A Bacterial Genetic Screen Identifies Functional Coding Sequences of the Insect *mariner* Transposable Element *Famar1* Amplified From the Genome of the Earwig, *Forficula auricularia*

Elizabeth G. Barry,[*,1] **David J. Witherspoon**[†,1] **and David J. Lampe**[*,2]

*Department of Biological Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282 and †Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah 84112

ABSTRACT

Transposons of the *mariner* family are widespread in animal genomes and have apparently infected them by horizontal transfer. Most species carry only old defective copies of particular *mariner* transposons that have diverged greatly from their active horizontally transferred ancestor, while a few contain young, very similar, and active copies. We report here the use of a whole-genome screen in bacteria to isolate somewhat diverged *Famar1* copies from the European earwig, *Forficula auricularia*, that encode functional transposases. Functional and nonfunctional coding sequences of *Famar1* and nonfunctional copies of *Ammar1* from the European honey bee, *Apis mellifera*, were sequenced to examine their molecular evolution. No selection for sequence conservation was detected in any clade of a tree derived from these sequences, not even on branches leading to functional copies. This agrees with the current model for *mariner* transposon evolution that expects neutral evolution within particular hosts, with selection for function occurring only upon horizontal transfer to a new host. Our results further suggest that *mariners* are not finely tuned genetic entities and that a greater amount of sequence diversification than had previously been appreciated can occur in functional copies in a single host lineage. Finally, this method of isolating active copies can be used to isolate other novel active transposons without resorting to reconstruction of ancestral sequences.

*M*ARINER transposable elements are a large and diverse family of small eukaryotic transposons. These elements are ~1.3 kb in length and encode a single protein, the *mariner* transposase, that allows them to mobilize their DNA in virtually any eukaryotic genetic background through a cut-and-paste mechanism. When expressed with the appropriate promoters, these elements are active in Bacteria and Archaea as well (RUBIN *et al.* 1999; ZHANG *et al.* 2000).

In Eukarya, organisms as diverse as Hydra and humans contain *mariners* and often more than one type (where "type" refers to all copies descended from a single horizontally transferred ancestral element in a population of host organisms). The human genome, for example, contains two distinct types of *mariner* elements (ROBERTSON and MARTOS 1997; ROBERTSON and ZUMPANO 1997) while that of *Caenorhabditis elegans* contains nine (WITHERSPOON and ROBERTSON 2003). When these and similar elements are isolated through PCR, screens of

genomic libraries, or genomic sequencing projects, their transposase genes are almost always found to be defective in some way, due to the presence of either numerous indels or in-frame stop codons that would inactivate them. Thus, although we know of many hundreds of different inactive *mariners* through their sequences, only two have been demonstrated to be functional. The first of these is *Mos1*, a particularly active copy of the original *mariner* element isolated from *Drosophila mauritiana* (MEDHORA *et al.* 1991). The second is *Himar1*, reconstructed as a consensus from many copies of a type of *mariner* element found in the horn fly, *Haematobia irritans* (LAMPE *et al.* 1996). Both of these elements have found wide use as genetic tools in a broad variety of species in all domains of life (LOHE and HARTL 1996b; GUEIROS-FILHO and BEVERLEY 1997; COATES *et al.* 1998, 2000; RUBIN *et al.* 1999; ZHANG *et al.* 2000).

Despite our knowledge of many different *mariners*, we do not yet fully understand how they evolve or persist as functional genetic entities. Any model of their evolution must explain their activity, distribution, and diversity. A general model for *mariner* evolution has been developed and modified by several research groups (ROBERTSON and LAMPE 1995; HARTL *et al.* 1997c; LAMPE *et al.* 2003). This model posits that a single copy of a *mariner* transposon invades the germline of an organism by an unknown mechanism, perhaps by hitchhiking on viruses (HARTL

---

et al. 1997a). This founder element increases in copy number through transposition and spreads through the population via sexual reproduction. As copy number increases, the rate of transposition may slow, as intrinsic (*e.g.,* "overproduction inhibition"), emergent (*e.g.,* dominant-negative inhibition by defective transposases; HARTL *et al.* 1997c), or host-mediated (*e.g.,* transcriptional repression of repeated sequences) regulatory mechanisms take effect or evolve *de novo*. Random mutations accumulate in all element lineages, and since the transposase produced by functional elements acts at random on any recognizable copies in the same genome, there is no effective selection against nonfunctional yet recognizable copies. Eventually, mutations inactivate all copies, and it is this final deteriorated state in which we find most *mariner* sequences.

Only by undergoing repeated horizontal transfer can some *mariners* escape this fate and stay ahead of mutation, drift, and emergent regulatory mechanisms. These horizontal transfers are selection events, since only active elements will be able to establish new lineages in new populations (ROBERTSON and LAMPE 1995; WITHERSPOON 1999; SILVA and KIDWELL 2000; LAMPE *et al.* 2003). Nonsense or missense mutations in typical genes are immediately tested by natural selection, but several such mutations may accumulate with no consequence in the transposase gene of a *mariner* lineage before being tested by the next horizontal transfer event. The divergence between successive founder elements could be significant, so this process of drift interspersed with selection events may accelerate the generation of *mariner* diversity (LAMPE *et al.* 2001).

Some organisms contain *mariners* that, although diverged in sequence from their putative ancestral element, may still be active because at least some copies appear intact (GOMULSKI *et al.* 1997; LAMPE *et al.* 2003; WITHERSPOON and ROBERTSON 2003). The European earwig, *Forficula auricularia*, a common household and garden insect in North America, is one such organism (LAMPE *et al.* 2003). *F. auricularia* contains many copies of *Famar1*, an element in the mellifera subfamily of *mariners* and very closely related (~98% amino acid identity for their ancestral sequences) to the *Ammar1* of the honeybee, *Apis mellifera*. Individual copies of *Famar1* differ from each other on average by ~2.3% at the amino acid level (LAMPE *et al.* 2003). Although some of these copies appear intact, it is not obvious which are active and which contain inactivating point mutations. This situation is ideal for studying the later steps of the evolution of a *mariner* transposon in a host population, because it represents neither a very recent colonization where all of the copies are nearly identical (*e.g.,* CAPY *et al.* 1992; GARCIA-FERNANDEZ *et al.* 1995) nor a terminal stage in which all copies are inactive.

We report here the application of bacterial genetic methods to discriminate between functional and non-

functional coding sequences of *Famar1* obtained via genomic PCR. By sequence analyses we find that the transposase genes of *Famar1*, including those still encoding a functional transposase, evolve under no selection for the conservation of their function (*i.e.,* neutrally) within *F. auricularia*.

MATERIALS AND METHODS

Media and antibiotics: Bacteria were grown at the temperatures indicated in Luria broth (LB) or on agar plates as described (SAMBROOK *et al.* 1989). Papillation assays were performed on thick (~50 ml) MacConkey lactose agar plates. Antibiotic concentrations were ampicillin (Amp), 100 μg; gentamicin (Gen), 10 μg; tetracycline (Tet), 15 μg; naladixic acid (Nal), 20 μg; chloramphenicol (Cam), 34 μg; streptomycin (Str), 75 μg; and apramycin (Apr), 80 μg/ml, respectively, unless otherwise noted.

Plasmids and bacterial strains: Plasmids and bacterial strains used in this study are listed and described briefly in Table 1. A *Famar1* "minimariner" with an open reading frame (ORF) through one inverted terminal repeat (ITR) was constructed by amplifying pminiAm (LAMPE *et al.* 2001) with the primers T7 (5′-TAATACGACTCACTATAGGG) and AmORF-R2 (5′-CAGATCTGAGTGAAATCCGCAATTAC) in one PCR and the primers SP6 (5′-CGATTTAGGTGACACTATAG) and Am1212f (5′-AAATGGGCAACACATTACAGAA) in another PCR. The PCR products were kinased and ligated together as described previously (LAMPE *et al.* 2001). The ligation reaction was used in another PCR using the primers T7 and SP6 to recover a product that had two ITRs that were distinct. This product was cut with the restriction enzymes *Xho*I and *Hin*dIII and cloned into the identical sites of pCDNAII, resulting in the plasmid pFaORF.

To make a *Famar1* plasmid capable of mediating papillation in *Escherichia coli*, pFaORF was cut with *Bgl*II and the *Bam*HI/*Bgl*II fragment of pRZ1495 containing the *lacZYA* and *Tet*^R genes ligated to it. This ligation creates a fusion of the open reading frame in one ITR of FaORF and *lacZ* from pRZ1495 at amino acid coding position 8 of the *lacZ* gene. This fusion created pFaLacTet.

A single-copy plasmid containing the *lacZ*/FaORF fusion was created in two steps. First, pFaLacTet was digested with *Not*I, the overhanging ends made blunt with the Klenow fragment of *E. coli* DNA polymerase I, and the DNA cut again with *Bam*HI. This fragment was ligated to the plasmid pACMarKan that had been digested with *Bst*EII, made blunt as above, and digested again with *Bgl*II. The resulting plasmid, pACFaLacTet, contains the FaLacTet papillation element inside the ITRs of the *Himar1* transposon on a p15a origin of replication. *Himar1* carrying FaLacTet was mobilized onto a matable F plasmid by cotransformation of RZ212 *E. coli* with pACFaLacTet and pBADC9. The cells were selected with Amp/Tet/Gen. The transposase encoded by pBADC9 can mobilize the *Himar1* carrying FaLacTet off pACFaLacTet and onto the F plasmid, which is matable. These cells were mated to DH5α *E. coli* and recipient cells containing an insertion of *Himar1* on the F selected with Nal/Tet. The resulting strain, DL13, was used in a papillation assay to detect functional and nonfunctional coding sequences of *Famar1* and *Ammar1*.

Mating-out strains of *E. coli* for both *Famar1* and *Himar1* were constructed as follows. The *Apr*^R cassette from pOJ427 was inserted as a *Bam*HI/*Xba*I blunt fragment into the *Bgl*II site of pFaORF to make pEB17. The *Bam*HI restriction fragment from pEB17 containing the ITRs and *Apr*^R was moved

TABLE 1

Bacterial plasmids and strains used in this study

Plasmid or strain	Description	Reference
Plasmid		
pFaOrf	Inverted terminal repeat sequences of *Famar1*, open reading frame through one end	This study
pFaLacTet	pFaOrf containing lacZYA and Tet^R gene from pRZ1495	This study
pACFaLacTet	FaLacTet *Famar1* transposon cloned onto pACYC184	This study
pmmORF	Inverted terminal repeat sequences of *Himar1* with internal cloning sites	Lampe *et al.* (1999)
pminiAM	A mini-*Ammar1* plasmid	Lampe *et al.* (2001)
pACMarKan	Kan^R *Himar1* on p15a ori plasmid	Lampe *et al.* (1999)
pBADC9	Hyperactive version of *Himar1* transposase	Lampe *et al.* (1999)
pRZ1495	Tn5 papillation factor: the source of *lacZYA*-Tet^R	Makris *et al.* (1988)
pACYC184	p15a ori cloning vector	Chang and Cohen (1978)
pCDNAII	ColE1 ori cloning vector	Invitrogen (San Diego)
pOJ427	Source of the *acc(3)IV* (*Apr*^{R}) gene cassette	B. Schoener (unpublished results); see also Kieser *et al.* (2000)
pEB17	*Apr*^{R} cassette from pOJ427 inserted as *Bam*HI/*Xba*I blunt fragment into pFaORF at *Bgl*II site	This study
pEB18	*Apr*^{R} cassette from pOJ427 inserted as *Bam*HI/*Xba*I blunt fragment into pmmORF at *Bgl*II site	This study
pEB19	ITRs and *Apr*^{R} from pEB17 inserted as *Bam*HI fragment into *Bam*HI site of pACYC184; mating-out plasmid for *Famar1*	This study
pEB20	ITRs and *Apr*^{R} from pEB18 inserted as *Bam*HI fragment into *Bam*HI site of pACYC184; mating-out plasmid for *Himar1*	This study
pBAD24	*E. coli* expression vector	Guzman *et al.* (1995)
pBADFa1-pBADFa40	Coding sequences of *Famar1* cloned into the *Nde*I/*Pst*I restriction sites of pBAD24	This study
pBADAm1-pBADAm53	Coding sequences of *Ammar1* cloned into the *Nde*I/*Pst*I restriction sites of pBAD24	This study
pDL129	Ancestral *Famar1* coding sequence in pBAD24	This study
Strain		
DH5α	*endA1 hsdR17* (r$_k^-$ m$_k^+$) *supE44 thi-1 recA1 gyrA*(Nal^R) *relA1* Δ(*lacIZYA-argF*) *U169 deoR* (φ*80dlacΔ(lacZ)M15*)	Hanahan (1983)
RZ212	Δ*(lac-pro) ara str recA56 srl thi/* pOX38-Gen	Johnson and Reznikoff (1984)
RZ221	*polA* Δ*(lac-pro) ara str nal*	Johnson *et al.* (1982)
EB8	RZ212/pOX38-Gen(F′)/pEB19; *Famar1* mating-out strain	This study
EB9	RZ212/pOX38-Gen(F′)/pEB20; *Himar1* mating-out strain	This study
DL13	DH5α F′::miniFamar1LacTet from pFaLacTet; *Famar1* papillation strain	This study

into pACYC184 at the *Bam*HI site to make pEB19. EB8 *E. coli* cells conferring Gen^R, Cam^R, and Apr^R were produced by transforming pEB19 into electrocompetent RZ212 *E. coli* containing pOX38-Gen (F) (Johnson and Reznikoff 1984).

Identical versions of the above plasmids were constructed for *Himar1*. The *Apr*^{R} cassette was inserted into a blunt *Bgl*II site of pmmOrf (Lampe *et al.* 1999) as above to produce pEB18, and the ITRs and *Apr*^{R} were moved onto pACYC184 by inserting a *Bam*HI fragment of pEB18 into the *Bam*HI site to make pEB20. The *E. coli* strain EB9 was produced by transforming pEB20 into electrocompetent RZ212 cells containing pOX38-Gen (F′) as described above.

An *E. coli* genetic screen for *Famar1* sequences encoding functional transposase: The genome of *F. auricularia* was

screened for the presence of functional and nonfunctional coding sequences of *Famar1* using the PCR and a *Famar1*-specific *E. coli* papillation screen similar to the one developed for *Himar1* (Lampe *et al.* 1999). Copies of *Famar1* were amplified from genomic *F. auricularia* DNA using the primers Famar7F-2 (5′-TTGAACCATGGAAAATCAAAAGGAAC) and Famarstop (5′-TTCTGCAGTCATGGAACTAAATAACTTTA) based on the consensus sequence of *Famar1* (Lampe *et al.* 2003). Copies of *Ammar1* were amplified from the *A. mellifera* genome using the primers Am24cod-f (5′-GTTGAACCATG GAAAATCAAAAGGAACATTA) and Am349R(Pst) (5′-TTCT GCAGAATAACTTTATTCTGTAATGTG). The forward primers extend into the coding sequences by seven and eight codons, respectively, due to the necessity of introducing a *Nco*I restric-

tion site at the start codon, so we were unable to determine if these first few amino acids varied between copies. Amplified coding sequences were cloned into the *Nco*I and *Pst*I sites of pBAD24 (Guzman *et al.* 1995). This ligation was transformed into electrocompetent DL13 *E. coli* cells, plated onto MacConkey lactose agar plates containing Amp and Tet, and grown at 32° for no more than 4 days. Papillation normally occurs after ∼52 hr under these conditions for *Himar1* (Lampe *et al.* 1999). The total number of colonies was recorded as were the number of colonies that were papillating. Papillating colonies were deemed to contain functional copies of the *Famar1* coding sequence. Plasmids from papillating and nonpapillating colonies were sequenced. An identical procedure was used to screen the genome of *A. mellifera* to search for functional and nonfunctional coding sequences of the closely related *Ammar1* transposon (Robertson and MacLeod 1993).

Quantification of relative transpositional activity of copies of *Famar1*: The relative activity of each functional *Famar1* coding sequence was measured by using a *Famar1*-specific mating-out assay, similar to one developed for *Himar1* transposase (Lampe *et al.* 1999). Each individual functional *Famar1* coding sequence in pBAD24 was transformed into the Apr^R/Gen^R/Str^R *E. coli* strain EB8 (Table 1), and for comparison *HimarI* was transformed into the Apr^R/Gen^R/Str^R *E. coli* strain EB9. The cells were plated on Gen, Amp, and Apr LB-agar plates. Three colonies from each plate were then grown in culture overnight using the same antibiotics. The next day, 10 μl of these cells was mixed with 30 μl of an overnight culture of RZ221 *E. coli* cells in 1 ml of LB broth for mating. The cells were grown at 37° with slow rotation on a roller drum for 6 hr. Suitable dilutions of matings were plated on Nal-Gen plates to select total exconjugates and Apr-Gen to select exconjugate transposition products.

The proportion of all exconjugant bacteria (Gen^R colonies, see above) whose F plasmids carried transposition products was used to estimate the activity of each *mariner* copy. This "Apr^R/Gen^R" ratio was measured at least three times for each *Famar1* assayed. To control for experimental variations, the *Himar1* transposase was also assayed in triplicate alongside each set of *Famar1* experiments. The ratios were log$_{10}$ transformed to normalize their distribution (the variance of the ratios increases with their magnitude; not shown). The average of the appropriate *Himar1* control measurements was subtracted from the *Famar1* measurements to yield activity values relative to *Himar1*. The average of the transformed, then normalized measurements was used as the estimated relative activity for each *Famar1* copy.

The activity of some *Famar1* copies was measured in more than one experiment, and some separately cloned copies (Fa6, Fa7, and Fa8) were identical in sequence. To simplify analysis and avoid incorrectly inflating the number of independent measures, only one set of measurements (the first) was used for each unique (by sequence) *Famar1*. The choice of sets does not significantly affect any results (not shown.)

Statistical analysis to determine whether *Famar1* transposase activity is affected as amino acid changes accumulate is problematic. Due to phylogenetic structure, some *Famar1* copies have accumulated nearly identical sets of mutations and so cannot be treated as independent trials. For each such set of nearly identical *Famar1*, only one member of the set was used in the analysis. This eliminates much of the nonindependence, although some phylogenetic correlations and homoplasies remain. A Spearman rank regression test (Zar 1999) was then applied to the reduced data set to detect any correlation of activity with amino acid divergence from the ancestral sequence. The test was applied to all possible reduced data sets; set choice does not affect the results.

To estimate the statistical power of this method, new activity values were generated for each observed divergence value by either (a) resampling the observed *Famar1* activities or (b) generating random activity values with the observed variances. Linear trends of varying size were then added to the simulated data sets and the above testing method was applied. This procedure was repeated over all possible combinations of eliminated *Famar1* copies (as above). To produce a conservative estimate of power, results from the activity-generating method and *Famar1* subset that resulted in the least power are reported.

Construction of the *Famar1* "ancestral" sequence: To compare the activities of clones isolated from the genomic screen to those of the element that invaded the genome of *F. auricularia*, an "ancestral" *Famar1* element was constructed by a PCR-ligation-PCR technique (Ali and Steinkasserer 1995). The ancestral sequence was inferred by maximum likelihood and agreed with that inferred previously when genomic copies of this element were isolated (Lampe *et al.* 2003). Clones isolated from the activity screen were examined to determine which parts overlapped the ancestral coding sequence, and these were used to "stitch" together a sequence encoding the ancestral transposase. Clones pBADFa5 and pBADFa4 cover the ancestral sequence up to amino acid 270 when spliced together at amino acid 77 (Fa5, amino acids 1–77; Fa4, amino acids 78–269). pBADFa6 covers the ancestral sequence from amino acid 270 on. The final product of this PCR-ligation-PCR is an ancestral sequence, which was cloned into pBAD24 as a *Nco*I/*Pst*I restriction fragment to give pDL129.

Phylogenetic and statistical analyses: Sequences were aligned using ClustalX 1.8 with default settings (Jeanmougin *et al.* 1998). Minor modifications were made by hand near small insertions in some sequences. Phylogenetic analyses were done with PAUP*4.0b8 (Swofford 2001). The equilibrium base frequencies and the transition/transversion ratio (equivalently, 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 and -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 (LRT; Edwards 1992) as described (Lampe *et al.* 2003). Since the *Famar1* clade of *mariner*s is nested within the *Ammar1* clade, the sequence at the root node of the *Famar1* clade represents the *Famar1* ancestral sequence. The node nearest the root inferred for the entire data set (by midpoint rooting with a molecular clock assumed) was used as the ancestral *Ammar1*.

Estimating the genome size of *F. auricularia* and the copy number of *Famar1*: The genome size of *F. auricularia* was kindly determined by Spencer Johnson at Texas A&M University. Tissue samples from *F. auricularia* were run on an Epic (New York) Elite cytometer with excitation from an argon laser tuned to 514 nm at 0.4 W. The samples were stained with propidium iodide (50 ppm). The head and thorax of a *D. melanogaster* Canton-S strain was coprepared and costained with each head from the earwig to provide a reference value for a genome of known size. Red fluorescence from isolated stained nuclei was read using a 610-nm high pass filter to exclude reflected laser light at 514 nm. All samples were stained 40–60 min prior to analysis and 6000–8000 nuclei were scored for each sample.

The copy number of *Famar1* in *F. auricularia* was determined essentially as described in Robertson and Lampe (1995). The

copy number of *Famar1* in *F. auricularia* was determined by using a quantitative genomic slot blot and the radiolabeled full-length coding sequence of pBadFa2. The probe was radiolabeled with dATP (50 μCi, 3000 Ci/mmol) using 5× labeling buffer (Promega, Madison, WI) and the Klenow fragment of *E. coli* DNA polymerase. Genomic *F. auricularia* DNA prepared using a QIAGEN (Chatsworth, CA) genomic DNA purification kit was blotted onto nitrocellulose at concentrations of 1000, 100, and 10 ng and plasmid DNA from pBadFa2 was blotted at concentrations of 10, 3, 1, and 0.3 ng. DNA to be blotted was added to 150 μl of 2× SSPE (300 mм NaCl, 20 mм NaH$_2$PO$_4$, pH 7.4, 2 mм EDTA) and 100 ng/ml of polydI/dC, boiled for 5 min, cooled on ice, and blotted onto nitrocellulose. The samples were hybridized to the probe in an overnight reaction in 5 ml of hybridization solution containing 5× SSPE, 2× Denhardt's solution, 150 μg/ml herring sperm DNA, 0.25% SDS, and 50% formamide at 45°. The samples were washed at 65° for 30 min (each wash) in solutions with 1% SDS and declining concentrations of SSPE from 2× to 1× to 0.5×. A standard curve of hybridization intensity *vs.* mass was constructed on the basis of cloned *Famar1* pBADFa2 using Molecular Imager software (Bio-Rad, Richmond, CA). Hybridization intensities of known masses of *Famar1* genomic DNAs were compared to the standard curve and the copy number was estimated from this value and from the genome size determination.

RESULTS

A large fraction of the *F. auricularia* genome consists of copies of the *Famar1* transposon: The genome size of *F. auricularia* was estimated by flow cytometry of propidium-iodide-stained nuclei compared to similarly treated *D. melanogaster*. This method showed that *F. auricularia* has a relatively large diploid genome of 3.37 ± 0.13 pg (2C), which is ~7.65 times as large as the genome of *D. melanogaster*. Using 180 Mb for the *D. melanogaster* genome size (Adams *et al.* 2000), the haploid genome size of *F. auricularia* is thus estimated to be ~1.377 Gb. Slot-blot analysis of *Famar1* in the genome of *F. auricularia* indicates that copies of it account for ~4% of the total genome, or 57 Mb (Figure 1). Since *Famar1* is 1287 bp, the copy number of this element is thus ~44,000. This value is even larger than that for *Himar1*, which exists in ~17,000 copies in *H. irritans*. It seems likely that the large genome size of the earwig is due to the presence of these and other different kinds of transposable elements, each present in many copies, as is the case with many organisms, including humans (Lander *et al.* 2001).

F. auricularia contains copies of *Famar1* that encode a functional transposase: Some copies of *Famar1* were known to be intact from a previous analysis of a *F. auricularia* genomic library (Lampe *et al.* 2003). We reasoned that some of these copies might encode functional transposase, although we did not know which ones, since none of them differed from the reconstructed ancestral amino acid sequence by fewer than seven changes. We screened >1900 PCR-amplified genomic copies of *Famar1* coding sequences in the *E. coli* papillation screen to find functional ones. Our PCR
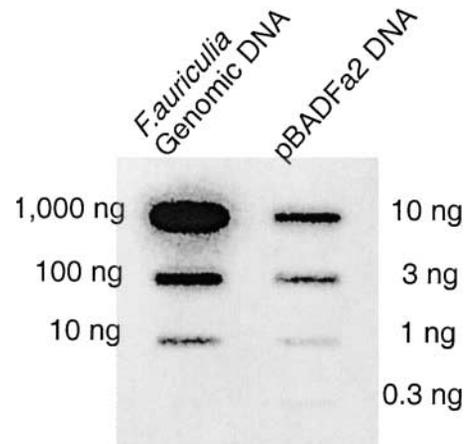

Figure 1.—Slot-blot analysis *F. auricularia* genomic DNA to determine the copy number of *Famar1*. A slot blot of *F. auricularia* genomic DNA probed with a full-length coding sequence of *Famar1*. Genomic DNA is in the left column. Control *Famar1* plasmid DNA is in the right column. Densities were compared by phosphorimaging, a standard curve was constructed, and a value for percentage of the genome occupied by *Famar1* copies was calculated as described in the RESULTS. We estimate that ~44,000 copies of *Famar1* are in the genome of *F. auricularia* and that this element accounts for ~4% of the genome of this insect.

procedure may have excluded a minority of copies whose ITRs had diverged somewhat because we used exact primers; however, there is no reason to believe that the excluded copies differ from those that amplified in any significant respect.

The papillation assay, used extensively to study prokaryotic transposons, relies on the ability of a specially engineered *Famar1* transposon to convert lac(−) *E. coli* to a lac(+) phenotype via transposition (Huisman and Kleckner 1987; Krebs and Reznikoff 1988; Lampe *et al.* 1999). This transposon carries a *lacZ* gene deleted for the first eight amino acids. If this transposon can be mobilized off its plasmid and into an *E. coli* gene in frame, a fusion protein that has β-galactosidase activity may result. These events can be visualized easily by the appearance of red bumps, or papillae, on an otherwise colorless colony when plated on MacConkey lactose agar (for details, see Figure 2). The number of papillae present after a certain period of time is a rough measure of the transpositional activity of the element. A typical colony of cells carrying a functional *Famar1* developed hundreds of papillae.

In this screen we identified 45 copies of *Famar1* that encoded a functional transposase (~2.4% of the total copies examined). Twenty of these were sequenced. We also sequenced multiple nonfunctional copies, of both *Famar1* and the related *Ammar1* from the honey bee (Lampe *et al.* 2003). It is important to note that the copies that we identified as functional are only *potentially* active in *F. auricularia* itself. There is no screen for the
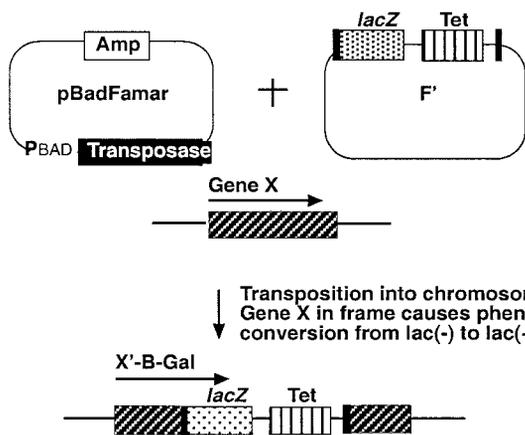
FIGURE 2.—An *E. coli* papillation screen to isolate functional and nonfunctional copies of *Famar1* and *Ammar1*. A papillation screen in *E. coli* to distinguish potentially active *vs.* inactive copies of *Famar1* and *Ammar1* by PCR amplication of genomic copies of coding sequences from these elements. A strain of lac(−) *E. coli* is constructed that contains a modified F′ plasmid. This F′ contains a promoterless *lacZ* gene that also lacks a translational start site. This *lacZ* gene is fused to one ITR of *Famar1/Ammar1* (these elements have identical ITRs) via an open reading frame in the ITR. *E. coli* cells containing this construct are still lac(−). Genomic copies of the coding sequences of *Famar1* and *Ammar1* are amplified and cloned into the expression vector pBAD24. Ligation reactions with these products are transformed into the papillation strain and plated on MacConkey agar. After ∼50 hr at 32°, small red bumps (papillae) form on colonies containing potentially active copies of the transposase sequences. These papillae are due to the conversion of some of the cells in the colony from lac(−) to lac(+) via transposition of the *Famar1-lacZ* transposon off F′ and into a nonessential *E. coli* gene, creating a gene fusion that provides β-galactosidase activity. Cells undergoing this conversion can utilize the lactose in the medium and thus grow faster than their lac(−) neighbors. Cells not showing papillae after 4 days were deemed to contain inactive copies of transposase.
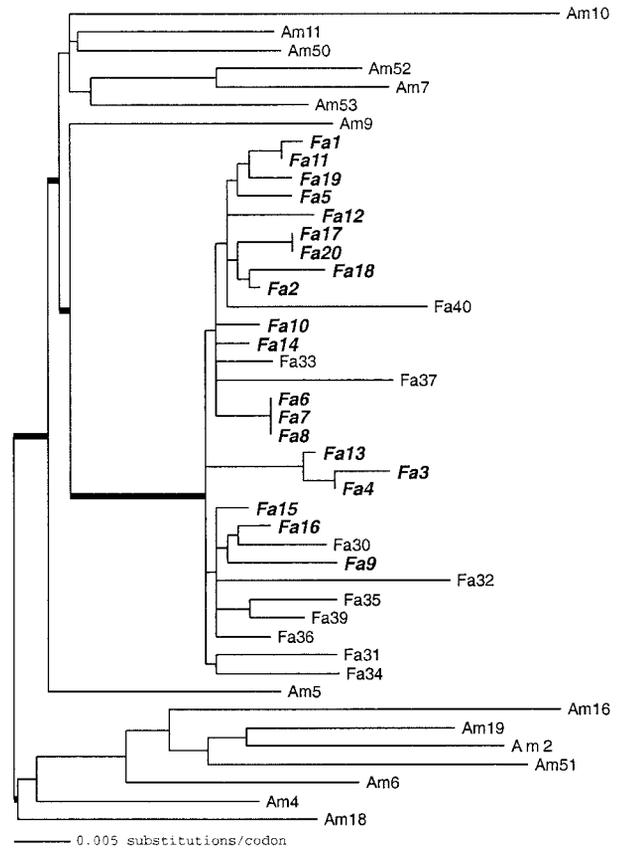

FIGURE 3.—Phylogenetic analysis of the *Famar1* and *Ammar1* coding sequences isolated in this study. The maximum-likelihood phylogenetic tree is based on the coding sequences of all elements isolated in this study. This tree was produced with PAUP* and is midpoint rooted. Functional copies of the *Famar1* transposase coding sequence are shown in boldface italic type. Branches are scaled in units of expected nucleotide changes per codon, according to the scale bar, and were derived from the distances calculated by CODEML (YANG 1997). The branches of the horizontal transfer lineage are thicker than the other branches.

activity of *Famar1 in vivo*. Mutated, transpositionally inactive *mariner*s can repress transposition through dominant-negative effects (LOHE *et al.* 1996; HARTL *et al.* 1997b), so transposition in this insect may be repressed by some of the large fraction (>97%) of apparently nonfunctional coding sequences. Of the ∼2000 copies of *Ammar1* that we screened from the genome of the European honey bee in a similar fashion, none appeared to encode functional transposase. This result is in keeping with our observation that all of the *Ammar1* copies examined, in both this screen and sequenced genomic clones isolated from a bacteriophage library (LAMPE *et al.* 2003), contain obvious inactivating mutations of one sort or another (*i.e.*, indels or nonsense mutations) and are more diverged from their presumed

ancestor than are those of *Famar1*. The sequences of these coding regions from both species can be found in GenBank (accession nos. AY226463–AY226507).

Sequence analysis of *Famar1* and *Ammar1* coding sequences: A phylogenetic tree based on maximum likelihood of all of the clones examined in this study is shown in Figure 3. The tree is consistent with data produced by sequencing entire genomic clones of *Ammar1* and *Famar1*, in that *Ammar1* and *Famar1* are closely related and copies of *Ammar1* are much more diverged from their presumed ancestor compared with those of *Famar1* (LAMPE *et al.* 2003). As many as nine amino acid changes separate functional *Famar1* from the ancestral *Famar1* (*e.g.*, in Fa3). The amino acid distance between each *Famar1* and the ancestor is the number of changes inferred on branches connecting the two. Conceptual translations of all copies can be obtained from D. Lampe upon request.

Current models of *mariner* transposable element evolution predict that all copies, whether active or not, should evolve neutrally in a host population; that is, there should be no significant lack of nonsynonymous codon mutations compared to synonymous, presumably neutral mutations. However, we do expect to find evidence of selection on branches in a phylogeny that contain one or more horizontal transfer events (WITHERSPOON 1999; LAMPE *et al.* 2003). Therefore, we looked for evidence of selection in three sets of branches in the phylogeny, chosen *a priori*: branches within the *Famar1* clade, the five branches connecting the ancestors of the *Famar1* and *Ammar1* clades (thicker branches in Figure 3), and the remaining branches, which reflect evolution within *A. mellifera*. For each set of branches, a separate value of the parameter ω (essentially equal to d_N/d_S, which is the rate of nonsynonymous evolution relative to synonymous evolution) was estimated by maximum likelihood (CODEML, PAML 3.0a; YANG 1997). Since a mutational bias against transversions can be incorrectly interpreted as evidence of selection, κ (the strength of the mutational bias against transversions) was estimated simultaneously to account for that effect. This model yields κ = 4.82, $ω_A$ (*Ammar1*) = 0.91, $ω_T$ (transfer) = 0.55, $ω_F$ (*Famar1*) = 0.97. The smaller the value of ω, the more stringent selection is against nonsynonymous changes. Values near 1 imply no selection.

The likelihood of this model was compared to three separate submodels in which one of the three ω parameters, $ω_F$, $ω_A$, or $ω_T$, was fixed at one and all other parameters (base frequencies, branch lengths, other ω values, and κ) were reestimated. These are submodels in which no selection is allowed in the respective set of branches. The likelihoods of these submodels were then compared to the likelihood of the model with three freely estimated ω values by likelihood ratio tests (1 d.f.; EDWARDS 1992) to determine whether ω differed significantly from one in that set of branches. The appropriate significance cutoff criterion for three tests is α = 0.05/3 = 0.017 (Bonferroni correction, SOKAL and ROHLF 1981). No significant evidence of selection was obtained in any of the three tests (P = 0.87, 0.47, and 0.25 for *Famar1*, *Ammar1*, and transfer branches, respectively.) Thus *mariner*s within *A. mellifera* and *F. auricularia* evolve neutrally.

There is slight evidence of selection in the horizontal transfer lineage (represented by the thick branches connecting the ancestral *Ammar1* with the ancestral *Famar1* in Figure 3), where at least one horizontal transfer event must have occurred. Relatively little evolution has occurred on these branches (only 18 mutations with a ratio of 7:11 synonymous:nonsynonymous, instead of the expected 4.7:13.3; reconstructed by CODEML with one free ω per branch), so the statistical test has little power for detecting selection. With only 18 changes, $ω_T$ would have to be ∼0.28 to yield a significant result,

instead of the observed 0.55. If such lax selection is typical of horizontal transfer events, it will be necessary to examine numerous such events simultaneously to rigorously determine whether or not selection is acting there.

Just as the ancestral, founding *Famar1* element was selected for its functionality, the *Famar1* copies identified as functional by the papillation screen are a nonrandom subset of the mostly nonfunctional *Famar1* family. We therefore examined the evolution on branches of the phylogeny (Figure 3) connecting the ancestor to functional *Famar1* copies by specifying two ω parameters within the *Famar1* clade: one for branches leading to at least one functional element and the other for branches leading only to nonfunctional elements. As expected, there is slight evidence for selection in branches leading to functional *Famar1* copies, but it is not significant (LRT, P > 0.05). If the within-population evolution of functional copies does differ from that of nonfunctional copies, it will require a larger data set to establish that fact.

Activity of individual *Famar1* copies: We quantified the transpositional activity of each active *Famar1* copy in a bacterial mating-out assay. An outline of this assay is shown in Figure 4A and the results of the assay are shown in Figure 4B. The variances of the repeated activity measurements are not homogeneous (Bartlett's test, d.f. = 17, P < 0.05; SOKAL and ROHLF 1981), so nonparametric statistics are used in further analyses. The estimated activities of *Famar1* copies differ significantly (Kruskal-Wallis test of location, d.f. = 18, P < 0.005; SOKAL and ROHLF 1981). Activities span a 10-fold range, from roughly one-fifth as active as *Himar1* (*e.g.*, Fa15) to approximately twice as active (*e.g.*, Fa6).

Since *Famar1* copies do differ in their activities, we looked for any tendency of accumulating amino acid changes to decrease or increase activity. None is apparent in Figure 5. Activity decreases slightly with amino acid divergence, but the trend is not significant (Spearman rank regression test, P > 0.05, d.f. = 15; ZAR 1999). For this analysis, one member of each of the pairs (Fa4, Fa13), (Fa2, Fa18), and (Fa1, Fa11) was removed from the data, since they have largely redundant phylogenetic histories (Figure 5). According to a conservative estimate of the power of this test, a linear trend of ±0.13 relative log_{10} activity units would have been detected with at least 90% probability. This corresponds to a 35% change in activity per amino acid change or an ∼15-fold change expected over the nine-change distance observed in the data.

Some amino acid changes undoubtedly have dramatic effects on element activity (*e.g.*, LOHE *et al.* 1997), but no such changes could be unambiguously identified in our data. All the inactive elements sequenced here contained stop codons and frameshifting indels that would prevent expression of most of the transposase ORF. There are 55 variable sites among the active *Fa-*

mar1, and the changes we observed at these sites do not have dramatic effects on activity. Thus, among *Famar1* copies with measurable activity, our data suggest that most amino acid changes have modest effects, if any, and that they do not predominantly decrease nor increase activity.

DISCUSSION

Copies of *Famar1* encoding functional transposase are detected by a bacterial genetic screen: This study reports the construction and use of a bacterial genetic screen to isolate coding sequences of functional copies of the *Famar1 mariner* transposase from the earwig, *F. auricularia*. *Famar1*, a member of the mellifera subfamily of *mariner* transposons, is only the third known active

mariner element, in addition to *Mos1* and *Himar1* (Jacobson *et al.* 1986; Lampe *et al.* 1996). The fraction of copies remaining in the *F. auricularia* genome that appear to encode a functional transposase is small (2.4%), yet this translates into a total of ~1000 copies encoding a transposase capable of mediating transposition. This begs the question of whether *Famar1* is actually still transposing in this earwig species. It is known that some mutations of transposase can be dominant negative in their phenotype and thus can repress transposition (Lohe *et al.* 1997). Moreover, *mariner* transposons seem to exhibit a gene dosage effect, which leads to "overexpression inhibition," a phenomenon that can lead to wild-type transposase repressing transposition as well (Lohe and Hartl 1996a; Lampe *et al.* 1998). Finally, hosts contain endogenous systems to silence transposons and viruses, such as RNAi (Waterhouse *et al.* 2001). The complex interplay of these factors is still poorly understood and remains an area in need of further research.

Sequence analysis of *Famar1* shows that even remaining functional coding sequences have evolved neutrally: The current model of *mariner* evolution requires that functional elements evolve neutrally within a population (reviewed by Eickbush and Malik 2002). Clearly, elements that are initially nonfunctional cannot evolve under selection. Neutral evolution is the expected result in that case, so it is interesting only if the elements in question can be shown to be, or to have been, functional. Most previous work (*e.g.*, Robertson and Lampe 1995; Lampe *et al.* 2003) has used genomic clones or sequence to demonstrate neutral evolution and relies on indirect arguments to establish the functionality of at least the founding *mariner* copy. It is argued that the ancestor and its immediate descendants must have been functional; otherwise, they could not have created the large number of descendant copies we now observe. This is compelling, but several *mariner* subfamilies in *C.*
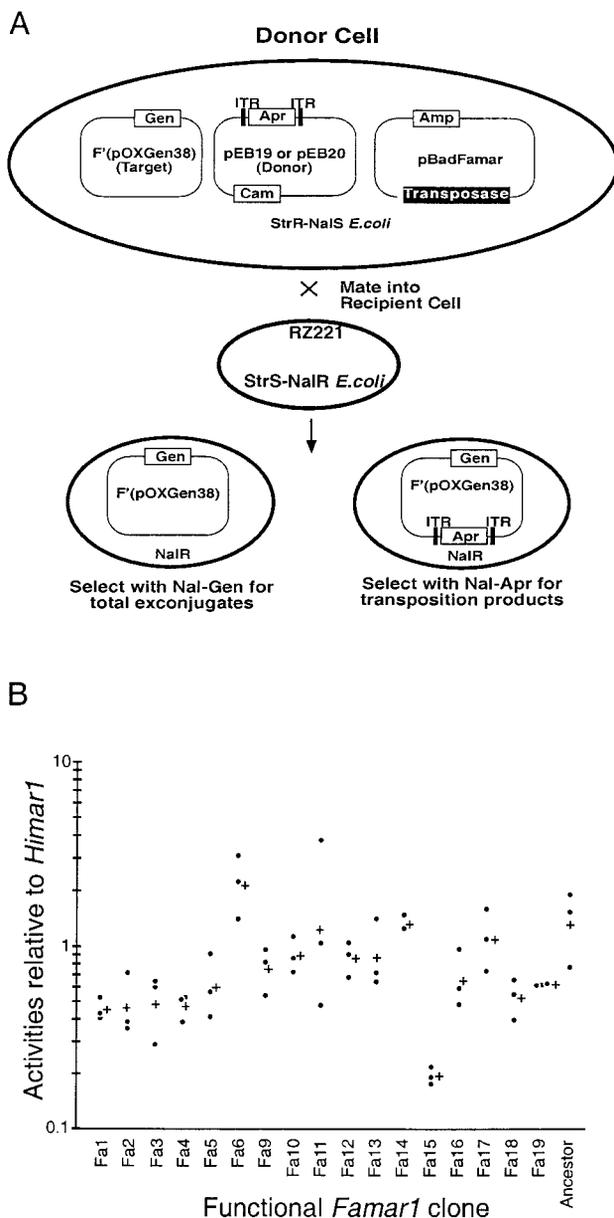

Figure 4.—Quantification of individual functional *Famar1* copies. (A) An overview of the mating-out assay to quantify the activity of individual active *Famar1* copies. The mating-out assay involves the mobilization of a *Famar1* transposon carrying an apramycin-resistance gene from a donor plasmid to a target F′ plasmid. F′ plasmids are subsequently transferred to a different strain of *E. coli* by mating. A determination of the total number of F′ plasmids mated is made by plating the mating mixture on naladixic acid (Nal) and gentamycin (Gen), which selects for all F′ plasmids in the recipient strain. A determination of what fraction of F′ plasmids contain a transposon insertion is made by plating the mating mixture on Nal and apramycin (Apr), which selects only those F′ plamids that contain a transposon insertion in the recipient strain. Activity is expressed as the ratio of Apr^R colonies/Gen^R colonies ($\times 10^5$). (B) Activity measurements (transposition rates relative to *Himar1*; see materials and methods) for unique *Famar1* copies that encode a functional transposase. Individual measurements are shown as circles; crosses mark the estimated activity of each *Famar1*.
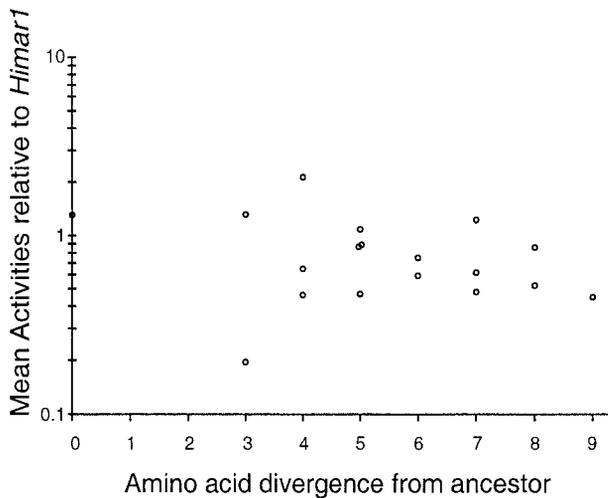
FIGURE 5.—Activities of *Famar1* copies encoding functional transposase as a function of amino acid divergence from the ancestral sequence. Activity *vs.* number of amino acid differences from the ancestral *Famar1*. No correlation is evident (see text). The vertical axis is $\log_{10}$ scaled to allow easier comparisons of relative rates. Measurements for Fa7, Fa8, and Fa20 are not shown or used, since these copies are identical in sequence to others in the data set.

elegans appear to thrive even though their ancestral (and modern) transposase genes contain multiple ORF-interrupting stop codons and indels (WITHERSPOON and ROBERTSON 2003). The functionality of numerous *mariner* copies in Drosophila was demonstrated by CAPY *et al.* (1992), but these copies were too similar to each other to allow statistical tests to distinguish between neutral and selected evolution.

Our data are unique in that they directly demonstrate the functionality of transposases from certain *mariner* copies that are also shown to have diverged neutrally, thus confirming a basic presumption of the current model of *mariner* evolution. Our results are in keeping with those of previous studies (ROBERTSON and LAMPE 1995; LAMPE *et al.* 2003; WITHERSPOON and ROBERTSON 2003) and those of most other DNA-mediated transposons so far examined (WITHERSPOON 1999; SILVA and KIDWELL 2000). No statistically significant evidence of selection was detected in any part of the sequence data set, even for functional coding sequences. A small degree of selection was observed between the ancestors of *Famar1* and *Ammar1*, in which at least one horizontal transfer event must have happened. Depending on the typical stringency of selection at horizontal transfers and the number of mutations that accumulate between them, the predicted selection may become statistically significant only after several transfers have occurred in a lineage (WITHERSPOON 1999).

No clustering of functional transposase sequences in the tree was detected, with one caveat. In two cases, identical sequences were isolated: Fa6, Fa7, and Fa8 are

identical, as are Fa17 and Fa20. This might indicate a greater abundance of these particular sequence types. A clustering of functional sequences in one part of the tree has been predicted to indicate that a subset of the elements might be evolving into a new *mariner* subtype (LAMPE *et al.* 2001). These data are not extensive enough to test that hypothesis, nor can they provide the critical predicted link between ITR and coding sequence coevolution.

Sequence divergence between horizontal transfers— effects on *mariner* diversity and implications for mode of selection: It has been suggested that *mariner*s may diverge by very slowly accumulating mutations during repeated horizontal transfers, each involving few, if any, changes (HARTL *et al.* 1997a; ROBERTSON 2002). However, in the *F. auricularia* genome we find copies of *Famar1* that have diverged substantially from their founder under no apparent selection yet still encode a functional transposase (*e.g.*, Fa3, which differs from the *Famar1* ancestor by nine amino acid changes, or 2.6%). This suggests that *mariner*s may diversify by leaps of a half-dozen or more amino acid changes per horizontal transfer and that fewer horizontal transfer events than previously thought are needed to explain *mariner* sequence diversity.

This combination of infrequent selection and divergence by sizable doses of random mutations is apparently enough to maintain minimal *mariner* functionality: the ability to transpose at a moderate rate, that is, frequently enough to allow their proliferation, but not so often as to severely reduce their hosts' fitness. However, it may not be enough to create and maintain more precise or complex adaptations such as optimal transposition rates or self-regulatory mechanisms. Thus *mariner*s would be coarsely adapted entities at best, whose sequence variability would be matched by functional variability in transposition rate and its regulation. This in turn might help explain the wide variation in *mariner* copy number across host species.

Practical considerations: Several practical applications of our results are apparent. First, as the third known functional *mariner* element, *Famar1* thus expands the choices of those using *mariner*s as genetic tools.

Second, it has been shown that *mariner*s from different subfamilies do not interact (LAMPE *et al.* 2001). Thus a *mariner* from one subfamily can be used as a genetic tool in the presence of another, considerably broadening their utility as a group. We used *Himar1* in just this way in this study to create the *E. coli* papillation strain DL13 for *Famar1*, in which the *Famar1* indicator transposon was carried between the ITRs of *Himar1* for the purpose of mobilizing it onto an F plasmid.

Finally, our genomic screen eliminates the need to reconstruct consensus or ancestral sequences to obtain a functional element. Previously, we reconstructed the *Himar1* transposon by *in vitro* mutagenesis to ensure its

activity (LAMPE *et al.* 1996) and the *Sleeping Beauty* Tc1-like element was reconstructed by heroic means (IVICS *et al.* 1997). These approaches are no longer necessary if one screens a genome that contains at least some elements that, although divergent from the consensus, still appear to be intact. Finding other novel functional *mariner*s (or any DNA-mediated transposon that is functional in *E. coli*) should now be facile.

LITERATURE CITED

ADAMS, M. D., S. E. CELNIKER, R. A. HOLT, C. A. EVANS, J. D. GOCAYNE *et al.*, 2000   The genome sequence of Drosophila melanogaster. Science **287:** 2185–2195.

ALI, S. A., and A. STEINKASSERER, 1995   PCR-ligation-PCR mutagenesis: a protocol for creating gene fusions and mutations. Biotechniques **18:** 746–750.

CAPY, P., J. R. DAVID and D. L. HARTL, 1992   Evolution of the transposable element mariner in the Drosophila melanogaster species group. Genetica **86:** 37–46.

CHANG, A. C., and S. N. COHEN, 1978   Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134:** 1141–1156.

COATES, C. J., N. JASINSKIENE, L. MIYASHIRO and A. A. JAMES, 1998   Mariner transposition and transformation of the yellow fever mosquito, Aedes aegypti. Proc. Natl. Acad. Sci. USA **95:** 3748–3751.

COATES, C. J., N. JASINSKIENE, D. MORGAN, L. R. TOSI, S. M. BEVERLEY *et al.*, 2000   Purified mariner (Mos1) transposase catalyzes the integration of marked elements into the germ-line of the yellow fever mosquito, Aedes aegypti. Insect Biochem. Mol. Biol. **30:** 1003–1008.

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

EICKBUSH, T. H., and H. S. MALIK, 2002   *Origin and Evolution of Retrotransposons*. ASM Press, Washington, DC.

GARCIA-FERNANDEZ, J., J. R. BAYASCAS-RAMIREZ, G. MARFANY, A. M. MUNOZ-MARMOL, A. CASALI *et al.*, 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.

GOMULSKI, L. M., C. TORTI, A. R. MALACRIDA and G. GASPERI, 1997   Ccmar1, a full-length mariner element from the Mediterranean fruit fly, Ceratitis capitata. Insect Mol. Biol. **6:** 241–253.

GUEIROS-FILHO, F. J., and S. M. BEVERLEY, 1997   Trans-kingdom transposition of the Drosophila element mariner within the protozoan Leishmania. Science **276:** 1716–1719. (erratum: Science **277:** 753).

GUZMAN, L. M., D. BELIN, M. J. CARSON and J. BECKWITH, 1995   Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. **177:** 4121–4130.

HANAHAN, D., 1983   Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. **166:** 557–580.

HARTL, D. L., A. R. LOHE and E. R. LOZOVSKAYA, 1997a   Modern thoughts on an ancyent marinere: function, evolution, regulation. Annu. Rev. Genet. **31:** 337–358.

HARTL, D. L., A. R. LOHE and E. R. LOZOVSKAYA, 1997b   Regulation of the transposable element mariner. Genetica **100:** 177–184.

HARTL, D. L., E. R. LOZOVSKAYA, D. I. NURMINSKY and A. R. LOHE,

1997c   What restricts the activity of mariner-like transposable elements. Trends Genet. **13:** 197–201.

HUISMAN, O., and N. KLECKNER, 1987   A new generalizable test for detection of mutations affecting Tn10 transposition. Genetics **116:** 185–189.

IVICS, Z., P. B. HACKETT, R. H. PLASTERK and Z. IZSVAK, 1997   Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell **91:** 501–510.

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

JEANMOUGIN, F., J. D. THOMPSON, M. GOUY, D. G. HIGGINS and T. J. GIBSON, 1998   Multiple sequence alignment with Clustal X. Trends Biochem. Sci. **23:** 403–405.

JOHNSON, R. C., and W. S. REZNIKOFF, 1984   Copy number control of Tn5 transposition. Genetics **107:** 9–18.

JOHNSON, R. C., J. C. YIN and W. S. REZNIKOFF, 1982   Control of Tn5 transposition in Escherichia coli is mediated by protein from the right repeat. Cell **30:** 873–882.

KIESER, T., M. J. BIBB, M. J. BUTTNER, K. F. CHATER and D. A. HOPWOOD, 2000   *Practical Streptomyces Genetics*. John Innes Centre, Norwich, United Kingdom.

KREBS, M. P., and W. S. REZNIKOFF, 1988   Use of a Tn5 derivative that creates lacZ translational fusions to obtain a transposition mutant. Gene **63:** 277–285.

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. (erratum: EMBO J. **16:** 4153).

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., 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., 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.

LAMPE, D. J., D. J. WITHERSPOON, F. N. SOTO-ADAMES and H. M. ROBERTSON, 2003   Recent horizontal transfer of mellifera subfamily *mariner* transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. Mol. Biol. Evol. **20:** 554–562.

LANDER, E. S., L. M. LINTON, B. BIRREN, C. NUSBAUM, M. C. ZODY *et al.*, 2001   Initial sequencing and analysis of the human genome. Nature **409:** 860–921.

LOHE, A. R., and D. L. HARTL, 1996a   Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol. Biol. Evol. **13:** 549–555.

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

LOHE, A. R., D. T. SULLIVAN and D. L. HARTL, 1996   Subunit interactions in the mariner transposase. Genetics **144:** 1087–1095.

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.

MAKRIS, J. C., P. L. NORDMANN and W. S. REZNIKOFF, 1988   Mutational analysis of insertion sequence 50 (IS50) and transposon 5 (Tn5) ends. Proc. Natl. Acad. Sci. USA **85:** 2224–2228.

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

ROBERTSON, H. M., 2002   *Evolution of DNA Transposons in Eukaryotes*. ASM Press, Washington, DC.

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., 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 *et al.*, 1999 *In vivo* transposition of *mariner*-based elements in enteric bacteria and mycobacteria. Proc. Natl. Acad. Sci. USA **96:** 1645–1650.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

SOKAL, R. R., and F. J. ROHLF, 1981 *Biometry.* W. H. Freeman, San Francisco.

SWOFFORD, D. L., 2001 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).* Sinauer Associates, Sunderland, MA.

WATERHOUSE, P. M., M. B. WANG and T. LOUGH, 2001 Gene silencing as an adaptive defence against viruses. Nature **411:** 834–842.

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

WITHERSPOON, D. J., and H. M. ROBERTSON, 2003 Neutral evolution of ten types of mariner transposons in the genomes of Caenorhabditis elegans and C. briggsae. J. Mol. Evol. **56:** 751–769.

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

ZAR, J. H., 1999 *Biostatistical Analysis.* Prentice-Hall, Upper Saddle River, NJ.

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.

Communicating editor: M. J. SIMMONS