J Mol Evol (1998) 47:282–291

JOURNAL OF MOLECULAR EVOLUTION

Molecular Evolution of *P* Transposable Elements in the Genus *Drosophila.* II. The *obscura* Species Group

Javier García-Planells,[1] Nuria Paricio,[1] Jonathan B. Clark,[2] Rosa de Frutos,[1] Margaret G. Kidwell[2]

[1] Departamento de Genética, Universitat de Valéncia, 46100-Burjasot, Spain
[2] Department of Ecology and Evolutionary Biology, The University of Arizona, 310 BioSciences West, Tucson, AZ 85721 USA

Abstract. A phylogenetic analysis of *P* transposable elements in the *Drosophila obscura* species group is described. Multiple *P* sequences from each of 10 species were obtained using PCR primers that flank a conserved region of exon 2 of the transposase gene. In general, the *P* element phylogeny is congruent with the species phylogeny, indicating that the dominant mode of transmission has been vertical, from generation to generation. One manifestation of this is the distinction of *P* elements from the Old World *obscura* and *subobscura* subgroups from those of the New World *affinis* subgroup. However, the overall distribution of elements within the *obscura* species group is not congruent with the phylogenetic relationships of the species themselves. There are at least four distinct subfamilies of *P* elements, which differ in sequence from each other by as much as 34%, and some individual species carry sequences belonging to different subfamilies. *P* sequences from *D. bifasciata* are particularly interesting. These sequences belong to two subfamilies and both are distinct from all other *P* elements identified in this survey. Several mechanisms are postulated to be involved in determining phylogenetic relationships among *P* elements in the *obscura* group. In addition to vertical transmission, these include retention of ancestral polymorphisms and horizontal transfer by an unknown mating-independent mechanism.

Key words: Phylogenetic analysis — Transposable elements — Horizontal transfer — *Drosophila* — *Sophophora*

Correspondence to: M.G. Kidwell; *e-mail:* kidwell@azstarnet.com

Introduction

P elements were first described in *Drosophila melanogaster* because of their association with a syndrome of abnormal genetic traits known as hybrid dysgenesis (Kidwell et al. 1977; Bingham et al. 1982). *P* elements are now among the best-characterized eukaryotic transposable elements (for a review, see Engels 1989). Sequences belonging to the *P* family have been detected in many *Drosophila* species and are particularly common in the four principal species groups comprising the subgenus *Sophophora* (Daniels et al. 1990). These include the Old World *melanogaster* and *obscura* groups and the New World *saltans* and *willistoni* groups. *P* elements have also been described in a few species outside of the genus *Drosophila* (Anxolabéhère and Périquet 1987; Simonelig and Anxolabéhère 1991, 1994; Perkins and Howells 1992).

Although *P* elements are widespread in species of the subgenus *Sophophora,* several discontinuities can be detected. For example, *P* elements are apparently absent from those species most closely related to *D. melanogaster* (Brookfield et al. 1984). Nucleotide sequence comparisons suggest that *P* elements were transferred horizontally from *D. willistoni* to *D. melanogaster* (Daniels et al. 1990). Horizontal transfer can also explain the unexpected sequence similarity of *P* elements isolated from *D. bifasciata* and *Scaptomyza pallida* (Simonelig and Anxolabéhère 1991; Hagemann et al. 1992). An earlier reconstruction of a *P* element phylogeny suggested that *P* element dynamics during species evolution is complex, possibly involving horizontal transfer between species and the coexistence of multiple indepen-

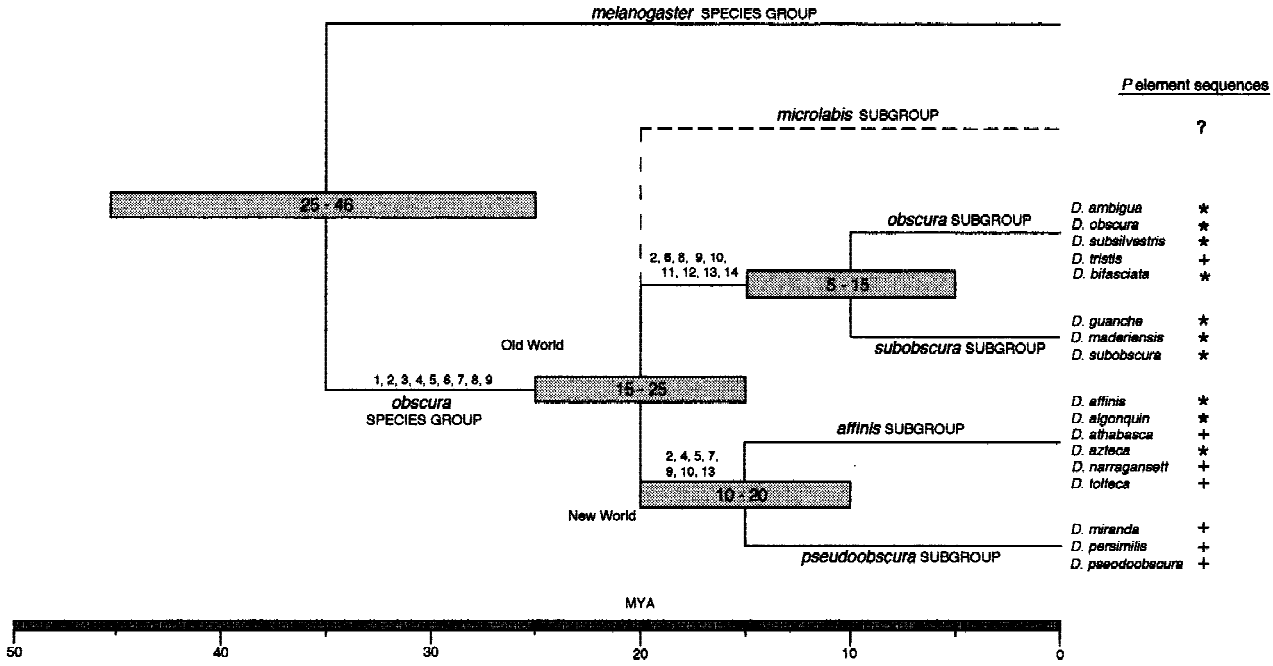

Fig. 1. Phylogeny of the *obscura* species group, showing estimates for divergence times. *Numbers* refer to particular studies, listed in Table 1, that support the branching patterns and time estimates depicted here. The distribution of *P* element sequences is denoted in two ways. An *asterisk* identifies those taxa that were sampled in this study. A *plus sign* indicates that, although *P* elements have been found in this species, they were not amplified in this study, because they either are internally deleted or have diverged to the point where the primers employed here do not work. The *dashed line* leading to the *microlabis* subgroup indicates that the relationship of this recently discovered subgroup to the others has not been investigated in detail.

dent *P* element subfamilies within the same genome (Clark et al. 1994).

Sequences homologous to *P* elements have been detected in all members of the *obscura* species group examined (Anxolabéhère et al. 1985; Daniels et al. 1990; De Frutos et al. 1992), but very little is known about the molecular structure of these sequences. The first *P* sequences characterized from this group were those of *D. bifasciata* (Hagemann et al. 1990). This species contains both internally deleted and complete *P* elements, and two types of complete elements have been described, the M and O types (Hagemann et al. 1992, 1994). The M-type elements are closely related to the *P* elements of *S. pallida* (Simonelig and Anxolabéhère 1991), while the O-type elements are nearly identical to a separate subfamily of *S. pallida* sequences (Hagemann et al. 1996). *P* sequences isolated from *D. subobscura* (Paricio et al. 1991), *D. guanche* (Miller et al. 1992), and *D. madeirensis* (Paricio et al. 1996) have an unusual structure and genomic rearrangement. These elements are clustered in tandem in a single chromosome region and are truncated at the 5′ and 3′ ends. They have lost the ability to transpose because they lack inverted repeats and exon 3 but can encode a repressor-like protein (Paricio et al. 1991; Miller et al. 1992). In addition, *P* sequences which differ in structure and nucleotide sequence from the truncated elements were found in *D. subobscura* heterochromatin (Paricio et al. 1994).

Phylogenetic relationships among species within the *obscura* group have been studied extensively using morphological (Sturtevant 1942; Buzzati-Traverso and Scossiroli 1955), biogeographical (Throckmorton 1975), electrophoretic (Lakovaara et al. 1976; Lakovaara and Saura 1982; Cabrera et al. 1983; Loukas et al. 1984; Cariou et al. 1988; Acosta et al. 1995), chromosome (Brehm and Krimbas 1992; Moltó et al. 1992; Segarra and Aguadé 1992) DNA/DNA hybridization (Goddard et al. 1990), mitochondrial DNA (Latorre et al. 1988; González et al. 1990; Beckenbach et al. 1993; Barrio et al. 1994), and nuclear DNA (Barrio and Ayala 1997) data. Although the phylogeny of the *obscura* species group is by no means completely resolved, there is general support for the existence of five subgroups: *obscura, subobscura,* and *microlabis* in the Old World and *affinis* and *pseudoobscura* in the New World (Gleason et al. 1997; O'Grady 1998). A phylogenetic tree of the *obscura* species group based on the above studies is presented in Fig. 1 (see also Table 1). It provides a basis for comparison of the results of the PCR-based phylogenetic analysis of *P* element sequences reported in this paper.

Materials and Methods

Drosophila *Stocks.* Sixteen species from the *obscura* group were examined in this study. The subgroup classification and sources of these species are given in Table 2. Also included in the analysis were several *P* element sequences obtained from the literature: *D. subobscura* G2, A1, and A2 (Paricio et al. 1991), *D. madeirensis* (Paricio et al. 1996),

Table 1. List of phylogenetic studies supporting various nodes in the phylogeny presented in Fig. 1

No.[a]	Type of study	Reference
1	Biogeography	Throckmorton (1975)
2	Enzyme electrophoresis	Lakovaara and Saura (1982)
3	Enzyme electrophoresis	Cairou et al. (1988)
4	mtDNA RFLP	Latorre et al. (1988)
5	DNA–DNA hybridization	Goddard et al. (1990)
6	mtDNA RFLP	Barrio et al. (1994)
7	mtDNA gene sequences	Beckenbach et al. (1993)
8	mtDNA gene sequences	Barrio et al. (1994)
9	Nuclear gene sequences	Barrio and Ayala (1997)
10	Enzyme electrophoresis	Cabrera et al. (1983)
11	Enzyme electrophoresis	Loukas et al. (1984)
12	mtDNA RFLP	González et al. (1990)
13	Enzyme electrophoresis	Acosta et al. (1995)
14	Molecular	Watabe et al. (1997)

[a] Citation numbers correspond to those given in Fig. 1

D. guanche G1 (Miller et al. 1992), *D. nebulosa* N10 (Lansman et al. 1987), *D. bifasciata* M and O types (Hagemann et al. 1992, 1994), *D. melanogaster* (O'Hare and Rubin 1983), and *Scaptomyza pallida* 2 and 18 (Simonelig and Anxolabéhère 1991).

PCR Amplification. Genomic DNA was prepared from adult flies of each species following the method described by Junakovic et al. (1984). Approximately 100 ng of this DNA was used as template in PCR amplifications with two degenerate primers: 2016 (5′-CGWRACCAITAYGTKGAITCCGG-3′), complementary to positions 1305–1327, and 2017 (5′-CCWTCMAGGGAWGCATTRTTSAC-3′), complementary to positions 1758–1780. These primers (see Fig. 2) amplify a 450-bp fragment within exon 2 of the canonical *D. melanogaster P* element [the first *P* element sequenced and the standard frame of reference for sequence comparisons (O'Hare and Rubin 1983)]. If amplification was not successful, an alternative primer, 2015 (5′-TGGTTTASCCATCCWRCRGAYG-3′), was substituted for primer 2016. Primer 2015 is complementary to *D. melanogaster P* element positions 1230–1251 and primers 2015 and 2017 amplify a 550-bp fragment. PCR was generally carried out in a 50-μl reaction using 200 μ*M* dNTPs, 5 pmol of each primer, and 2.5 U of Taq polymerase. Temperature cycling was done for 30 cycles, each consisting of 95°C for 1 min, 58°C for 1 min, and 72°C for 1 min; a final extension was done at 72°C for 10 min.

Cloning and Sequence Analysis. Purified PCR fragments were subcloned into pCRII (Invitrogen, San Diego, CA) or pCR-Script (Stratagene, La Jolla, CA) plasmid vectors and sequenced with T7 DNA polymerase (Amersham–US Biochemical, Pharmacia) using the dideoxy chain termination method (Sanger et al. 1977). For each species, between 4 and 14 individual clones were sequenced. DNA multiple alignments were obtained with the CLUSTAL V program (Higgins and Sharp 1988) and adjusted by eye to conform to codon assignments. Parsimony analysis of the aligned data matrix was performed (Fig. 3), with all characters weighted equally. Bootstrapping of 100 replicates was performed on a reduced data set of 35 sequences that included representative taxa from each of the major clades shown in Fig. 3. Neighbor-joining analysis (Saitou and Nei 1987) was performed on the same data set. In separate analyses, trees were constructed from distance matrices using the Kimura (1980) two-parameter model for substitution and the model of Tajima and Nei (1984), which compensates for unequal nucleotide base frequencies, and uncorrected distances. All taxa were included in the bootstrap analysis of 100 replicates.

Results

Amplification of P *Homologous Sequences*

Primers 2016 and 2017 were used for almost all of the amplifications in this study. In the few instances in which amplification was not achieved with this primer pair, primer 2015 was substituted for primer 2016. The primer pairs used were chosen based on conserved regions of available published *P* element sequences (see Fig. 2). The primers were designed to preserve amino acid identities in these conserved regions, but codon usage preferences were also considered so that the degeneracy of the primers could be minimized. It is possible that certain, more divergent, subfamilies of *P* elements were missed with these primer combinations. Thus, the results presented here should be viewed not as an exhaustive survey but, rather, as a sample of the diversity that may exist among the *P* elements in a genome.

Using one or the other of the two combinations of degenerate oligonucleotide primers, amplification of *P* element sequences was obtained in all but six of the species analyzed. *P* sequences were determined for 10 species in the *obscura* species group, representing the *obscura*, *subobscura*, and *affinis* but not the *pseudoobscura* or *microlabis* subgroups. In six species (*D. tolteca* of the *affinis* subgroup, *D. tristis* of the *obscura* subgroup, *D. narragansett* of the *affinis* subgroup, and *D. persimilis*, *D. miranda*, and *D. pseudoobscura* of the *pseudoobscura* subgroup) no amplification was obtained with these primer pairs. However, using *P* element inverted-repeat primers, PCR products were obtained with template DNA from the three species of the *pseudoobscura* subgroup. For *D. pseudoobscura*, the amplified fragment corresponds to a non-*P*-specific DNA sequence flanked by two *P* element inverted repeats (data not shown).

Loss or sequence divergence might account for the failure of amplification in the other five species using internal *P* element primer pairs. If the locations of primers 2015, 2016, and 2017 overlap with regions that are deleted, certain *P* elements may have been missed in this PCR survey. Table 3 shows that 30 of 79 (38%) of the sequences obtained from the *obscura* group have insertions or deletions within the region represented by the PCR fragment. The 27-bp deletion shared by clones 1, 3, 11, and 12 from *D. guanche* occurs just one nucleotide downstream of primer 2016. Had this deletion been shifted upstream a few nucleotides, there would have been no amplification in this species using this particular primer.

Phylogenetic Analysis of P *Element Sequences*

Regardless of which of the two primer pairs was used for the PCR amplification, the phylogenetic analysis was confined to the 448-bp fragment flanked by primers 2016

Table 2. List of *Drosophila* species used in this study and their sources

Subgroup	Species[a]	Source
obscura	*D. ambigua**	Dr. C.B. Krimbas (University of Athens, Athens, Greece)
	*D. bifasciata**	Dr. C.B. Krimbas (University of Athens, Athens, Greece)
	D. narragansett	National Drosophila Species Resource Center, Bowling Green, OH, USA (NDSRC-BG)
	*D. obscura**	NDSRC-BG
	*D. subsilvestris**	NDSRC-BG
	D. tristis	NDSRC-BG
subobscura	*D. guanche**	Dr. V.M. Cabrera (University of La Laguna, La Laguna, Spain)
	*D. madeirensis**	Dr. D. Sperlich (University of Tübigen, Tübingen, Germany)
	*D. subobscura**	H271 laboratory strain
affinis	*D. affinis**	NDSRC-BG
	*D. algonquin**	NDSRC-BG
	*D. azteca**	NDSRC-BG
	D. tolteca	NDSRC-BG
pseudoobscura	*D. miranda*	NDSRC-BG
	D. persimilis	NDSRC-BG
	D. pseudoobscura	Captured on Mt. Lemmon (near Tucson, AZ, USA)

[a] *P* element sequences obtained from those species marked with an asterisk

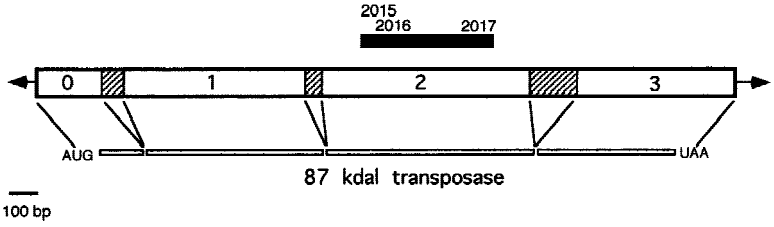

Fig. 2. Schematic representation of the *P* element showing the location of primers used for phylogenetic analysis. The canonical *P* element from *D. melanogaster* is 2907 bp in length and is flanked by perfect 31-bp inverted repeats. Exons 0–3 (*open boxes*) encode the transposase necessary for *P* element mobility. The relative locations of the primers (2015, 2016, and 2017) used to amplify the DNA fragment used for phylogenetic analysis are indicated above exon 2.

and 2017. Phylogenetic trees were constructed using maximum parsimony and neighbor-joining. For both methods, the *P* sequence from the blowfly, *L. cuprina*, was used as an outgroup. Parsimony searches were also performed using sequences from *D. melanogaster*, *D. willistoni*, and *D. nebulosa* as outgroups. This had no effect on the branching patterns within the ingroup. Of 448 characters, 169 were constant, 40 uninformative, and 239 parsimony informative. Since both neighbor-joining and parsimony analyses gave identical branching patterns for the major groups, only the parsimony tree is shown. However, bootstrap values for both methods are given in this tree, which is shown in Fig. 3. In order to provide a frame of reference for discussion, the major clades of *P* element sequences are designated *A–K*.

Clade *F* consists of two main lineages, clades *G* and *H*, and comprises sequences isolated only from species with Old World distributions. Sequences from four species, *D. guanche, D. madeirensis*, and *D. subobscura* of the *subobscura* subgroup and *D. obscura* of the *obscura* subgroup, are represented in clade *H*. Clade *K* includes four sequences from *D. subobscura* described previously (Paricio et al. 1991) and one from *D. guanche* described previously (Miller et al. 1992), along with several new sequences from these two species and from *D. madeirensis*. Divergence among these sequences is very low (maximum, 4.2%), likely reflecting the close relationships among these three species. Four sister lineages to clade *K* are represented by sequences *D. subobscura* 2, 7, 8, and 10. These differ from the sequences in clade *K* by up to 17% and may represent ancestral polymorphisms that have been retained (and detected) in *D. subobscura* (see Discussion).

Clade *J* comprises four closely related sequences from *D. guanche*. These sequences differ from *D. guanche* G1 (clade *K*) by up to 17.1%, a value similar to the maximum divergence among the various sequences from *D. subobscura*. Thus, *D. guanche* may have also retained ancestral polymorphic sequences in its genome. Clade *I* includes nine sequences from *D. obscura* of the *obscura* subgroup. The maximum divergence among these sequences is 5.3%, probably reflecting the divergence among multiple copies in the genome following transposition.

Clade *G*, the other major group of sequences from Old World species, includes *P* sequences detected in only three species from the *obscura* subgroup, *D. subsilvestris, D. obscura*, and *D. ambigua*. There are two distinct *P* element types isolated from *D. ambigua*, one represented by *D. ambigua* 5 and 3, which is closely related to sequences from *D. subsilvestris* and *D. obscura*, and the other by *D. ambigua* 14 and 20. The two *P* element types

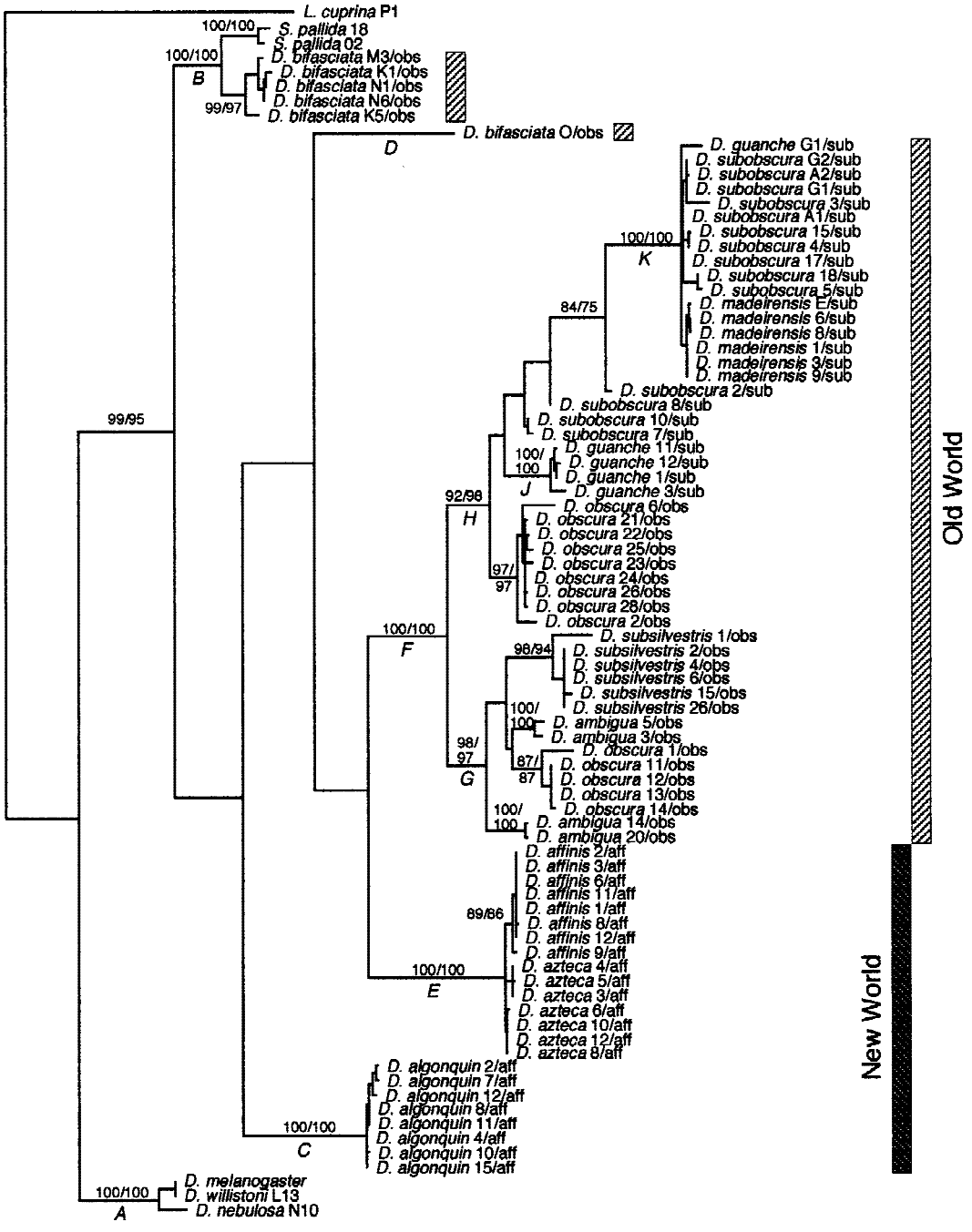

Fig. 3. Phylogenetic analysis of *P* element nucleotide sequences using parsimony. Comparisons were limited to the region between primer 2016 and primer 2017. This cladogram was generated by PAUP 3.1.1 (Swofford 1993) using the heuristic search algorithm with TBR branch swapping and random stepwise addition of taxa. This is an arbitrarily chosen representative of 1000 equally parsimonious trees, each requiring 1012 steps. (The topology of the major lineages was identical in all of the trees obtained.) The consistency index is 0.570, and the retention index 0.916. The bootstrap values, shown *at the nodes*, are percentages for 100 replicates and were determined by parsimony analysis (*before the slash*) and neighbor-joining analysis (*after the slash*). Bootstrap values of 50% or greater are shown only for the major groups. *Italic capital letters* (*A–K*) refer to clades that are discussed in the text. Species names are given in *italics*, followed by a *numbered* clone designation. Subgroup affiliations for the *obscura* species group are indicated *after the slash:* aff, *affinis;* obs, *obscura;* sub, *subobscura. Hatched* and *stippled bars* are used to distinguish sequences obtained from species with Old World and New World distributions.

from *D. ambigua* differ from one another by 9.7% and a likely explanation is the retention of an ancient polymorphism in this species. It is possible that *D. subsilvestris* and *D. obscura* also possess sequences similar to *D. ambigua* 14 and 20 but they were missed in this PCR survey. The maximum divergence among *P* elements from these three species is 14.2%, a value similar to the divergence of *P* elements from *D. guanche* and *D. subobscura* of clade *K* (discussed above). Compared to the variation between species, there is more modest variation

Table 3. Insertions and deletions among *P* element sequences from the *obscura* species group

Subgroup	Element (indel)[a]	Clade[b]
obscura	*D. ambigua* 5 (−35)	*G*
	D. obscura 1 (−29)	*G*
	D. obscura 11 (−29)	*G*
	D. obscura 12 (−29)	*G*
	D. obscura 13 (−29)	*G*
	D. obscura 14 (−29)	*G*
subobscura	*D. guanche* G1 (+3)	*K*
	D. guanche 1 (−27) (−2) (+1)	*J*
	D. guanche 3 (−27) (−2) (+1)	*J*
	D. guanche 11 (−27) (−2) (+1)	*J*
	D. guanche 12 (−27) (−2) (+1)	*J*
	D. madeirensis E (+3)	*K*
	D. madeirensis 1 (+3)	*K*
	D. madeirensis 3 (+3)	*K*
	D. madeirensis 6 (+3)	*K*
	D. madeirensis 8 (+3)	*K*
	D. madeirensis 9 (+3)	*K*
	D. subobscura 7 (−3)	*H*
	D. subobscura A1 (+3)	*K*
	D. subobscura A2 (+3)	*K*
	D. subobscura G1 (+3)	*K*
	D. subobscura G2 (+3)	*K*
	D. subobscura 2 (+3)	*H*
	D. subobscura 4 (+3)	*K*
	D. subobscura 5 (+3)	*K*
	D. subobscura 8 (+3)	*H*
	D. subobscura 15 (+3)	*K*
	D. subobscura 17 (+3)	*K*
	D. subobscura 18 (+3)	*K*
affinis	*D. algonquin* 2 (−3) (−12)	*C*
	D. algonquin 4 (−3)	*C*
	D. algonquin 5 (−3)	*C*
	D. algonquin 7 (−3) (−12)	*C*
	D. algonquin 8 (−3)	*C*
	D. algonquin 10 (−3)	*C*
	D. algonquin 11 (−3)	*C*
	D. algonquin 12 (−3) (−12)	*C*

[a] The size and nature of the insertion (+) or deletion (−) are given in parentheses. Noncontiguous indels are denoted in separate parentheses
[b] Letters correspond to clades designated in Fig. 3.

among the *P* elements sampled from *D. subsilvestris* (up to 5.2%) and *D. obscura* (up to 5.1%), reflecting differentiation within a given species.

Adjacent to clade *F* is clade *E,* which includes several sequences from *D. affinis* and *D. azteca* of the New World *affinis* subgroup. Despite their modest differentiation (2.2%), sequences from clade *E* do form well-supported, species-specific monophyletic groups. Their modest differentiation probably reflects the recent divergence of these two species from one another. The sequences obtained from another member of the *affinis* subgroup, *D. algonquin* (clade *C*), are clearly distinct from those of *D. affinis* and *D. azteca,* showing up to 30% sequence divergence. It is possible that clade *E* sequences are present in *D. algonquin* but were missed in the PCR survey. Similarly, clade *C* sequences may be present in *D. affinis* and *D. algonquin* but, again, were not amplified. Thus sequences in clades *C* and *E* may be extant representatives of ancestral polymorphisms in the *affinis* subgroup that have been differentially retained (or sampled by PCR) in the three species. Another explanation for the distinction between sequences from *D. algonquin* and those from *D. azteca* and *D. affinis* is that they have distinct evolutionary origins. This implies that these sequences were transferred horizontally, in two independent events, to the respective species from different donor species outside of the *obscura* species group. This could explain why both of these clades are distinct from clade *F,* which includes all of the sequences obtained from the Old World *obscura* and *subobscura* subgroups.

Sequences from *D. bifasciata* of the *obscura* subgroup are the exception to the monophyly of most sequences from Old World species (clade *F*). The M-type elements (Hagemann et al. 1992) are closely related to those of *S. pallida* (clade *B*), while the single O-type sequence (Hagemann et al. 1994) from this same species (clade *D*) shows no apparent similarity to any of the sequences in Fig. 3. M-type *P* elements, or their progenitors, were likely transferred horizontally from the genus *Scaptomyza* to *D. bifasciata* (Hagemann et al. 1992; Clark et al. 1994). We have sequenced four additional clones (K1, K5, N5, N6) from the *D. bifasciata* genome and all of them cluster with the M-type sequences of this species (clade *B*). The O-type sequence, which was obtained from a genomic library (Hagemann et al. 1994), shows a high degree of divergence (approximately 30%) from the other sequences isolated from Old World species. The *D. bifasciata* O-type sequence could represent an ancient type of *P* element within the *obscura* group, but most likely its origin in this species is a result of a second horizontal transfer from *S. pallida* (Hagemann et al. 1996).

In general, within the *obscura* group, the *P* elements of a single clade tend to show strong sequence similarity within a given species. This is most likely explained by a recent transposition resulting in identical copies which subsequently diverged from one another. There is strong evidence for this pattern in clade *G. D. obscura* 1, which differs by up to 4.5% from the other sequences from this species, shares with them a 29-bp deletion, indicative of a common evolutionary origin (see Table 3). An alternative explanation for strong sequence conservation within a species is homogenization of sequences by gene conversion. However, there is very little evidence for gene conversion among individual members of the *P* element family. The simplest explanation for identical (or nearly so) sequences obtained from the same species is that they represent duplicate samples from the same genomic *P* element that were cloned and sequenced more than once, an occurrence expected in a PCR-based survey. Minor sequences differences (ca. 1% or less) among clones isolated from the same species could be due to errors introduced by Taq polymerase during PCR.

Discussion

The phylogenetic analysis described here provides the first broad description of the relationships among *P* elements of the *obscura* species group. The resulting *P* element phylogeny (Fig. 3) appears to differ in several respects from the phylogeny of the species themselves (Fig. 1). In order to understand better the events that may have been involved in the evolution of these elements, the results are discussed in the context of the evolutionary history and geographical distribution of the species within which they reside.

The *obscura* group was originally subdivided into the *obscura* subgroup and the *affinis* subgroup (Sturtevant 1942; Buzzati-Traverso and Scossiroli 1955). Subsequently, the Nearctic species of the *obscura* subgroup were placed in two subgroups, *pseudoobscura* and *affinis* (Lakovaara and Saura 1982). Within the Nearctic region the *pseudoobscura* subgroup is found in western continental areas of the New World from British Columbia to Mexico and Colombia. Some species of the *affinis* subgroup are distributed in western areas, from Oregon to Bolivia and in Haiti; others are found in eastern areas from southern Canada to Florida. Distinctions within the Old World members of the *obscura* subgroup were subsequently recognized, leading to a division of this subgroup into the *obscura* and *subobscura* complexes (Lakovaara and Saura 1982). Recent molecular studies suggest that these complexes are themselves distinct subgroups (Barrio et al. 1994; Acosta et al. 1995; Barrio and Ayala 1997); that proposal is followed here (see Fig. 1). A fifth subgroup, *microlabis*, which includes African species, has been described recently (Cariou et al. 1988). Its exact relationship to the *obscura*, *subobscura*, *affinis*, and *pseudoobscura* subgroups has not yet been determined (Gleason et al. 1997; O'Grady 1998).

Species from the Old World are distributed in overlapping areas throughout the Palearctic region from western Europe to eastern Asia. The range of distribution of the *obscura* and *subobscura* subgroups is fairly well-known for western Europe but poorly known for both eastern Europe and Asia. *D. bifasciata* is the most widely distributed species and is found throughout Europe, including Russia, and into northern Asia and Japan. One species, *D. subobscura*, distributed in Europe and North Africa, has been found in North and South America (Prevosti et al. 1985, 1988). However, this species appears to have colonized the Americas only relatively recently.

The widespread distribution of *P* sequences within the *obscura* species group is consistent with the results of earlier surveys based on Southern blots (Daniels et al. 1990; De Frutos et al. 1992). This broad distribution suggests that ancestral *P* element sequences were present at, or soon after, the divergence of the *obscura* species group from the proto-*melanogaster* lineage in the Old World. This proto-*obscura* lineage may have split into two lineages before any major migration occurred, one of them (*microlabis*) spreading in East Africa and the other (ancestral to the *affinis*, *pseudoobscura*, *subobscura*, and *obscura* subgroups) extending to temperate regions (Cariou et al. 1988). Diversification occurred in the Palearctic region and subsequently ancestors of the *pseudoobscura* and *affinis* subgroups spread to North America prior to the mid-Miocene, about 15 MYA (Throckmorton 1975).

Considerable diversity exists among the various subfamilies of *P* elements from the *obscura* species group described here. Although there is strong evidence for short-term vertical transmission, the overall phylogeny of the *P* sequences is not consistent with the phylogenetic relationships among the species themselves. Most of the species carry in their genomes sequences belonging to different *P* element clades, suggesting that the coexistence of divergent *P* sequences in the genomes of single species is a common feature in the *obscura* group. The most plausible explanation for this distribution is the existence of multiple *P* element polymorphisms whose origin predated species divergence. However, rare horizontal transfer events, which introduced new elements from other species, also seem to have been involved.

It is interesting that four sequences from *D. subobscura* (2, 7, 8, 10) have diverged significantly (up to 17%) from all of the other sequences in clade *K*. As mentioned above, two kinds of *P* sequences have been described in *D. subobscura*, truncated elements located in a euchromatic region of the O chromosome and heterochromatic elements (Paricio et al. 1991, 1994). A partial sequence (exons 0 and 1) from a heterochromatic *P* element indicates that it is degraded and highly divergent from the euchromatic elements (Paricio et al. 1994). Although direct sequence comparisons of this heterochromatic element to the four divergent *P* elements from *D. subobscura* are not possible because sequences were obtained from two regions of the element, each shows a similar degree of nucleotide divergence from the known euchromatic elements. Thus a plausible explanation for the relatively high degree of differentiation of these four sequences from *D. subobscura* from the rest of the sequences in clade *K* is that they are heterochromatic elements.

The detailed structure, function, and genomic location of most of the *P* elements sampled in this survey are unknown. The elements could be complete or defective, functional or nonfunctional, euchromatic or heterochromatic. However, clade *K* includes the unusual tandemly arrayed *P* sequences from *D. subobscura* (A1 and A2) (Paricio et al. 1991), *D. guanche* (G1) (Miller et al. 1992), and *D. madeirensis* (Paricio et al. 1996), which have previously been characterized in detail and shown to lack exon 3. In contrast, sequences homologous to exon 3 have been detected by Southern blot analysis in *D. ambigua*, also of the *obscura* subgroup (data not shown). This suggests that the truncation occurred after

the divergence of the *subobscura* subgroup from the ancestor of *D. ambigua* but before the diversification of *D. subobscura, D. guanche,* and *D. madeirensis.* There is a recent description of similarly truncated *P* elements from the *melanogaster* species group but it is not yet clear if this truncation was the result of an independent event or if it occurred in the common ancestor of the *melanogaster* and *obscura* species groups (Nouaud and Anxolabéhère 1997).

P elements from the three species, *D. affinis, D. algonquin,* and *D. azteca,* are expected to be monophyletic because they belong to the same subgroup. The radiation of the *affinis* subgroup is thought to have occurred within the past 2 million years or so (Beckenbach et al. 1993). Most phylogenies indicate that *D. affinis* is more closely related to *D. algonquin* than either is to *D. azteca* (Lakovaara and Saura 1982; Beckenbach et al. 1993; Barrio et al. 1992, 1994; Gleason et al. 1997; O'Grady 1998). However, the *P* elements from *D. algonquin* are clearly distinct from those of *D. affinis* and *D. azteca.* As discussed above, this could be attributed to PCR sampling. Another explanation for this pattern of distribution is that the sequences in clades *C* and *E* represent ancestral *obscura* group polymorphisms that were differentially retained in these three species but were not retained in any of the species with Old World distributions. If this is the case, we would expect to see more sequence differentiation of the sequences within the genome of a given species. An alternative explanation for the origin of these sequences is that they were transferred horizontally to the respective species from an unidentified donor species outside of the *obscura* species group. If this is the case, then two separate transfers would have been necessary, one giving rise to *P* elements in clade *C* and one to those in clade *E.* Until donor species can be identified, however, this remains only a possibility.

The most interesting *P* element sequences from the *obscura* species group are those isolated from *D. bifasciata,* previously referred to as the O- and M-types (Hagemann et al. 1992, 1994). It is possible that both O- and M-type sequences represent ancestral *P* elements, present since the origin of the *obscura* species group, that have been retained only in *D. bifasciata.* This explanation implies that at least four *P* sequences may have been present in the ancestor to the *subobscura* and *obscura* species groups. In this scenario, two of these sequences (represented by sequences in clades *G* and *H*) would have remained in the genomes of most of the Old World species of the *obscura* group, while the other two (O- and M-types) would have been lost in all species except *D. bifasciata.* The opposite but complementary pattern (i.e., loss of sequences represented by clades *G* and *H* and retention of the M- and O-type sequences) could have occurred in the *D. bifasciata* genome. It seems unlikely that a single species, *D. bifasciata,* would be the lone exception from the Old World members of the *obscura* group. In addition, this explanation does not explain the similarity of *P* elements from *D. bifasciata* to those from *S. pallida.*

An alternative explanation is that the M- and O-type *P* elements in *D. bifasciata* were transferred horizontally from another species. The close relationship of the M-type elements with those of *S. pallida* provides a strong argument for the recent introduction of these elements into the *D. bifasciata* genome by this means (Hagemann et al. 1992). Similarly, *S. pallida* possesses O-type *P* elements that are very similar in sequence to those from *D. bifasciata* (Hagemann et al. 1996). [The region of the *P* element compared by Hagemann et al. (1996) does not overlap with the PCR fragment used in this study so the *S. pallida* O-type element is not shown in Fig. 3.] Thus, there seems to be little doubt that *D. bifasciata* acquired the *P* elements included in this analysis by horizontal transfer from another species. Sequence comparisons indicate that this probably occurred in two independent horizontal transfer events (Hagemann et al. 1996).

D. bifasciata has the most extensive geographical distribution of any species in the *obscura* species group (Lakovaara and Saura 1982). The affiliation of *D. bifasciata* with the other Old World members of the *obscura* group is somewhat uncertain (see Fig. 1). In some analyses it clusters with other subgroups (see Barrio et al. 1994; Barrio and Ayala 1997; O'Grady 1998), and in some it is an independent lineage (Cariou et al. 1988). An early branching of *D. bifasciata* from the other species in the *obscura* group could provide an argument for the divergence of the *D. bifasciata P* sequences from those of the rest of the *obscura* species group. However, although there is some disagreement about the exact placement of *D. bifasciata* in the species phylogeny, it is clearly a member of the *obscura* species group. An early divergence of *D. bifasciata* from the other species in the *obscura* species group cannot account for the similarity between both the M- and O-type sequences and the sequences from *S. pallida.* However, its broad geographical distribution relative to other species in the *obscura* group is consistent with *D. bifasciata* being the recipient in these proposed horizontal transfers.

Finally, the *P* element phylogeny of the *obscura* species group can be integrated with that of the *saltans, willistoni,* and *melanogaster* species groups (Clark et al. 1995, 1998; Clark and Kidwell 1997). Together these species groups represent the four main lineages of the *Drosophila* subgenus *Sophophora.* Whereas all of these species groups share a common ancestor, the *saltans–willistoni* lineage diverged early from the Old World *obscura–melanogaster* lineage and diversified exclusively in the New World (Throckmorton 1975). When the *melanogaster, obscura, saltans,* and *willistoni P* sequences are combined into a single phylogenetic analysis, the overall structure of the phylogeny of *obscura P* elements depicted in Fig. 3 is maintained. Sequences

from clade *F*, which represent the Old World species, are part of a larger clade (*E–K*) that includes some sequences from the *melanogaster, saltans,* and *willistoni* species groups. This clade may be equivalent to the T-type subfamily identified by Paricio et al. (1996), and its members may be extant representatives of an ancestral *Sophophora P* element lineage. Sequences from *D. bifasciata,* from *D. affinis* and *D. azteca,* and from *D. algonquin* remain as distinct lineages. Most prevalent in the *saltans* and *willistoni* species groups are the so-called canonical *P* elements (Clark et al. 1995), which form a clade that includes the active *P* element described from *D. melanogaster.* With the exception of sequences from *D. melanogaster,* canonical *P* elements have not been detected in the *melanogaster* or *obscura* species groups.

Although there are some differences among the phylogenies of *P* elements in the *melanogaster, obscura, saltans,* and *willistoni* species groups, similar evolutionary processes can explain the patterns observed in all four groups. First, multiple, independent *P* element subfamilies can coexist in the genome of a species for extended periods of time. If such multiple sequences are viewed as paralogous, this could provide one explanation for the incongruence between species and *P* element phylogenies (Goodman et al. 1979). Second, *P* elements are deleted or absent in some species, as was observed in this study for the *pseudoobscura* subgroup. Assuming that *P* elements were present in the ancestor to each species group, it is not clear why they become degenerate or lost in some species but not in others. Third, the evolution of *P* elements is dominated by vertical transmission from generation to generation. Fourth, horizontal transfer provides a plausible explanation for some relationships that are incongruent with species phylogenies. For example, it is the best explanation for the similarity of *P* elements from *D. bifasciata* and *S. pallida,* and those from *D. willistoni* and *D. melanogaster.* Horizontal transfer may also explain the origin of canonical *P* elements in the *saltans* and *willistoni* species groups and their absence in the *melanogaster* and *obscura* species groups.

Similar conclusions of vertical transmission, sequence degeneration, sequence loss, and horizontal transfer have been reached for the *mariner* transposable element family (Lohe et al. 1995). Both *P* and *mariner* elements transpose by means of a ''cut-and-paste'' mechanism using a transposase. In contrast, transposable elements that move by means of reverse transcriptase appear to be more stable and less likely to transfer horizontally between species. For example, there is good evidence that the *R2* retrotransposable elements that insert specifically in the 28S rRNA genes of many insects have been transmitted strictly vertically since the divergence of the genus *Drosophila* about 60 MYA (Lathe and Eickbush 1997). The relationships of *R2* elements from 23 *Drosophila* species are congruent with the species phylogeny. Thus the emerging picture is complex; the evolution of a particular transposable element seems to be strongly influenced by its mode of transposition and other factors, such as the degree of host specificity.

Acknowledgments. This work was supported by Grants PB-90426 and PB-0050 from Dirección General de Investigatión Científica y Tecníca and DEB-9119349 from the U.S. National Science Foundation and a fellowship to J.B.C. from the NSF-sponsored Research Training Grant in the Analysis of Biological Diversification at the University of Arizona. Some of the sequence analyses were done at Servicio de Bioinformática de la Universitat de Valéncia. We are grateful to Dr. J.S. Yoon and his staff at the National Drosophila Species Resource Center (Bowling Green State University) and to others mentioned in Table 2 for providing many of the fly stocks used in this study. We thank Dr. Wilhelm Pinsker for comments on the manuscript.

References

Acosta T, Pinto F, Hernandez M, Gonzalez AM, Cabrera VM, Larruga LM (1995) Phylogeny of the *Drosophila obscura* group as inferred from one- and two-dimensional protein electrophoresis. J Zool Syst Evol Res 33:101–108

Anxolabéhère D, Périquet G (1987) *P*-homologous sequences in Diptera are not restricted to the Drosophilidae family. Genet Iber 39: 211–222

Anxolabéhère D, Nouaud D, Periquet G (1985) Séquences homologues a l'élément *P* chez des espèces de *Drosophila* du groupe *obscura* et chez *Scaptomyza pallida* (Drosophilidae). Genet Sel Evol 17:579–584

Barrio E, Ayala FJ (1997) Evolution of the *Drosophila obscura* species group inferred from *Gpdh* and *Sod* genes. Mol Phylogenet Evol 7:79–93

Barrio E, Latorre A, Moya A (1994) Phylogeny of the *Drosophila obscura* species group deduced from mitochondrial DNA sequences. J Mol Evol 39:478–488

Beckenbach AT, Wei V, Liu H (1993) Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. Mol Biol Evol 10:619–634

Bingham PM, Kidwell MG, Rubin GM (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

Brehm A, Krimbas CB (1992) The phylogeny of nine species of the *Drosophila obscura* group inferred by the banding homologies of chromosomal regions. III. Element D. Genome 35:1075–1085

Brookfield JFY, Montgomery E, Langley C (1984) Apparent absence of transposable elements related to the *P* elements of *D. melanogaster* in other species of *Drosophila.* Nature 310:330–332

Buzzati-Traverso AA, Scossiroli RE (1955) The ''obscura group'' of the genus *Drosophila.* Adv Genet 7:47–92

Cabrera VM, Gonzalez AM, Larruga JM, Gullon A (1983) Genetic distance and evolutionary relationships in the *Drosophila obscura* group. Evolution 37:675–689

Cariou ML, Lachaise D, Tsacas L, Sourdis J, Krimbas CB, Ashburner M (1988) New African species in the *Drosophila obscura* species group: genetic variation, differentiation and evolution. Heredity 61: 73–84

Clark JB, Kidwell MG (1997) A phylogenetic perspective on *P* transposable element evolution in *Drosophila.* Proc Natl Acad Sci USA 94:11428–11433

Clark JB, Maddison WP, Kidwell MG (1994) Phylogenetic analysis supports horizontal transfer of *P* transposable elements. Mol Biol Evol 11:40–50

Clark JB, Altheide TK, Schlosser MJ, Kidwell MG (1995) Molecular evolution of *P* transposable elements in the genus *Drosophila.* I.

The *saltans* and *willistoni* species groups. Mol Biol Evol 12:902–913

Clark JB, Kim P, Kidwell MG (1998) Molecular evolution of *P* transposable elements in the genus *Drosophila.* III. The *melanogaster* species group. Mol Biol Evol 15:746–755

Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A (1990) Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics 124:339–355

De Frutos R, Peterson KR, Kidwell MG (1992) Distribution of *Drosophila melanogaster* transposable element sequences in species of the *obscura* group. Chromosoma 101:293–300

Engels WR (1989) *P* elements in Drosophila. In: Berg D, Howe M (eds) Mobile DNA. American Society for Microbiology, Washington, DC, pp 437–484

Gleason JM, Caccone A, White KP, Powell JR (1997) Mitochondrial DNA phylogenies for the *Drosophila obscura* group. Evolution 51:433–440

Goddard K, Caccone A, Powell JR (1990) Evolutionary implications of DNA divergence in the *Drosophila obscura* group. Evolution 44: 1656–1670

González AM, Hernández M, Volz A, Pestano J, Larruga JM, Sperlich D, Cabrera VM (1990) Mitochondrial DNA evolution in the *obscura* species subgroup of *Drosophila.* J Mol Evol 31:122–131

Goodman M, Czelusnick JC, Moore EW, Romero-Herrera AE, Matsuda G (1979) Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. Syst Zool 28:132–163

Hagemann S, Miller WJ, Pinsker W (1990) *P*-related sequences in *Drosophila bifasciata:* a molecular clue to the understanding of *P*-element evolution in the genus *Drosophila.* J Mol Evol 31:478–484

Hagemann S, Miller WJ, Pinsker W (1992) Identification of a complete *P* element in the genome of *Drosophila bifasciata.* Nucleic Acids Res 20:409–413

Hagemann S, Miller WJ, Pinsker W (1994) Two distinct P element subfamilies in the genome of *Drosophila bifasciata.* Mol Gen Genet 244:168–175

Hagemann S, Haring E, Pinsker W (1996) Repeated horizontal transfer of *P* transposons between *Scaptomyza pallida* and *Drosophila bifasciata.* Genetica 98:43–51

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Junakovic N, Caneva R, Ballario P (1984) Genomic distribution of *copia*-like elements in laboratory stocks of *Drosophila melanogaster.* Chromosoma 90:378–382

Kidwell MG, Kidwell JF, Sved JA (1977) Hybrid dysgenesis in *Drosophila melanogaster:* a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 36:813–833

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative analysis of nucleotide sequences. J Mol Evol 16:111–120

Lakovaara S, Saura A (1982) Evolution and speciation in the *Drosophila obscura* group. In: Ashburner M, Carson HL, Thompson JN (eds) The genetics and biology of *Drosophila.* Academic Press, New York, pp 2–59

Lakovaara S, Saura A, Lankinen P, Pohjola L, Lokki J (1976) The use of isoenzymes in tracing evolution and in classifying Drosophilidae Zool Scripta 5:173–179

Lansman RA, Shade RO, Grigliatti TA, Brock HW (1987) Evolution of *P* transposable elements: sequences of *Drosophila nebulosa* P elements. Proc Natl Acad Sci USA 84:6491–6495

Lathe WC, Eickbush TH (1997) A single lineage of R2 retrotransposable elements is an active, evolutionarily stable component of the *Drosophila* rDNA locus. Mol Biol Evol 14:1232–1241

Latorre A, Barrio E, Moya A (1988) Mitochondrial DNA evolution in the *Drosophila obscura* group. Mol Biol Evol 5:717–728

Lohe AR, Moriyama EN, Lidholm DA, Hartl DL (1995) Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. Mol Biol Evol 12:62–72

Loukas M, Krimbas CB, Vergini Y (1984) Evolution of the *obscura* group *Drosophila* species. II. Phylogeny of ten species based on electrophoretic data. Heredity 53:483–493

Miller WJ, Hagemann S, Reiter E, Pinsker W (1992) *P* element homologous sequences are tandemly repeated in the genome of *Drosophila guanche.* Proc Natl Acad Sci USA 89:4018–4022

Moltó MD, Pascual L, Martínez-Sebastián MJ, Frutos RD (1992) Genetic analysis of heat shock response in three *Drosophila* species of the *obscura* group. Genome 35:870–880

Nouaud D, Anxolabéhère D (1997) *P* element domestication: a stationary truncated *P* element may encode a 66 kDa repressor-like protein in the *Drosophila montium* species subgroup. Mol Biol Evol 14:1132–1144

O'Grady PM (1998) Reevaluation of phylogeny in the *Drosophila obscura* species group based on combined analysis of nucleotide sequences. Mol Phylo Genet Evol (in press)

O'Hare K, Rubin GM (1983) Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell 34:25–35

Paricio NM, Perez-Alonso M, Martínez-Sebastián MJ, de Frutos R (1991) *P* sequences of *Drosophila subobscura* lack exon 3 and may encode a 66 kd repressor-like protein. Nucleic Acids Res 19:6713–6718

Paricio N, Martínez-Sebastián MJ, de Frutos R (1994) A heterochromatic *P* sequence in the *Drosophila subobscura* genome. Genetica 92:177–186

Paricio N, Miller WJ, Martínez-Sebastián MJ, Hagemann S, de Frutos R, Pinsker W (1996) Structure and organization of the *P* element-related sequences in *Drosophila madeirensis.* Genome 39:823–829

Perkins HD, Howells AJ (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

Prevosti A, Serra L, Ribo G, Aguadé M, Sagarra E, Monclus M, García MP (1985) The colonization of *D. subobscura* in Chile. II. Clines in the chromosomal arrangements. Evolution 39:73–81

Prevosti A, Ribo G, Serra L, Aguadé M, Balaña J, Monclus M, Mestres F (1988) Colonization of America by *D. subobscura:* experiments in natural populations that support the adaptive role of chromosomal-inversion polymorphism. Proc Natl Acad Sci USA 85:5597–5600

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sanger FS, Nicklen S, Coulson AR (1977) DNA sequencing with chain-termination inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Segarra C, Aguadé M (1992) Nucleotide divergence of the *rp49* gene region between *D. melanogaster* and two species of the *obscura* group of Drosophila. J Mol Evol 36:243–248

Simonelig M, Anxolabéhère D (1991) A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster.* Proc Natl Acad Sci USA 88:6102–6106

Simonelig M, Anxolabéhère D (1994) *P* elements are old components of the *Scaptomyza pallida* genome. J Mol Evol 38:232–240

Sturtevant AH (1942) The classification of the genus *Drosophila,* with descriptions of nine new species. Univ Texas Publ 4213:6–51

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, Version 3.1.1. Smithsonian Institution, Washington, DC

Tajima F, Nei M (1984) Estimation of evolutionary distances between nucleotide sequences. Mol Biol Evol 1:269–285

Throckmorton LH (1975) The phylogeny, ecology and geography of *Drosophila.* In: King RC (ed) Handbook of genetics. Plenum, New York, pp 421–469

Watabe H, Bachmann L, Haring E, Sperlich D (1997) Taxonomic and molecular studies on *D. sinobscura* and *D. hubeiensis,* two sibling species of the *D. obscura* group. J Zool Syst Evol Res 35:81–94