Molecular Evolution of *P* Transposable Elements in the Genus *Drosophila*. III. The *melanogaster* Species Group

Jonathan B. Clark, Pyong C. Kim, and Margaret G. Kidwell

Department of Ecology and Evolutionary Biology, University of Arizona

Phylogenetic relationships were determined for 76 partial *P*-element sequences from 14 species of the *melanogaster* species group within the *Drosophila* subgenus *Sophophora*. These results are examined in the context of the phylogeny of the species from which the sequences were isolated. Sequences from the *P*-element family fall into distinct subfamilies, or clades, which are often characteristic for particular species subgroups. When examined locally among closely related species, the evolution of *P* elements is characterized by vertical transmission, whereby the *P*-element phylogeny traces the species phylogeny. On a broader scale, however, the *P*-element phylogeny is not congruent with the species phylogeny. One feature of *P*-element evolution in the *melanogaster* group is the presence of more than one *P*-element subfamily, differing by as much as 36%, in the genomes of some species. Thus, *P* elements from several individual species are not monophyletic, and a likely explanation for the incongruence between *P*-element and species phylogenies is provided by the comparison of paralogous sequences. In certain instances, horizontal transfer seems to be a valid alternative explanation for lack of congruence between species and *P*-element phylogenies. The canonical *P*-element subfamily, which represents the active, autonomous transposable element, is restricted to *D. melanogaster*. Thus, its origin clearly lies outside of the *melanogaster* species group, consistent with the earlier conclusion of recent horizontal transfer.

Introduction

P elements comprise 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, Kidwell, and Sved 1977). Since their molecular characterization (Bingham, Kidwell, and Rubin 1982; Rubin, Kidwell, and Bingham 1982), *P* elements have become invaluable as vectors for germ line transformation in *D. melanogaster* (Spradling and Rubin 1982). The fs hav-nd

Mol. Biol. Evol. 15(6):746–755. 1998

Table 1
List of *Drosophila* Species Surveyed and Their Geographical Distributions and Sources

Subgroup	Species	Distribution[a]	Source	*P* Element[b]	
				Southern	PCR
ananassae	*D. ananassae*	Circumtropical	BG 14024-0371.0	−	−
	D. bipectinata	East Palearctic; oriental; Australasian; Afrotropical	BG 14024-0381.0	F	+
	D. malerkotliana	Oriental; Afrotropical; Neotropical	BG 14024-0391.0	F	+
elegans	*D. elegans*	Oriental; Australasian	BG 14027-0461.0	−	−
eugracilis	*D. eugracilis*	Oriental; Australasian	BG 14026-0451.0	−	−
ficusphila.	*D. ficusphila*	Oriental; east Palearctic	BG 14025-0441.0	+	+
melanogaster	*D. mauritiana*	Afrotropical	BG 14021-0241.0	−	−
	D. melanogaster	Cosmopolitan	Harwich	+	+
	D. simulans	Cosmopolitan	BG 14021-0251.[c]	−	−
	D. yakuba	Afrotropical	BG 14021-0261.0	−	−
montium	*D. auraria*	East Palearctic; oriental	BG 14028-0471.0	+	+
	D. biauraria	East Palearctic	BG 14028-0501.0	+	+
	D. lacteicornis	Oriental (Okinawa)	M. Watada	+	+
	D. nikananu	Afrotropical	BG 14028-0601.0	+	+
	D. quadraria	Oriental (Taiwan)	BG 14028-0651.0	+	+
	D. rufa	Oriental; east Palearctic	M. Watada	+	+
	D. triauraria	East Palearctic; oriental	BG 14028-0691.0	+	+
	D. tsacasi	Afrotropical	BG 14028-0701.0	+	+
suzukii	*D. lucipennis*	Oriental	BG 14023-0331.0	+	+
	D. mimetica	Oriental	BG 14023-0341.0	+	+
	D. pulchrella	East Palearctic; oriental	BG 14023-0351.0	+	−
takahashii	*D. lutescens*	East Palearctic	BG 14022-0271.0	F	−
	D. prostipennis	Oriental	BG 14022-0291.0	F	−
	D. takahashii	East Palearctic; oriental; Australasian	BG 14022-0311.0	−	−

[a] Information on distribution obtained from Lemeunier et al. (1986).

[b] Presence (+) or absence (−) of *P*-element sequences as determined by Southern blotting (Daniels et al. 1990) or PCR (this study). F = faint signal. Sequences were obtained from those species designated PCR (+).

[c] Multiple strains surveyed.

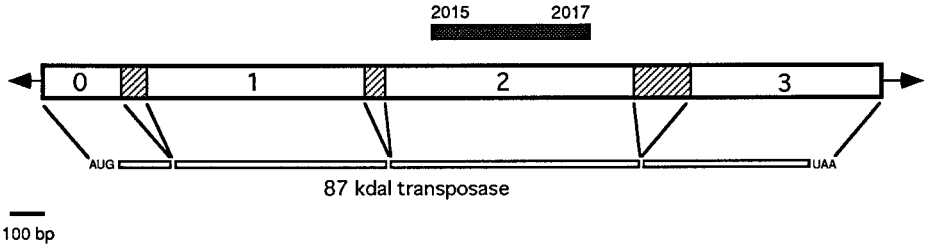

FIG. 1.—Schematic representation of the *P* element showing the locations of primers used for phylogenetic analysis. The canonical *P* element from *D. melanogaster*

Calif.) or pCR 2.1 (Invitrogen, Carlsbad, Calif.). Following transformation into *Escherichia coli*, plasmid DNA from individual clones was isolated and sequenced using Sequenase 2.0 (United States Biochemical, Cleveland,

Table 2
Insertions and Deletions Among *P*-Element Sequences from the *melanogaster* Species Group

Subgroup	Element (indel)	Clade
ananassae . .	*D. bipectinata* 5 (−12)	
	D. bipectinata 7 (−12)	
	D. bipectinata 9 (−12)	
	D. bipectinata 10 (−8)	
	D. bipectinata 13 (−12)	
	D. malerkotliana 1 (−5) (	

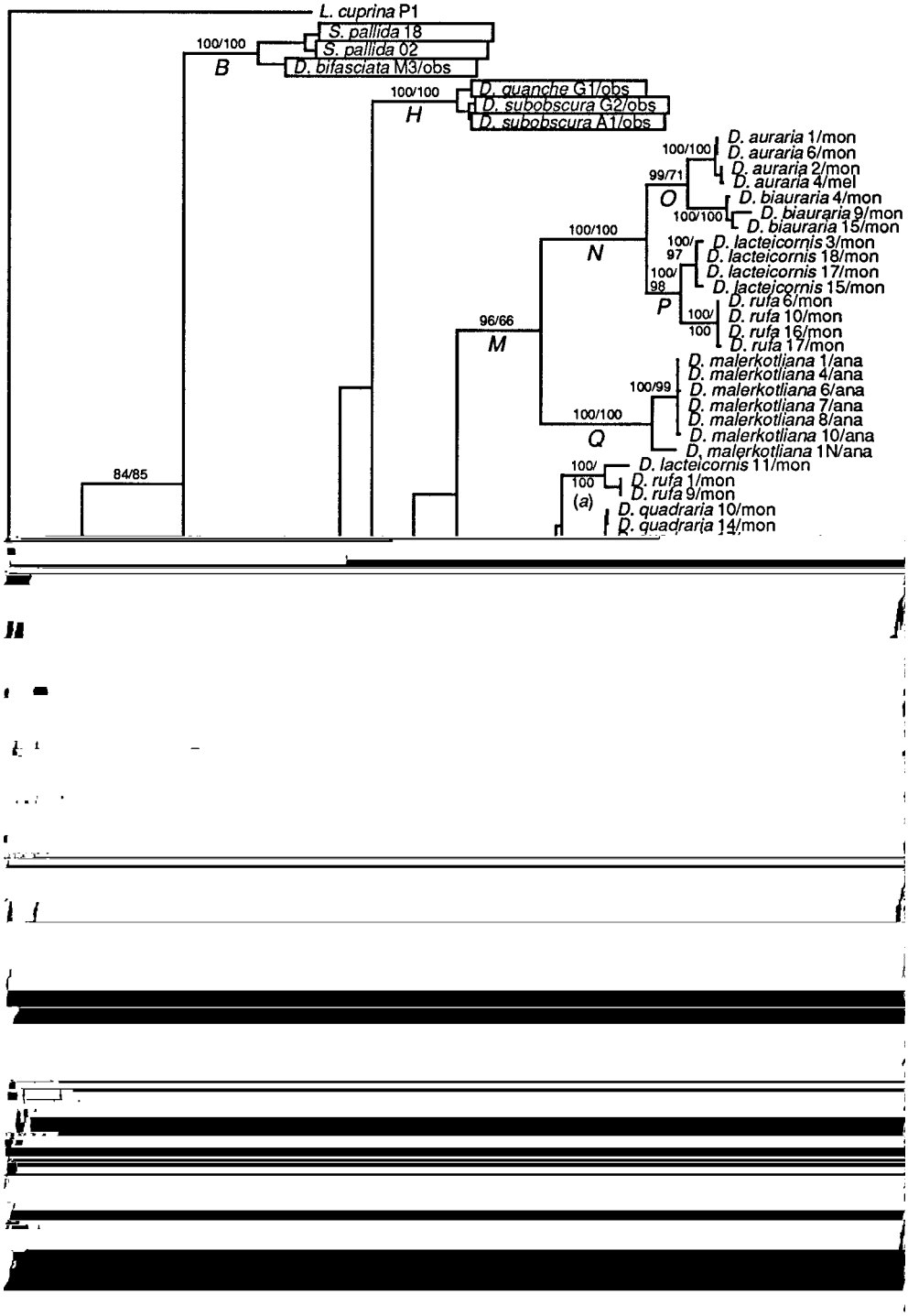

FIG. 2.—Phylogenetic analysis of *P*-element nucleotide sequences from the *melanogaster* species group. Comparisons were limited to 449 bp between primers 2015 and 2017, and the sequence from *L. cuprina* was used as an outgroup. This cladogram was generated by parsimony analysis as implemented by PAUP 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 1,000 equally parsimonious trees, each requiring 1,253 steps. The consistency index is 0.551, and the retention index is 0.911. Numbers above the branches are bootstrap percentages; those before the slashes are values for parsimony analysis (100 replicates); and those after the slashes are values for neighbor-joining (500 replicates). Only values of at least 50% are shown. Letters refer to clades that are discussed in the text. The *a* and *b* labels in clade *L* are also used in the discussion. Species names are given in italics followed by a numbered clone designation. Ten reference sequences are boxed, and species group designations are given after the slashes (obs, *obscura* species group; wil, *willistoni* species group). For species from the *melanogaster* species group, subgroup affiliations are given after the slashes (ana, *ananassae*; fic, *ficusphila*; mel, *melanogaster*; mon, *montium*; suz, *suzukii*).

perhaps serving to mitigate the deleterious effects of transposition of full-length active elements that exist elsewhere in the genome. The truncated sequence is present in eight other members of the *montium* subgroup; however, none of these other species was included in the survey described here.

Clade *I*, another subfamily of *P* elements from the *melanogaster* species group, comprises five nearly identical sequences from *D. ficusphila* of the *ficusphila* subgroup. This species is the only geographically widespread member of this small subgroup, and the distinction of its *P* elements suggests that it is indeed a separate subgroup within the *melanogaster* species group. Clade *G* includes *P*-element sequences from two species of the *suzukii* subgroup, *D. lucipennis* and *D. mimetica*, and from *D. malekotliana* of the *ananassae* subgroup. Sequences from the two species of the *suzukii* subgroup differ from each other by an average of 11%, while sequences from *D. malerkotliana* differ from those of the *suzukii* subgroup by an average of 14%. This implies that the *suzukii* and *ananassae* subgroups are closely related; however, alternative explanations, such as horizontal transfer, are considered in the *Discussion.*

Clade *D* comprises sequences that were isolated from the *ananassae* and *montium* subgroups. *P* elements from the *auraria* complex of the *montium* subgroup are again represented here (clade *F*). Clade *E* is another monotypic clade, comprising closely related sequences from *D. bipectinata* of the *ananassae* subgroup. It is noteworthy that these sequences are distinct from those of *D. malerkotliana*, also of the *ananassae* subgroup (clades *G* and *Q*). The most divergent sequence, *D. bipectinata* 10, has a unique 8-bp deletion and does not share a 12-bp deletion with the other sequences in clade *E* (see table 2).

Clade *A* includes the final subfamily of *P* elements from the *melanogaster* species group. The phylogeny of these sequences, referred to as canonical *P* elements, is unusual in several respects. First, sequences in clade *A* clearly have a distinct evolutionary origin from the other sequences from the *melanogaster* group, which fall into clade *C*. As mentioned above, *D. melanogaster* is the only member of the *melanogaster* subgroup with detectable *P* elements, and the only member of the *melanogaster* species group that possesses the canonical element. Canonical *P* elements are common, however, among species in the New World *saltans* and *willistoni* species groups (Clark et al. 1995). The sequences of the canonical elements from *D. melanogaster* are identical over this 500-bp region to those isolated from *D. willistoni* of the *willistoni* species group. This confirms the results of an earlier study and is consistent with the conclusion that the canonical *P* element was transferred horizontally from *D. willistoni* to *D. melanogaster* (Daniels et al. 1990; Kidwell 1994).

Although good evidence exists for the short-term vertical transmission of *P* elements in the *melanogaster* species group, the *P*-element phylogeny is not congruent with that of the species themselves. For example, as mentioned above, sequences from the *ananassae* subgroup are not monophyletic, falling into three distinct subfamilies (clades *E*, *G*, and *Q*). Another example is the three distinct subfamilies of *P* elements from the *auraria* complex of the *montium* subgroup (clades *F*, *L*, and *O*). For single-copy nuclear genes, sequences from the *montium* and *ananassae* subgroups are expected to be monophyletic. The genomes of at least five species within the *melanogaster* species group possess multiple *P*-element subfamilies. These are listed in table 3, along with the maximum percentages of divergence among the *P*-element subfamilies. The presence of multiple subfamilies of *P* elements in some species indicates that in some cases, the sequences compared are not orthologous. Thus, comparison of paralogous sequences is one explanation for the lack of congruence between *P*-element phylogeny and species phylogeny.

Table 3
Maximum Divergence Among *P*-Element Sequences from the Same *Drosophila* Species

Species	% Divergence	Clades
D. biauraria	30.5	*F, O*
D. lacteicornis	33.6	*L, P*
D. malerkotliana	35.8	*G, Q*
D. quadraria	32.0	*F, L*
D. rufa	23.8	*L, P*

NOTE.—Maximum percentages of divergence (uncorrected) are shown for only those species with multiple *P*-element subfamilies. Clade designations correspond to letters in figure 2.

Discussion

Phylogeny of the *melanogaster* Species Group

It is believed that the *Drosophila* subgenus *Sophophora* originated in the Old World tropics, probably between 40 and 55 MYA (Throckmorton 1975). The two extant Old World species groups, *melanogaster* and *obscura*, are believed to have diverged approximately 30–45 MYA, and the origin of the *melanogaster* species group was probably in southeast Asia (Throckmorton 1975; Bock 1980; Ashburner, Bodmer, and Lemeunier 1984). From here, secondary radiations occurred into adjacent regions, consistent with the current distribution throughout the eastern Palearctic, oriental, Australasian, and Afrotropical regions.

Relationships among the eight extant principle subgroups are uncertain. However, evidence from a number of studies supports three major lineages, one represented by the *ananassae* subgroup, a second by the *montium* subgroup, and a third comprising the *elegans*, *eugracilis*, *ficusphila*, *melanogaster*, *suzukii*, and *takahashii* subgroups (summarized in Lemeunier et al. 1986). This view is supported by the limited number of molecular studies that have included species from the *melanogaster* group (Tsakas and Tscas 1984; Pélandakis, Higgins, and Solignac 1991; Pélandakis and Solignac 1993). A general phylogeny, synthesized from a number of morphological, biogeographical, and molecular studies, is presented in figure 3. This species phylogeny is intended to guide and inform discussion of the *P*-element phylogeny that follows.

certain. They could represent different *P*-element invasions of the genome at different times, or ancient polymorphisms that have been retained in a particular lineage. *P* elements from different subfamilies are not detected in all species (e.g., *D. ficusphila,* clade *I*; *D. bipectinata,* clade *E*; *D. tsacasi* and *D. nikananu,* clade *K*). Assuming that different subfamilies were at one time present in the ancestor of these species, this could be due to the stochastic loss of a particular subfamily or the retention of only a single member of an ancestral polymorphism. Alternatively, multiple subfamilies may indeed exist in these species but were simply not detected in this PCR-based survey, either because of limited sampling or because the primers used were not successful in amplifying all divergent sequence variants that were present.

Among the more puzzling aspects of this *P*-element phylogeny are the sequences from *D. bipectinata* (clade *E*) and *D. malerkotliana* (clades *G* and *Q*), both of the *bipectinata* complex of the *ananassae* subgroup (see fig. 3). These two species are very similar morphologically and can be hybridized in the laboratory to produce fertile females and sterile males (Bock 1978). This implies that these two species diverged from one another only relatively recently, perhaps within the past few million years, yet their corresponding *P* elements are clearly distinct both in sequence and in the presence of characteristic deletions that are not shared between them (see table 2). It is interesting that some sequences from *D. malerkotliana* are affiliated with sequences from the *montium* subgroup (clade *N*), while sequences from *D. bipectinata* are affiliated with a different subfamily of *montium* sequences (clade *F*). One explanation for this pattern is that these two species have retained in their genomes different ancestral *melanogaster* group *P* elements. However, this seems unlikely, since hybridization studies indicate that these two species have only recently become reproductively isolated (Bock 1978). *P*-element evolution in the *ananassae* subgroup is further complicated by a second subfamily of sequences from *D. malerkotliana,* which are affiliated with sequences from the *suzukii* subgroup (clade *G*).

Examination of the degree of divergence between sequences from the *ananassae* subgroup and related sequences from other subgroups can be used to identify other explanations for this unusual distribution of *P* elements in *D. bipectinata* and *D. malerkotliana.* Within clade *D,* the average divergence among sequences from *D. bipectinata* (clade *E*) and those from the *montium* subgroup (clade *F*) is 19.7%. Within clade *M,* the average divergence between sequences from *D. malerkotliana* (clade *Q*) and those from various species in the *montium* subgroup (clade *N*) is 29.3%. This compares to maximum divergence of 16.6% among the 15 sequences isolated from any of the four species of the *montium* subgroup (clade *N*). This pattern of relatively modest differentiation among sequences from the same subgroup and greater differentiation among sequences from different subgroups is what would be expected for vertical transmission.

The situation in the third *ananassae P*-element subfamily is different. Within clade *G,* the average divergence between sequences from *D. malerkotliana* of the *ananassae* subgroup and either *D. lucipennis* or *D. mimetica* of the *suzukii* subgroup is 13.7%. However, the divergence of sequences between the two *suzukii* species is 11.2%. Thus, the divergence between *P* elements from the same subgroup is similar to the divergence of *P*-element sequences from different subgroups. One explanation for this pattern is that the *ananassae* and *suzukii* subgroups are in fact closely related. Although phylogeny among the various subgroups within the *melanogaster* species group is by no means resolved, these two subgroups appear to be only distantly related (see Lemeunier et al. 1986). Thus, the phylogeny of the flies provides no support for this explanation. Another explanation is that the *P* elements from *D. malerkotliana* originated in the *suzukii* subgroup and were transferred horizontally to *D. malerkotliana* at some time in the past. This raises the possibility that horizontal transfer is responsible for some aspects of the unusualin theTJ -17 (mean%TD [(1

Horizontal Transfer and the Origin of *P* Elements in *D. melanogaster*

As seen in figure 2, the canonical *P* element (clade

transmission of the *P* transposable element between *Drosophila* species. Genetics **124**:339–355.

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.

ENGELS, W. R. 1996. *P* elements in *Drosophila.* Pp. 103–123 *in* H. SAEDLER and A. GIERL, eds. Transposable elements. Springer-Verlag, Berlin.

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

HAGEMANN, S., W. J. MILLER, and W. PINSKER. 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.

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. Natl. Acad. Sci. USA **80**:1655–1659.

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

———. 1994. The evolutionary history of the P family of transposable elements. J. Hered. **85**:339–346.

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

KIM, B. K., T. AOTSUKA, and O. KITAGAWA. 1993. Evolutionary genetics of the *Drosophila montium* subgroup. II. Mitochondrial DNA variation. Zool. Sci. **10**:991–996.

LANSMAN, R. A., S. N. STACEY, T. A. GRIGLIATTI, and H. W. BROCK. 1985. Sequences homologous to the P mobile element of *Drosophila melanogaster* are widely distributed in the subgenus Sophophora. Nature **318**:561–563.

LEMEUNIER, F., J. R. DAVID, L. TSACAS, and M. ASHBURNER. 1986. The *melanogaster* species group. Pp. 147–256 *in* M. ASHBURNER, H. L. CARSON, and J. N. THOMPSON, eds. The genetics and biology of Drosophila. Vol. 3e. Academic Press, London.

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.

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. **89**:4018–4022.

NOUAUD, D., and D. ANXOLABÉHÈRE. 1997. *P* element domestication: a stationary truncated *P* element may encode a 66 kD repressor-like protein in the *Drosophila montium* species subgroup. Mol. Biol. Evol. **14**:1132–1144.

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.

PÉLANDAKIS, M., D. G. HIGGINS, and M. SOLIGNAC. 1991. Molecular phylogeny of the subgenus Sophophora of *Drosophila* derived from large subunit of ribosomal RNA sequences. Genetica **84**:87–94.

PÉLANDAKIS, M., and M. SOLIGNAC. 1993. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. J. Mol. Evol. **37**:525–543.

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., and D. J. LAMPE. 1995. Distribution of transposable elements in arthropods. Annu. Rev. Entomol. **40**:333–357.

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.

RUSSO, C. A., N. TAKEZAKI, and M. NEI. 1995. Molecular phylogeny and divergence times of drosophilid species. Mol. Biol. Evol. **12**:391–404.

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

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.

———. 1994. P-elements are old components of the *Scaptomyza pallida* genome. J. Mol. Evol. **38**:232–240.

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. Plenum Press, New York.

TSAKAS, S. C., and L. TSCAS