D. lusaltans,* clade *I* elements differ by as much as 8%, while clade *D* elements differ at most by 2%. Although this pattern is consistently observed among the elements from the *saltans-willistoni* species, it could simply be a reflection of a sampling bias (or small sample size) in this PCR-based survey.

Identical sequences from different clones isolated from the same species are found primarily in clade *D* (*D. melanogaster* and *D. willistoni* 13; *D. sturtevanti* 13, 25, 26, 27; *D. austrosaltans* 21 and 22; *D. lusaltans* 4 and 37). These are delineated by the absence of horizontal branch lengths in fig. 2. The only other identical sequences are *D. nebulosa* 15 and 16 found in clade *I*. The sequences could be identical because the clones chosen were the result of amplification of the same genomic copy, or because the clones represent different *P* element copies in the genome that, for whatever reasons, are identical. Regardless of which explanation is true, it is clear from the preceding paragraph and figure 2 that variation among the elements within a species of clade *D* is less than that seen in the noncanonical clades from the same species.

Clade *K* is interesting because it contains *P* element sequences from three of the four species groups included in *Sophophora.* Although these molecules have diverged considerably from one another, the overall branching, with the *willistoni* and *saltans* species groups forming a

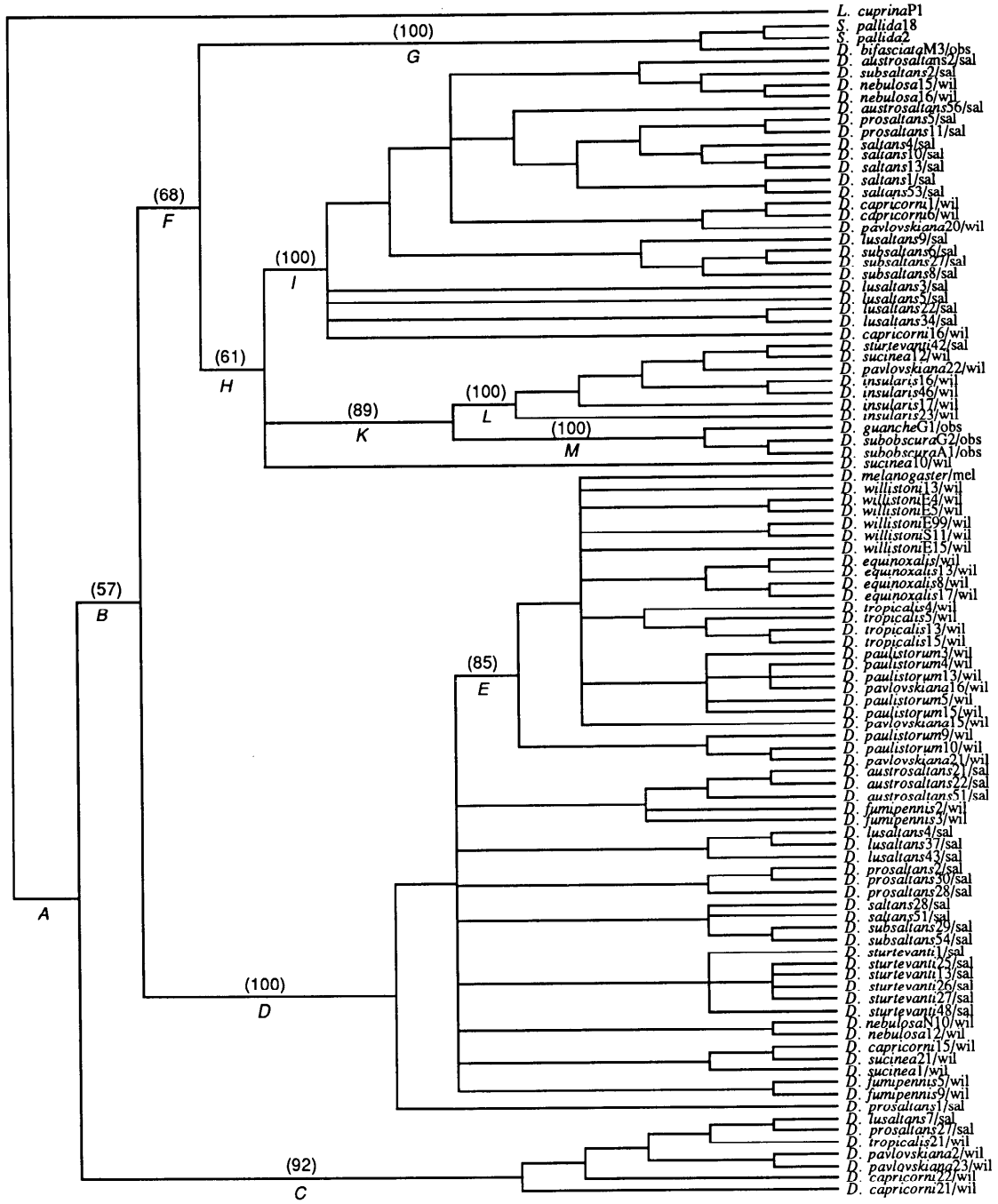

FIG. 3.—Consensus tree of *P* element phylogenetic analysis. This is a strict consensus tree computed from 1,000 equally parsimonious trees of 1,689 steps. Major clades and element names are designated as in fig. 2. The numbers in parentheses are the percent bootstrap values (out of 100 replicates) for the major clades.

New World clade (*L*) and the *obscura* species group an Old World clade (*M*), is consistent with the phylogeny of the species themselves. Clade *K* also contains an unusually divergent *P* element, *D. sucinea* 10, which differs by between 33% and 46% from other elements in this clade. This could be the only representative sampled of another widely distributed noncanonical clade, or simply a divergent *P* element peculiar to *D. sucinea*. As seen in figure 3, the exact placement of *D. sucinea* 10 in clade *J* is not well supported, and a parsimony tree of 1,690 steps (versus 1,689) places this element in clade *I* instead of clade *J*.

Table 2
Uncorrected Nucleotide Differentiation among *P* Element Sequences

Clade	Proportion of Different Sites
Lucilia cuprina P1	0.428 − 0.557 ($\bar{X}$ = 0.476)
A (Drosophilidae)	<0.484
B	<0.471
C	<0.202
D (canonical)	<0.105
E	<0.041
F	<0.466
G	<0.075
H	<0.471
I	<0.170
J	<0.457
K	<0.406
L	<0.305
M	<0.025

NOTE.—The region compared is the 450-bp PCR fragment in exon 2, flanked by primers 2016 and 2017. Letters correspond to the clades designated in fig. 2. For those clades with more than one taxon, the diversity is expressed as a maximum value. Within that clade, nucleotide sequence differentiation could range from 0 (identical sequences) to the maximum value listed. For the *L. cuprina P* element, used as an outgroup, the range of the divergence (and mean) from all other sequences is given.

Integrity of *P* Element Coding Regions

In spite of the large size of the data set used in this study, there are surprisingly few gaps in the alignment. These fall into two classes: single extra or missing nucleotides and larger insertions and deletions (indels). The occurrence of these indels is summarized in table 3. Since the nucleotide sequence alignment was based on amino acid assignment, the occurrence of termination codons was also examined. Table 4 lists those sequences that have termination codons interrupting the transposase reading frame.

There is an equal frequency of indels among the 92 sequences from the *saltans* and *willistoni* groups. Indels are found in 13/41 (32%) of the *saltans* elements and 17/51 (33%) of the *willistoni* elements. Termination codons are similarly evenly distributed among the two species groups: 9/41 (22%) *saltans* elements and 12/51 (24%) *willistoni* elements contain stop codons. Although canonical sequences account for nearly 60% of the *saltans-willistoni* elements identified (table 5), fewer canonical elements than expected contain either indels or termination codons in this region of exon 2. For the *saltans* group, 4/13 (31%) of the indel-containing elements are clade *D* (canonical) elements, while 5/17 (29%) of the *willistoni* elements that contain indels fall into clade *D*. Clade *D* elements with termination codons represent 3/9 (33%) of the *saltans* elements and 2/12 (17%) of the *willistoni* elements.

The analysis of the distribution of indels and termination codons in this region of exon 2 is useful for several reasons. First, it permits the identification of those elements that are assumed to be nonautonomous because of reading-frame interruptions. In 4/14 *saltans* and 4/17 *willistoni* sequences the indels are exclusively multiples of three (table 3). Although the transposase reading frame could theoretically be preserved, five of these are deletions that would remove three or more contiguous amino acids, almost certainly resulting in a defective *P* element transposase. As shown above, a disproportionate number of the elements with indels are found in the noncanonical clades of figure 2, and canonical elements are less likely to have reading-frame interruptions in this exon. With the exception of the *P* element from *S. pallida* (Simonelig and Anxolabéhère 1991) and, possibly, *D. bifasciata* (Hagemann et al. 1992), only canonical *P*

Table 3
Insertions and Deletions among *P* Element Nucleotide Sequences from the *saltans* and *willistoni* Species Groups

Species Group	Element (indel)	Clade
saltans	*D. lusaltans* 5 (−3)	*I*
	D. lusaltans 7 (−1)	*C*
	D. prosaltans 1 (−40)	*D*
	D. prosaltans 2 (+1) (+2)	*D*
	D. prosaltans 5 (−45)	*I*
	D. prosaltans 28 (+3)	*D*
	D. saltans 1 (+1) (−1)	*I*
	D. saltans 51 (−20)	*D*
	D. saltans 53 (−45)	*I*
	D. sturtevanti 42 (−3) (+1)	*L*
	D. subsaltans 2 (−2)	*I*
	D. subsaltans 6 (−1)	*I*
	D. subsaltans 27 (−1)	*I*
willistoni	*D. capricorni* 15 (−8) (+1) (−2)	*D*
	D. capricorni 16 (−1)	*I*
	D. insularis 16 (−3)	*L*
	D. insularis 17 (−3) (−1)	*L*
	D. insularis 23 (−3)	*L*
	D. insularis 46 (−9) (−3)	*L*
	D. nebulosa N10 (+4) (−15) (−11)	*I*
	D. nebulosa 12 (−6) (+1)	*D*
	D. nebulosa 15 (−1) (−1)	*I*
	D. nebulosa 16 (−1) (−1)	*I*
	D. pavlovskiana 2 (−1) (−1)	*C*
	D. pavlovskiana 20 (−1)	*I*
	D. sucinea 1 (−7) (−8)	*D*
	D. sucinea 10 (−12) (−30)	*J*
	D. sucinea 12 (−2) (+3)	*L*
	D. tropicalis 13 (+1)	*D*
	D. tropicalis 15 (+1)	*D*
	D. willistoni S11 (+2)	*D*

NOTE.—The size and nature of the insertion (+) or deletion (−) is given in parentheses. Noncontiguous indels in the same sequence are denoted with separate parentheses. Letters correspond to the clades designated in fig. 2.

Table 4
Termination Codons among *P* Element Nucleotide Sequences from the *saltans* and *willistoni* Species Groups

Species Group	Element (codon)	Clade
saltans	*D. prosaltans* 1 (TGA)	*D*
	D. saltans 4 (TAA)	*I*
	D. saltans 28 (TGA)	*D*
	D. saltans 51 (TGA)	*D*
	D. sturtevanti 42 (TGA)	*L*
	D. subsaltans 2 (TAA)	*I*
	D. subsaltans 6 (TAA)	*I*
	D. subsaltans 8 (TGA)	*I*
	D. subsaltans 27 (TAA)	*I*
willistoni ...	*D. capricorni* 16 (TAA)	*I*
	D. capricorni 21 (TAG) (TGA) (TGA)	*C*
	D. capricorni 22 (TAG) (TGA)	*C*
	D. insularis 46 (TAA)	*L*
	D. nebulosa 15 (TGA)	*I*
	D. nebulosa 16 (TGA)	*I*
	D. pavlovskiana 2 (TAG)	*C*
	D. pavlovskiana 22 (TAA)	*L*
	D. sucinea 12 (TAA) (TAG)	*L*
	D. sucinea 21 (TGA)	*D*
	D. tropicalis 21 (TAA)	*C*
	D. willistoni S11 (TAG)	*D*

NOTE.—The reference reading frame is that for the canonical *P* element transposase (O'Hare and Rubin 1983). Stop codon identities are given in parentheses. Clade designations correspond to those in fig. 2.

elements are known to be autonomous transposable elements.

Second, the co-occurrence of interrupted elements from the same species indicates that, rather than always sampling separate elements from a genome, the same element may have occasionally been sampled more than once. This is to be expected in a PCR-based survey. For example, *D. subsaltans* 6 and 27 share deletions and a TAA stop codon, while *D. nebulosa* 15 and 16 share two separate deletions and a TGA stop codon. An alternative explanation is that these elements are in fact different genomic copies which share a common ancestry. Even though these elements are expected to be nonautonomous in theory, they could have been mobilized (and duplicated) by an autonomous element somewhere else in the genome. However, *D. nebulosa* may not have had any active elements for some time (Lansman et al. 1987). Third, elements with one type of reading-frame interruption are apparently more likely to have the other type. In addition to the two examples given above, the canonical element *D. prosaltans* 1 has a TGA stop codon and a 40-bp deletion; *D. saltans* 51 has a TGA stop codon and a 20-bp deletion. In all, there are 12 elements listed in tables 3 and 4 which have reading-frame interruptions of both kinds.

Fourth, among the more rapidly evolving *P* element sequences (as evidenced by longer branches in fig. 2), indels and termination codons are common (e.g., *D. sucinea* 12 [clade *J*]; *D. capricorni* 15 [clade *D*]; *D. pavlovskiana* 22 [clade *L*]). This is consistent with the idea that the ultimate fate of many transposable elements may be inactivation (Kaplan et al. 1985) followed by neutral molecular evolution.

Discussion

Evolution of *P* Elements in *Sophophora*

The phylogenetic analysis presented here provides a sample of the diversity of *P* elements in the *saltans* and *willistoni* species groups. The sequences amplified by these primers fall into four distinct subfamilies or clades. The divergence among the different subfamilies is striking. The challenge is to try to explain the existence of these different *P* elements and, in particular, why the degree of differentiation among the canonical elements is less than that observed in the other subfamilies.

The sophophoran radiation of the genus *Drosophila* occurred approximately 50-60 Mya (Throckmorton 1975; Beverley and Wilson 1984), resulting in the *melanogaster* lineage in the Old World tropics, the *obscura* lineage in the holoarctic temperate zone, and the New World tropical lineages of *willistoni* and *saltans*. Based on their distribution in the genus *Drosophila* (Daniels et al. 1990) and their identification in other Diptera (Perkins and Howells 1992), *P* elements may have been present when the sophophoran radiation began. If so, it is likely that they were passed vertically to the lineage that gave rise to the *melanogaster* and *obscura* species groups. It is shown here that the *P* elements of the *saltans-willistoni* clade *L* are more similar to those of the *obscura* species group (clade *M*) than to the canonical elements from the *saltans-willistoni D* clade. Thus, the *P* elements within clade *K* could represent an ancestral *P* element present at the beginning of the sophophoran radiation. Preliminary analysis (unpublished data) indicates that at least some of the *P* element sequences from the *melanogaster* species group fall into clade *K* as well. Thus, the diversity of sequences within this clade, which can be as great as 40%, could be a reflection of

Table 5
Distribution of *saltans-willistoni P* Element Nucleotide Sequences

Clade	Number of Elements/Total	Percentage
C ...	7/92	7.6
D ...	52/92	56.5
I	24/92	26.3
J	8/92	8.7

NOTE.—Clade designations are given in fig. 2.

50-60 Myr of *P* element evolution within *Sophophora*. This idea finds support in the results presented in tables 3 and 4 which show that at least six of the seven clade *L* elements are probably not autonomous due to reading frame interruptions. In addition, all of the clade *M* elements are apparently nonautonomous (Paricio et al. 1991; J. García-Planells, N. Paricio, J. B. Clark, R. de Frutos, and M. G. Kidwell, unpublished manuscript), suggesting that these elements have been inactive for quite some time.

Evolution of the *saltans-willistoni* Lineage

It is believed that the ancestral flies of the *saltans-willistoni* lineage became sequestered in tropical North America during the Eocene epoch (Throckmorton 1975). Based on current biogeography and ecology, the divergence of the two groups is thought to have occurred in tropical North America. The ancestor of the *willistoni* species group probably crossed the body of water that at that time separated North and South America, and the diversification of the *willistoni* species group took place exclusively in South America. The *saltans* lineage continued to diversify in North America, giving rise to the ancestors of the primitive *cordata* (represented here by *D. neocordata*) and *elliptica* (represented by *D. emarginata*) subgroups. Subsequently, it is proposed that a progenitor from North America crossed into South America, where further diversification produced the more derived *sturtevanti, parasaltans,* and *saltans* subgroups (Throckmorton 1975). There, both the *saltans* and *willistoni* species groups diversified, the most recent event being the advent of the sibling species clusters. However, the formation of the sibling species in each species group did not occur at the same time: the diversification of the *willistoni* sibling species was probably completed at a time when the *saltans* sibling species were still actively evolving (Throckmorton 1975).

Origin of Canonical *P* Elements

A comprehensive understanding of *P* element evolution in *Sophophora* must account for the absence of canonical *P* elements in the *melanogaster* (excluding *D. melanogaster*) and *obscura* species groups, and their presence in the *saltans* and *willistoni* groups. The results of this analysis show that the *saltans-willistoni P* elements of clade *D* are very similar to one another. Furthermore, there are relatively few reading-frame interruptions among these canonical elements. Because there are divergent *P* elements found in other *saltans-willistoni* clades (e.g., clades *C* and *K*), the extreme sequence conservation of the elements in clade *D* must be of some significance. Either the canonical elements are evolving relatively slowly because they alone are under some kind of strong selection, or they have been transferred relatively recently to the *saltans-willistoni* lineage. Assuming the latter, the 10% divergence among the clade *D* elements would reflect the differentiation of the sequences during their descent in the species in the *saltans* and *willistoni* species groups. In addition, as seen in figure 2, there has been a relatively modest differentiation among different canonical *P* elements within a given species, again consistent with a recent introduction. These results are in contrast to the diversity among *P* elements in the other subfamilies (e.g., clades *C* and *K*).

Horizontal Transfer from *D. willistoni* to *D. melanogaster*

As seen in figure 2, most of the canonical *P* elements from the *willistoni* species group form a single clade (*E*), in which nucleotide sequence differentiation ranges from 0% to 4.1% (see table 2). It is obvious that these elements are descended from a single ancestral *P* element copy and have diverged from one another only modestly during the descent of the species lineages. Because of the monophyly of the Old World *melanogaster* and *obscura* species groups and the fact that *P* elements from the latter fall into clade *M* only (J. García-Planells, N. Paricio, J. B. Clark, R. de Frutos, and M. G. Kidwell, unpublished manuscript), it is striking that the *D. melanogaster P* element sequence is part of the canonical *saltans-willistoni* species group clade. These results provide additional convincing support for a horizontal transfer event in which a canonical *P* element was passed from *D. willistoni* to *D. melanogaster* (Daniels et al. 1990).

Although horizontal transfer can never be unequivocally demonstrated, a number of purported cases have been reviewed recently (Kidwell 1992*a*, 1993). It has been argued that *P* and other transposable elements possess a number of properties that would allow them to cross species barriers with greater success than nonmobile DNA sequences (Kidwell 1992*b*). For example, class II transposable elements encode the enzyme necessary for their own mobility so that the integration of the element into the genome is a self-catalyzed process. With *P* elements, rates of transposition (and spread) can be very high following transfer to a naive genome, in spite of any deleterious effects caused by transposition. Both viruses (Miller and Miller 1982; Friesen and Nissen 1990) and parasitic mites (Houck et al. 1991) have been suggested as potential vectors for the horizontal transfer of transposable elements between insects, but hard evidence is not yet available for their implication in any specific case.

Vertical Transmission and Stochastic Loss

Another explanation to account for the distribution of *P* elements in the subgenus *Sophophora* is the vertical

transmission and coexistence of different *P* element subfamilies in the same genome. Because *P* elements are present in multiple copies in the genome, it seems plausible that the ancestral sophophorans had multiple *P* element copies. Independent evolution of these multiple copies could eventually result in a genome with both canonical and noncanonical elements. During the course of subsequent evolution, some lineages may have lost the canonical *P* element subfamily (e.g., *D. insularis*) or all *P* elements (e.g., *D. neocordata, D. emarginata*). In some lineages (e.g., the derived *saltans* species and the *willistoni* species group) both canonical and noncanonical subfamilies may have been retained. Although this scenario has the advantage of being able to account for any distribution of *P* elements, its use to explain some unexpected sequence similarities would require drastically different rates of evolution for different *P* element subfamilies (Clark et al. 1994). For example, the rate of nucleotide sequence evolution among the canonical subfamily of clade *D* would have had to have been much less than that of sequences in clades *C* or *L* to explain the small degree of differentiation among elements from the former clade.

The loss of a particular *P* element subfamily can perhaps best explain the situation in *D. insularis*. This species has a limited geographical distribution, confined as it is to a few islands in the Lesser Antilles (Spassky et al. 1971). Genetically, it is the most distinct species of the *willistoni* subgroup (Ayala et al. 1975) and may have been the first of the sibling species to diverge. Unlike the other members of the *willistoni* subgroup, *D. insularis* apparently does not possess canonical *P* elements. This assessment is based primarily on the failure to detect *P* elements in the *D. insularis* genome using a canonical *P* element probe against Southern blots (Daniels et al. 1990). Canonical *P* elements are present, however, in all species in the *bocainensis* subgroup (see table 1). An explanation for the presence of canonical *P* elements in the *willistoni* species group in terms of horizontal transfer requires that there were at least two separate transfer events, one to the *willistoni* subgroup after *D. insularis* became isolated, and a second to the *bocainensis* subgroup after it diverged from the *willistoni* subgroup.

The existence of multiple ancestral sophophoran *P* elements followed by the loss of certain subfamilies could also account for the absence of detectable *P* elements in the *cordata* and *elliptica* subgroups of the *saltans* species group. At first glance it would seem an unlikely coincidence that in two separate instances the most geographically isolated species in two different species groups independently lost one or more *P* element subfamily. This loss could, however, reflect some biological phenomenon associated with the isolation of these species from their respective species groups, such as passage through a population bottleneck when stochastic forces are expected to have a greater influence than in larger populations. Small population size or a low element copy number have been shown in theoretical studies to lead to the loss of an element family by genetic drift (Kaplan et al. 1985).

An Emerging Picture of Transposable Element Evolution

Finally, these results can be compared with those from *mariner* which, like *P*, is a Class II (Finnegan 1989) transposable element. This element was originally described from *D. mauritiana* and has been shown to be widely distributed among insects and other arthropods (Robertson 1993; Robertson and MacLeod 1993). Within a species, some *mariner* elements have been observed to be identical or nearly so, while others differ by as much as 50% at the nucleotide level. On a broader phylogenetic scale, *mariner* elements fall into distinct clades or subfamilies. Among the five subfamilies of *mariner* (Robertson and MacLeod 1993), nucleotide divergence is between 40% and 56%, values similar to those distinguishing the major *P* element clades described here. Direct comparisons of the results of the two studies are not easy because the *P* element survey is confined to a single family, Drosophilidae, whereas the *mariner* study includes taxa from several insect orders and extends to different arthropod classes as well. Furthermore, the different methods of PCR screening used in the two studies may in fact directly determine the degree of diversity of the elements identified. The *mariner* primers were designed to correspond to amino acids conserved across a broad phylogenetic spectrum of insects (Robertson and MacLeod 1993), while the *P* element primers were based on available *Drosophila* sequences. Thus, the *P* element primers would be expected to identify a less diverse sample of elements than the *mariner* primers.

When viewed together, the phylogenetic studies of *P* and *mariner* provide the first relatively extensive view of eukaryotic transposable element evolution. It appears that a genome can support multiple, divergent subfamilies of the same family (i.e., *P* or *mariner*) of transposable elements and that such subfamilies can indeed coexist for relatively long periods of time. The data strongly suggest that transposable element lineages have diverged and subsequently evolved independently for periods of time that may exceed the age of a particular species. However, given the possibility of horizontal transfer, it is difficult to infer for how long the different subfamilies of a particular element may have coexisted in a single species.

In conclusion, the accumulating evidence suggests strongly that horizontal transfer is the most likely ex-

planation for certain unexpected phylogenetic affiliations of both *P* and *mariner*. For example, Robertson and MacLeod (1993) have identified a number of cases in which the nucleotide sequence identity for *mariner* elements isolated from taxa separated by 100 to 200 MY exceeds 90%. Two parallel situations are the virtual identity of *P* elements from *D. melanogaster* and *D. willistoni*, two species which diverged 50–60 Mya, and the 93% identity of the *P* elements from *D. bifasciata* and *S. pallida*. Thus, horizontal transfer appears to have played some role in the evolution of both *P* elements and *mariner*. Introgression of transposable elements by horizontal transfer may be counterbalanced by their loss through stochastic processes and natural selection acting to minimize the potentially deleterious effects of transposition. The relative extent to which these mechanisms have been involved in *mariner* evolution has been examined recently (Lohe et al. 1995). Although it remains to be seen if horizontal transfer is widespread among eukaryotic transposable elements, it seems plausible that even a low frequency may have an important influence on the long-term evolutionary dynamics of these sequences.

Acknowledgments

This work was supported by National Science Foundation grant number DEB9119349. T.K.A. was supported by the Undergraduate Biology Research Program at the University of Arizona and National Science Foundation Graduate Research Fellowship 92-55671. J.B.C. and T.K.A. acknowledge support from the Research Training Group in the Analysis of Biological Diversification at the University of Arizona. We thank Dr. David Maddison for helpful suggestions about the phylogenetic analysis and Dr. Rosa de Frutos for sharing her laboratory's results on the *obscura* species group with us. We thank Allan Dickerman, Patrick O'Grady, Joana Silva, and Marty Wojciechowski for comments on the manuscript. We are grateful to Dr. J. S. Yoon and his staff at the National Drosophila Species Resource Center (Bowling Green State University) for providing most of the flies used in this study. We thank two anonymous reviewers for their helpful comments.

LITERATURE CITED

ANXOLABÉHÈRE, D., M. G. KIDWELL, and G. PÉRIQUET. 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. Mol. Biol. Evol. **5**: 252–269.

AYALA, F. J., M. L. TRACEY, D. HEDGECOCK, and R. C. RICHMOND. 1975. Genetic differentiation during the speciation process in *Drosophila*. Evolution **28**:576–592.

BEVERLEY, S. M., and A. C. WILSON. 1984. Molecular evolution in *Drosophila* and the higher Diptera. II. A time scale for fly evolution. J. Mol. Evol. **21**:1–13.

BINGHAM, P. M., M. G. KIDWELL, and G. M. RUBIN. 1982. The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P strain-specific transposon family. Cell **29**:995–1004.

BROOKFIELD, J. F. Y., E. MONTGOMERY, and C. LANGLEY. 1984. Apparent absence of transposable elements related to the *P* elements of *D. melanogaster* in other species of *Drosophila*. Nature **310**:330–332.

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.

DANIELS, S. B., K. R. PETERSON, L. D. STRAUSBAUGH, M. G. KIDWELL, and A. CHOVNICK. 1990. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics **124**:399–355.

DANIELS, S. B., and L. D. STRAUSBAUGH. 1986. The distribution of *P* element sequences in *Drosophila:* the *willistoni* and *saltans* species groups. J. Mol. Evol. **23**:138–148.

FINNEGAN, D. J. 1989. Eukaryotic transposable elements and genome evolution. Trends Genet. **5**:103–107.

FRIESEN, P. D., and M. S. NISSEN. 1990. Gene organization and transcription of *ted*, a lepidopteran retrotransposon integrated within the baculovirus genome. Mol. Cell. Biol. **10**:3067–3077.

HAGEMANN, S., W. J. MILLER, and W. PINSKER. 1992. Identification of a complete P element in the genome of *Drosophila bifasciata*. Nucleic Acids Res. **20**:409–413.

HIGGINS, D. G., A. J. BLEASBY, and R. FUCHS. 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8**:189–191.

HOUCK, M. A., J. B. CLARK, K. R. PETERSON, and M. G. KIDWELL. 1991. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. Science **253**:1125–1129.

KAPLAN, N., T. DARDEN, and C. LANGLEY. 1985. Evolution and extinction of transposable elements in Mendelian populations. Genetics **109**:459–480.

KIDWELL, M. G. 1983. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. Proc. Nat. Acad. Sci. USA **80**:1655–1659.

———. 1992*a*. Horizontal transfer. Curr. Opin. Genet. Develop. **2**:868–873.

———. 1992*b*. Horizontal transfer of P elements and other short inverted repeat transposons. Genetica **19**:913–916.

———. 1993. Lateral transfer in natural populations of eukaryotes. Ann. Rev. Genet. **27**:235–256.

KIDWELL, M. G., J. F. KIDWELL, and J. A. SVED. 1977. Hybrid dysgenesis in *Drosophila melanogaster:* a syndrome of aberrant traits including mutation, sterility & male recombination. Genetics **36**:813–833.

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

LANSMAN, R. A., S. N. STACEY, T. A. GRIGLIATTI, and H. W. BROCK. 1985. Sequences homologous to the P mobile ele-

ment of *Drosophila melanogaster* are widely distributed in the subgenus *Sophophora.* Nature **318**:561–563.

LOHE, A. R., E. N. MORIYAMA, D.-A. LIDHOLM, and D. L. HARTL. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. Mol. Biol. Evol. **12**:62–72.

MADDISON, W. P., and D. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.01. Sunderland, Mass., Sinauer.

MILLER, W. J., S. HAGEMANN, E. REITER, and W. PINSKER. 1992. P element homologous sequences are tandemly repeated in the genome of *Drosophila guanche.* Proc. Natl. Acad. Sci. USA **89**:4018–4022.

MILLER, D.W., and L. K. MILLER. 1982. A virus mutant with an insertion of a *copia*-like transposable element. Nature **299**:562–564.

MISRA, S., and D. C. RIO. 1990. Cytotype control of *Drosophila P* element transposition: the 66 kd protein is a repressor of transposase activity. Cell **62**:269–284.

O'HARE, K., and G. M. RUBIN. 1983. Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34**:25–35.

PARICIO, N., M. , M. PEREZ-ALONSO, M. J. MARTÍNEZ-SEBASTIÁN, and R. DE FRUTOS. 1991. *P* sequences of *Drosophila subobscura* lack exon 3 and may encode a 66 kd repressor-like protein. Nucleic Acids Res. **19**:6713–6718.

PERKINS, H. D., and A. J. HOWELLS. 1992. Genomic sequences with homology to the *P* element of *Drosophila melanogaster* occur in the blowfly *Lucilia cuprina.* Proc. Natl. Acad. Sci. USA **89**:10753–10757.

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.

RUBIN, G. M., M. G. KIDWELL, and P. M. BINGHAM. 1982. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell **29**:987-994.

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

SAKOYAMA, Y., T. TODO, S. ISHIWA-CHIGUSA, T. HONJO, and S. KONDO. 1985. Structures of defective P transposable elements prevalent in natural Q and Q-derived M strains of *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **82**: 6236–6239.

SIMONELIG, M., and D. ANXOLABÉHÈRE. 1991. A P element of *Scaptomyza pallida* is active in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **88**:6102–6106.

SPASSKY, B. S., R. C. RICHMOND, S. PERÉZ-SALAS, O. PAVLOVSKY, C. A. MOURAO, A. S. HUNTER, H. HOENIGSBERG, T. DOBZHANSKY, and F. J. AYALA. 1971. Geography of sibling species related to *D. willistoni* and semispecies of the *D. paulistorum* complex. Evolution **25**:129–143.

SPRADLING, A. C., and G. M. RUBIN. 1982. Transposition of cloned P elements into *Drosophila* germ line chromosomes. Science **218**:341–347.

SWOFFORD, D. 1993. PAUP: Phylogenetic analysis using parsimony. Version 3.1.1. Smithsonian Institution, Washington, D.C.

THROCKMORTON, L. H. 1975. The phylogeny, ecology and geography of *Drosophila.* Pp. 421–469 *in* R. C. KING, ed. Handbook of genetics. Vol. **3.** Invertebrates of genetic interest. Plenum, New York.

DANIEL L. HARTL, reviewing editor

Received December 16, 1994

Accepted April 11, 1995