



Pergamon

Insect Biochemistry and Molecular Biology 33 (2003) 381–387

*Insect
Biochemistry
and
Molecular
Biology*

www.elsevier.com/locate/ibmb

Rapid Communication

P elements are found in the genomes of nematoceran insects of the genus *Anopheles*

A. Sarkar^a, R. Sengupta^a, J. Krzywinski^a, X. Wang^a, C. Roth^b, F.H. Collins^{a,*}

^a Center for Tropical Disease Research and Training, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556-0369, USA

^b Institut Pasteur, 75724 Paris, Cedex 15, France

Received 13 May 2002; received in revised form 7 January 2003; accepted 8 January 2003

Abstract

We report the identification of genomic sequences in various anopheline mosquitoes (family Culicidae: suborder Nematocera: order Diptera) showing homology to the class II, short inverted-terminal-repeat (ITR) transposable element *P* from *Drosophila melanogaster* (family Drosophilidae; suborder Brachycera: order Diptera). *Anopheles gambiae* appears to have at least six distinct *P* elements. Other anopheline species, including four additional members of the *An. gambiae* species complex (*An. arabiensis*, *An. merus*, *An. melas* and *An. quadriannulatus*), *Anopheles stephensi* (all subgenus *Cellia*), *An. quadrimaculatus* (subgenus *Anopheles*) and *Anopheles albimanus* (subgenus *Nyssorhynchus*) also possess *P* elements similar to those found in *An. gambiae*. Ten distinct *P* element types were identified in the genus *Anopheles*. At least two of the *An. gambiae* elements appears to be intact and potentially functional. Phylogenetic analysis of the anopheline *P* elements reveals them to belong to a distinctly different clade from the brachyceran *P* elements.

© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: *P* element; Transposable elements; Mosquito; *Anopheles*; Nematocera; Brachycera

1. Introduction

Short ITR transposable elements (Class II transposons) are common features of eukaryotic genomes. They have been studied extensively and their ability to move around in genomes has been used to develop them as genetic engineering tools (Atkinson et al., 2001). The original *P* element was identified and characterized from *Drosophila melanogaster*. Related transposons are present in the genomes of many members of the dipteran family Drosophilidae (Loreto et al., 1997). The full length, transpositionally autonomous *P* element of *D. melanogaster* is 2907 bp long, has 31 bp ITRs and generates an 8 bp target site duplication upon insertion. Its 87 kDa (766 amino acids, hereafter a.a.) *P* element transposase protein is required for the movement of the element and is encoded by a transposase

coding region with four exons (exons 0, 1, 2 and 3) (Fig. 1). The development of the *D. melanogaster P* element as a transformation vector and tool for molecular analysis (e.g. enhancer trapping) of *Drosophila* has revolutionized eukaryotic molecular biology (Wilson et al., 1989). Attempts to use the *D. melanogaster P* element as a genetic engineering tool for medically and economically important non-drosophilid insects has failed because of the lack of mobility of the *P* element outside the family Drosophilidae (O'Brochta et al., 1991). The *P* element has also been extensively studied to understand the genetics and evolution of class II transposable elements in eukaryotic genomes; for example, the role of selection pressure and horizontal transfers in maintaining *P* elements in populations (Silva and Kidwell, 2000).

DNA sequence elements with similarity to the drosophilid *P* elements have been isolated from only two other insects, the Australian sheep blowfly, *Lucilia cuprina* (family Calliphoridae) (*Lu-P1* and *Lu-P2*) (Perkins and Howells, 1992) and the housefly, *Musca domestica* (family Muscidae) (*Md-P1*) (Lee et al., 1999). Drosophi-

* Corresponding author. Tel.: +1-574-631-8045; fax: +1-574-631-3996.

E-mail address: frank.h.collins.75@nd.edu (F.H. Collins).

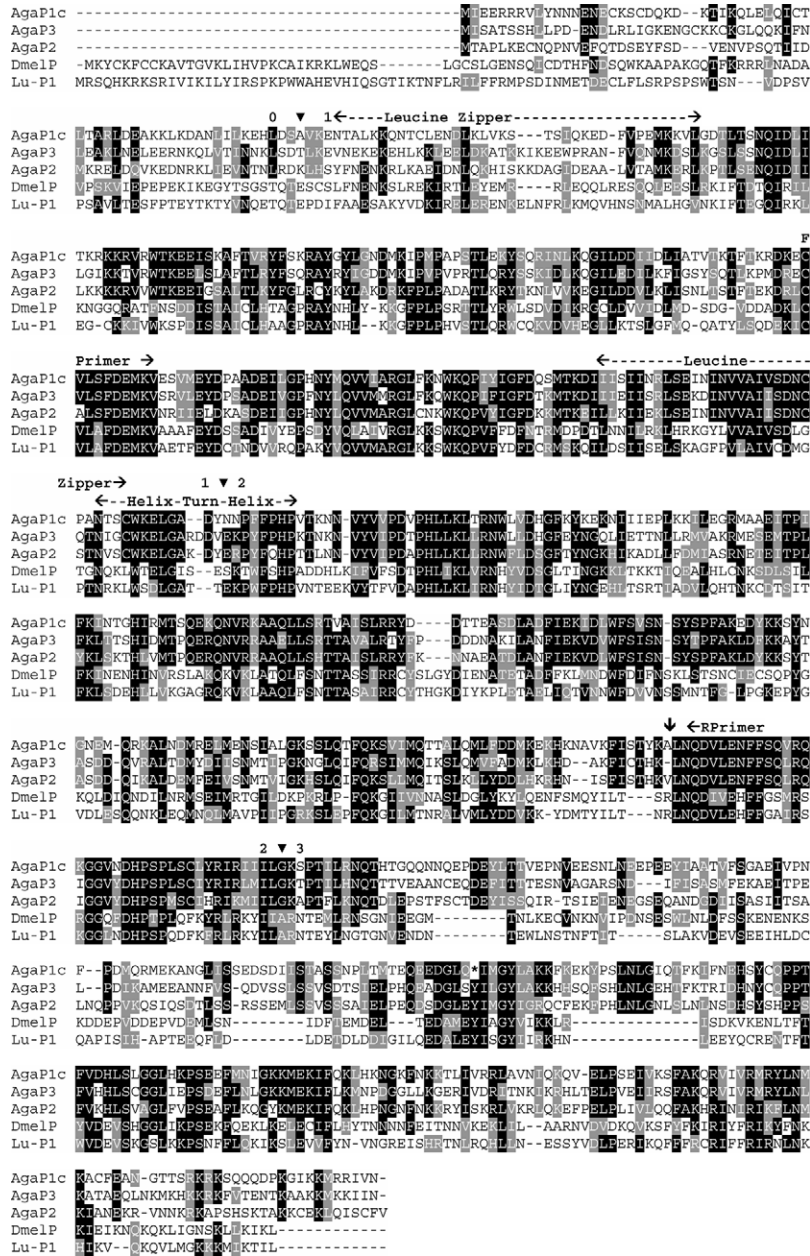


Fig. 1. Amino acid alignment of the conceptual translation of three putative full-length *An. gambiae* P elements' transposase coding regions with the corresponding regions of P elements from *D. melanogaster* and *Lucilia cuprina*. See Table 1 for nomenclature and source of the P elements. The labels spanning the top of the amino acids in the alignment include the *D. melanogaster* leucine zippers, the *D. melanogaster* helix-turn-helix motif (as per Lee et al., 1999), the forward primer sequence (FPrimer) and the reverse primer sequence (RPrimer). The anopheline P elements' intron is indicated by down arrow (↓). Exon boundaries of the *DmelP* element are indicated by filled triangle (▼). *AgaP1c* has a stop codon (*) in its coding region.

lidae, Calliphoridae and Muscidae are closely related families in the dipteran suborder Brachycera. We here report the isolation and analysis of transposons homologous to the brachyceran P elements from the genomes of the lower dipterans of the genus *Anopheles*.

2. Materials and methods

Strains used in this study include *An. gambiae* PEST and Suakoko, *An. arabiensis* AHERO, *An. merus* V12,

An. melas BAL, *An. quadriannulatus* CHIL, *An. quadrimaculatus* Orlando, and *An. stephensi* Delhi (<http://klab.agsci.colostate.edu/~mbenedic/>). DNA of *An. albimanus* was from an unspecified colony (Table 1). Sequence similarity analysis using the BLAST family of programs (Altschul et al., 1990) of the end sequences of clones of an *An. gambiae* PEST BAC genomic library (ND-1; vector pBeloBacII) revealed the presence of sequences with similarity to the brachyceran P elements (ND-1 BAC 28j12; AL154806). The *An. gambiae* BAC

Table 1

Nomenclature and original source of dipteran *P* elements in this study. *P* elements belonging to the same family have the same numeral suffix. The Genbank accession number for *Lu-P1* is A46361, *DmelP* is A24786 and *DguaP* is AAA73165.1

Element	Species	Strain
<i>AgaP1a</i>	<i>An. gambiae</i>	PEST
<i>AgaP2</i>	<i>An. gambiae</i>	PEST
<i>AgaP3</i>	<i>An. gambiae</i>	Suakoko
<i>AgaP4</i>	<i>An. gambiae</i>	Suakoko
<i>AgaP1b</i>	<i>An. gambiae</i>	PEST
<i>AgaP6</i>	<i>An. gambiae</i>	PEST
<i>AgaP1c</i>	<i>An. gambiae</i>	PEST
<i>AgaP8</i>	<i>An. gambiae</i>	PEST
<i>AarP3</i>	<i>An. arabiensis</i>	Ahero
<i>AmlP1</i>	<i>An. melas</i>	BAL
<i>AmrP1</i>	<i>An. merus</i>	V12
<i>AmrP6</i>	<i>An. merus</i>	V12
<i>AqaP8</i>	<i>An. quadriannulatus</i>	CHIL
<i>AstP5</i>	<i>An. stephensi</i>	Delhi
<i>AqmP7</i>	<i>An. quadrimaculatus</i>	Orlando
<i>AalP9</i>	<i>An. albimanus</i>	
<i>AalP10</i>	<i>An. albimanus</i>	
<i>Lu-P1</i>	<i>L. cuprina</i>	
<i>DmelP</i>	<i>D. melanogaster</i>	
<i>DguaP</i>	<i>D. guanche</i>	

library was screened by PCR using primers (P028F-GCAAATACGAGCTGTTGGAAAGAAC and P028R-CTCCGTAATGAAATTGCCACTGTAC designed from an alignment of the published brachyceran and putative anopheline *P* element transposase coding regions identified by sequence similarity searches. This led to the identification of multiple clones for one member of the putative *P* element family (named *AgaP1*). One BAC clone representing this family, ND-1 BAC 01f17 (AL140039, AL140040), was selected for further analysis; the element contained within this clone was called *AgaP1a*. The BAC clone 01f17 was digested with *Bam* HI or *Hind* III separately, run on a 1% TBE-agarose gel and Southern blotted onto a nylon membrane (Duralon; Stratagene) using standard protocols (Sambrook et al., 1989). The membrane was probed with the radioactive PCR amplification products of the primers P028F and P028R using the BAC clone 01f17 as template. The largest DNA bands on the agarose gel hybridizing to the radioactive amplification PCR products were gel extracted (Qiagen Gel Extraction Kit), ligated into the *Bam* HI site of the pBlueScriptSKII(+) (Stratagene) vector and transformed into *E. coli* by electroporation (strain DH10B; Gibco-BRL). The resulting subclones were analyzed by restriction digestion and the presence of the putative *P* element sequences reconfirmed by PCR. A plasmid subclone was selected and sequenced by a primer walking strategy, using the Big-Dye chemistry for the ABI377 automated DNA sequencer (Applied Biosystems).

In order to find more anopheline *P* elements, genomic DNA from various anopheline species (Table 1) was used as a template for the PCR amplification of putative *P* element sequences by the degenerate primers (Pele1130F- TGYGTIYTIICITTYGAYGARATGAARGT) and (Pele2332R- AARAARTKYTCIARIACRTCCT GRIT) (see Table 1). The amplified region corresponds to the sequence between nt 1130 and 2332 of the *D. melanogaster* *P* element. This region was chosen because it is flanked by the peptides CVLAFDEMKV and NQDVLENFF, which were conserved in the published brachyceran and putative anopheline *AgaP1a* element transposase coding regions (hereafter called the conserved region). The amplified PCR products were run on a 1% TBE-agarose gel and the expected 1 kbp products gel extracted (Qiagen Gel Extraction Kit), ligated into the pGEM-T Easy vector (Promega) and transformed into *E. coli* by electroporation (strain DH10B; Gibco-BRL). The DNA sequence of the resulting cloned PCR products was determined using the vector's M13 universal forward and reverse primers flanking the cloning site by utilizing the BigDye chemistry for the ABI377 automated DNA sequencer (Applied Biosystems).

P element sequences cloned from the genomic DNA of anopheline mosquitoes were used as the initial query sequences in BLASTN homology searches at high sequence identity to identify and extend more *P* elements from the genome of *An. gambiae* (PEST strain) using the sequence data from the trace archives of the *An. gambiae* Genome Project (<http://www.ncbi.nlm.nih.gov/blast/mmtrace.html>). These data were then used for the phylogenetic analyses of the anopheline *P* elements. The recent release of the assembled sequence of *An. gambiae* also facilitated the search for the inverted repeats of the elements (<http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/agambiae.html>). The chromosomal location of BAC clones from the *An. gambiae* genomic library that contained putative *P* elements were determined by in situ hybridization to polytene chromosomes of ovarian nurse cells of blood-fed PEST females using standard protocols (Kumar et al., 1997). Sequence analysis was carried out using the Sequencing Analysis and Sequence Navigator suites (ABI/Perkin-Elmer; MacOS) and the Wisconsin Package 10.1 (GCG; Solaris). Similarity searches were carried out using the BLAST family of programs (NCBI; <http://www.ncbi.nlm.nih.gov/Blast.html>) (Altschul et al., 1990). Sequences were aligned using ClustalX 1.8 (Windows 98) (Thomson et al., 1997). Phylogenetic trees were constructed by the maximum parsimony and maximum likelihood methods using PAUP*4.0b4b (Mac OS) (Swofford, 2002). The F81+G model of DNA substitution used in the maximum likelihood analysis as best fitting the data was determined by a likelihood ratio test using MODELTEST 2.0 (Posada and Crandall, 1998). Bootstrapping (1000 replicates)

was used to test the reliability of the branches in the parsimony analysis. *P* elements were classified according to the nucleotide evolutionary distance of their coding sequences of the conserved fragment calculated in PAUP using the F81+G distance matrix model. Elements with sequence distance scores less than 0.1 from each other have been classified for purposes of discussion into the same type; they carry the same numeral suffix in their nomenclature. The first three letters in the nomenclature of these elements indicate the species in which they were identified.

3. Results and discussion

Ten different *P* element types were identified from *Anopheles* spp. (see Table 1). Six were present in members of the *An. gambiae* complex (*P1*, *P2*, *P3*, *P4*, *P6*, and *P8*), *An. albimanus* had two (*AalP9* and *AalP10*), and one was found in each of *An. stephensi* (*AstP5*) and *An. quadrimaculatus* (*AqmP7*). Representatives of all *P* element types found in members of the *An. gambiae* complex were present in the PEST genome. The most divergent element pairs were largely *An. gambiae* (subgenus *Cellia*) and *An. albimanus* (subgenus *Nyssorhynchus*), where distance measures ranged from a low of 0.67944 between *AgaP2* and *AalP10* to a high of 0.86098 between *AgaP8* and *AalP10*. However, even within *An. gambiae*, distance measures as high as 0.74649 were observed between elements *P1* and *P8* (see Table 1). This relatively limited difference between element-type divergence in the *An. gambiae* (subgenus *Cellia*) genome and element-type divergence between *Anopheles* species from other subgenera is probably due to a biased sampling of elements from members of the *An. gambiae* complex, but it also suggests a possible long term persistence in these diverged *Anopheles* species of polymorphic element types present in the ancestral anopheline.

Potentially full length copies of *AgaP1*, *AgaP2*, *AgaP3*, *AgaP4*, *AgaP6*, and *AgaP8* were examined either from the sequenced subclone of BAC 01f17 or from contigs assembled around the sequence in the trace archive most similar to the PCR sequence of the type. A 3748 bp stretch of sequence from BAC 01f17 revealed the *P* element designated *AgaP1a*. When corrected for 3 frame shifts, this element appeared to have an open reading frame capable of encoding a 772 amino acid transposase. The transposase gene consists of two exons. Elements of the *AgaP1* strain appear to be present in low copy numbers (less than ten) in the PEST genome, as estimated by Southern analysis (data not shown) and also by the total number of BAC clones containing the element (ten) identified by screening the ND-1 BAC library (5× genome coverage) using PCR. The BAC clones 28j12 and 01f17, both containing the *AgaP1a*

sequence, localized to the heterochromatic division 6 (which includes the centromere) of the X chromosome. Six of seven other BAC clones containing *AgaP1* sequences also mapped to the centromeric heterochromatin of the polytene chromosomes, one mapped additionally to band 33B on the right arm of the third chromosome, a region also believed to be rich in repeat sequences (M. Sharakhov and F. H. Collins, unpublished data). *AgaP1b* (assembled from the trace archive starting with a sequence most similar to the *AmlP1* PCR clone) and *AgaP1c* (assembled from the trace archive based on similarity to *AmrP1*, Fig. 1) were found to be highly similar to the *AgaP1a* element and had only minor differences at the nucleotide level. Thus they appear to be essentially slightly diverged copies of the same element, *AgaP1*. Like *AgaP1a*, the assembled *AgaP1b* and *AgaP1c* sequences contained 1–3 frameshifts and/or stop codons in the potential transposase coding regions. Furthermore, no indications of ITRs were found in any of these 3 *AgaP1* elements, suggesting that this strain is unlikely to be active in the genome of *An. gambiae*.

Assembled trace file sequences for *AgaP8*, *AgaP6*, *AgaP4*, originally isolated by PCR from the Suakoko strain of *An. gambiae*, were present in the genome of PEST. The transposases potentially encoded by these three elements (after corrections for frame shifts and stop codons) were similar in length and relatively easy to align in the conserved region with that of *P1* elements. All also contained a single intron at the same site in the transposase gene. We were unable to identify ITRs in any of the examined representatives of these three element types.

The *AgaP2* element, originally identified in the PEST strain by PCR and subsequently analyzed using the trace archives (and recently confirmed using the whole genome assembly data), appears to be the only complete and potentially active *P* element in *An. gambiae*. The full-length element is 4397 bp long. Like *P1*, it has a coding sequence consisting of two exons separated by a 61 bp intron and has the potential to encode a 776 amino acid transposase (Fig. 1). The assembled copy *AgaP2* has 29 bp identical ITRs (CAAAGTGAATGAAAGGGA GGTGAGCTTAT) and like the *D. melanogaster P* element, it creates an 8 bp target site duplication upon insertion. There appear to be at least three copies of the *AgaP2* element in the genome, two of them on the second chromosome and another on the third chromosome.

Analysis of the trace archive data revealed that *AgaP3*, originally isolated from the Suakoko strain of *An. gambiae*, was also present in the genome of the PEST strain. The full-length element is apparently 4947 bp long. The *AgaP3* element potentially encodes a 774 a.a. transposase (Fig. 1). The ITRs of *AgaP3* were imperfect. The 5' end sequence is CAAGGTTAA-TAGACTGTATACAGGTTACGACA while the 3' end

sequence is CAAGGTTATTATATATTCTACTGAG-GTTAAGACA; it is flanked by a 8 bp target site duplication.

The *Anopheles* *P* elements are clearly members of the *P* family of transposons. The ITRs of the anopheline *P2* and *P3* elements are roughly comparable in size to the 31 bp perfect ITRs of the *P* element from *D. melanogaster* (CATGATGAAATAACATAAGGTGGTCCCGTCG), and like the *Drosophila* *P* element, they produce an 8 bp target site duplication on insertion. However, all of the *Anopheles* *P* element types have a transposase apparently encoded by two exons separated by an intron at a site not found in the transposase coding sequences of *Drosophila* spp. and the other brachyceran *P* elements (*Musca* and *Lucilia*).

Our failure to identify terminal inverted repeats for *P* elements other than *AgaP2* and *AgaP3* is probably a consequence of their residence in heterochromatin or regions populated by many other transposable elements that disrupt or confound the sequence analysis of these elements. Some transposons immobilized by the loss of ITRs may be “recruited” by the host genome for encoding for novel cellular functions (Miller et al., 1999). Recently, sequences with homology to *P* elements have been reported from the genomes of *Homo sapiens* (*Phsa*), as well as *Bos taurus* and *Gallus gallus* (Hagemann and Pinsker, 2001); it was hypothesized that the human *P* element is a nuclear gene of unknown function because of the identification of a *P* transcript in the human EST database. This could also be true for insect *P* elements outside the insect family Drosophilidae, as well as some drosophilid *P* elements.

When the *AgaP1*, *AgaP2* and *AgaP3* elements were used to search the Genbank EST database, an *An. gambiae* cDNA clone (gi:18944571) was found that was nearly identical to *AgaP1c* when the EST was corrected for three frameshifts. Although the corrected EST is capable of encoding ~160 amino acids, this cDNA did not have the features of a full-length transcript. It may be a partial *AgaP1* transcript truncated during cloning, although whether this transcript is functional or simply a transcript relic is not known.

Phylogenetic analysis of the conserved amino acid fragment of the anopheline *P* element sequences was carried out using the maximum parsimony criterion for 303 characters, of which 223 are parsimony informative (Fig. 2; see Genbank Accession Nos. AY230474-AY230490 for nucleic acid alignment of the conserved region). It revealed that all the anopheline *P* elements are distinct from the brachyceran *P* elements (bootstrap value 100%). Within the anopheline mosquitoes, while *P* element sequence phylogeny does not correspond strictly to the species phylogeny (Krzywinski et al., 2001), the relationship among the *P* elements is not significantly different from the species phylogeny. An analysis of the nucleotide data using the maximum likeli-

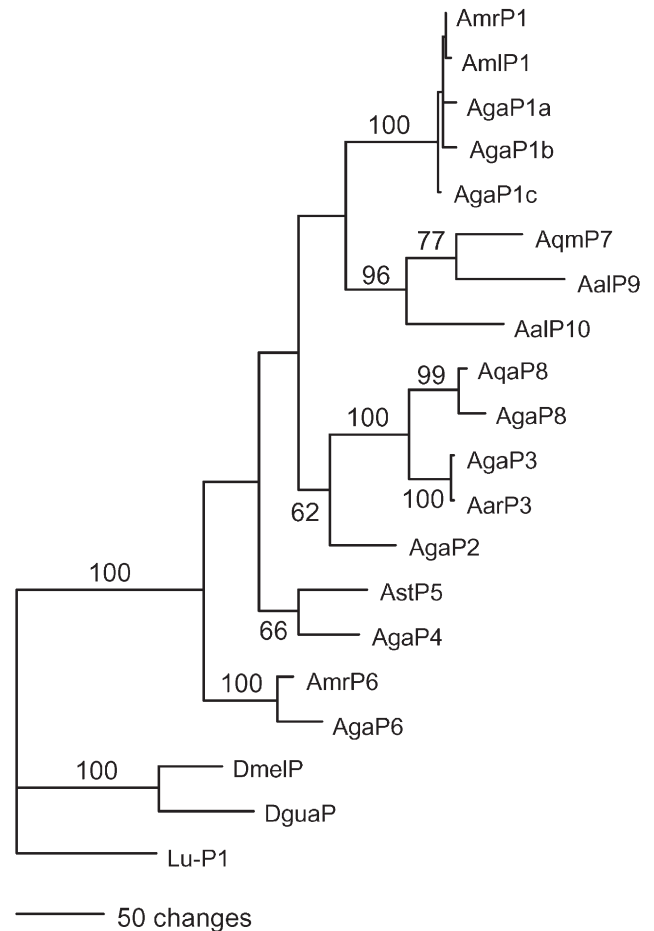


Fig. 2. Maximum parsimony based phylogenetic relationships of nematoceran *P* elements inferred from the amino acid sequences of the conserved region of the *P* element transposase. This is a tree from the maximum parsimony analysis using the amino acids of the conserved region of *P* element transposases (303 included characters, 223 parsimony informative). Sequences representing the degenerate primers (see Fig. 1) used to amplify the conserved region were excluded from the analysis (see Material and Methods for description of the conserved region). The brachyceran *P* elements represented by *Lu-P1* and the drosophilid *P* elements (*DmelP* and *DguaP*) were used as outgroups. The numbers represent the bootstrap support for each branch. The scale bar at the bottom represents 50 changes at the amino acid level. See Table 1 for the nomenclature and source of the *P* elements. See Genbank Accession Nos. AY230474-AY230490 for the nucleic acid alignment of the conserved region of *P* element transposases. Sequence data used for this analysis was obtained using PCR and from the *An. gambiae* Genome Project Trace Archives.

hood approach resulted in trees of similar topology (data not shown). While the occurrence of mixed clades, especially within the members of the *An. gambiae* species complex, is suggestive of gene flow or horizontal transfer of the elements between the species, we believe the data are not robust enough to preclude the possibility of lineage sorting by retention or loss of ancestral elements or (even more likely) insufficient sampling of many species. The conservative interpretation is that all the anopheline *P* elements have evolved within the ano-

pheline lineage. Since the breakup of continents are thought to have played a major role in the radiation of anopheline subgenera, the first appearance of *P* in the *Anopheles* ancestral genome must have preceded the breakup of South America and Africa ~95 million years ago.

A nucleotide distance matrix analysis of the different types of anopheline *P* elements reveals that the intron sequences appear to have diverged more than the coding sequences (exons) in the conserved fragment (Table 2). This would indicate that selection pressure existed for the production of transposase even after the introduction of the element in a novel lineage. This could be due to a selective advantage in encoding a repressor to prevent the newly introduced *P* element from overwhelming or disrupting the host genome while the terminal repeats of newly invading *P* elements accumulated mutations, thereby immobilizing them (Witherspoon, 1999).

Drosophilid P elements are known to be mobile only within the family *Drosophilidae* and are inactive in non-*drosophilid* insects, even when the transposase is overexpressed (O'Brochta and Atkinson, 1996). The mobility of the *D. melanogaster P* element is suppressed when it is introduced in *D. melanogaster* strains containing endogenous copies of the element in the genome. It is likely that the mobility of both endogenous *P* elements and also heterologous *P* elements introduced from other species as gene vectors would be constrained in the genome of the anopheline mosquitoes. *AgaP2* appears to be

intact, possessing terminal inverted repeats and a potentially open transposase. *AgaP3* also appears to encode a full-length transposase and it has terminal inverted repeats, which, even though imperfect, might facilitate its transposition. Thus *AgaP2* and *AgaP3* have the potential for development as gene vectors in mosquitoes and other insects, although they are unlikely to be useful in anopheline mosquitoes unless strains are identified that lack these *P* elements in their genomes.

We here report the isolation of *P* elements from the genomes of many anopheline mosquitoes (family *Culicidae*). This is the first report of the presence of *P* elements in nematoceran insects and the first structurally complete *P* element (transposase gene plus ITRs) identified outside the dipteran family *Drosophilidae*. This suggests that the *P* element family of Class II transposons has a more widespread distribution than previously realized. Efforts are underway to identify *P* elements from other taxa to better understand the evolution of this transposon.

Acknowledgements

We would like to thank Chuong Huynh at the National Center for Biotechnology Information for his help with the trace archives of *An. gambiae* Genome Project. The work described here was supported by NIH grant P01AI45123 (P.I. Frank Collins), NIH cooperative

Table 2

Nucleotide sequence distance matrix (F81+G) for the coding region of the conserved sequence of the anopheline *P* element transposases. Numbers are the range of distance scores (minimum–maximum) for all members of the given family. Numbers in bold face, in parentheses, are the range of intra-family distance scores for the introns of each family for families with more than one member element. Sequence data used for this analysis was obtained using PCR and from the *An. gambiae* Genome Project trace archives

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	0.02459 (0.03263–0.08681)									
P2	0.63126–0.65337	0.0								
P3	0.64229–0.69196	0.53874–0.54669	0.01049 (0.01835)							
P4	0.6448–0.65572	0.58556	0.58443–0.59274	0.0						
P5	0.58385–0.6128	0.59305	0.63279–0.64545	0.39917	0.0					
P6	0.61923–0.69196	0.64968–0.72387	0.62571–0.67487	0.49648–0.54607	0.63062–0.68466	0.09939 (0.19182)				
P7	0.65147–0.66903	0.60755	0.65567–0.66941	0.62358	0.62398	0.64662–0.74841	0.0			
P8	0.69011–0.74649	0.59013–0.62981	0.34819–0.3736	0.58867–0.60908	0.6479–0.67088	0.59372–0.66832	0.65334–0.66864	0.03392 (0.13822)		
P9	0.7522–0.77613	0.74884	0.68784–0.71604	0.7473	0.72329	0.72066–0.83103	0.3987	0.78618–0.82307	0.0	
P10	0.78367–0.81374	0.67944	0.71675–0.74082	0.72067	0.79327	0.73493–0.85365	0.56199	0.82452–0.86098	0.64186	0.0

agreement U01AI48846 (P.I. Frank Collins) and NIH cooperative agreement U01AI50687-01 (Large scale sequencing and assembly of the *An. gambiae* genome; P.I. Robert A. Holt, Celera Genomics).

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.
- Atkinson, P.W., Pinkerton, A.C., O'Brochta, D.A., 2001. Genetic transformation systems in insects. *Annual Review of Entomology* 46, 317–346.
- Hagemann, S., Pinsker, W., 2001. *Drosophila P* transposons in the human genome. *Molecular Biology and Evolution* 18, 1979–1982.
- Krzywinski, J., Wilkerson, R.C., Besansky, N.J., 2001. Towards understanding Anophelinae (Diptera, Culicidae) phylogeny: Insights from nuclear single-copy genes and weight of evidence. *Systematic Biology* 50, 540–556.
- Kumar, V., Cornel, A.J., Mukabayire, O., 1997. In: Crampton, J.M., Beard, C.B., Louis, C. (Eds.), *The Molecular Biology of Insect Disease Vectors*. Chapman and Hall, London, pp. 337–345.
- Lee, S.H., Clark, J.B., Kidwell, M.G., 1999. A *P* homologous sequence in the house fly, *Musca domestica*. *Insect Molecular Biology* 8, 491–500.
- Loreto, E.L., da Silva, L.B., Zaha, A., Valente, V.L., 1997. Distribution of transposable elements in neotropical species of *Drosophila*. *Genetica* 101, 153–165.
- Miller, W.J., McDonald, J.F., Nouaud, D., Anxolabehere, D., 1999. Molecular domestication—more than a sporadic episode in evolution. *Genetica* 107, 197–207.
- O'Brochta, D.A., Atkinson, P.W., 1996. Transposable elements and gene transformation in non-drosophilid insects. *Insect Biochemistry and Molecular Biology* 26, 739–753.
- O'Brochta, D.A., Gomez, S.P., Handler, A.M., 1991. *P* element excision in *Drosophila melanogaster* and related drosophilids. *Molecular and General Genetics* 225, 387–394.
- Perkins, H.D., Howells, A.J., 1992. Genomic sequences with homology to the *P* element of *Drosophila melanogaster* occur in the blowfly, *Lucilia cuprina*. *Proceedings of the National Academy of Sciences USA* 89, 10753–10757.
- Posada, D., Crandall, K.A., 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. In: *Sambrook, J., Fritsch, E.F., Maniatis, T. (Eds.), Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York, 9.
- Silva, J.C., Kidwell, M.G., 2000. Horizontal transfer and selection in the evolution of *P* elements. *Molecular Biology and Evolution* 17, 1542–1557.
- Swofford, D.L., 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b2. Sinauer, Sunderland, MA.
- Thomson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24, 4876–4882.
- Wilson, C., Pearson, R.K., Bellen, H.J., O'Kane, C.J., Grossniklaus, U., Gehring, W.J., 1989. *P*-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes and Development* 3, 1301–1313.
- Witherspoon, D.J., 1999. Selective constraints on *P* element evolution. *Molecular Biology and Evolution* 16, 472–478.