Genomic and Evolutionary Analysis of Feilai, a Diverse Family of Highly Reiterated SINEs in the Yellow Fever Mosquito, Aedes aegypti

Zhijian Tu

Department of Entomology and Center for Insect Science, University of Arizona

Five short interspersed repetitive elements (SINEs) were found fortuitously in the introns of a steroid hormone receptor AaHR3-2 gene of the yellow fever mosquito, Aedes aegypti, constituting a novel family of tRNA-related SINEs named Feilai. In addition, nine other Feilai elements were found in currently available sequences in Ae. aegypti, six of which were also near genes. Approximately 5.9 \times 10^4 copies of Feilai were present in Ae. aegypti, equivalent to 2% of the entire genome. An additional 35 Feilai elements were isolated from a genomic library. Of the total 49 Feilai elements, 20 were full-length. Sequence comparisons and phylogenetic analyses of the full-length elements strongly suggest that there are at least two subfamilies within the Feilai family. There is a high degree of conservation within the two subfamilies. However, sequence divergence between the subfamilies, along with the presence of highly degenerate Feilai elements, suggests that Feilai is likely a diverse family of SINEs that has existed in Ae. aegypti for a long time. Many Feilai elements were closely associated with other transposons, especially with fragments of non-LTR retrotransposons and miniature inverted-repeat transposable elements. The 500-bp sequences immediately flanking a Feilai element were highly A+T-rich, which is consistent with the fact that no Feilai has been found in the coding regions of genes. It is likely that the highly reiterated and interspersed Feilai elements are partially responsible for the pattern of short-period interspersion of the Ae. aegypti genome. The evolutionary relationship between Feilai and the Ae. aegypti genome is likely complex.

Introduction

There are two classes of transposable elements, categorized by the mechanisms of their transposition (Finnegar 1992; Robertson and Lampe 1995). Class II elements such as P and mariner transpose directly from DNA to DNA. Class I elements transpose via an RNA intermediate. The RNA transcripts of the original elements are reverse-transcribed to cDNA molecules. The cDNAs are then inserted elsewhere in the genome. Class I elements can be further divided into three groups, including long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons, and short interspersed repetitive elements (SINEs). While both LTR and non-LTR retrotransposons code for reverse transcriptase and other activities that are essential for retrotransposition, SINEs do not have any coding potential. Recent evidence suggests that they may have been borrowing the retrotransposition machinery of autonomous non-LTR retrotransposons (Ohshima et al. 1996; Okada and Hamada 1997; Terai, Takahashi, and Okada 1998). SINEs can be further divided into two groups based on their sequence similarities to small RNAs. Elements such as the primate Alu family share sequence similarities with 7SL RNA (e.g., Jurka 1995), while most other SINEs belong to a different group that shares sequence similarities to tRNA molecules (Okada and Ohshima 1995).

Some families of SINEs are highly repetitive in the genome. For example, the human Alu family consists of more than 600,000 copies, equivalent to more than 5% of the entire genome (Jurka 1995). Although it is believed that the majority of SINEs may be parasites in their host genomes, there are cases in which SINEs are providing gene regulatory sequences and are sometimes involved in the modification of gene expression (Murane and Morales 1995; Britten 1996; Shimamura et al. 1998). SINEs may also impact the host genomes in other ways, such as in recombination and modulation of chromatin structures (Makatowsky 1995; Marais 1995).

Polymorphic insertions of SINEs are potentially rich sources of genetic markers for population and genetic mapping studies, because SINEs are often interspersed and highly repetitive. The power of such markers has been demonstrated in several recent phylogenetic and population studies of fish, whales, and humans (e.g., Batzer et al. 1994; Hammer 1994; Batzer et al. 1996; Shimamura et al. 1997; Hamada et al. 1998; Novick et al. 1998; Takahashi et al. 1998).

Several families of transposable elements have been found in the yellow fever mosquito, Aedes aegypti, as listed in Tu, Isoe, and Guzova (1998). Here describe the discovery, genomic characterization, and evolutionary analysis of Feilai, which is the first family of SINEs reported in Ae. aegypti. The analysis of a large number of genomic clones that contain Feilai elements also provides insights into the distribution of Feilai elements and the organization of the Ae. aegypti genome. The possibility of developing genetic markers based on the potentially polymorphic insertions of Feilai elements is also discussed.

Materials and Methods

Mosquitoes

Mosquitoes used in this study were from the Rock strain of Ae. aegypti.
Construction of a λ ZapExpress Genomic Library

A genomic library that contains inserts from 1.3 to 2.5 kb was prepared using DNA from the Rock strain of *Ae. aegypti* and a λ ZapExpress vector kit from Stratagene Cloning Systems (La Jolla, Calif.). The vector was predigested with *Bam*HI. The genomic DNA was partially digested with *Sau*3AII. The digestion conditions were optimized to produce mostly 1–3-kb fragments. The digested fragments were separated on an agarose gel. Bands between 1.3 and 2.5 kb were purified using the Sephaglas Bandprep Kit from Amersham Pharmacia Biotech (Arlington Heights, Ill.). An approximately 1:1 ratio of insert and vector was used to minimize double inserts in a clone. The primary library has 2.8 × 10^9 original plaque-forming units, with a 1.7% background. A total of 2.0 × 10^8 original plaque-forming units were amplified and stored. Aliquots of the rest of the primary library were used in the screening experiments described below.

Screening of the λ ZapExpress Genomic Library

The above unamplified library was screened using a probe covering 138 bp of the tRNA-unrelated conserved region in *Feilai-Aa-B1*. It also contained a 36-bp stretch in the tRNA-related region which did not affect the specificity of the probe. The template for the labeling reaction was a PCR product obtained using a plasmid that contains *Feilai-Aa-B1* and the primers AG-GATCTTTTCTGAATGGGA and TCACTGGGACAAA-GGCC. The PCR product was then labeled using only the first primer. The labeling condition was the same as that described by Tu and Hagedorn (1997), using a digoxigenin antibody and the two phosphatase substrates X-phosphate and nitroblue tetrazolium salt, following the protocol of Boehringer Mannheim, Biochemicals (Indianapolis, Ind.). Elements

Estimation of the Copy Number of *Feilai* Elements

The copy number of the *Feilai* elements in the *Ae. aegypti* genome was estimated during one of the above screening experiments under the following conditions: hybridization at 55°C, as described above, and the final two washes at 0.5 × SSC. The following formula was used:

\[
N (\text{copy number}) = \frac{P'}{G} \times (1 - f/\% \text{false positives}) \times I \times (1 - b/\% \text{background})
\]

where \( P' \) = \( P \) (number of positive clones), \( G \) (size of the genome), \( I \) (average size of the inserts), \( S' = S \) (number of clones screened), and \( b \) is the background of the library.

As described above, the background of the library (b) is 1.7%. The average insert size of the genomic library (I) is 1.900 bp. The percentage of false positives (f) was estimated in an experiment shown in Table 3. At the washing stringency of 55°C with 0.5 × SSC, \( f = (64 + 48)/(64 + 60) = 9\% \). The size of the haploid genome of *Ae. aegypti* Rock strain (G) has been estimated to be 8 × 10^9 bp by Rao and Rai (1987). Therefore, the copy number of the *Feilai* elements can be estimated based on the number of positive clones among the total number of clones screened.

In Vivo Excision and DNA Sequencing

Inserts in λ ZapExpress clones were excised in vivo into the pBKT-CMV phagemid vector, using the ExAssist helper phage from Stratagene Cloning Systems. Sequencing was done by the Sequencing Facility at the University of Arizona with either T3/T7 primers or custom synthetic primers, using an automatic sequencer (model 377) from Applied Biosystem Intl. (Foster City, Calif.).

Sequence Analysis and Phylogenetic Inference

Searches for matches of either nucleotide or amino acid sequences in the database (nonredundant GenBank + EMBL + DDBJ + PDB) were done using FASTA of GCG (Genetics Computer Group, Madison, Wis.; version 9.0, 1996) and BLAST (Altschul et al. 1997). Pairwise comparisons were done by Gap and Bestfit of GCG. Multiple sequences were aligned by Pileup, which is a progressive, pairwise method from GCG (gap weight = 1, gap length weight = 0). Consensus of the multiple-sequence alignment was obtained using Pretty of GCG. Phylogenetic trees were constructed using the neighbor-joining, minimum-evolution, and maximum-parsimony methods of PAUP* 4.0 b1 (Swofford 1998). Specific parameters used in the phylogenetic analyses are described in the figure legends. Five hundred bootstrap resamplings were used to assess the confidence in the grouping (Felsenstein and Kishino 1993).

Statistical Test

The F-test was used to analyze the probability of equal variance between two data populations. Because the assumption of equal variance was not violated in any of the comparisons, a simple t-test was used to compare the means. These tests were performed using Excel, version 5.0, from Microsoft (Seattle, Wash.).

Results

Discovery of a Family of SINEs Named *Feilai* in the Noncoding Regions of Genes and Other Sequences of *Ae. aegypti*

The first *Feilai* element, *Feilai-Aa-A1*, was discovered fortuitously in the introns of a newly isolated steroid hormone receptor gene, AaHR3-2 (GenBank AF106703). Both the AaHR3-2 gene and the previously analyzed AaHR3-1 gene (GenBank U87543) are mosquito homologs of the *Drosophila melanogaster* steroid...
FIG. 1.—Discovery of five copies of Feilai elements in the AaHR3-2 gene of Aedes aegypti. A, Evidence of insertion of Feilai-Aa1 in AaHR3-2. The sequence comparison shown here is between two genomic sequences of the AaHR3 genes. The 9-bp direct repeats caused by the insertion are underlined. B, Distribution of the five Feilai elements in the AaHR3-2 gene. These five elements belong to two subfamilies, as shown in figures 2 and 3. Elements in subfamily A are shown in black boxes. Elements in subfamily B are shown in striped boxes. The putative exons of the AaHR3-2 gene are shown as open boxes. Dashed lines represent regions for which sequences are not available. Note that Feilai-Aa-B3 is inserted in Feilai-Aa-A2. C, Common structural characteristics of the five Feilai elements.

hormone receptor gene DHR3 (Koelle, Segraves, and Hogness 1992). A 286-bp insertion was identified in the AaHR3-2 sequence when it was compared with AaHR3-1. As shown in figure 1A, the insertion sequence is flanked by 9-bp target duplications. This insertion sequence was inferred to be a mobile element and was designated Feilai-Aa-A1. In addition, four other Feilai elements were discovered in the 19.8-kb AaHR3-2 sequence, as shown in figure 1B. Pairwise comparisons of these five Feilai elements showed 67%–86% identity. All five elements are full-length and are flanked by putative target duplications, as discussed below. Common structural characteristics of the five elements suggest that Feilai is a family of tRNA-related SINEs. As shown in figure 1C, they share a composite structure similar to that of other tRNA-related SINEs (Okada and Ohshima 1995), where a tRNA-related sequence is followed by a tRNA-unrelated conserved region and a \((\text{GAA})_n\) repeat. For example, a 89-bp fragment at the 5' end of Feilai-Aa-A1 showed 69% identity to a Caenorhabditis elegans Leu-tRNA gene (GenBank U50191). This region contained the A and B boxes of the polymerase III promoter.

In addition to the five Feilai elements found in AaHR3-2, nine copies of Feilai elements were also found in other sequences in Ae. aegypti, as shown in table 1. Six of the nine copies were found in the 5'- and 3'-flanking regions of four genes. Three of the nine copies were associated with two different retrotransposons. Among the total of 14 Feilai elements, eight are full-length, including five in the introns of the AaHR3-2 gene, as shown in figure 2. The rest are truncated at either the 5' or 3' ends as indicated in table 1.

The Feilai Family of SINEs is Highly Reiterated and Interspersed in the Ae. aegypti Genome

The copy number of the Feilai elements in the Ae. aegypti genome was estimated by screening an unamplified \(\lambda\) ZapExpress genomic library as described in Materials and Methods. A total of \(3.1 \times 10^4\) plaques were screened, and the final two washes were at 0.5 \(\times\) SSC. There were approximately 4,680 positive plaques. Thus, according to the formula shown in Materials and Methods, the total number of Feilai elements per \(\text{Ae. aegypti}\) haploid genome under the above screening conditions was approximately \(5.9 \times 10^4\), which is equivalent to approximately 2% of the entire genome. Because one in every six to seven clones in the above library contains a Feilai element, it is likely that Feilai elements are highly interspersed in the genome. Screening of a \(\lambda\) Dash II genomic library which has inserts averaging 16 kb in size showed that one in every two to three clones has at least one Feilai element under similar conditions (data not shown). This further indicates the interspersal of Feilai elements.

Sequence Comparisons and Phylogenetic Analyses of the Full-Length Elements Suggest the Presence of Subfamilies

As shown in table 2, an additional 35 Feilai elements were found in 34 positive clones, all of which were deposited in GenBank (accession numbers...
were two other nodes that were well supported by all bootstrap values all above 98%. In addition, there
lies A and B was well supported by all three methods, minimum parsimony were used. The division into subfamilies A and B was well supported by all three methods, including minimum evolution distance, neighbor joining, and maximum parsimony were used. The division into subfamilies A and B was well supported by all three methods with bootstrap values all above 98%. In addition, there were two other nodes that were well supported by all three methods: the node connecting Feilai-Aa-A4, Feilai-Aa-A8, and Feilai-Aa-A10, and the node connecting Feilai-Aa-A1 and Feilai-Aa-A3. There were other nodes that were supported by bootstrap analysis at a less confident level, as shown in figure 3A. Interestingly, the average branch length within the group A elements, excluding the three more diverged elements A2, A5, and A11, was shorter than the average branch length of the elements in group B, indicating different levels of sequence divergence within these two subfamilies. Because there is a possibility of recombination between different copies of the highly reiterated Feilai elements during evolution, a phylogenetic tree was also constructed using the 15 elements that are flanked by direct repeats to reduce the possible complications introduced by recombination. As shown in figure 3B, the basic pattern of the phylogenetic relationship among the 15 Feilai elements is the same as that in figure 3A, while some nodes are supported with slightly higher bootstrap values.

**Significant Rate of Truncation Among Feilai Elements**

As shown in tables 1 and 2, 21 of the 49 Feilai elements identified are truncated, 14 at the 5’ end, 6 at the 3’ end, and one at both ends. This does not include the 8 Feilai elements truncated during the cloning process. The elements that have a few bases missing at one or both ends are not counted as truncated. All 21 truncated copies were compared with the consensus sequences of groups A and B in a multiple-sequence alignment, as shown in figure 4. Some of the truncated copies are highly similar to the consensus sequences, while many others are much more divergent. Most of the truncated copies that have relatively long sequences in the more variable tRNA-related region are more similar to

---

**Table 1**

**Feilai Elements Identified in Currently Available Sequences of *Aedes aegypti***

<table>
<thead>
<tr>
<th>Feilai</th>
<th>Locus</th>
<th>Location</th>
<th>5’ A+T</th>
<th>3’ A+T</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feilai-Aa-A1...</td>
<td>AaHR3-2</td>
<td>Intron (fig. 1A)</td>
<td>0.534</td>
<td>0.537</td>
<td>This paper</td>
</tr>
<tr>
<td>Feilai-Aa-A2...</td>
<td>AaHR3-2</td>
<td>Intron (fig. 1A)</td>
<td>0.525</td>
<td>0.604</td>
<td>This paper</td>
</tr>
<tr>
<td>Feilai-Aa-A3...</td>
<td>Mospal-Aa5</td>
<td>Insertion$^c$</td>
<td>0.591</td>
<td>0.686</td>
<td>This paper</td>
</tr>
<tr>
<td>Feilai-Aa-A5...</td>
<td>VCP</td>
<td>−1223 to −986</td>
<td>0.619</td>
<td>0.686 (479)</td>
<td>Deitsch and Raikhel (1993)</td>
</tr>
<tr>
<td>Feilai-Aa-B1...</td>
<td>AaHR3-2</td>
<td>Intron (fig. 1A)</td>
<td>0.642</td>
<td>0.66</td>
<td>This paper</td>
</tr>
<tr>
<td>Feilai-Aa-B2...</td>
<td>AaHR3-2</td>
<td>Intron (fig. 1A)</td>
<td>0.524</td>
<td>0.644</td>
<td>This paper</td>
</tr>
<tr>
<td>Feilai-Aa-B3...</td>
<td>AaHR3-2</td>
<td>Intron (fig. 1A)</td>
<td>0.586</td>
<td>0.507</td>
<td>This paper</td>
</tr>
<tr>
<td>Feilai-Aa-U1$^e$</td>
<td>VCP</td>
<td>+569 to +765</td>
<td>0.677</td>
<td>0.625</td>
<td>Deitsch and Raikhel (1993)</td>
</tr>
<tr>
<td>Feilai-Aa-U2$^e$</td>
<td>Vg B</td>
<td>−2257 to −2039</td>
<td>0.586</td>
<td>0.545</td>
<td>Isoe and Hagedorn (personal communication)</td>
</tr>
<tr>
<td>Feilai-Aa-U3$^e$</td>
<td>Vg C</td>
<td>+1023 to +1212$^f$</td>
<td>0.55</td>
<td>0.551 (154)</td>
<td>Isoe and Hagedorn (personal communication)</td>
</tr>
<tr>
<td>Feilai-Aa-U4$^e$</td>
<td>15a-3</td>
<td>−1912 to −1807</td>
<td>0.614 (136)</td>
<td>0.688</td>
<td>Edwards and Hagedorn (1999)</td>
</tr>
<tr>
<td>Feilai-Aa-U5-$^e$</td>
<td>15a-3</td>
<td>−575 to −446</td>
<td>0.723</td>
<td>0.59</td>
<td>Edwards and Hagedorn (1999)</td>
</tr>
<tr>
<td>Feilai-Aa-U6$^e$</td>
<td>Zebedeel</td>
<td>−846 to −736</td>
<td>0.515</td>
<td>0.662</td>
<td>Warren, Hughes, and Crampton (1997)</td>
</tr>
</tbody>
</table>

$^a$ A+T contents of 5′-flanking sequences of Feilai elements.

$^b$ A+T contents of 3′-flanking sequences of Feilai elements.

$^c$ 5′ truncations.

$^d$ 3′ truncations.

$^e$ An insertion in Mospal-Aa5.

$^f$ Counting from the putative poly-A addition signal.

Note.—The system used to organize Feilai elements is as follows: the first letter of the genus name and the first letter of the species name are used to identify the organism in which the element is found. The capital letter after that (A or B) represents the subfamily to which the element belongs. The letter U stands for the organism in which the element is found. The capital letter after that (A or B) represents the subfamily to which the element belongs. The letter U stands for the organism in which the element is found.
Fig. 2.—Sequence comparisons of full-length Feilai elements in Aedes aegypti. A, A multiple-sequence alignment of 20 full-length Feilai elements. Names of the elements were abbreviated by omitting the common prefix “Feilai-Aa.” The system used to name the Feilai elements is explained in the footnote to table 1. The alignment was done by Pileup of GCG and was then slightly modified manually at the 3' end to improve the alignment in the region of GAA repeats. The consensus sequence was created by Pretty (plurality = 10, threshold = 1) of GCG. Dots indicate sequences that are identical to the consensus. Lowercase letters indicate sequence variation. Dashed lines indicate gaps. The tRNA-related region is shown in Table B.

B

<table>
<thead>
<tr>
<th>tRNA-related region</th>
<th>tRNA-related conserved region</th>
<th>GAA repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cons A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cons B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu-tRNA A box</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu-tRNA B box</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cons A

Cons B

tRNA-related conserved region

GAA repeats

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Fig. 2.—Sequence comparisons of full-length Feilai elements in Aedes aegypti. A, A multiple-sequence alignment of 20 full-length Feilai elements. Names of the elements were abbreviated by omitting the common prefix “Feilai-Aa.” The system used to name the Feilai elements is explained in the footnote to table 1. The alignment was done by Pileup of GCG and was then slightly modified manually at the 3' end to improve the alignment in the region of GAA repeats. The consensus sequence was created by Pretty (plurality = 10, threshold = 1) of GCG. Dots indicate sequences that are identical to the consensus. Lowercase letters indicate sequence variation. Dashed lines indicate gaps. The tRNA-related region is shown in Table B.

B

<table>
<thead>
<tr>
<th>tRNA-related region</th>
<th>tRNA-related conserved region</th>
<th>GAA repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cons A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cons B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu-tRNA A box</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu-tRNA B box</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cons A

Cons B

tRNA-related conserved region

GAA repeats

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Fig. 2.—Sequence comparisons of full-length Feilai elements in Aedes aegypti. A, A multiple-sequence alignment of 20 full-length Feilai elements. Names of the elements were abbreviated by omitting the common prefix “Feilai-Aa.” The system used to name the Feilai elements is explained in the footnote to table 1. The alignment was done by Pileup of GCG and was then slightly modified manually at the 3' end to improve the alignment in the region of GAA repeats. The consensus sequence was created by Pretty (plurality = 10, threshold = 1) of GCG. Dots indicate sequences that are identical to the consensus. Lowercase letters indicate sequence variation. Dashed lines indicate gaps. The tRNA-related region is shown in Table B.
the group B consensus than to that of group A. However, because of the difficulty in classifying the truncated elements using a phylogenetic approach, all truncated copies are currently grouped in the unclassified category. No obvious direct repeats could be identified flanking any of the truncated elements. Moreover, there are more 5′ truncations than 3′ truncations. Interestingly, 7 of the 14 truncations at the 5′ end occurred within 5 bp of an adenine base, as shown by the arrow in figure 4.

Relative Abundance of Feilai Elements at Different Levels of Divergence

The 34 sequenced clones listed in table 2 were obtained from a multiple-lift screening experiment at three different stringencies, as shown in table 3. The probe used here covers most of the conserved region and matches very well (94% identity) the consensus sequence of all full-length Feilai elements shown in figure 2. Therefore, the similarity of a particular element to the probe is a good measure of its similarity to the general consensus. The relative number of positive clones identified under different conditions can be used to estimate the number of Feilai elements at these stringencies, which may reflect the relative number of Feilai elements with different levels of sequence similarities to the probe and to the consensus sequence. As shown in table 3, there were $2.3 \times 10^3$ Feilai elements per genome when screening was done at $0.1 \times$ SSC. There were an additional $1.7 \times 10^4$ Feilai elements when the stringency was lowered to $0.5 \times$ SSC. Only $4.7 \times 10^3$ more copies of Feilai were identified when the stringency was further lowered to $2 \times$ SSC. Thus, a large portion of the Feilai elements were highly similar to the probe and the consensus, while a significant number of Feilai elements had a relatively high sequence divergence from the consensus.

Association with Other Transposable Elements

As shown in table 2, 33 of the 34 clones isolated contained at least one Feilai element. Twenty-two of the 33 Feilai clones had at least one additional transposable or repetitive element. Feilai elements and these repetitive elements are very close physically, as the average insert size of these clones is less than 2 kb. The elements found in these Feilai clones are predominantly fragments of non-LTR retrotransposons and miniature inverted-repeat transposable elements (MITEs). For example, six of the Feilai elements are associated with fragments of the Juan-A retrotransposon (Mouches, Bensaadi, and Salvado 1992). Most of the fragmentation of the retrotransposons seemed to have happened quite some time ago, because most of the fragments showed signs of degeneration, such as relatively high sequence divergence from the full-length copies. The associations between Feilai and some of the transposons are likely to be nonrandom. For example, the number of Juan-A elements expected to be found in the above 33 clones is less than 0.02, assuming random distribution. This is based on the estimation that there are approximately 200 copies of Juan-A per genome, full-length and truncated (Mouches, Bensaadi, and Salvado 1992). However, as mentioned above, six Juan-A elements were found in these Feilai clones, far exceeding the expectation of 0.02.

Highly Biased Distribution Toward A+T-Rich Regions

No obvious consensus was observed in the short direct repeats flanking 15 of the 20 full-length Feilai elements. However, 9 of the 15 direct repeats were highly A+T-rich. The 5′ halves of 7 of the 15 direct repeats were preceded by the AA dinucleotide. Furthermore, the extended 5′- and 3′-flanking regions of Feilai elements, approximately 500 bp at each end, were also highly A+T-rich. As shown in table 4, the average A+T content in the 5′- and 3′-flanking regions of the Feilai elements found near genes are 0.597 ± 0.065 and 0.609 ± 0.063, respectively. The average A+T content of the 5′- and 3′-flanking regions of the Feilai elements found in other regions of the genome are 0.616 ± 0.065 and 0.617 ± 0.054, respectively. However, the average A+T content of the open reading frames (ORFs) of Ae. aegypti genes is approximately 0.472 ± 0.054, based on a calculation of 34 genes found during a database search performed in August 1998. As shown in table 4, the flanking sequences of Feilai elements are highly significantly more A+T-rich than are the ORFs in all comparisons based on the t-test analyses. However, there are no significant differences between the sequences flanking Feilai elements near genes and the sequences flanking Feilai elements in other regions. There is no significant difference between the 5′- and 3′-flanking sequences.

Discussion

The study of Feilai described here provides an addition to the limited number of SINEs analyzed in insects (Bradfield, Locke, and Wyatt 1985; Adams et al. 1986; Sun et al. 1991; Lampe and Willis 1994), and therefore expands our current knowledge of SINEs, which is based mainly on studies of vertebrate genomes.

Evolution of Feilai in Ae. aegypti

Feilai is a family of highly reiterated tRNA-related SINEs in the genome of Ae. aegypti. There are a sig-
### Table 2
Characteristics and Distribution of Feilai Elements Identified During the Screening of an Unamplified Genomic Library of *Aedes aegypti*

<table>
<thead>
<tr>
<th>Feilai-Aa-A6</th>
<th>Clone</th>
<th>Stringency (SSC)</th>
<th>Nearby Repetitive Elements</th>
<th>5' A+T&lt;sup&gt;a&lt;/sup&gt;</th>
<th>3' A+T&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Accession Number(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feilai-Aa-A7</td>
<td>435&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.5</td>
<td>An insertion&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.578</td>
<td>0.642 (127)</td>
<td>AF107691</td>
</tr>
<tr>
<td>Feilai-Aa-A8</td>
<td>440</td>
<td>0.1</td>
<td></td>
<td>0.584 (462)</td>
<td>0.673</td>
<td>AF107695</td>
</tr>
<tr>
<td>Feilai-Aa-A9</td>
<td>431</td>
<td>0.1</td>
<td>2. Feilais</td>
<td>0.576 (35)</td>
<td>0.577</td>
<td>AF107688</td>
</tr>
<tr>
<td>Feilai-Aa-A10</td>
<td>401&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.1</td>
<td>Kaizoku-Aa2</td>
<td>0.653 (233)</td>
<td>0.704</td>
<td>AF107689</td>
</tr>
<tr>
<td>Feilai-Aa-A11</td>
<td>436</td>
<td>0.1</td>
<td>Juan-A</td>
<td>0.678</td>
<td>0.404 (63)</td>
<td>AF107692</td>
</tr>
<tr>
<td>Feilai-Aa-A12</td>
<td>419&lt;sup&gt;f&lt;/sup&gt;</td>
<td>2</td>
<td>An insertion&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.562 (105)</td>
<td>0.64</td>
<td>AF107679</td>
</tr>
<tr>
<td>Feilai-Aa-A13</td>
<td>403</td>
<td>0.1</td>
<td>Juan-A and Kaizoku-Aa1</td>
<td>0.604</td>
<td>0.616</td>
<td>AF107666</td>
</tr>
<tr>
<td>Feilai-Aa-A14</td>
<td>439</td>
<td>0.1</td>
<td>0.6</td>
<td>0.576 (206)</td>
<td>0.53</td>
<td>AF107693</td>
</tr>
<tr>
<td>Feilai-Aa-A15</td>
<td>441</td>
<td>0.1</td>
<td>Microsatellite</td>
<td>0.714</td>
<td>0.572</td>
<td>AF107696, AF107697</td>
</tr>
<tr>
<td>Feilai-Aa-A16</td>
<td>405</td>
<td>0.1</td>
<td>2. Feilais and 1 Juan-A</td>
<td>0.659 (279)</td>
<td>0.584</td>
<td>AF107667</td>
</tr>
<tr>
<td>Feilai-Aa-A17</td>
<td>405</td>
<td>0.1</td>
<td>2. Feilais and 1 Juan-A</td>
<td>0.633 (255)</td>
<td>0.66</td>
<td>AF107667</td>
</tr>
<tr>
<td>Feilai-Aa-B7</td>
<td>429</td>
<td>0.1</td>
<td>Bo-Aa3</td>
<td>0.702</td>
<td>0.53</td>
<td>AF107686</td>
</tr>
<tr>
<td>Feilai-Aa-U7</td>
<td>406</td>
<td>0.1</td>
<td>Juan-A</td>
<td>NA</td>
<td>0.601</td>
<td>AF107668</td>
</tr>
<tr>
<td>Feilai-Aa-U8</td>
<td>407</td>
<td>0.1</td>
<td>Juan-A, Lian</td>
<td>NA</td>
<td>0.559</td>
<td>AF107669</td>
</tr>
<tr>
<td>Feilai-Aa-U9</td>
<td>408</td>
<td>0.1</td>
<td>NA</td>
<td>NA</td>
<td>0.651</td>
<td>AF107670</td>
</tr>
<tr>
<td>Feilai-Aa-U10</td>
<td>438</td>
<td>0.1</td>
<td>Q-like</td>
<td>0.62</td>
<td>NA</td>
<td>AF107690</td>
</tr>
<tr>
<td>Feilai-Aa-U11</td>
<td>434</td>
<td>0.5</td>
<td></td>
<td>0.496</td>
<td>NA</td>
<td>AF107689</td>
</tr>
<tr>
<td>Feilai-Aa-U12</td>
<td>432</td>
<td>0.5</td>
<td>JAM1</td>
<td>0.496</td>
<td>NA</td>
<td>AF107689</td>
</tr>
<tr>
<td>Feilai-Aa-U13</td>
<td>410&lt;sup&gt;g&lt;/sup&gt;</td>
<td>0.5</td>
<td>NA</td>
<td>NA</td>
<td>0.594</td>
<td>AF107672</td>
</tr>
<tr>
<td>Feilai-Aa-U14</td>
<td>423</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>0.612</td>
<td>AF107683</td>
</tr>
<tr>
<td>Feilai-Aa-U15</td>
<td>402</td>
<td>0.1</td>
<td></td>
<td>0.696</td>
<td>0.684</td>
<td>AF107665</td>
</tr>
<tr>
<td>Feilai-Aa-U16</td>
<td>431</td>
<td>0.1</td>
<td>2. Feilais</td>
<td>0.604</td>
<td>0.5 (299)</td>
<td>AF107688</td>
</tr>
<tr>
<td>Feilai-Aa-U17</td>
<td>414</td>
<td>0.5</td>
<td>Juan-A</td>
<td>0.571 (49)</td>
<td>NA</td>
<td>AF107673, AF107674</td>
</tr>
<tr>
<td>Feilai-Aa-U18</td>
<td>409</td>
<td>0.5</td>
<td>JAM1</td>
<td>0.571</td>
<td>0.683 (65)</td>
<td>AF107671</td>
</tr>
<tr>
<td>Feilai-Aa-U19</td>
<td>427</td>
<td>0.5</td>
<td>JAM1</td>
<td>0.623</td>
<td>0.556</td>
<td>AF107685</td>
</tr>
<tr>
<td>Feilai-Aa-U20</td>
<td>442</td>
<td>0.5</td>
<td>Bo-Aa6</td>
<td>0.682</td>
<td>0.586</td>
<td>AF107698</td>
</tr>
<tr>
<td>Feilai-Aa-U21</td>
<td>422</td>
<td>2</td>
<td>Bo-Aa4</td>
<td>0.613</td>
<td>0.686</td>
<td>AF107682</td>
</tr>
<tr>
<td>Feilai-Aa-U22</td>
<td>430</td>
<td>2</td>
<td>Youzi-Aa2</td>
<td>0.582 (193)</td>
<td>0.561</td>
<td>AF107687</td>
</tr>
<tr>
<td>Feilai-Aa-U23</td>
<td>417</td>
<td>2</td>
<td></td>
<td>0.609</td>
<td>0.567</td>
<td>AF107677</td>
</tr>
<tr>
<td>Feilai-Aa-U24</td>
<td>425</td>
<td>0.5</td>
<td></td>
<td>0.454</td>
<td>0.682 (291)</td>
<td>AF107684</td>
</tr>
<tr>
<td>Feilai-Aa-U25</td>
<td>415</td>
<td>0.5</td>
<td></td>
<td>0.675 (123)</td>
<td>0.654</td>
<td>AF107675</td>
</tr>
<tr>
<td>Feilai-Aa-U26</td>
<td>418</td>
<td>2</td>
<td></td>
<td>0.656</td>
<td>0.643</td>
<td>AF107678</td>
</tr>
<tr>
<td>Feilai-Aa-U27</td>
<td>416</td>
<td>2</td>
<td></td>
<td>0.673</td>
<td>0.688</td>
<td>AF107676</td>
</tr>
<tr>
<td>Feilai-Aa-U28</td>
<td>420</td>
<td>2</td>
<td>Lian</td>
<td>0.636</td>
<td>0.568</td>
<td>AF107680</td>
</tr>
<tr>
<td>Feilai-Aa-U29&lt;sup&gt;e&lt;/sup&gt;</td>
<td>421</td>
<td>2</td>
<td></td>
<td>NA</td>
<td>0.533</td>
<td>AF107681</td>
</tr>
</tbody>
</table>

**Note:** The system used to organize Feilai elements is described in table 1. The fourth column shows other repetitive elements in a clone that contains a Feilai element. Only the relatively well characterized repetitive elements are listed. JAM1 (Warren, Hughes, and Crampton 1997), Juan-A (Mouches, Bensaadi, and Salvado 1992), Lian (Tu, Iose, and Guzova 1998), and Q (Besansky, Bedell, and Mukabayire 1994) are non-LTR retrotransposons found in mosquitoes. Elements of the above four families listed in the table are fragments. Kaizoku and Youzi are miniature inverted-repeat transposable elements. Bo is a family of short repetitive elements. Detailed listings of all putative elements in these clones and the references can be found in the GenBank entries. Artificial associations between repetitive elements and a Feilai element resulting from double inserts in the genomic clones are not listed. The same consideration was taken into account when calculating the A+T contents of the flanking sequences. NA = not available because of cloning limitations. Five-hundred-base-pair 5'- and 3'-flanking sequences were listed. The same consideration was taken into account when calculating the A+T contents of the flanking sequences. 5'- and 3'-flanking sequences were analyzed when possible. Numbers in brackets represent the lengths of the flanking sequences that are less than 500 bp.

| a | A+T contents of 5'-flanking sequences of Feilai elements. |
| b | A+T contents of 3'-flanking sequences of Feilai elements. |
| c | Feilai elements truncated at a Sau3AI site during the cloning process. |
| d | A1 truncations. |
| e | 3' truncations. |
| f | Insert longer than 2,500 bp, indicating possible double inserts in the clone. |
| g | A 319-bp insertion in Feilai-Aa-A6, resulting in 4-bp direct repeats. |
| h | A 277-bp insertion in Feilai-Aa-A11, resulting in 2-bp direct repeats. |
significant number of highly diverged and often truncated copies of *Feilai* in the genome. The mechanism of these truncations is not yet clear. One possibility is that the truncated copies are products of incomplete reverse transcription during retrotransposition. This is consistent with the observation that there are more 5' than 3' truncations. Furthermore, 7 of the 14 truncations at the 5' end occurred within 5 bp of an adenine base in the tRNA-related region (fig. 4), indicating a possible stalling point for reverse transcription. However, the lack of direct repeats flanking the truncated copies and the presence of 3'-truncated copies suggest that incomplete reverse transcription cannot account for all truncations. Another possibility is that some of the *Feilai* elements may have served as targets for extensive insertion by other transposons, which resulted in the fragmentation of these *Feilai* sequences. It is also possible that some of the truncated copies are caused by deletion or recombination. In any case, the presence of a significant number of truncated and often highly degenerate copies of *Feilai* elements indicates that *Feilai* may have existed in *Ae. aegypti* for a relatively long time. In this regard, it will be interesting to determine the distribution of *Feilai* in other related species of mosquitoes.

On the other hand, a large number of *Feilai* elements are full-length and are relatively conserved in sequences. Phylogenetic analyses and sequence comparisons of the 20 full-length *Feilai* elements suggest that there are at least two subfamilies, A and B. The tRNA-related 5' regions are more variable between subfamilies. It is not yet clear whether different subfamilies may be corresponding to different tRNAs. Fifteen of the 20 full-length elements are flanked by direct repeats, which is another indication of recent transposition. As shown in figure 3, the branch lengths of the *Feilai* elements in groups A and B are quite variable. This is likely a reflection of the relative times of the amplification events in these two groups, assuming a similar rate of mutation. Moreover, within group A, all but the three highly divergent elements form a single node that is supported

---

**Fig. 3.**—Phylogenetic analyses of the full-length *Feilai* elements. Naming of the elements is described in the legend to figure 2. All analyses were conducted using PAUP* 4.0 b1 (Swofford 1998). All trees were unrooted. The relative branch length was calculated by PAUP* 4.0 b1.

A, Phylogenetic relationship of 20 full-length *Feilai* elements. The alignment used here was the same as that in figure 2A except that the three nucleotides at the extreme 5' end and the four nucleotides at the extreme 3' end that are not shared by most of the *Feilai* elements were not included. The tree shown here was constructed using a minimum-evolution algorithm. The heuristic search was conducted using the tree bisection-reconnection branch-swapping algorithm. All characters are of equal weight and unordered. Neighbor joining and maximum parsimony were also used in the analyses. Confidence of the groupings was estimated using 500 bootstrap replications. The Arabic number at the base of a node is the bootstrap value which represents the percentage of times out of 500 bootstrap resamplings that branches were grouped together at a particular node. The first number is the bootstrap value derived from a minimum-evolution analysis, the second is that derived from a neighbor-joining analysis, and the third is that derived from a parsimony analysis. A dash indicates that the bootstrap value is below 50%. Groupings supported by two independent bootstrap analyses are marked as thicker branches. Groupings supported by all three bootstrap analyses are marked as the thickest branches.

B, Phylogenetic relationship of 15 full-length *Feilai* elements that are flanked by target duplications. The alignment was done as described for the alignment of the 20 full-length elements in figure 2A. Other methods and indications for the symbols are the same as in A.
Fig. 4.—Multiple-sequence alignment of the truncated Feilai elements and the consensus sequences of the two subfamilies A and B. Naming of the elements is described in the legend to figure 2. The alignment was done following these steps: (1) The consensus sequences of the two subfamilies, consA and consB, were aligned with truncated yet relatively long sequences. 2) The other short stretches of truncated sequences were added to the alignment based on pairwise comparisons between these sequences and the two consensus sequences. (3) the alignment was then slightly modified manually at the ends. The consensus sequence of the above alignment was created by Pretty (plurality = 3, threshold = 1) of GCG. ‘‘*’’ indicates a 5’ truncation, and ‘‘<’’ indicates a 3’ truncation. Other symbols are the same as in figure 2.

by two of the three bootstrap analyses. Elements in this node can be further divided into a few subgroups. Elements in these subgroups showed shorter branch lengths than other elements, indicating multiple recent amplifications. Therefore, it is possible that there is more than one source gene for the recent amplification of Feilai elements in the A subfamily. However, analysis of a large number of Feilai elements is necessary before any
Estimated number of
E. 3 D. Combining B and C ....................
C. 5 introns and other noncoding regions of genes as listed in table 1. Because a histone gene was found in clone 440 near
Feilai-Aa-A7, 
G. Combining E and F ....................
Ae. aegypti genes found during a database search performed in August 1998 were used as the baseline for comparison.

... to note that the question of whether a single master gene
can be resolved, because recent observations showed
that the noncoding regions of a large portion of charac-
terized genes in Ae. aegypti contained local regions
rich in repetitive elements (unpublished data). In fact,
many of the novel families of transposable elements
were first discovered during the analyses of the noncod-
ing regions of genes in Ae. aegypti (Tu 1997; Tu, Iose,
and Guzova 1998). A recent survey of sequences of the
total genome are highly A+T-rich, significantly more than
the average of the ORFs of Ae. aegypti. Furthermore,
the majority of the direct repeats flanking the full-length
elements are also A+T-rich. This may explain why no
Feilai was found in the coding regions of genes in Ae.
aegypti. The mechanism responsible for the above bias
is not yet clear. It is possible that either selection against
insertion in coding regions, preferential insertion into
A+T-rich noncoding regions, or both could provide
such bias. On the other hand, many Feilai elements were
found to be associated with other repetitive elements.
This distribution bias was first indicated when five Feilai
elements were found in the introns of the AaHR3-2
gene. In a subsequent analysis of 34 genomic clones,
the majority of the Feilai elements were found to be
closely associated with other repetitive elements, espe-
cially with fragments of non-LTR retrotransposons and
MITEs (as shown in table 2). Furthermore, there are
indications that such associations may be nonrandom,
as discussed in Results. It is possible that some of the Feilai
elements could be in the heterochromatic regions in
which repetitive elements concentrate. However, this ex-
planation is probably only part of the story. First of all,
as described above, a number of Feilai elements were
found in introns and flanking regions of genes. Further-
more, the association between Feilai and other repetitive
elements and the association between Feilai and the
noncoding regions of genes are not mutually exclusive.
In other words, a Feilai element could be found in single-
copy DNA such as the genomic regions while still asso-
ciated with repetitive elements. This “contradiction”
may be resolved, because recent observations showed
that the noncoding regions of a large portion of charac-
terized genes in Ae. aegypti contained local regions
rich in repetitive elements (unpublished data). In fact,
many of the novel families of transposable elements
were first discovered during the analyses of the noncod-
ing regions of genes in Ae. aegypti (Tu 1997; Tu, Iose,
and Guzova 1998). A recent survey of sequences of the
total genome of the baker’s yeast, Saccharomyces cer-
evisiae, showed highly biased distribution of the five families of LTR retrotransposons, Ty1–Ty5 (Kim et al.
1998; Sandmeyer 1998). Ty5 elements were found in

Table 3
Relative Abundance of Feilai Elements at Different Levels of Sequence Divergence

<table>
<thead>
<tr>
<th>Stringency</th>
<th>0.1 × SSC</th>
<th>0.5 × SSC</th>
<th>2 × SSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of positive clones</td>
<td>64</td>
<td>60</td>
<td>18</td>
</tr>
<tr>
<td>Percentage of false positives</td>
<td>0</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>Number of real positives</td>
<td>64</td>
<td>48</td>
<td>13</td>
</tr>
</tbody>
</table>

Feilai elements were found in one clone, only the ones with higher similarities
to the probe were used in the calculation. Some of the clones isolated at strin-
gencies of 0.5 and 2 SSC contained fragments of Feilai elements that
were highly similar to the probe, while the lengths of the fragments were too
short to be detected at a higher stringency.

Biased Distribution of Feilai in Ae. aegypti

A number of Feilai elements were found in the noncoding regions of genes (as listed in table 2). The sequences flanking Feilai elements found in the non-
coding regions of genes as well as other regions of the

gene are highly A+T-rich, significantly more than
the average of the ORFs of Ae. aegypti. Furthermore,
the majority of the direct repeats flanking the full-length
elements are also A+T-rich. This may explain why no
Feilai was found in the coding regions of genes in Ae.
aegypti. The mechanism responsible for the above bias
is not yet clear. It is possible that either selection against
insertion in coding regions, preferential insertion into
A+T-rich noncoding regions, or both could provide
such bias. On the other hand, many Feilai elements were
found to be associated with other repetitive elements.
This distribution bias was first indicated when five Feilai
elements were found in the introns of the AaHR3-2
gene. In a subsequent analysis of 34 genomic clones,
the majority of the Feilai elements were found to be
closely associated with other repetitive elements, espe-
cially with fragments of non-LTR retrotransposons and
MITEs (as shown in table 2). Furthermore, there are
indications that such associations may be nonrandom,
as discussed in Results. It is possible that some of the Feilai
elements could be in the heterochromatic regions in
which repetitive elements concentrate. However, this ex-
planation is probably only part of the story. First of all,
as described above, a number of Feilai elements were
found in introns and flanking regions of genes. Further-
more, the association between Feilai and other repetitive
elements and the association between Feilai and the
noncoding regions of genes are not mutually exclusive.
In other words, a Feilai element could be found in single-
copy DNA such as the genomic regions while still asso-
ciated with repetitive elements. This “contradiction”
may be resolved, because recent observations showed
that the noncoding regions of a large portion of charac-
terized genes in Ae. aegypti contained local regions
rich in repetitive elements (unpublished data). In fact,
many of the novel families of transposable elements
were first discovered during the analyses of the noncod-
ing regions of genes in Ae. aegypti (Tu 1997; Tu, Iose,
and Guzova 1998). A recent survey of sequences of the
total genome of the baker’s yeast, Saccharomyces cer-
evisiae, showed highly biased distribution of the five families of LTR retrotransposons, Ty1–Ty5 (Kim et al.
1998; Sandmeyer 1998). Ty5 elements were found in

Table 4
Statistical Analysis of the Flanking Sequences of Feilai Elements Showed that They Were Significantly More A+T-
Rich than the Open Reading Frames (ORFs) of Genes

<table>
<thead>
<tr>
<th>Average A+T Content</th>
<th>SD</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. ORFs in Ae. aegypti</td>
<td>0.472</td>
<td>0.054</td>
</tr>
<tr>
<td>B. 5’ of Feilai found in genes</td>
<td>0.597</td>
<td>0.065</td>
</tr>
<tr>
<td>C. 5’ of Feilai found in other regions</td>
<td>0.616</td>
<td>0.065</td>
</tr>
<tr>
<td>D. Combining B and C</td>
<td>0.609</td>
<td>0.065</td>
</tr>
<tr>
<td>E. 3’ of Feilai found in genes</td>
<td>0.609</td>
<td>0.063</td>
</tr>
<tr>
<td>F. 3’ of Feilai found in other regions</td>
<td>0.617</td>
<td>0.054</td>
</tr>
<tr>
<td>G. Combining E and F</td>
<td>0.616</td>
<td>0.056</td>
</tr>
</tbody>
</table>

NOTE.—The A+T contents of the flanking sequences are from tables 1 and 2. B and E are 5’- and 3’-flanking sequences of the Feilai elements found in introns and other noncoding regions of genes as listed in table 1. Because a histone gene was found in clone 440 near Feilai-Aa-A7, the flanking sequences of Feilai-Aa-A7 were also included in B and E. Sequences flanking Feilai elements found in Mosquito-Aa5 and Zebedee listed in table 1 are included in C and F, together with the rest of the sequences in table 2. Only flanking sequences longer than 300 bp are used in these calculations. The A+T contents of ORFs of 34 Ae. aegypti genes found during a database search performed in August 1998 were used as the baseline for comparison.
the silenced loci, such as telomeres and mating-type cassettes. On the other hand, 294 of the 324 Ty1–Ty4 elements were found within 750 bp of genes transcribed by polymerase III, which are interspersed in the genome. The distribution bias described above suggests close interactions between transposable elements and the host genome. Several authors have proposed that the genome is a complex ecological system in which both the lineage of the host and the lineage of the transposable elements operate (e.g., Brookfield 1995; Kidwell and Lisch 1997). Such perspective may help us to understand the evolution of the Feilai elements in the Ae. aegypti genome.

SINEs and Genome Organization

Feilai elements may constitute up to 2% of the entire genome in Ae. aegypti. More importantly, they were also shown to be highly interspersed in the genome, although they are often associated with other repetitive elements. In other words, Feilai does not seem to belong to the type of repetitive elements that are predominately found in heterochromatin such as centromeric or telomeric regions. It has been shown that up to 80% of the Ae. aegypti genome is organized in a “short-period interspersion” pattern in which the single-copy DNA is partitioned into small blocks by repetitive elements (Gale 1987; Warren and Crampton 1991). The presence of highly repetitive Feilai elements in Ae. aegypti and their locations in the noncoding regions of a significant portion of analyzed genes suggest that they may have contributed to the pattern of short-period interspersion in this species, similar to the highly repetitive MITEs recently found in Ae. aegypti (Tu 1997). In this regard, it is interesting to note that both D. melanogaster and Anopheles gambiae apparently lack any significant number of SINEs (Robertson and Lampe 1995), which is consistent with the fact that their genomes are organized in a “long-period interspersion” pattern in which single-copy DNA is less interrupted by repetitive elements (Davidson et al. 1975; Crain et al. 1976; Black and Rai 1988). Thus, the presence and absence of highly repetitive SINEs correlates well with the types of genomes of these three dipteran insects.

Potential Use

Mosquito-transmitted diseases such as malaria and dengue fever are on the rise. Traditional control methods have gradually become less effective due to increasing insecticide and drug resistance of both the mosquitoes and the pathogens. Current research has focused on using a genetic engineering approach to render the mosquitoes less effective hosts for the pathogens. Significant progress has been made in both the transformation of mosquitoes and the search for possible genes that may determine refractory traits in mosquitoes (e.g., Zheng et al. 1997; Coates et al. 1998; Jasinskiene et al. 1998). However, a key link is still missing in the current approach of genetic engineering. The population genetics of mosquitoes in the field, particularly the population structure and the movement or spread of alleles in populations, is poorly understood. As discussed in Introduction, polymorphic insertions of SINEs have been successfully used in several recent studies as powerful genetic markers for population studies. Because it is likely that there are relatively “young” subfamilies of Feilai elements in Ae. aegypti, there may be a good chance to find polymorphic insertions of the Feilai elements among different populations or different individual mosquitoes. If this is the case, these polymorphic insertions could be used to develop genetic markers for population studies that could greatly facilitate our understanding of the population genetics of this species of mosquito. These markers could also be used for mapping the genes that confer refractory traits in mosquitoes. Because Feilai elements are highly repetitive and interspersed in the genome, they could potentially be very good sources of such markers. The importance of developing Feilai-based markers is further underscored because microsatellites are, for reasons not currently understood, quite rare in Ae. aegypti.

Current study of a small sample of the vast number of Feilai elements has already indicated the complexity of the evolutionary relationship between Feilai and the Ae. aegypti genome. Further analysis of the Feilai elements will likely provide deeper insights into the basic genetic makeup and genomic organization of Ae. aegypti, as well as potentially powerful genetic tools for the control of mosquito-transmitted diseases.

Acknowledgments

I thank H. H. Hagedorn for excellent advice, constant interest, and generous encouragement. I thank H. H. Hagedorn and M. G. Kidwell for critical comments on the manuscript. I am grateful to A. A. James for the gift of a genomic library of Ae. aegypti. I thank Jun Isoe and Jennifer Hill for isolating the genomic DNA used to construct the Δ ZapExpress library. I thank Jun Isoe for providing the compiled ORFs of the Ae. aegypti genome. I also thank the Sequencing Facility of the University of Arizona for their excellent service. This work was supported by NIH grant AI42421 to Z.T., by NIH grant HD 24869 to H. H. Hagedorn and A. M. Fallon, and by a MacArthur Foundation grant to the Center for Insect Science of the University of Arizona.

LITERATURE CITED


MARAIA, R. J. 1995. The impact of short interspersed elements (SINES) on the host genome. R. G. Landes, Austin, Tex.


Swofford, D. L. 1998. PAUP*. Version 4.0 b1. (A commercial test version; completed version 4.0 to be distributed by Sinauer, Sunderland, Mass.)


Terai, Y., K. Takahashi, and N. Okada. 1998. SINE cousins: the 3′-end tails of the two oldest and distantly related families of SINES are descended from the 3′ ends of LINEs with the same genealogical origin. Mol. Biol. Evol. 15:1460–1471.


Pierre Capy, reviewing editor

Accepted February 18, 1999