Recent Horizontal Transfer of Mellifera Subfamily *Mariner* Transposons into Insect Lineages Representing Four Different Orders Shows that Selection Acts Only During Horizontal Transfer

David J. Lampe, David J. Witherspoon,† Felipe N. Soto-Adames,‡ and Hugh M. Robertson§*

*Department of Biological Sciences, Duquesne University; †Oncological Sciences, University of Utah; ‡Department of Biology, University of Vermont; and §Department of Entomology, University of Illinois at Urbana-Champaign

We report the isolation and sequencing of genomic copies of *mariner* transposons involved in recent horizontal transfers into the genomes of the European earwig, *Forficula auricularia*; the European honey bee, *Apis mellifera*; the Mediterranean fruit fly, *Ceratitis capitata*; and a blister beetle, *Epicauta funebris*, insects from four different orders. These elements are in the mellifera subfamily and are the second documented example of full-length *mariner* elements involved in this kind of phenomenon. We applied maximum likelihood methods to the coding sequences and determined that the copies in each genome were evolving neutrally, whereas reconstructed ancestral coding sequences appeared to be under selection, which strengthens our previous hypothesis that the primary selective constraint on *mariner* sequence evolution is the act of horizontal transfer between genomes.

Introduction

Most examples of horizontal gene transfer in eukaryotes involve transposable elements (e.g., Kidwell 1992; Syvanen and Kado 1998). Such transfers are usually recognized by the presence of very closely related elements in relatively distantly related hosts. The archetype of this phenomenon is the transfer of the P-element from a *Drosophila willistoni* group species into the genome of *D. melanogaster*, possibly during the last century (Daniels et al. 1990). That this was not a unique event is shown by the discovery of other P-element transfers within this family of flies (e.g., Clark, Maddison, and Kidwell 1994; Clark and Kidwell 1997).

Perhaps no other transposon family demonstrates the ability to undergo horizontal transfer as well as the *mariners*. Study of the *mariner* family of transposons has yielded evidence of widespread and frequent horizontal transfer across great taxonomic distances. Typical phylogenetic trees of *mariner* transposons show a characteristic lack of concordance between the host phylogeny and the *mariner* phylogeny, indicating the prevalence of horizontal transfer in this group (e.g., fig. 1). In our initial studies (Robertson 1993; Robertson and MacLeod 1993) of *mariners* we used a polymerase chain reaction (PCR) screen that revealed evidence of recent transfers across orders of insects. Other research groups have produced evidence of similar horizontal transfers (e.g., between a fly and a flea [Lohe et al. 1995]) and between insects and a flatworm, *Dugesia tigrina* (Garcia-Fernandez et al. 1995). Our subsequent work using this PCR screen has revealed additional instances of horizontal transfer across phyla of animals involving flatworms, hydras, and primates (Robertson 1997; Robertson and Zumpano 1997). These data and our view that such horizontal transfers to new hosts are an integral part of *mariner* "life cycles" are summarized in Robertson et al. (1998) (see also Hartl, Lohe, and Lozovskaya 1997; Hartl et al. 1997 for reviews of *mariner* life cycles).

To confirm these examples and examine their evolutionary dynamics we have studied two sets of particularly recent and dramatic horizontal transfers in more detail by cloning and sequencing multiple full-length elements from each host. The first example involved closely related *mariners* of the irritans subfamily in a green lacewing *Chrysoperla plorabunda*, the horn fly *Haematobia irritans*, the African malaria mosquito *Anopheles gambiae*, and a drosophilid *D. ananassae* (Robertson and Lampe 1995). Examination of the molecular evolution of these copies revealed that within their hosts they evolved essentially neutrally, accumulating incapacitating mutations. In contrast, their within-host consensus sequences are intact and comparison across hosts indicated that they were evolutionarily conserved through horizontal transfers. Indeed, it appeared that the primary selective constraint on the evolution of these *mariners* was their ability to undergo horizontal transfers to new hosts, presumably because this requires an active element, and likely involves single active copies. We have subsequently confirmed that the consensus sequence of the horn fly element, *Himar1*, is functional in that it encodes a functional transposase capable of catalyzing transposition of a cognate marked element in an in vitro assay (Lampe, Churchill, and Robertson 1996; Lampe, Grant, and Robertson 1998) and in several in vivo systems (Robertson et al. 1998; Zhang et al. 1998; Rubin et al. 1999; Zhang et al. 2000), and we anticipate that the other consensus sequences similarly represent functional *mariner* transposons.

Here we describe full-length copies of a second set of apparent recent horizontal transfers involving closely related mellifera subfamily *mariners* in four orders of insects. These were all first recognized in PCR screens from the genomes of the European honeybee *Apis mellifera*, the European earwig *Forficula auricularia*, the Mediterranean fruit fly (Med fly) *Ceratitis capitata*, and a blister beetle *Epicauta funebris* (*pestifera*) (Robertson 1993; Robertson and MacLeod 1993; Robertson et al. 1998). Figure 1 shows these elements in bold. We analyzed the evolution of these elements using maximum likelihood and were able to show that, as in the case of the

Key words: *mariner*, transposition, transposable element, horizontal transfer.

E-mail: lampe@duq.edu.

Mol. Biol. Evol. 20(4):554–562. 2003
DOI: 10.1093/molbev/msg069

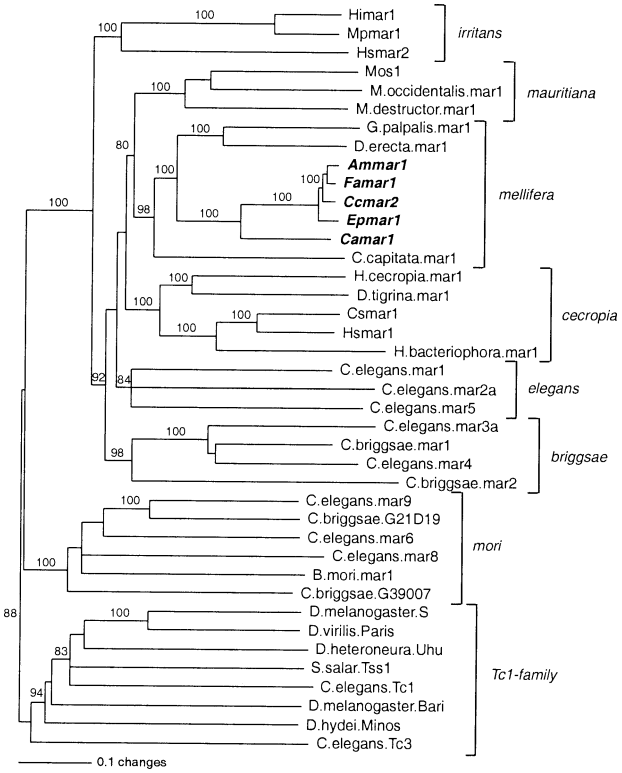

FIG. 1.—A phylogenetic tree of *mariners* based on aligned partial transposase sequences based on a genomic PCR screen. This tree uses the same data set as that of Lampe et al. (2001) except for the inclusion sequences for *Famar1*, *Ccmar2*, *Epmar1*, and *Camar1*. The tree was produced with PAUP* version 4.0b8a using Neighbor-Joining (Swofford 2001). Numbers above branches are bootstrap support values (from 1000 Neighbor-Joining runs) that exceed 75% support. Subfamily designations are shown to the right of the tree in italics. The elements discussed in this paper are in the mellifera subfamily and are highlighted in bold. The Tc1 family is used as an outgroup to root the tree.

irritans subfamily elements, transposons that underwent horizontal transfer also underwent selection, whereas the individual copies within genomes are evolving neutrally. These examples reinforce our previous results and those of other investigators. Together with earlier studies, the examples demonstrate that *mariners* representing three of the five major subfamilies are capable of horizontal transfer into, and functioning in, the host genomes of diverse insects and other animals.

Materials and Methods

Insects

Specimens of *Apis mellifera*, *Forficula auricularia*, and *Epicauta funebris* were collected in Urbana, Ill. Specimens of *Ceratitis capitata* were provided by Mori Tanaka from Hawaii; related *Apis* species were provided by Deborah Smith; related *Forficula* species were provided by Thierry Wirth; related tephritid fruit flies were supplied by Stewart Berlocher; and related blister beetles were collected in Urbana, Ill. and provided by Jim Carroll, Randall Higgins, and Michel Veuille. A strain of the drosophilid fly *Chymomyza amoena* was obtained from the Drosophila Species Stock Center because it has one of the most closely related mellifera subfamily *mariners* that serves as a reasonably close outgroup in phylogenetic analyses (Robertson and MacLeod 1993).

Polymerase Chain Reaction Screens

Polymerase chain reaction using specific primers designed to amplify all four of these closely related sequences was used to screen the genomic DNA samples of all ±700 animals in our previous survey studies. The specific primers are MAR-188F (5′ATCAAAAGCTGRTATTCATC) and MAR-251R (5′CAAAGATGTGTGTGGCCTTG), which end in codons corresponding to the 188th and 251st amino acids of the canonical *Mos1 mariner* from *D. mauritiana* (see fig. 2).

Genomic Libraries and Screens

DNA was extracted by standard phenol/chloroform methods (Robertson et al. 1998). Complete digestion genomic libraries were constructed using *Eco*R1 digestion and cloning in the Lambda ZAP II vector as described elsewhere (Robertson and Lampe 1995). They were screened with the 496 bp "Honeybee 4.2" PCR fragment labeled with ^{32}P, also as described (Robertson and Lampe 1995). The *A. mellifera* library was screened a second time using a 262 bp PCR product amplified with the primers 52f-5′CTCATAGWTGRCTTCAGTTG and 31R-5′CTGTTTATGGYGACGAAGC. These sequences were derived from a single intact genomic copy we obtained in the initial screen and five truncated copies missing a 350 bp region of the 5′ ends. We reasoned that a probe derived from the truncated region would recover only full-length copies. The *C. amoena* library was screened separately with the "C. amoena.35.4" PCR fragment as a probe (Robertson and MacLeod 1993) because this *mariner* is significantly different in sequence. Clones were transformed from phage into pBluescript clones in vivo and enzymatically sequenced manually using custom internal primers designed to work in all of these *mariners*. All sequences were obtained on both strands except for the first and last ±100 bp, which were only obtained on one strand using primers directed outwards into the flanking DNA.

Sequence and Phylogenetic Analysis

Exon sequences and encoded amino acid sequences were aligned by ClustalX (Jeanmougin et al. 1998) and modified by eye in the editor of PAUP*4.0b8 (Swofford 2001). The alignments of the amino acid sequences encoded by the ancestral sequences for these *mariners* and the other members of the mellifera subfamily used as outgroups are colinear, so there is no ambiguity in their alignments, and the alignment of DNA sequences followed the amino acid alignment. The 5′ and 3′ untranslated regions were first aligned using Clustal X using the default values for gap opening and gap extension penalties with minor manual modifications to increase matches between sequences. These alignments are available in the *Supplementary Material* online.

Phylogenetic analyses were done with PAUP*4.0b8 (Swofford 2001). The equilibrium base frequencies and

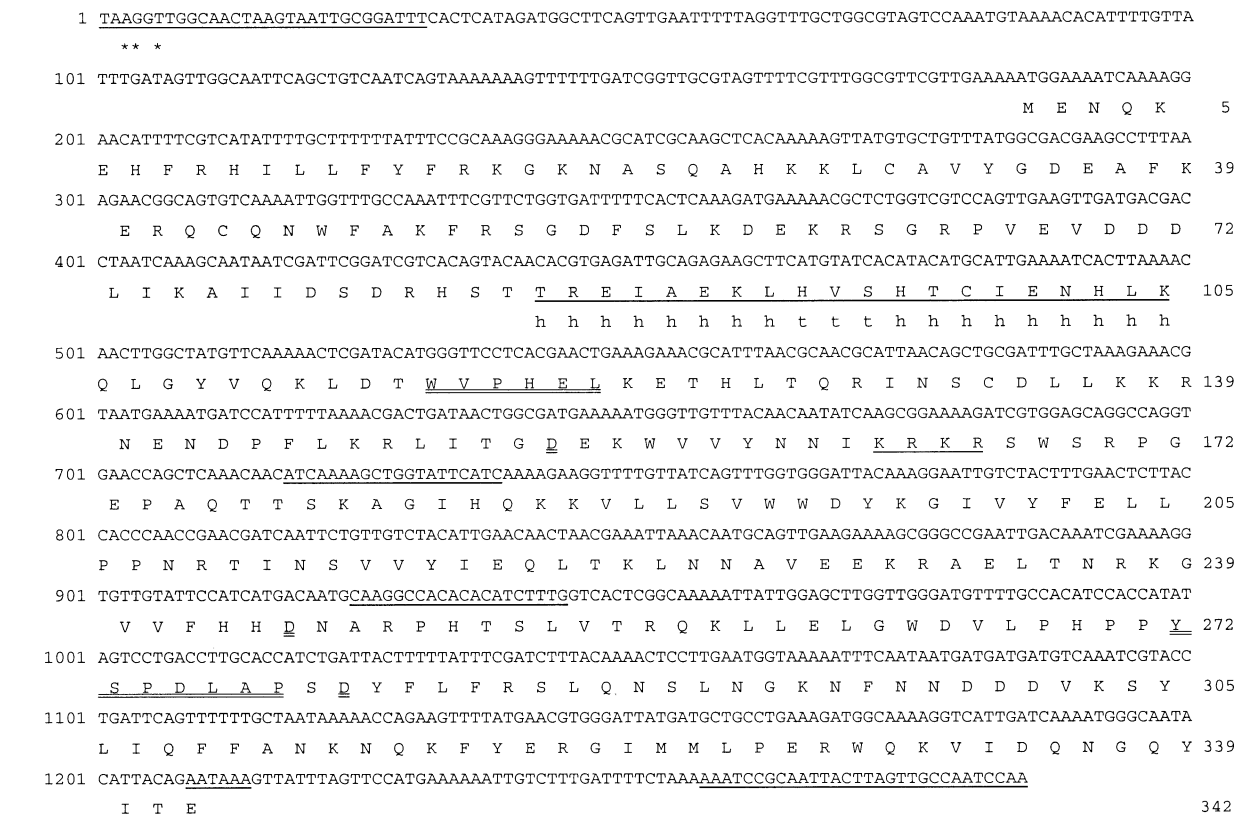

FIG. 2.—Nucleotide and encoded amino acid sequence of the *Famar1* transposable element. Numbers on the left indicate nucleotide position, those on the right indicate amino acid position. The amino acids to which the original fully degenerate PCR primers were designed are double-underlined (WVPHEL at aa positions 116–121 for primer MAR-124F and YSPDLAP at aa positions 272–278 for primer MAR-276R) (Robertson 1993). The nucleotides to which the specific primers MAR-188F and MAR-251R were designed are underlined (bp positions 718–737 and 924–943, respectively). The 30 bp inverted terminal repeats are underlined as is the putative polyadenylation signal (positions 1209–1215). The ITRs contain three mismatches indicated by asterisks under the 5′ ITR nucleotides. The 3′ ITR matches the overall consensus of the elements reported herein. The putative helix-turn-helix motif predicted by Pietrovski and Henikoff (1997) is single underlined (aa positions 84–105). The aspartic acids of the D,D34D catalytic domain are double-underlined (aa positions 153, 245, and 280). The putative nuclear localization signal (KRKR) is underlined (aa positions 163–166).

the transition/transversion ratio (equivalently, kappa or κ) were estimated by maximum likelihood (ML) using an arbitrarily selected most-parsimonious phylogeny resulting from an initial heuristic search (starting tree obtained by stepwise addition, followed by Tree Bisection-Reconnection branch swapping). With these parameters fixed, 10 heuristic ML searches were performed (with the above search settings) to find the most likely phylogeny. This phylogeny was used in all subsequent analyses.

To detect evidence of selection, the CODEML program of PAML 3.0a (Yang 1997) was used in conjunction with likelihood ratio tests (Edwards 1992). Various models of sequence evolution were evaluated by altering the value of the omega (ω essentially the dN/dS) of particular branches, depending on the model (see table 1). These algorithms were run on a Power Macintosh G4 and required 4 to 25 h per run, averaging ~7 h each. Ancestral coding sequences were inferred by ML using CODEML from a model where ω was allowed to freely vary for all branches in the tree (see model J, table 2).

Protein sequence motifs in the transposase amino acid sequence of *Famar1* were identified by using the Web-based programs SMART (Schultz et al. 2000) (http://smart.embl-heidelberg.de/) and PSORT II (Nakai and Horton 1999) (http://psort.ims.u-tokyo.ac.jp/) with the default settings for each. Putative promoter elements in the 5′ untranslated region were predicted by using the Web-based program MatInspector V2.2 (Quandt et al. 1995) (http://transfac.gbf.de/cgi-bin/matSearch/matsearch.pl) using the TRANSFAC 4.0 database of transcription factor binding sites (Wingender et al. 2000).

Results

A PCR Survey of Animals for the Presence of Closely Related Mellifera Subfamily *Mariners*

In earlier surveys, the PCR was used with degenerate primers to screen the genomes of many animal species (Robertson 1993, 1997; Robertson and MacLeod 1993; Robertson et al. 1998) for the presence of *mariner* elements in general. Cloned and sequenced PCR products from *A. mellifera*, *F. auricularia*, *C. capitata*, and *E. funebris* showed that they contained very closely related mellifera subfamily *mariner* sequences, and so these were chosen to obtain genomic clones. Specific primers for these *mariners* were designed from these sequences, and were then used in a PCR to screen all of the ±700 animal

Table 1
Pairwise Amino Acid and Coding Nucleotide Differences Between Ancestral Elements Reconstructed by Maximum Likelihood

Ancestral Element	*Ammar1*	*Camar1*	*Ccmar2*	*Efmar1*	*Famar1*
Amino acid differences					
Ammar1	—	25.1%	4.1%	6.1%	2.0%
Camar1	86	—	25.4%	26.5%	24.8%
Ccmar2	14	87	—	4.1%	2.6%
Efmar1	21	91	14	—	4.7%
Famar1	7	85	9	16	—
Coding sequence differences					
Ammar1	—	24.5%	1.9%	3.3%	1.2%
Camar1	252	—	24.4%	25.1%	24.2%
Ccmar2	20	251	—	2.1%	1.2%
Efmar1	34	258	22	—	2.7%
Famar1	2	249	12	28	—

NOTE.—Above diagonal = % difference; below diagonal = total character differences.

[a] These estimates were taken from model J (see table 2).

species in our earlier surveys. No other instances of closely related *mariners* were found beyond these four species.

Genomic Clones

Full-length genomic sequences from genomic DNA libraries were obtained for the honeybee, the European earwig, the Med fly, and *C. amoena*. The initial screen on the honeybee library yielded five clones with *mariner* copies that shared a 350 bp deletion near the 5′ end, and one seemingly intact copy. This deletion is apparently shared by most of the 450 copies in this genome (Ebert, Hileman, and Nguyen 1995). Therefore the library was screened again with a PCR fragment generated from the single intact clone from the region of the deletion to obtain full-length clones, and only those will be presented here. The sequences of these genomic clones confirm the previous PCR results in that the PCR fragments of Robertson (1993) and Robertson and MacLeod (1993) clearly derive from the *mariners* in these genomes and not some unknown source of contamination. Following the naming scheme proposed for *mariners* in Robertson and Asplund (1996), we name these elements *Ammar1*, *Famar1*, *Ccmar2*, and *Camar1*, respectively (*Ccmar1* is a basal mellifera subfamily *mariner* described by Gomulski et al. 1997), and these names will be applied to their ancestral sequences. Individual copies are indicated by an additional period and unique numeral, e.g., *Ammar 1.6*. Unfortunately, we failed to clone genomic copies of the *Epicauta funebris* elements despite five screenings of two libraries in which we screened over 400,000 phage. Nevertheless we are convinced that they are present in this genome for several reasons. First, full-length sequences can be amplified using a PCR primer designed to the inverted terminal repeats of the other elements. Second, the sequences of six of these full-length PCR fragments differ significantly and consistently from those of the other species, so we believe they represent *Epicauta* genomic copies rather than PCR contaminants from one of the other similar elements in the other species. If they were contaminants, the sequences would nest inside one of the other clades instead of forming a clade of their own (see below). Finally, these sequences agree with the sequences of internal PCR fragments isolated in a separate screen using degenerate primers (Robertson et al. 1998). It may be that these elements reside in a genomic region that is systematically underrepresented in our libraries (Lohe and Hartl 1996), or that the *Epicauta* genome is unusually large and thus we missed the copies by not screening enough phage.

The sequences for all of the genomic clones analyzed in this study have been deposited in GenBank

Table 2
Log Likelihood Values ($\ln\ell$) and Parameter Estimates Under Different Models of Sequence Evolution

Model	P	ω_0	ω_{other}	$\ln\ell$	$\hat{\kappa}$
A	54	0.534	$= \omega_o$	−6132.99	2.12
B	53	1.0	$= \omega_o$	−6163.53	2.51
D	54	0.756	$\omega C = 1.0$	−6158.83	2.37
E	54	0.537	$\omega A = 1.0$	−6133.86	2.13
F	54	0.537	$\omega E = 1.0$	−6134.89	2.14
G	55	0.534	$\omega C_{free} = 0.337$	−6132.94	2.12
H	55	0.536	$\omega A_{free} = 0.001$	−6132.40	2.12
I	55	0.540	$\omega E_{free} = 0.340$	−6132.54	2.12
J	105	ω_{0free}	NA	−6076.16	2.19
K	54	0.528	$\omega Ca_{clade} = 1.0$	−6132.80	2.13
L	54	0.523	$\omega F_{clade} = 1.0$	−6132.82	2.14
M	54	0.501	$\omega Am_{clade} = 1.0$	−6131.62	2.15
N	54	0.422	$\omega Cc_{clade} = 1.0$	−6129.08	2.21
O	54	0.521	$\omega Ef_{clade} = 1.0$	−6131.73	2.12
P	55	0.527	ωCa_{clade} = 1 free value	−6132.60	2.12
Q	55	0.500	ωAm_{clade} = 1 free value	−6130.79	2.13
R	55	0.522	ωFa_{clade} = 1 free value	−6132.31	2.12
S	55	0.417	ωCc_{clade} = 1 free value	−6125.80	2.13
T	55	0.521	ωEf_{clade} = 1 free value	−6131.73	2.12

NOTE.—P: number of parameters in the model; ω_0: background dN/dS ratio for the tree, estimated freely; ω_{other}: dN/dS ratio in the branch(es) under consideration, fixed or estimated separately from the background; $\ln\ell$: log likelihood values; $\hat{\kappa}$: effectively equal to the transition/transversion ratio estimated under the model.

and their accession numbers are listed in *Supplementary Material*.

The *Famar1* Element

The sequence for *Famar1* was reconstructed in two steps. The ancestral coding sequence was inferred by maximum likelihood (model J, see below), and the 5′ and 3′ untranslated regions were inferred by majority rule and comparison to other mellifera subfamily elements to decide ties. The majority rule coding sequence matched the inferred ML ancestral sequence, so the method of checking the majority rule consensus between elements to decide ties appears robust. Hereafter, references to ancestral sequences mean sequences reconstructed in this manner. The *Famar1* sequence thus reconstructed is shown in figure 2 and was chosen to highlight landmarks of these elements, because its five individual copies are the most intact and a confident sequence is readily obtained from them for most regions of the transposon. *Famar1* is 1287 bp in length with 30 bp inverted terminal repeats (ITRs) that contain three mismatches in the first five bp between the 5′ and 3′ ends. The ancestral sequences of the perfect *Ammar1* and *Ccmar2* ITRs have exactly the same sequence as the 3′ end of *Famar1*, so this sequence best represents the ITR for this group of *mariners* (the perfect *Camar1* ITRs differ from this sequence at another 4 bp). No convincing promoter elements were apparent in the 5′ untranslated region when examined with the program MatInspector V2.2 using the TRANSFAC 4.0 database of transcription factor biding sites, except for a very weak similarity to a TATA box at position 128. The stop codon and polyadenylation sequence overlap as they do in many other *mariners* (e.g., Robertson and Lampe 1995; Robertson and Martos 1997; Robertson and Zumpano 1997). The individual clones of *Famar1* vary in the length of a poly-A track near the 3′ end (positions 1253–1260), which may be due to slippage of DNA polymerase (e.g., Schlotterer and Tautz 1992). The copies of *Ammar1* and *Ccmar2* do not exhibit this variation, so the length of this poly-A tract in them was employed in the reconstructed *Famar1*.

The encoded 342 amino acid sequence of the *Famar1* transposase (fig. 2) exhibits features expected of a *mariner* transposase. An N-terminal helix-turn-helix motif predicted by computer models (Pietrokovski and Henikoff 1997) is evident and presumably mediates, at least in part, the binding of this transposase to the ITR sequences (D. Lampe, unpublished data) as similar motifs do for Tc3 transposase (van Pouderoyen et al. 1997). Using the program SMART (Schultz et al. 2000), we could not detect any other significant protein motifs in the sequence. *Famar1* transposase also contains the signature D,D34D catalytic motif, which is *mariner*-specific (Doak et al. 1994; Robertson 1995) and has been shown to be required for transposition activity (Lohe, De Aguiar, and Hartl 1997). Finally, *Famar1* transposase contains a nuclear localization signal (NLS) consisting of four consecutive basic amino acids of a type first described for the SV40 large T-antigen (Rihs, Peters, and Hobom 1991). This kind of NLS is also found in *Himar1* transposase, but not in *Mos1* transposase or in the Tc1-like elements (Lampe et al. 1999; Plasterk, Izsvak, and Ivics 1999), which contain a bipartite NLS.

Comparisons Within Species

To illustrate the changes in these *mariners*, the encoded amino acid sequences of the genomic copies from each species were compared. An alignment of these sequences can be found in the *Supplementary Material* online. Three of the five *Famar1* copies have full-length open reading frames, whereas the other two copies have single base deletions causing frameshifts, and *Famar1.4* also has an encoded stop codon. Each copy also differs from the ancestral sequence by 7–9 amino acids or 2%–2.6%, which may be sufficient to inactivate even those with open reading frames, although which changes might do that is unclear. The 5′ and 3′ untranslated regions of these *Famar1* copies are equally diverged from the ancestral sequence, with two clones exhibiting minor deletions, besides the length variation in the poly-A region overlapping the 3′ ITR mentioned above. The five copies differ from their ancestral sequence by 1.3% on average in full-length DNA sequence, excluding indels.

In contrast, the *Ammar1* copies, and to an even greater extent the *Ccmar2* copies, have suffered numerous incapacitating mutations, averaging 23 and 37 amino acid changes (7% and 12%) from their ancestral sequences, exhibiting an average of 4 and 6 indels that commonly cause frameshifts, and encoding many stop codons. Their 5′ and 3′ untranslated regions are similarly highly mutated from their ancestral sequences, and some copies have also suffered terminal truncations. The *Ammar1* copies differ from their ancestors by 3.4% and the *Ccmar2* copies by 5.8%, on average, in DNA sequence**.** Note that this is a small subset of these *mariner* copies in the honeybee, the majority sharing a 350 bp deletion near the 5′ end (Ebert, Hileman, and Nguyen 1995).

The clones of *Efmar1* from *Epicauta funebris* are not full length because they were obtained via PCR using an inverted repeat primer designed from the other *mariners* in the horizontal transfer clade. Repeated attempts at cloning copies of this element from genomic libraries failed. The individual copies from this species differ from the inferred ancestral sequence by only 0.5% to 2.1% at the DNA level in their coding sequences. Two of the copies, Efmar1.7 and Efmar1.10, have the same deletion near their 3′ ends, indicating that one might be derived from the other.

The four copies of *Camar1* from *Chymomyza amoena* were obtained to provide a reasonably closely related outgroup. They appear to be even younger in this genome than those in the earwig, in that three have full-length open reading frames and they differ from their ancestral sequences by 6 amino acids, or 1.7% on average. The four copies differ from their ancestor by 0.9% on average in DNA sequence.

Comparisons Between Species

The ancestral coding sequences of these earwig, honeybee, beetle, and Med fly *mariners* are remarkably

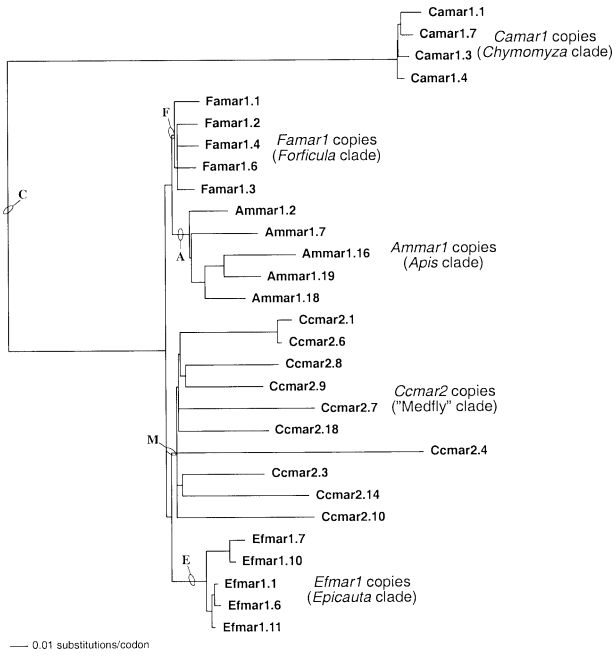

FIG. 3.—A phylogenetic tree of individual copies of the elements involved in a recent horizontal transfer in the mellifera subfamily. This tree was produced using the coding sequence for each copy and the maximum likelihood algorithm (heuristic search mode) of PAUP* v.4.0b8a (Swofford 2001). Branches are scaled according to the branch lengths produced under model J (see table 2) using the CODEML algorithm of PAML. Circled branches are those branches thought to represent evolution of the elements that underwent horizontal transfer into each of these species (i.e., the transfer lineages).

similar to one another, with the encoded transposase of *Famar1* differing from those of *Ammar1*, *Ccmar1*, and *Efmar1* by just 7, 9, and 16 amino acids, respectively. Pairwise comparisons of the ancestral DNA coding and amino acid sequences are listed in table 1. All four differ from the *Camar1* transposase by an average of 87 changes or 25%.

This extraordinary level of similarity extends to the DNA level, where *Famar1* differs from *Ammar1*, *Ccmar1*, and *Efmar1* by just 18, 20, and 28 changes, respectively, over their full length (about 1.5%), while *Ammar1* and *Ccmar1* differ at 28 positions or by 2.2%. They have therefore also diverged very little from one another, even in their third codon positions and untranslated regions. Again, all four differ from *Camar1* by ±29%.

Are There Related *Mariners* in Closely Related Species?

A prediction of the hypothesis of relatively recent horizontal transfer into each of these disparate insect lineages is that the closest relatives of each species should have the same *mariner* shared by vertical inheritance, whereas more distantly related congeners and non-congeners should not (e.g., Robertson and Lampe 1995), unless there have been additional horizontal transfers within a genus (e.g., Maruyama and Hartl 1991). In other words, the distribution of a particular *mariner* should be spotty, reflecting the consequences of the timing of both speciation and horizontal transfer. We attempted to document this spotty distribution by analyzing several species related to each of the four focus species by screening them qualitatively for the presence of these *mariners* by PCR using specific primers MAR-188F and MAR-251R.

In the case of *F. auricularia*, two other congeneric species, *F. lesnei* and *F. decipiens*, as well as *Euborelia* sp. were tested for the presence of mellifera-subfamily *mariners*. Of these three species, only *F. lesnei* produced amplification of the size expected based on the band observed in *F. auricularia*. This is somewhat unexpected because *F. decipiens* is as divergent from *F. auricularia* as is *F. lesni* lineage (Wirth, Le Guellec, and Veuille 1999), and may indicate that the *Famar1* element was lost from the *F. decipiens* lineage. The congeners of the honeybee, *A. andreniformis*, *A. cerana*, *A. dorsata*, and *A. flora* yielded several bands on agarose gels, but none of them corresponded to the expected size based on the amplification product in *A. mellifera*. This *mariner* therefore appears to have entered *A. mellifera* after its separation from these congeners; alternatively it has been lost from all the congeners, which is a less parsimonious explanation for the data. A third possibility is that the bands we detected represent internally deleted forms of the element, much as the predominant form in *A. mellifera* is internally deleted near the 5′ end. Because of the difficulty in obtaining specimens, close relatives of *C. capitata* were not tested for the presence of these *mariners*, but, with the specific PCR test, distant relatives in the genus *Rhagoletis* (*R. pomonella*, *R. mendax*, and *R. cornivora*) all produced negative results for the presence of closely related mellifera subfamily *mariners*. Congeners of the blister beetle *Epicauta funebris* were also examined. *E. pennsylvanicus* yielded no closely related *mariners*, but *E. vittata*, a close relative, showed positive results. Overall, the PCR results point to a spotty distribution of these particular *mariners*, in keeping with the predictions of the horizontal transfer hypothesis.

Phylogenetic Analysis

The relationships of the four recently horizontally transferred *mariners* are shown in the tree in figure 3. *Famar1*, *Ammar1*, *Ccmar2*, and *Efmar1* form a clade in which most of the changes are contained in individual terminal branches. Branches circled in the figure are those leading to the ancestral elements, and these branches are for the most part very short, reflecting the close relationships between these transposons.

Where Does Selection Act in *Mariner* Evolution?

We used likelihood ratio tests (LRT) to examine the molecular evolution of particular lineages in the phylogenetic tree shown in figure 3 after the method of Yang (1997). The data set we used can be found in *Supplementary Material* online. From the analysis of other *mariners*, we expected to find a mixture of neutral evolution and selection in the overall tree, depending on which lineage was examined. In particular, we hypothesized that selection would exist in the branches leading to the clades of each individual *mariner* (labeled branches in figure 3),

Table 3
Tests of Hypotheses Using Likelihood Ratio Tests

Null Hypothesis	Models Compared	$2\Delta\ell$	Conclusion
$\omega_0 = 1$	A & B	61.08	There is selection in the tree.
$\omega C = 1$	D & G	51.78	There is selection in branch C.
$\omega A = 1$	E & H	2.92	There is no selection in branch A.*
$\omega E = 1$	F & I	4.7	There is selection in branch E.
$\omega C_{clade} = 1$	K & P	0.4	Copies in the *Chymomyza* clade are evolving neutrally.
$\omega Fa_{clade} = 1$	L & R	1.02	Copies in the *Forficula* clade are evolving neutrally.
$\omega Am_{clade} = 1$	M & Q	1.66	Copies in the *Apis* clade are evolving neutrally.
$\omega Cc_{clade} = 1$	N & S	6.56	At least some lineages in the Med fly clade have been under selection ($0.01 < P < 0.05$).
$\omega Ef_{clade} = 1$	O & T	0.0	Copies in the Epicauta clade are evolving neutrally.

NOTE.—The models compared are from table 2. The likelihood ratio statistic ($2\Delta\ell$) is approximated by the χ-squared distribution.

* $0.10 > P > 0.05$. This may indicate evidence of selection that is biologically relevant but statistically insignificant. There are very few changes in this branch (10.5 changes), but the tendency is toward an excess of synonymous changes.

and that individual copies within clades would be evolving neutrally. These hypotheses reflect our current understanding of how *mariners* evolve, namely neutrally within a population of organisms, with selection occurring only during a horizontal transfer event (Lohe et al. 1995; Robertson and Lampe 1995), a property these elements share with other DNA mediated elements (Witherspoon 1999; Silva and Kidwell 2000). Likelihood methods differ significantly from our previous methods to examine the molecular evolution of *mariners* because they do not rely on the reconstruction of consensus sequences and can also take into account the transition/transversion rate bias and nonuniform codon usage. Log-likelihood values for different models of sequence evolution are shown in table 2. Tests of hypotheses using these values are shown in table 3.

We hypothesized first that there should be evidence of selection in the entire tree. This hypothesis was tested by comparing models of sequence evolution where a single background level of ω ($=\omega_0$) was estimated for the entire tree (model A) with a model that fixed ω_0 for all branches in the tree at 1.0, which is the value expected for neutral evolution (model B). A LRT comparing these two models indicates that by allowing one more free parameter (ω_0 not fixed) the fit of the model to the data is significantly improved, so we can decisively reject model B ($P \ll 0.001$; table 3). This indicates that there is selection in the tree.

We next tested the hypothesis that elements that undergo horizontal transfer will show evidence of selection. These elements are the ancestral transposons that give rise to each clade of elements within each species (represented by the node at the base of each species clade in figure 3). Branches that lead to these elements are those that were tested for evidence of selection and are circled and labeled in the tree in figure 3. A problem arises in testing this hypothesis, however, because several of these branches have very few codon changes of any kind. For example, branch F (leading to the *F. auricularia* clade) is estimated to contain only one nonsynonymous change and one synonymous change. Similarly, branch M (leading to the Med fly clade) contains a total of only three nonsynonymous changes. Indeed, even branch A (leading to the *Apis* clade) contains only a total of 10.5 changes, the bare minimum number for a test of significance (with fewer changes, the assumptions underlying the LRT are not valid). We therefore restricted the tests of this hypothesis to branches C, A, and E (containing an estimated 408.9, 10.5, and 20 changes, respectively). Selection was easily detected in branch C ($P \ll 0.01$). No selection could be detected in branch A that is statistically significant ($0.1 > P > 0.05$). As noted earlier, this branch only contained an estimated 10.5 changes, so the selection signal might be biologically relevant even if it is not statistically significant. Finally, selection was also detected in branch E ($0.05 > P > 0.01$).

We also looked for evidence of selection within each clade. Within each clade all branches were fixed such that $\omega = 1.0$ (models K-O, table 2). These models were compared to ones where a single ω was freely estimated for the entire clade, and that value was applied to each branch (models P-T, table 2). The results of these tests are shown in table 3. No selection could be detected (i.e., individual copies were evolving neutrally) in the *Chymomyza*, *Forficula*, *Apis*, and *Epicauta* clades. However, selection was detected within the Med fly clade ($0.05 > P > 0.01$), which was unexpected.

To determine which branches were likely to be responsible for the selection signal we detected in the Med fly clade, we calculated the binomial cumulative probability of observing zero up to the number of nonsynonymous changes estimated under a model where ω for each branch was allowed to vary freely for the entire tree (model J), given the proportion of such changes that would be expected from random nucleotide changes to the sequence. We did this in preference to performing likelihood ratio tests branch by branch within the clade because of the amount of computer time necessary to estimate parameters for each model (there are 16 branches within the Med fly clade). Five branches within the Med fly clade produced binomial probabilities below 0.05, but only one of these can be considered significant by the multiple tests criterion. This is the branch leading to the Ccmar2.3 copy. Other branches have "low" dN/dS ratios and, even though these are not significant, they could

Table 4
Models of Sequence Evolution in a Subset of the Tree Shown in Figure 3 (the Med fly clade)

Model	P	ω_0	κ	$\ln\ell$
A′	18	0.745	2.12	−3427.45
B′	17	1.0	2.28	−3430.28

NOTE.—The parameters are those described in table 2. Model B′ fixes ω at 1.0 which constrains the model to one of neutral evolution.

contribute to the selection signal we detected in the LRT in the overall clade.

One explanation for the unexpected evidence of selection in the Med fly clade is that there is something "different" about Med fly sequence evolution that is providing a false selection signal (e.g., a different κ for this clade in contrast to the overall data set). To determine if the selection signal was real, we removed the Medfly clade from the overall data set, retaining the tree topology from figure 3, and reran the new data set to test for selection in the Medfly treelet (models A′ and B′, table 4). Using these models in a LRT, a very small selection signal can still be detected ($2\Delta l = 2.82$), but it is not statistically significant ($0.10 > P > 0.05$).

In general, it seems that as with most other *mariners* examined (e.g., Robertson and Lampe 1995; Robertson and Martos 1997; Robertson and Zumpano 1997), most evolution of copies within species is neutral, with rapid accumulation of incapacitating mutations, after which they must evolve as pseudogenes.

Discussion

We present a detailed look at the characteristics and evolution of five different transposable elements in one of the major subfamilies in the *mariner* family of transposable elements, the mellifera subfamily. Four of these elements (*Famar1*, *Ammar1*, *Ccmar2*, and *Efmar1*) are very closely related to each other and represent instances of horizontal transfer of virtually identical transposons into very different insect hosts. Each copy of these elements is evolving neutrally within its hosts (perhaps with the exception of a single *Ccmar2* copy), but ancestral lineages (i.e., those undergoing horizontal transfer) do show evidence of selection where there are enough codon changes to test for it. This scenario reinforces our current hypothesis about the evolution of *mariners*, namely that it is dominated by neutral evolution and drift, with selection intervening only when a copy of an element manages to undergo horizontal transfer. That selection acts on *mariners* only when they undergo horizontal transfer means that this feature of their life cycle is absolutely essential to their survival as genetic entities. How often they manage to undergo this phenomenon, and which copies are actually able to make the jump is uncertain (and probably largely stochastic), but some insight into this can be gained by looking at a genome that may contain a mix of active and inactive copies. Our data set indicates that *F. auricularia* at least fits this description because some of its copies, although divergent from the ancestral sequence, appear intact. Data to be published elsewhere demonstrate that some of these do indeed encode functional transposases. Thus it is possible that most of the differences we saw between the ancestral sequence of these four *mariners* in four different insects were in fact accumulated before they underwent horizontal transfers, in which case the rate of horizontal transfer might not be as high as is sometimes speculated (e.g., Robertson et al. 1998). This possibility fits somewhat better with our observation that closely related *mariners* were only discovered in these four insects among approximately 700 animal species examined.

Supplementary Material

Accession numbers for sequences reported here are as follows: *Camar1* copies AY154747–AY154750; *Ammar1* copies AY154751–AY154755; *Efmar1* copies AY154756–AY154760; *Famar1* copies AY154761–AY154765; *Ccmar2* copies AY154766–AY154775). Alignments used in this study are available as Supplementary Material online.

Acknowledgments

We thank M. Tanaka, D. Smith, T. Wirth, J. Carroll, and S. Berlocher for the gift of insect specimens, M. Asplund for assistance in sequencing clones, and two anonymous reviewers for comments. This work was supported by National Institutes of Health grant 2R01AI33586 to D.J.L. and H.M.R. and National Science Foundation grant MCB-0091044 to D.J.L.

Literature Cited

Clark, J. B., and M. G. Kidwell. 1997. A phylogenetic perspective on P transposable element evolution in Drosophila. Proc. Natl. Acad. Sci. USA **94**:11428–11433.

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**:339–355.

Doak, T. G., F. P. Doerder, C. L. Jahn, and G. Herrick. 1994. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc. Natl. Acad. Sci. USA **91**:942–946.

Ebert, P. R., J. P. T. Hileman, and H. T. Nguyen. 1995. Primary sequence, copy number, and distribution of *mariner* transposons in the honey bee. Insect. Mol. Biol. **4**:69–78.

Edwards, A. W. F. 1992. Likelihood. Johns Hopkins University Press, Baltimore.

Garcia-Fernandez, J., J. R. Bayascas-Ramirez, G. Marfany, A. M. Munoz-Marmol, A. Casali, J. Baguna, and E. Salo. 1995. High copy number of highly similar *mariner*-like transposons in planarian (Platyhelminthe): evidence for a trans-phyla horizontal transfer. Mol. Biol. Evol. **12**:421–431.

Hartl, D. L., A. R. Lohe, and E. R. Lozovskaya. 1997. Modern thoughts on an ancient *mariner*: function, evolution, regulation. Annu. Rev. Genet. **31**:337–358.

Hartl, D. L., E. R. Lozovskaya, D. I. Nurminsky, and A. R. Lohe. 1997. What restricts the activity of *mariner*-like transposable elements. Trends Genet. **13**:197–201.

Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998. Multiple sequence alignment with ClustalX. Trends Biochem. Sci. **23**:403–405.

Kidwell, M. G. 1992. Horizontal transfer. Curr. Opin. Genet. Dev. **2**:868–873.

Lampe, D. J., B. J. Akerley, E. J. Rubin, J. J. Mekalanos, and H. M. Robertson. 1999. Hyperactive transposase mutants of the Himar1 *mariner* transposon. Proc. Natl. Acad. Sci. USA **96**: 11428–11433.

Lampe, D. J., M. E. Churchill, and H. M. Robertson. 1996. A purified *mariner* transposase is sufficient to mediate transposition in vitro. EMBO J. **15**:5470–5479.

Lampe, D. J., T. E. Grant, and H. M. Robertson. 1998. Factors affecting transposition of the Himar1 *mariner* transposon in vitro. Genetics **149**:179–187.

Lampe, D. J., K. K. O. Walden, and H. M. Robertson. 2001. Loss of transposase-DNA interaction may underlie the divergence of *mariner*-family transposable elements and the ability of more than one *mariner* to occupy the same genome. Mol. Biol. Evol. **18**:954–961.

Lohe, A. R., D. De Aguiar, and D. L. Hartl. 1997. Mutations in the *mariner* transposase: the D,D(35)E consensus sequence is nonfunctional. Proc. Natl. Acad. Sci. USA **94**:1293–1297.

Lohe, A. R., and D. L. Hartl. 1996. Germline transformation of *Drosophila virilis* with the transposable element *mariner*. Genetics **143**:365–374.

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.

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

Nakai, K., and P. Horton. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. **24**:34–36.

Pietrokovski, S., and S. Henikoff. 1997. A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. Mol. Gen. Genet. **254**:689–695.

Plasterk, R. H., Z. Izsvak, and Z. Ivics. 1999. Resident aliens: the *Tc1/mariner* superfamily of transposable elements. Trends Genet. **15**:326–332.

Quandt, K., K. Frech, H. Karas, E. Wingender, and T. Werner. 1995. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. **23**:4878–4884.

Rihs, H. P., R. Peters, and G. Hobom. 1991. Nuclear localization of budgerigar fledgling disease virus capsid protein VP2 is conferred by residues 308–317. FEBS Lett. **291**:6–8.

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

———. 1995. The *Tc1-mariner* superfamily of transposons in animals. J. Insect Physiol. **41**:99–105.

———. 1997. Multiple *mariner* transposons in flatworms and hydras are related to those of insects. J. Hered. **88**:195–201.

Robertson, H. M., and D. J. Lampe. 1995. Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. Mol. Biol. Evol. **12**:850–862.

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.

Robertson, H. M., and R. Martos. 1997. Molecular evolution of the second ancient human *mariner* transposon, *Hsmar2*, illustrates patterns of neutral evolution in the human genome lineage. Gene **205**:219–228.

Robertson, H. M., F. N. Soto-Adames, K. K. O. Walden, R. M. P. Avancini, and D. J. Lampe. 1998. The *mariner* transposons of animals: horizontally jumping genes. Pp. 268–284 *in* M. Syvanen and C. I. Kado, eds. Horizontal gene transfer. Chapman and Hall, London.

Robertson, H. M., and K. L. Zumpano. 1997. Molecular evolution of an ancient *mariner* transposon, *Hsmar1*, in the human genome. Gene **205**:203–217.

Rubin, E. J., B. J. Akerley, V. N. Novik, D. J. Lampe, R. N. Husson and J. J. Mekalanos. 1999. In vivo transposition of *mariner*-based elements in enteric bacteria and mycobacteria. Proc. Natl. Acad. Sci. USA **96**:1645–1650.

Schlotterer, C., and D. Tautz. 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Res. **20**:211–215.

Schultz, J., R. R. Copley, T. Doerks, C. P. Ponting, and P. Bork. 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. **28**:231–234.

Silva, J. C., and M. G. Kidwell. 2000. Horizontal transfer and selection in the evolution of P elements. Mol. Biol. Evol. **17**:1542–1557.

Swofford, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Mass.

Syvanen, M., and C. Kado, eds. 1998. Horizontal gene transfer. Chapman and Hall, London.

van Pouderoyen, G., R. F. Ketting, A. Perrakis, R. H. Plasterk, and T. K. Sixma. 1997. Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. Embo J. **16**:6044–6054.

Wingender, E., X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. Pruss, I. Reuter, and F. Schacherer. 2000. TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. **28**:316–319.

Wirth, T., R. Le Guellec, and M. Veuille. 1999. Directional substitution and evolution of nucleotide content in the cytochrome oxidase II gene in earwigs (Dermapteran insects). Mol. Biol. Evol. **16**:1645–1653.

Witherspoon, D. J. 1999. Selective constraints on P-element evolution. Mol. Biol. Evol. **16**:472–478.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**: 555–556.

Zhang, J. K., M. A. Pritchett, D. J. Lampe, H. M. Robertson, and W. W. Metcalf. 2000. In vivo transposon mutagenesis of the methanogenic archaeon Methanosarcina acetivorans C2A using a modified version of the insect *mariner*-family transposable element *Himar1*. Proc. Natl. Acad. Sci. USA **97**: 9665–9670.

Zhang, L., U. Sankar, D. J. Lampe, H. M. Robertson, and F. L. Graham. 1998. The *Himar1 mariner* transposase cloned in a recombinant adenovirus vector is functional in mammalian cells. Nucleic Acids Res. **26**:3687–3693.

Thomas Eickbush, Associate Editor

Accepted November 26, 2002