Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 7475–7480, July 1997
Genetics

Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*

(interspersed repeats/transposons/genome/Insecta)

ZHIJIAN TU*

Department of Entomology and Center for Insect Science, University of Arizona, Tucson, AZ 85721

Communicated by Margaret G. Kidwell, University of Arizona, Tucson, AZ, May 12, 1997 (received for review February 19, 1997)

ABSTRACT **Three novel families of transposable elements, *Wukong*, *Wujin*, and *Wuneng*, are described in the yellow fever mosquito, *Aedes aegypti*. Their copy numbers range from 2,100 to 3,000 per haploid genome. There are high degrees of sequence similarity within each family, and many structural but not sequence similarities between families. The common structural characteristics include small size, no coding potential, terminal inverted repeats, potential to form a stable secondary structure, A+T richness, and putative 2- to 4-bp A+T-biased specific target sites. Evidence of previous mobility is presented for the *Wukong* elements. Elements of these three families are associated with 7 of 16 fully or partially sequenced *Ae. aegypti* genes. Characteristics of these mosquito elements indicate strong similarities to the miniature inverted-repeat transposable elements (MITEs) recently found to be associated with plant genes. MITE-like elements have also been reported in two species of *Xenopus* and in *Homo sapiens*. This characterization of multiple families of highly repetitive MITE-like elements in an invertebrate extends the range of these elements in eukaryotic genomes. A hypothesis is presented relating genome size and organization to the presence of highly reiterated MITE families. The association of MITE-like elements with *Ae. aegypti* genes shows the same bias toward noncoding regions as in plants. This association has potentially important implications for the evolution of gene regulation.**

Recently, several families of short interspersed elements with terminal inverted repeats have been found in maize and other plants (1–5). These elements, named MITEs (miniature inverted-repeat transposable elements) by Wessler *et al.* (6), are grouped into different families which share many structural, but not sequence, similarities. Common features include small size, no coding potential, conserved terminal inverted repeats, A+T richness, A+T-biased specific target sites, and in many cases the potential to form stable secondary structures. These families, such as *Tourist* and *Stowaway*, are highly reiterated in the genome; all have thousands or more copies. Multiple families of highly repetitive elements similar to the plant MITEs have also been found in two species of *Xenopus* (7, 8) and in *Homo sapiens* (9–11). The mechanism of transposition of MITE-like elements has not yet been clearly elucidated (6, 12). However, a DNA-mediated mechanism seems likely because a few MITE-like elements have been found to have terminal sequences almost identical to those of some DNA-mediated elements that have coding potentials (9, 10, 13). On the other hand, MITE-like elements seem to differ from other nonautonomous deletion derivatives of DNA-mediated elements in being present in high copy numbers and in being relatively homogeneous in size within each family or subfamily (8, 12), indicating that they are units of highly successful transposition.

Recent evidence suggests that some transposable elements may have contributed to the evolution of gene regulation (reviewed in refs. 6 and 14–17), as previously proposed by Britten and Davidson (18, 19). Many plant MITEs are associated with genes, where more than 90% are found in the noncoding regions, mostly in the 5′ and 3′ flanking regions (1–3, 5). There are several cases where *Tourist, Stowaway*, and other MITEs overlap previously identified cis-acting regulatory sequences and poly(A)-addition sites of wild-type plant genes, indicating a potential involvement in gene-regulatory evolution (3, 5, 6).

Here I report the discovery and analysis of three novel families of MITE-like elements in the yellow fever mosquito, *Aedes aegypti*. The presence of these elements in the noncoding regions of a large fraction of characterized mosquito genes is described. The possible evolutionary implications of this association are explored.

MATERIALS AND METHODS

Genomic Library Screening. λ-Dash-II genomic libraries prepared from the Rock strain of *Ae. aegypti* that were used in this study were the gifts of A. A. James of the Department of Molecular Biology and Biochemistry of the University of California at Irvine. One of these libraries was custom made by Stratagene Cloning Systems (La Jolla, CA). Both libraries were amplified only once after packaging. The libraries were screened using digoxigenin (DIG)-labeled single-stranded DNA probes. Single-stranded DNA probes were made from double-stranded DNA template by asymmetric PCR in which only one primer was used. Five microliters of a DIG-dUTP labeling mixture (1 mM dATP, 1 mM dCTP, 1 mM dGTP, 0.65 mM dTTP, and 0.35 mM DIG-dUTP) was used in a 100-μl reaction mixture. MagnaGraph nylon membrane (Micron Separations, Westborough, MA) was used to lift the plaques. The prehybridization solution was 5× SSC with 2% nonfat milk, 0.1% *N*-lauroylsarcosine, and 0.02% SDS. Approximately 20 ng of probe per ml of prehybridization solution was used for hybridization. Hybridization was carried out at 55°C. Prehybridization, hybridization, and washings were all performed in a Gene Roller from Savant. The washing stringencies were calculated according to Meinkoth and Wahl (20), allowing approximately 20% or less mismatches in all screenings.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: MITE, miniature inverted-repeat transposable element.
Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. U87544–U87551 for the *Wukong* elements, U88302–U88304 for the *Wuneng* elements, and U88305–U88307 for the *Wujin* elements).
*e-mail: jaketu@ag.arizona.edu.

Phage DNA Purification and Southern Blotting. Phage DNAs were purified according to Sambrook *et al.* (21). The restriction digestion and DNA separation conditions were as described by Lin *et al.* (22). DNA blotting was carried out using a VacuGene XL vacuum blotting system (Pharmacia). Preparation of the probe and hybridization conditions were the same as described above for the screening.

DNA Sequencing. Positive fragments from the phage DNA Southern blot were subcloned into pBluescript SK (−) plasmid from Stratagene Cloning Systems. Sequencing was done by the Sequencing Facility of the University of Arizona with synthetic primers, using an automatic sequencer (model 373) from Applied Biosystems. Sequences were determined from both strands.

Sequence Analysis. Searches for matches of either nucleotide or amino acid sequences in the current database (Nonredundant GenBank +EMBL +DDBJ +PDB) were done using BLAST (23). Pairwise comparisons were done using COMPARE, DOTPLOT, GAP, and BESTFIT from GCG (Genetics Computer Group, Madison, WI, version 8.1, 1995). Multiple sequences were aligned by PILEUP, a progressive, pairwise method from GCG. The parameters for PILEUP were 3.0 for gap creation weight and 0.1 for gap length weight. The potential of sequences to form stable secondary structure was analyzed by using FoldRNA (24) of the GCG package, where the base pairing and stacking energies and the loop destabilizing energies were from Freier *et al.* (25).

RESULTS

Discovery and Analysis of *Wukong*, a Novel Family of MITE-like Elements. The discovery of *Wukong* was an unexpected result from the analysis of a genomic clone that contains an open reading frame (ORF) of an *Ae. aegypti* AaHR3–1 gene (GenBank accession no. U87543). BLAST database search revealed a putative repetitive element 3′ to the ORF that showed 91% identity ($P = 3.6\ e^{-45}$ as calculated by BLAST) to a 167-bp fragment of the 5′ flanking region of an *Ae. aegypti* late trypsin gene (ref. 26, GenBank L17023). The putative element in the AaHR3–1 genomic clone was named *Wukong-Aa1*, and the 167-bp fragment in the late trypsin gene was named *Wukong-Aa2*. Screening of two independent genomic libraries of the *Ae. aegypti* Rock strain using *Wukong-Aa1* as the probe indicated that there were approximately 2,200–3,000 copies of *Wukong* elements per haploid genome. The above numbers were calculated on the basis of the known size of the haploid genome of *Ae. aegypti* Rock strain (800 Mbp; ref. 27) and the 16-kbp average insert size of the genomic libraries. Seven additional *Wukong* elements were sequenced after purifying and subcloning the positive clones obtained from the above genomic screenings.

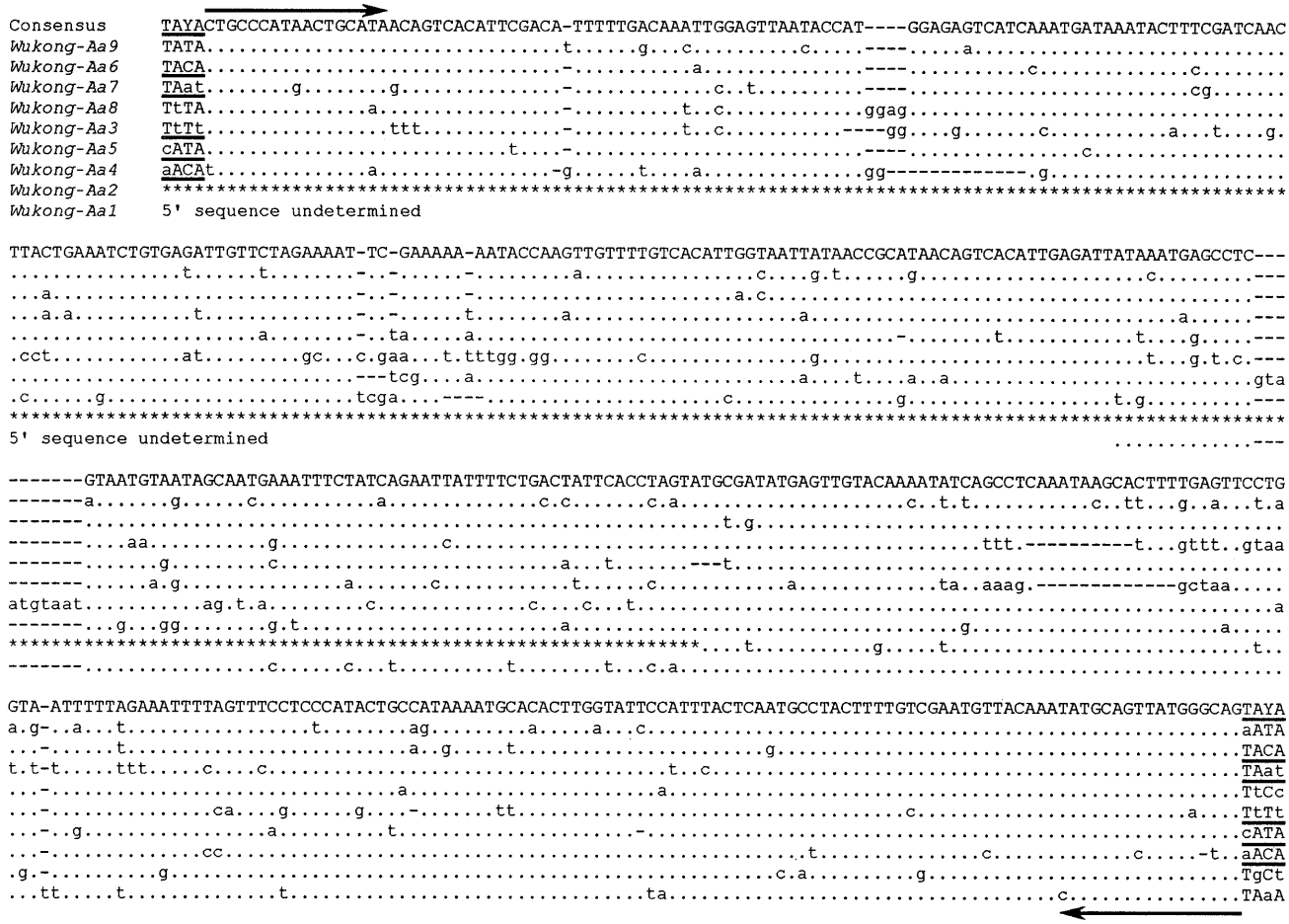

FIG. 1. Multiple sequence alignment of the *Wukong* elements of *Ae. aegypti*. The consensus is based on simple majority rule. In cases where there is no majority base, a base that occurs no less than any other base is chosen arbitrarily. Dots indicate sequences that are identical to the consensus. Dashed lines indicate gaps. Lowercase letters indicate sequence variation. The nonhomologous sequences 5′ of *Wukong-Aa2* are marked by ∗. The 5′ end of *Wukong-Aa1* was not determined because it was beyond the genomic clone. The putative 4-bp target duplications flanking five of the seven complete elements are underlined. The consensus for the target duplication was generated from the above five sequences. Y represents the pyrimidines cytosine (C) and thymine (T). The terminal inverted repeats are marked by arrows. Locations of *Wukong-Aa1* and *Wukong-Aa2* are shown in Table 2. The rest of *Wukong* elements were isolated by screening a genomic library.

Table 1. Structural characteristics of *Wukong*, *Wujin*, and *Wuneng* in comparison with *Tourist* and *Xbr*

Element	Size, bp	TIR, bp	Target site consensus	A+T content, %	$-\Delta G$,* kcal/mol	Copies per genome	Organism
Wukong	420–440	17	TAt/cA	64–69	90.9–109.7	2,200–3,000	*Ae. aegypti*
Wujin	185	23	TA	57–58	51.5–73.1	2,100	*Ae. aegypti*
Wuneng	256–257	19	TTAA/t	63	54.9–70.3	2,700	*Ae. aegypti*
Tourist	113–299	14	TAA	53–82	20.1–84.6	>10,000	*Zea mays*
Xbr	462	42	TTAA	67	156.2	5,000	*X. laevis*

Only complete *Wukong*, *Wujin*, and *Wuneng* elements were used to calculate their sizes, A+T contents, and ΔG values. Data for *Tourist* elements were from Bureau and Wessler (1, 2). *Xbr* is a family of MITE-like elements found in *Xenopus* (8). Size, A+T content, and ΔG value were calculated according to the consensus sequence shown in ref. 8. The length of the element and the terminal inverted repeat does not include the putative target duplications. TIR, terminal inverted repeats.

*The ΔG values for *Wukong*, *Wujin*, *Wuneng*, and *Xbr* were calculated using the default settings of FoldRNA of the GCG package (24), where the base pairing and stacking energies and the loop destabilizing energies were from Freier *et al.* (25) (1 kcal = 4.18 kJ). The ΔG values for *Tourist* were from Bureau and Wessler (1, 2) and were calculated using FoldRNA with the base pairing and stacking energies modified for DNA (28). Analysis of several sequences using both methods showed only 1–16% differences in the ΔG values.

A multiple sequence alignment of the *Wukong-Aa1* in the AaHR3–1 gene, *Wukong-Aa2* in the late trypsin gene, and the seven additional *Wukong* elements is shown in Fig. 1. The positions of the 5′ and 3′ ends of these elements were deduced from the alignment and subsequently confirmed by locating the target duplications flanking *Wukong-Aa5* as described below. The length of the seven complete elements is conserved, ranging from 420 to 440 bp. The 5′ half of *Wukong-Aa1* was not determined because it is beyond the end of the genomic clone. *Wukong-Aa2* is a truncated element containing only a 167-bp sequence at the 3′ end. There is a 300-bp incomplete *Wukong* sequence immediately upstream of *Wukong-Aa8.* This 300-bp sequence is 85% identical to the 5′ region of *Wukong-Aa8* (GenBank U87550). Five of the seven complete *Wukong* elements are flanked by perfect 4-bp putative target site duplications, with TAY(t/c)A as the consensus sequence. A preference for A+T-rich sequence as the site for insertion is apparent. The termini of the *Wukong* elements are defined by 17-bp conserved inverted repeats as shown in Fig. 1. The sequence identity among the seven complete *Wukong* elements is quite high, ranging from 81% to 94%. No coding potential was found in the *Wukong* elements. As shown in Table 1, they are highly A+T rich (64–69%). All of the complete elements had the potential to form stable secondary structures as indicated by low ΔG values ranging from -90.9 to -109.7 kcal/mol. The above characteristics strongly suggest that *Wukong* is a novel family of MITEs.

Past mobility of the *Wukong* elements was indicated by the presence of putative target duplications. More direct evidence came from the analysis of *Wukong-Aa5*. As shown in Fig. 2, *Wukong-Aa5* interrupts *Wujin-Aa1*, another repetitive element described below, at a CATA site. Comparison of *Wujin-Aa1* with other *Wujin* elements that have no *Wukong* interruption revealed the insertion of *Wukong-Aa5* in the *Wujin-Aa1* sequence resulting in a duplication of the CATA insertion target.

Discovery and Analysis of *Wujin* and *Wuneng*, Two Additional Families of MITE-like Elements. A second family of repetitive elements, *Wujin*, was discovered when analyzing a genomic clone that contains *Wukong-Aa5*. *Wujin-Aa1*, which flanks *Wukong-Aa5*, was recognized as a putative repetitive element because it showed 89% identity ($P = 9.2\,e^{-48}$) to a 185-bp fragment in the 5′ flanking region of an *Ae. aegypti* Maltase-like I gene (ref. 29, GenBank M30443). The fragment in the Maltase-like I gene was then named *Wujin-Aa2*. The recognition of *Wuneng* as a third family of putative short repetitive elements was based on the discovery of a 256-bp fragment, near a retrotransposon *Lian-Aa4* (Z.T., J. Isoe, J. Guzova, and H. H. Hagedorn, unpublished work), which showed 91% identity ($P = 8.6\,e^{-77}$) to a 257-bp fragment in the 5′ flanking region of an *Ae. aegypti* D7 gene (ref. 30, GenBank M33156). These two fragments were named *Wuneng-Aa1* and *Wuneng-Aa2*, respectively. On the basis of screenings of a genomic library of *Ae. aegypti*, genomic copy numbers of *Wujin* and *Wuneng* were estimated to be 2,100 and 2,700, respectively.

Sequences of additional copies of *Wujin* and *Wuneng* elements were obtained by the same method as described for *Wukong* elements. Multiple sequence alignments of four *Wujin* elements and five *Wuneng* elements are shown in Figs. 3 and 4 respectively. A high degree of sequence conservation within each family is apparent. Although *Wukong*, *Wujin*, and *Wuneng* share no sequence similarity, they share many structural characteristics, as shown in Table 1. Similar to *Wukong* elements, complete *Wujin* and *Wuneng* elements are short, are A+T-rich, and have no coding potential. They also have the potential to form secondary structures and have conserved terminal inverted repeats. Putative TA target duplications flank each of the complete *Wujin* elements, and the putative target sequences for the three complete *Wuneng* elements are

```
Wujin-Aa1  TGAgGGaACCATCATA Wukong-Aa5 CATAGTAACCATGAAAT
Wujin-Aa2  TGAAGGGACCATCATA----------------GTAACCATGAAAT
Wujin-Aa3  TctAGGGACCATCATA----------------GTAACCATGAAAT
Wujin-Aa4  TctAGGGACCATCATA----------------GTAACCATGAAAT
```

FIG. 2. Evidence of previous mobility of *Wukong* elements in *Ae. aegypti*. The sequence comparison shown here is a segment of Fig. 3. Note the CATA duplications in *Wujin-Aa1* compared with other *Wujin* elements.

```
Consensus  TACAGTGAAACCTCCATGAGTCGATAT-TGAAGGGACCATCGACT
Wujin-Aa2  TA.......................-.................
Wujin-Aa1  TA....c...t..............a..........a.....
Wujin-Aa3  TA.......................-.................
Wujin-Aa4  TA..............a.........-......a..........

CATGGAAATATCGAGTCATGGAACAGCAATCCTTTGGAAAGCTGTTTGAAGGGACC
...........................................................
.................g.-...a.cg...g.c.............g..a...
...........................................c..ct.......
...........----------------..?.........c..ct.......

ATCATAGTAACCATGAAATTTTGTTTTTAGTATGGTTCCATGAGTCGATATCGAGT
.........................cc.....a.....................
.......................ag.............a.............
...........................................................
..............................................t............

CATGGAACATCGACTCATGGAGGGATCACTGTA
.............................TA
a.......c.............ct......TA
.....................t.......TA
a....................**********
```

FIG. 3. Multiple sequence alignment of the *Wujin* elements of *Ae. aegypti*. Symbols are as shown in Fig. 1. *Wujin-Aa1* is interrupted by *Wukong-Aa5* as shown in Fig. 2. The 4-bp target sequence of *Wukong-Aa5* insertion is in boldface. Location of *Wujin-Aa2* is shown in Table 2. The rest of the *Wujin* elements were isolated by screening a genomic library.

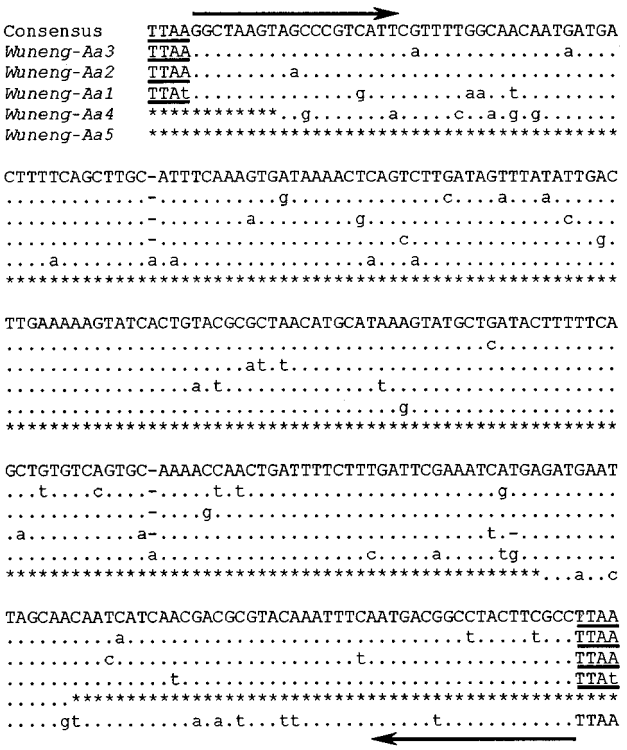

FIG. 4. Multiple sequence alignment of the *Wuneng* elements of *Ae. aegypti*. Symbols are as shown in Fig. 1. *Wuneng-Aa1* was found near a retrotransposon *Lian-Aa4* (Z.T., J. Isoe, J. Guzova, and H. H. Hagedorn, unpublished results). Locations of *Wuneng-Aa2*, *Wuneng-Aa4*, and *Wuneng-Aa5* are shown in Table 2. *Wuneng-Aa3* was isolated by screening a genomic library.

TTAT, TTAA, and TTAA. A preference for A+T-rich sequences as insertion targets is also evident. All these properties indicate that *Wujin* and *Wuneng* are two additional families of MITE-like elements in *Ae. aegypti*.

MITEs Are Located in the Noncoding Regions of a Large Fraction of Characterized *Ae. aegypti* Genes. There are a total of 16 fully or partially sequenced genes in *Ae. aegypti* that have a reasonable length of noncoding sequences available. In addition to the 7 genes shown in Table 2, they include two vitelline envelope genes, *15-a1* and *15-a3* (ref. 31, GenBank U91680 and U91682); three vitellogenin genes, *VgA1* (ref. 32, GenBank L41842), *VgB*, and *VgC* (J. Isoe and H. H. Hagedorn, personal communication); a vitellogenic carboxypeptidase gene (ref. 33, GenBank L46594); an *abd-A* gene (ref. 34, GenBank X67132); a γ-aminobutyric acid receptor gene (ref. 35, GenBank L44606); and an apyrase gene (ref. 36, GenBank L41391). Seven of these 16 genes were found to be associated with a MITE-like element from one of the above three families, during a comprehensive analysis using BLAST, COMPARE, DOTPLOT, GAP, and BESTFIT. As shown in Table 2, four of these elements are found in the 5′ flanking regions (within 1,000 bp from the transcription start site), and two in an intron. *Wuneng-Aa4* was found in the 3′ flanking region of the *AaE74–1* gene, which is an *Ae. aegypti* homolog (Z.T. and H. H. Hagedorn, unpublished data) of the *Drosophila melanogaster E74* gene (37). No elements from the three MITE families were found in the coding regions of these 16 gene sequences or any other sequences in the database.

Table 2. MITE-like elements associated with *Ae. aegypti* genes

Element*	*Ae. aegypti* gene†	Location in the gene‡
Wukong-Aa1	*AaHR3-1*	Intron§
Wukong-Aa2	Late trypsin	5′ (−603 to −437)
Wujin-Aa2	Maltase-I	5′ (−522 to −338)
Wujin-Aa5	Ferritin	Intron
Wuneng-Aa2	*D7*	5′ (−906 to −650)
Wuneng-Aa4	*AaE74-1*	3′¶
Wuneng-Aa5	*15a-2*	5′ (−740 to −682)

*All elements described here are shown in Figs. 1, 3, and 4 except *Wujin-Aa5*, which was found in the Ferritin gene (D. Pham and J. Law, personal communication). *Wujin-Aa5* is truncated at the 3′ end by 8 bp and it showed 85% identity ($P = 1.0\ e^{-27}$) to *Wujin-Aa2*.

†References or GenBank accession numbers for the gene sequences are *AaHR3-1*, U87543; late trypsin, L17023; Maltase-I, M30443; Ferritin, D. Pham and J. Law, personal communication; *D7*, M33156; *AaE74-1*, Z.T. and H. H. Hagedorn, unpublished data; and *15a-2*, U91681.

‡Numbers in parentheses are relative to the transcription start site.

§Position of *Wukong-Aa1* is deduced on the basis of its location relative to the ORF coding for the DNA-binding domain of the *AaHR3-1* gene.

¶There are approximately 4.2 kb between the *Wuneng-Aa4* and the stop codon of *AaE74*. The distance between *Wuneng-Aa4* and the poly(A) addition site of *AaE74* could be much shorter if *AaE74* has a long 3′ untranslated region as does *D. melanogaster E74* (37).

DISCUSSION

Studies presented here represent the first (to my knowledge) published characterization of highly repetitive MITE-like elements in any invertebrate genome. These results, together with the analyses in plants and vertebrates (1–11), suggest that MITE-like elements may have a broad host range in eukaryotic genomes.

In addition to *Wukong*, *Wujin*, and *Wuneng*, a putative fourth MITE-like element has also been identified in *Ae. aegypti* in a preliminary analysis. This element (>500 bp) was found in the 5′ flanking region of a vitelline envelope gene, *15-a2* (ref. 31, GenBank U91681), on the basis of its structural characteristics. It showed 74% identity ($P = 1.8\ e^{-7}$) to a 69-bp fragment in the 5′ flanking region of a vitellogenic carboxypeptidase gene (ref. 33, GenBank L46594). More elements of this family need to be analyzed to verify their structural characteristics and to determine whether they represent a fourth family of MITE-like elements in *Ae. aegypti*. Furthermore, a 185-bp sequence, which is 66–67% identical to the three complete *Wujin* elements ($P = 4.7\ e^{-22}$ in comparison to *Wujin-Aa2*), was found in *Ae. aegypti* in a preliminary study. This sequence had similar terminal-inverted repeats and was flanked by the same putative TA target duplications as the other *Wujin* elements. However, it did not give a positive signal in a dot blot under the stringency used in the genomic screening, which allows approximately 20% or less mismatch. Thus there could be many additional copies of more divergent elements within each of the MITE-like element families in *Ae. aegypti.* Therefore these highly reiterated MITE-like elements are likely to constitute a significant fraction of the *Ae. aegypti* genome.

Possible Mechanism of Transposition of MITE-Like Elements. The first 5 bp of the terminal inverted-repeat sequence in *Wujin* elements, CAGTG, is identical to that of the *Tc-1* and *Tc-3* elements in *Caenorhabditis elegans* (38, 39). *Wujin*, *Tc-1*, and *Tc-3* also seem to have the same TA target duplication sequence. It is therefore possible that *Wujin* may have borrowed an autonomous *Tc-1* or *Tc-3*-like class II element from within the *Ae. aegypti* for its transposition, as indicated in the case of some human MITE-like elements (9–11). It will be interesting to see if, or how, MITE-like elements achieved high copy number via the cut and paste mechanism used by other DNA-mediated elements and what effects such massive cut and paste events may have had on chromosome structures (11, 12).

MITE-Like Elements and the Size and Organization of Eukaryotic Genomes. Multiple families of highly reiterated MITEs have been found in cereal grasses, which have large genomes and a high level of repetitive sequences (1–5).

However, an extensive database search failed to identify any MITE-Like element in *Arabidopsis*, which has the smallest genome known in higher plants (145 Mbp) and a very low level of interspersed repetitive elements (5). Similarly, highly repetitive MITE families have been found in *Ae. aegypti*, which has a relatively large genome (800 Mbp) with a high level of repetitive sequences (27, 40). In contrast, no MITEs have been reported in the most extensively studied insect, *D. melanogaster*, which has a small genome (130–140 Mbp) and a low level of repetitive sequences (41–43). In addition, the genomes of *Xenopus* and *H. sapiens*, in which MITE-Like elements have been found in high copy numbers, are also large in size and rich in repetitive elements (8, 11, 43). The distribution of MITE-like elements in these various genomes suggests that proliferation of MITE-Like elements may be associated with large and more repetitive genomes in both plant and animal kingdoms. In this regard, it is interesting to note that Besansky *et al.* (44) have found a family of small transposable elements named *Pegasus* in *Anopheles gambiae*. This species has a small genome similar to that of *D. melanogaster* (45, 46). The *Pegasus* elements have features like MITEs, such as size homogeneity, an 8-bp terminal inverted repeat, and no coding potential. However, they lack the potential to form stable secondary structures, and there is no consensus within their 8-bp target duplication sequences (44). In contrast to the MITE-Like elements described in large genomes such as *Ae. aegypti*, there are only 30 copies of *Pegasus* in *An. gambiae*. In addition, a 354-bp insertion, flanked by TAA duplications, was identified in *Pegasus-27* (44). This insertion sequence has 77-bp terminal inverted repeats and the potential to form stable secondary structure, indicating a possible novel MITE-Like element in *An. gambiae.* It will be interesting to see if this insertion sequence has a copy number similar to that of the *Pegasus* family.

In addition to the differences in size and relative amount of repetitive elements, the genomes of *Ae. aegypti* and *D. melanogaster* also show distinctly different patterns of organization. Up to 80% of the *Ae. aegypti* genome is organized in a "short period interspersion pattern" in which the single-copy DNAs are partitioned into small blocks by repetitive elements (40, 47). In contrast, the majority of the *D. melanogaster* genome is organized in a "long period interspersion pattern" (43, 48) in which single-copy DNAs are less interrupted by the interspersion of long repetitive elements. The presence of highly repetitive MITE-Like elements in *Ae. aegypti* and their locations in the noncoding regions of a large portion of analyzed genes indicate that they may have contributed to the pattern of short-period interspersion in this species.

Association of MITE-Like Elements with Genes. A total of 16 *Ae. aegypti* genes are available for analysis of their associations with repetitive elements. The expected number of the three families of MITE-Like elements to be found near or in the 16 genes is 1.5, assuming a random distribution and an average gene size of 10 kbp. This is calculated on the basis of the total copy number of these three elements in the genome. Not counting *Wuneng-Aa5,* which may be too short to be detected by the method used to determine copy numbers, there are six MITE-Like elements associated with six genes (Table 2). The discrepancy between the observed number and the expected number of MITEs in the 16 genes indicates a possible nonrandom association. However, a larger set of randomly selected genes needs to be analyzed to test this hypothesis further.

There are several examples indicating preferential insertion of transposable elements in genic regions (49, 50) and other examples of transposable elements avoiding genes (51). Phenomena such as preferential insertion of transposable elements into DNase I-hypersensitive sites (52) may provide a basis for nonrandom association of certain families of transposable elements with genes. Moreover, as in plants, the MITE-Like elements in *Ae. aegypti* also appear to show a bias against coding regions. The preference for A+T-rich sequences as target sites may provide one explanation for frequent insertion into the noncoding regions that are A+T-rich. It is possible there may be a higher order of insertion preference, such as chromatin accessibility as discussed above. There also could be selection pressure against insertions in coding regions because they may eliminate gene function. Regardless, the results presented here underline the similar associations of MITE-Like elements with the noncoding regions of genes in cereal grasses (5, 6) and *Ae. aegypti*.

Evolutionary Implications of the Association Between MITE-Like Elements and Genes. Transposable elements have generally been regarded as "selfish" DNA since the early 1980s (53–55) because of their "parasitic" nature. However, the question of whether transposable elements are just "junk" DNA to the host, or whether they can play important and even adaptive roles in organismal evolution, is currently actively debated (e.g., refs. 6, 12, 14–17, and 55–58).

It has been proposed that changes in the regulation of gene expression may be important for the evolution and variation of morphological and behavioral characters (e.g., refs. 15 and 19). Increasing number of cases have been identified in which transcriptional control of genes has been modified by preserved insertions of transposable elements (e.g., refs. 14–17). As discussed above, MITE-like elements are frequently found in the flanking regions of genes in both cereal grasses and the mosquito. MITEs have been shown to overlap previously identified cis-acting regulatory domains and poly(A) addition sites of wild-type plant genes (3, 5, 6). It has not been determined if any of the MITE-like sequences in *Ae. aegypti* genes are involved in gene regulation because no cis-regulatory elements have been determined for the genes shown in Table 2. However, based on the importance of chromatin structure in gene regulation (e.g., refs. 59–62), it is possible that the insertion and fixation of a short inverted-repeat element, adjacent to the transcription start site of a gene, could modify its transcription. It will be interesting to see whether or not, and to what extent, these MITE-like elements contribute to the evolution of gene regulation in plants and animals. The rapid ongoing progress in large-scale genomic studies of a few model organisms, as well as the molecular analysis of a diverse range of eukaryotic organisms (63–65) will undoubtedly facilitate our understanding of the potential importance of transposable elements in the regulatory evolution of host genes.

I am indebted to Dr. Henry H. Hagedorn, whose constant support made this work possible. I thank Dr. Margaret G. Kidwell for helpful advice and critical comments on the manuscript. I also thank Mr. Jun Isoe and Mrs. Julia A. Guzova for valuable technical assistance and the Sequencing Facility of the University of Arizona for their service. I am grateful to Dr. Anthony A. James for the kind gifts of *Ae. aegypti* genomic libraries. I also thank Drs. Nora J. Besansky, Henry H. Hagedorn, Margaret G. Kidwell, John H. Law, Damon Lisch, Daphne Pham, and Mr. Jun Isoe for sharing unpublished information. This work was supported by National Institutes of Health Grant HD 24869 to Drs. Henry H. Hagedorn and Ann M. Fallon and by a MacArthur Foundation grant to the Center for Insect Science of the University of Arizona.

1. Bureau, T. E. & Wessler, S. R. (1992) *Plant Cell* **4,** 1283–1294.
2. Bureau, T. E. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1411–1415.
3. Bureau, T. E. & Wessler, S. R. (1994) *Plant Cell* **6,** 907–916.
4. Tenzen, T., Matsuda, Y., Ohtsubo, H. & Ohtsubo E. (1994) *Mol. Gen. Genet.* **245,** 441–448.
5. Bureau, T. E., Ronald, P. C. & Wessler, S. R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 8524–8529.
6. Wessler, S. R., Bureau, T. E. & White, S. E. (1995) *Curr. Opin. Genet. Dev.* **5,** 814–821.
7. Morgan, G. T. & Middleton, K. M. (1990) *Nucleic Acids Res.* **18,** 5781–5786.

8. Ünsal, K. & Morgan, G. T. (1995) *J. Mol. Biol.* **248,** 812–823.
9. Morgan, G. T. (1995) *J. Mol. Biol.* **254,** 1–5.
10. Smit, A. F. A. & Riggs, A. D. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 1443–1448.
11. Smit, A. F. A. (1996) *Curr. Opin. Genet. Dev.* **6,** 743–748.
12. Flavell, A. J., Paerce, S. R. & Kumar A. (1994) *Curr. Opin. Genet. Dev.* **4,** 838–844.
13. MacRae, A. F. & Clegg, M. T. (1992) *Genetica* **86,** 55–66.
14. McDonald, J. F (1993) *Curr. Opin. Genet. Dev.* **3,** 855–864.
15. McDonald, J. F (1995) *Trends Ecol. Evol.* **10,** 123–126.
16. Britten, R. J. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 9374–9377.
17. Kidwell, M. G. & Lisch, D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** in press.
18. Britten, R. J. & Davidson, E. H. (1969) *Science* **165,** 349–358.
19. Britten, R. J. & Davidson, E. H. (1971) *Q. Rev. Biol.* **46,** 111–133.
20. Meinkoth, J. & Wahl, G. (1984) *Anal. Biochem.* **138,** 267–284.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
22. Lin, Y., Hamblin, M. T., Edwards, M. J., Barillas-Mury, C., Kanost, M. R., Knipple, D. C., Wolfner, M. F. & Hagedorn, H. H. (1993) *Dev. Biol.* **155,** 558–568.
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
24. Zuker, M. (1989) *Methods Enzymol.* **180,** 262–288.
25. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9373–9377.
26. Barillas-Mury, C. & Wells, M. A. (1993) *Insect Mol. Biol.* **2,** 7–12.
27. Rao, P. S. & Rai, K. S. (1987) *Heredity* **59,** 253–258.
28. Breslauer, K. J., Frank, R., Blocker, H. & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3746–3750.
29. James, A. A., Blackmer, K. & Racioppi, J. V. (1989) *Gene* **75,** 73–83.
30. James, A. A., Blackmer, K., Marinotti, O., Ghosn, C. R. & Racioppi, J. V. (1991) *Mol. Biochem. Parasitol.* **44,** 245–253.
31. Edwards, M. J. (1996) Ph.D. dissertation (Univ. of Arizona, Tucson).
32. Romans, P., Tu, Z. J., Ke, Z. X. & Hagedorn, H. H. (1995) *Insect Biochem. Mol. Biol.* **25,** 939–958.
33. Deitsch, K. W. & Raikhel, A. S. (1993) *Insect Mol. Biol.* **2,** 205–213.
34. Eggleston, P. (1992) *Nucleic Acids Res.* **20,** 4095.
35. Shoykoski, F., Morris, A. C., James, A. A. & ffrench-Constant, R. H. (1996) *Gene* **168,** 127–133.
36. Smartt, C. T., Kim, A. P., Grossman, G. L. & James, A. A. (1995) *Exp. Parasitol.* **81,** 239–248.
37. Burtis, K. C., Thummel, C. S., Jones, C. W., Karim, F. D. & Hogness, D. S. (1990) *Cell* **61,** 85–99.
38. Rosenzweig, B., Liao, L. W. & Hirsh, D. (1983) *Nucleic Acids Res.* **11,** 4201–4209.
39. Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R. & Waterston, R. (1992) *Nature (London)* **356,** 37–41.
40. Warren, A. M. & Crampton, J. M. (1991) *Genet. Res.* **58,** 225–232.
41. Laird, C. D. (1971) *Chromosoma* **32,** 378–406.
42. Fristrom, J. W. & Yund, M. A. (1973) *CRC Crit. Rev. Biochem.* **1,** 537–570.
43. Davidson, E. H., Galau, G. A., Angerer, R. C. & Britten, R. J. (1975) *Chromosoma* **51,** 253–259.
44. Besansky, N. J., Mukabayire, O., Bedell, J. A. & Lusz, H. (1996) *Genetica* **98,** 119–129.
45. Besansky, N. J. & Powell, J. R. (1992) *J. Med. Entomol.* **29,** 125–128.
46. Besansky, N. J. & Collins, F. H. (1992) *Parasitol. Today* **8,** 186–192.
47. Gale, K. R. (1987) Ph.D. dissertation (Univ. of Liverpool, Liverpool, U.K.).
48. Crain, W. R., Eden, F. C., Pearson, W. R., Davidson, E. H. & Britten, R. J. (1976) *Chromosoma* **56,** 309–326.
49. Plasterk, R. H. A. (1995) in *Mobile Genetic Elements,* ed. Sherratt, D. J. (Oxford Univ. Press, Oxford), pp. 18–37.
50. Cresse, A. D., Hulbert, S. H., Brown, W. E., Lucas, J. R. & Bennetzen, J. L. (1995) *Genetics* **140,** 315–324.
51. SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z. & Bennetzen, J. L. (1996) *Science* **274,** 765–768.
52. Craigie, R. (1992) *Trends Genet.* **8,** 187–190.
53. Doolittle, W. F. & Sapienza, C. (1980) *Nature (London)* **284,** 601–603.
54. Orgel, L. E. & Crick, F. H. (1980) *Nature (London)* **284,** 604–607.
55. Brookfield, J. F. Y. (1995) in *Mobile Genetic Elements*, ed. Sherratt, D. J. (Oxford Univ. Press, Oxford), pp. 130–153.
56. Charlesworth, B. & Langley, C. H. (1989) *Annu. Rev. Genet.* **23,** 251–287.
57. von Sternberg, R. M., Novick, G. E., Gao, G.-P. & Herrera, R. J. (1992) *Genetica* **86,** 215–246.
58. Shapiro, J. A. (1995) in *Mobile Genetic Elements*, ed. Sherratt, D. J. (Oxford Univ. Press, Oxford), pp. 1–17.
59. Wolffe, A. P. (1994) *Trends Biochem. Sci.* **19,** 240–244.
60. Hendrich B. D. & Willard, H. F. (1995) *Hum. Mol. Genet.* **4,** 1765–1777.
61. Orlando, V. & Paro, R. (1995) *Curr. Opin. Genet. Dev.* **5,** 174–179.
62. Kornberg, R. D. & Lorch, Y. (1995) *Curr. Opin. Cell Biol.* **7,** 371–375.
63. Hartl, D. L., Kafatos, F. C. & O'Brien, S. J. (1995) *Curr. Opin. Genet. Dev.* **5,** 705–708.
64. Leipe, D. D. (1996) *Curr. Opin. Genet. Dev.* **6,** 686–691.
65. Boguski, M. S., Cox, D. R. & Myers, R. M. (1996) *Curr. Opin. Genet. Dev.* **6,** 683–685.