Intragenomic rDNA ITS2 Variation in the Neotropical Anopheles (Nyssorhynchus) albitarsis Complex (Diptera: Culicidae)

Cong Li and Richard C. Wilkerson

From the Department of Entomology, Walter Reed Army Institute of Research, 503 Robert Grant Avenue, Silver Spring, MD 20910-7500.

Address correspondence to Dr. R. C. Wilkerson, Smithsonian Institution, Museum Support Center, 4210 Silver Hill Road, Suitland, MD 20746, or e-mail: wilkersonr@si.edu.

Abstract

We cloned and sequenced the rDNA internal transcribed spacer 2 (ITS2) of 4 species belonging to the neotropical Anopheles (Nyssorhynchus) albitarsis complex, that is, A. albitarsis, A. albitarsis B; Anopheles marajoara, a proven malaria vector; and Anopheles dannenstroem, a suspected vector. Even though the ITS2 sequences of these species were very similar (≤1.17% divergence), we found differences suitable for species identification and intragenomic variation of possible consequence in phylogenetic reconstruction. Variation came from 2 microsatellite regions and a number of indels and base substitutions. The existence of partially correlated subsets of clones in A. albitarsis is hypothesized either to be separate rDNA loci or to be semi-independently evolving portions of a single rDNA locus. No differences were found between males and females, suggesting that similar rDNA arrays exist on both the X and Y chromosomes. In addition, highly variant clones, possibly pseudogenes, were found in A. marajoara from Venezuela.

Concerted evolution is the process where all members of a multigene family are converted to the same sequence. The mechanism of concerted evolution has been attributed to either unequal crossing over or gene conversion (Smith 1976; Zimmer et al. 1980; Dover 1982). The rDNA is a multigene family that exists as one or more tandem arrays of many transcriptional units per cell (Gerbi 1985), where concerted evolution rapidly spreads mutations to all members of the gene family, even if arrays are located on different chromosomes (Dover 1982; Gerbi 1985; Tautz et al. 1988). In mosquitoes, each rDNA transcriptional unit is composed of an external transcribed spacer, an 18S subunit, an internal transcribed spacer 1 (ITS1), a 5.8S subunit, an ITS2, and a 28S subunit. Each rDNA unit within an array is linked to each other by an intergenic spacer (IGS). The transcribed spacers are thought to contain conserved structures important in forming the mature ribosomal amplon (Gerbi 1985; Thweatt and Lee 1990; Wesson et al. 1992; Parshwitz et al. 1993; van Nues et al. 1995). The rDNA sequence is a valuable source of information because the functional regions that produce the ribosomes are highly conserved but the transcribed and nontranscribed spacers have high interspecific and low intraspecific variability, making them useful for explaining relationships of recently diverged species and also useful as a basis for polymerase chain reaction (PCR) identification of morphologically similar species. As such, ITS1 and ITS2 have been used extensively in phylogenetic reconstruction of closely related and cryptic species complexes, as well as in the development of diagnostic species-specific PCR-based markers. However, because PCR can amplify all sequences of ITS present within the genome, variation among ITS sequences within individuals or species could result in inaccurate phylogenies and erroneous markers for species diagnostics. Consequently, identifying and quantifying levels of intragenomic and intraspecific variation among ITS sequences are of real importance.

The mosquito genus Anopheles (443 formally named species) contains all the vectors of human malaria parasites. Because many of the primary vectors belong to cryptic species complexes, it is necessary to have accurate phylogenetic reconstructions and species diagnostics for the study of malaria transmission and its relation to Anopheles evolution. Sequences of ITS1 and ITS2 are an excellent source for such information. However, in Anopheles, there are examples of rDNA intragenomic variation (Wilkerson et al. 2004; Fairley et al. 2005), but its prevalence and magnitude is not well studied. A consideration in search of an explanation for Anopheles intragenomic rDNA sequence variation is the possibility that
Intragenomic rDNA ITS2 Variation in the Neotropical Anopheles (Nyssorhynchus) albisetis Complex (Diptera: Culicidae)

1. REPORT DATE  
2. REPORT TYPE  
3. DATES COVERED  
00-00-2006 to 00-00-2006

4. TITLE AND SUBTITLE

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Walter Reed Army Institute of Research, Department of Entomology, Silver Spring, MD, 20910

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:

a. REPORT
   unclassified

b. ABSTRACT
   unclassified

c. THIS PAGE
   unclassified

17. LIMITATION OF ABSTRACT
   Same as Report (SAR)

18. NUMBER OF PAGES
   9

19a. NAME OF RESPONSIBLE PERSON

Standard Form 298 (Rev. 8-98)  
Prescribed by ANSI Std Z39-18
rDNA arrays are linked to different sex chromosomes, that is, they have been found only on the X of some species (Collins et al. 1989) or on both the X and Y of others (Marchi and Pili 1994). Consequently, rDNA arrays on sex chromosomes that exhibit limited recombination could result in incomplete homogenization.

In this study, we examine ITS sequences from multiple individuals of 4 closely related species of the neotropical Anopheles albimanus complex. These include Anopheles marajoara Galvão & Damasceno (Brazil, Venezuela, Colombia, and southern Central America), a known carrier of malaria (Conn et al. 2002); Anopheles deaneorum Rosa-Freitas (northern Argentina to western Brazil), a suspected malaria vector (Klein, Lima, and Tada 1991; Klein, Lima, Tada, and Miller 1991); and 2 other species Anopheles albimanus (southern Brazil, northern Argentina, and Paraguay) and Anopheles albimanus B (south, central, and eastern Brazil), whose role in malaria transmission is unclear. The 4 species can be reliably separated by random amplified polymorphic DNA (RAPD) (Wilkerson, Gaffigan, and Lima 1995; Wilkerson, Parsons, et al. 1995) and white gene (Mertz et al. 2005). We initially sought to examine the phylogenetic relationships among these species employing a number of genes, including ITS2 and cytochrome oxidase I (COI) (Wilkerson et al. 2005). We observed ambiguous results from direct sequencing; thus, we sought to clone and sequence ITS2 sequences from these species to quantify the magnitude and prevalence of intragenomic ITS2 variation and determine its effect on phylogenetic reconstruction. In addition, we sampled both male and female individuals within each species to investigate possible rDNA gender differences.

**Materials and Methods**

**Taxon Sampling**

Morphological identification of Anopheles albimanus s.l. was carried out using characters found in Linthicum (1988) and Peyton et al. (1992). Specimens used for cloning and sequencing are given in Table 1. They represent examples from progeny broods reported in Wilkerson, Gaffigan, and Lima (1995) and Wilkerson, Parsons, et al. (1995) from widely separated parts of the species ranges, including type localities of the 3 named species, and both sexes. In addition, a larger sample of Anopheles marajoara is represented because of its wide distribution and the possibility of a cryptic species, Anopheles albimanus E (Lehr et al. 2005). For brevity, letter designations are sometimes used that follow those in Wilkerson, Gaffigan, and Lima (1995) and Wilkerson, Parsons, et al. (1995): A = Anopheles albimanus, B = Anopheles albimanus B, C = Anopheles marajoara, D = Anopheles deaneorum, for example, in Tables 1–3. The ITS2 sequence reported here was used to design diagnostic primers (Li and Wilkerson 2005) that correctly identified all specimens first recognized with RAPD markers (Wilkerson, Gaffigan, and Lima 1995; Wilkerson, Parsons, et al. 1995) as follow: Anopheles albimanus, n = 56; Anopheles marajoara, n = 407; Anopheles deaneorum, n = 41; and Anopheles albimanus B, n = 56. Because there was complete concordance of data sets for a relatively large sample from many locations, we were able to base our conclusions on a much smaller number of cloned individuals.

**DNA Processing**

DNA was isolated from individual adult mosquitoes by phenol–chloroform extraction as described in Wilkerson et al. (1993). The ITS2 region was amplified using PCR primers based on conserved sequences in the 5.8S and 28S ribosomal subunits of Anopheles quadrimaculatus Say (Cornel et al. 1996). The boundaries of the ITS2 were determined as in Cornel et al. (1996, Figure 1A). PCRs were carried out as described in Li and Wilkerson (2005). Amplified PCR products were cleaned using QIAquick PCR purification kit (Promega, Madison, WI). About 200 ng of each purified PCR product was ligated into pCR-TOPO plasmid (Invitrogen, Carlsbad, CA). Two microtiter of the ligation reaction mixture was then transformed into competent One Shot cells (TOPO TA Cloning Kit, Invitrogen). Transformed cultures were plated on Luria-Bertani plates containing 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside, isopropyl-beta-D-thiogalactopyranoside, and 50 µg/ml ampicillin. Successful insertions are confirmed by PCR. Plasmids were extracted by the mini-prep method (Sambrook et al. 1989). Sequencing and alignment were as described in Li et al. (2005). Sequence statistics were obtained using PAUP version 4.0b4 (Swofford 1998). GenBank accession numbers are given on Table 2.

**Genetic Distance and Phylogenetic Analysis**

Uncorrected and pairwise distances were calculated by PAUP version 4.0b10 (Swofford 1998). The aligned ITS2 sequences were analyzed by maximum parsimony (MP) as implemented in PAUP and Bayesian analysis carried out using MRBAYES 3.1 (Huelsenbeck and Ronquist 2001). The parsimony and Bayesian analyses were chosen because gap information can be incorporated into both. Each gap was treated as a single character regardless of the length of the gap, under the assumption that a given gap is a result from one mutational event (Simmons and Ochoterena 2000). Single unique mutations were disregarded because of the possibility that they were the result of Taq replication error. Parsimony analysis was conducted using the heuristic search option with TBR (tree-bisection-reconnection) branch-swapping algorithm. Parsimony bootstrapping was done with 1000 pseudoreplicates with 10 random taxon addition replicates per pseudoreplicate. For Bayesian analysis, we used MrModelTest 2.2 (Nylander 2004) to choose an input evolutionary model. Markov chain Monte Carlo runs were 2 x 10^6 generations long with sampling every 5 x 10^5 generations, for a total of 4001 samples. Of these, the first 1001 were discarded as burn-in, which is well past the point where the likelihood plot reached a plateau.

**RNA Secondary Structure**

The putative secondary structure of the ITS2 was estimated using MFOLD (Zuker et al. 1999). A R-distance was
Table 1. Collection localities, number of clones, and GenBank accession numbers for specimens used in cloning of rDNA ITS2 of species belonging to the Anopheles (Nyssorhynchus) albitalis complex

<table>
<thead>
<tr>
<th>Species (M = male, F = female)</th>
<th>Code</th>
<th>Country</th>
<th>State</th>
<th>Locality</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>albitalis</em> (A1)</td>
<td>BR504(8)</td>
<td>Brazil</td>
<td>Paraná</td>
<td>Near Guaira</td>
<td>24°04'S, 54°15'W</td>
</tr>
<tr>
<td><em>albitalis</em> (AM2, AF2)</td>
<td>AR7(8)</td>
<td>Argentina</td>
<td>Buenos Aires</td>
<td>Baradero (type locality)</td>
<td>33°48'S, 59°30'W</td>
</tr>
<tr>
<td><em>albitalis</em> B (B1)</td>
<td>BR019(12)</td>
<td>Brazil</td>
<td>Ceará</td>
<td>Parapaba</td>
<td>3°25'S, 39°13'W</td>
</tr>
<tr>
<td><em>albitalis</em> B (BM2, BF2)</td>
<td>BR/SP 503(1)</td>
<td>Brazil</td>
<td>São Paulo</td>
<td>Near Registro</td>
<td>24°37'S, 47°53'W</td>
</tr>
<tr>
<td>marajoara (C1)</td>
<td>BR0236(12)</td>
<td>Brazil</td>
<td>Amazonas</td>
<td>Manaus</td>
<td>2°53'S, 60°15'W</td>
</tr>
<tr>
<td>marajoara (CF2, CM2)</td>
<td>BR/R001(10)</td>
<td>Brazil</td>
<td>Pará</td>
<td>Marajó Island (type locality)</td>
<td>1°00'S, 49°30'W</td>
</tr>
<tr>
<td>marajoara (C3)</td>
<td>COJ9</td>
<td>Venezuela</td>
<td>Cojedes</td>
<td>Finca &quot;Rosa Blanca&quot;</td>
<td>Not known</td>
</tr>
<tr>
<td>marajoara (C4)</td>
<td>COJ10</td>
<td>Venezuela</td>
<td>Cojedes</td>
<td>Finca &quot;Rosa Blanca&quot;</td>
<td>Not known</td>
</tr>
<tr>
<td>marajoara (C5)</td>
<td>BR4</td>
<td>Brazil</td>
<td>Roraima</td>
<td>Boa Vista</td>
<td>2°45'28&quot;N, 60°42'18&quot;W</td>
</tr>
<tr>
<td>marajoara (C0)</td>
<td>PIS9</td>
<td>Brazil</td>
<td>Amapá</td>
<td>North of Amapá</td>
<td>Not known</td>
</tr>
<tr>
<td>marajoara (C7)</td>
<td>ITB13763</td>
<td>Brazil</td>
<td>Pará</td>
<td>Near Iainha</td>
<td>Not known</td>
</tr>
<tr>
<td>deaneorum (D1)</td>
<td>BR/R007(15)</td>
<td>Brazil</td>
<td>Rondônia</td>
<td>Guajará Mirim</td>
<td>10°50'S, 65°20'W</td>
</tr>
<tr>
<td>deaneorum (DF2, DM2)</td>
<td>AR3(4)</td>
<td>Argentina</td>
<td>Corrientes</td>
<td>Corrientes</td>
<td>27°28'S, 59°50'W</td>
</tr>
<tr>
<td>deaneorum (DF3)</td>
<td>AR2(3)</td>
<td>Argentina</td>
<td>Corrientes</td>
<td>90 km West of Posadas</td>
<td>Not known</td>
</tr>
</tbody>
</table>

Letter designations, A1, A2, etc., correspond to Table 3.

calculated to compare the minimum free energy levels of all clones given by MFOLD (Sokal and Rohlf 1981).

**Results**

We cloned ITS2 PCR products from each sex of *A. albitalis* (n = 3), *A. albitalis* B (n = 3), *A. deaneorum* (n = 4), and *A. marajoara* (n = 8). Individuals from widely separated localities, including type localities, were used as described above and in Table 1. The larger sample of *A. marajoara* served to test for consistency of sequence in this widely distributed species and to test the hypothesis of the fifth species (Lehr et al. 2005). The number of clones from the 18 total individuals ranged from 3 to 28 (Table 2), giving a total of 217 clones. Alignment of sequences was straightforward because there was little sequence variation. Unless otherwise stated, the following description does not apply to 2 variant *A. marajoara* clones, C3.1 and C4.1, from individuals COJ9 and COJ10 from Cojedes, Venezuela (Tables 1 and 2), which we discuss separately.

**Inter- and Intragenomic Variation**

Total length of the ITS2 ranged from 344 to 365 bp. There were 4 microsatellite regions, (GT)<sub>7</sub> at position 118, (GA)<sub>3,11</sub> at 273, (CT)<sub>4</sub> at 147, and (GG)<sub>3</sub> at 345, all of which were common to all 4 species (Table 2). The first 2 regions were variable and contributed to all the length and intragenomic variations of ITS2 within *A. albitalis* B and *A. deaneorum*. However, repeat number was not species specific. There were 3 interspecific and/or intraspecific 2- or 3-base indels at positions 34, 271, and 236 and 8 single-base substitutions at positions 30, 43, 80, 248, 260, 268, 276, and 328. The polymorphic ACC and GC indels in *A. albitalis* occurred concordantly in clones from 2 individuals in about equal proportions, 5 of 15 in specimen A1 and 5 of 20 in specimen AM2. The third individual of *A. albitalis* (AF2) also had a low proportion of ACC indel clones (1 of 21), but the GC indel was not present in our sample. There was no indication of any obvious correlation of the other 2 polymorphic sites (positions 43 and 328) in this species with each other or the ACC and GC indels.

**Phylogenetic Analysis**

MP and Bayesian analyses were carried out for clones from 3 individuals of each species with the microsatellite regions removed and indels coded as 0 or 1 (Simmons and Ochoterena 2000). For clarity, because the combined and separate results were nearly identical, we present results from a single individual (Figure 1) of each species. Tree topology was the same for both analyses, but branch support was better with Bayesian analysis (support for both shown in Figure 1). *Anopheles deaneorum*, *A. marajoara*, and *A. albitalis* B all clustered into separate groups. However, *A. albitalis* clones separated into 2 groups corresponding to the correlated and partially correlated ACC and GC indels described above. Variation among clones was slight, with intragenomic base differences ranging from 0.0% to 0.57%, intraspecific variation ranging from 0.0% to 0.60%, and interspecific variation ranging from 0.28% to 1.17% (Table 3).

**Additional A. marajoara Clones**

Twenty-six clones from 3 individuals representing *A. albitalis* E of Lehr et al. (2005) were sequenced, 1 from Boa Vista in northern Brazil and 2 from Venezuela. Except for rare mutations, sequences of these clones matched sequences from other collection sites, including the type locality of *A. marajoara*, Marajó Island, Brazil (Table 2).

**Variant A. marajoara Clones**

Significant divergence was seen in a single clone from 2 individuals, COJ9 (clone C3.1) and COJ10 (clone C4.1), from Cojedes, Venezuela. These sequences were similar to each other but quite different from all other clones (Table 2).
<table>
<thead>
<tr>
<th>Specimen No.</th>
<th>Accession</th>
<th>Initial ΔG (kcal/mol)*</th>
<th>Allele name</th>
<th>GenBank Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1.1</td>
<td>AY828321</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
<tr>
<td>A1.2</td>
<td>AY828322</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
<tr>
<td>A1.3</td>
<td>AY828323</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
<tr>
<td>A1.4</td>
<td>AY828324</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
<tr>
<td>A1.5</td>
<td>AY828325</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
<tr>
<td>A1.6</td>
<td>AY828326</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
<tr>
<td>A1.7</td>
<td>AY828327</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
<tr>
<td>A1.8</td>
<td>AY828328</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
<tr>
<td>A1.9</td>
<td>AY828329</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
<tr>
<td>A1.10</td>
<td>AY828330</td>
<td>180.0</td>
<td>AJ310824</td>
<td>AJ310824</td>
</tr>
</tbody>
</table>

Species are as follows: A = Anopheles albimanus, B = A. albimanus, C = A. maculatus, D = A. salinarius. Column 1 and 2 are specimen number, clone number, and number of clones/total, for example, A1.1 is Anopheles albimanus specimen 1, not of clone number 1, which was found in 3 of 15 total clones. M and F denote male and female, if known. Only variable bases are shown. Single mutations that appeared only once were not considered because of the possibility they were due to Taq errors.

* The initial ΔG is the energy level of folded RNA. The region for the rest includes 91 bases in the 5.8S subunit and 43 bases in the 28S.
Clone C3.1 differed from other conspecific clones by 5.6-6.6% and C4.1 differed by 3.5-4.1%. Genetic difference between the 2 variant clones was 2.0%.

Secondary Structure of rRNA

The secondary structures of rDNA ITS2 were predicted by MFOLD (Zuker et al. 1999). Minimum free energies in kilocalories/mole were -181.5 to -185.5 for An. albitoris, -176.7 to -180.0 for An. albitoris B, -175.6 to -179.0 for An. marajoara, and -178.6 to -183.4 for An. deaneorum. The structures of the 2 variant An. marajoara clones, C3.1 and C4.1, have significantly lower energy (-167.4 and -168.9 kcal/mole; \(P < 0.01\) in the Student's \(t\)-test) and presumably lower stability than other An. marajoara. Figure 2A shows the predicted folding structures of all clones in the Albitoris complex except O.1 (Figure 2B) and C4.1 (Figure 2C). Note that the stem and loop near the presumptive ITS2 excision site (Fritz et al. 1994) (next to the right arrow in Figure 2A) is missing in the 2 variant clones (Figure 2B,C).

Discussion

A basic assumption about multigene families, such as rDNA, is that the processes collectively referred to as concerted evolution (gene conversion and unequal crossing over) maintain homogeneity of all copies (Hood et al. 1975; Smith 1976; Zimmer et al. 1980; Dover 1982). Mutations rapidly spread to all members of the gene family even if there are arrays located on different chromosomes (Dover 1982; Amheim 1983; Gerbi 1985; Tautz et al. 1988). In the case of noncoding regions, such as ITS2, this can lead to fixed interspecific differences and intraspecific homogeneity. The efficiency of homogenization of rDNA is usually high (Liao 1999), as exemplified by its common use as a marker for mosquito identification, most of which are derived from ITS2 (examples given in Wilkerson et al. 2004). However, as our results show, when mutation rates are higher than rates of homogenization, then variation within individuals may be greater than that observed between populations (see also Fritz et al. 1994; Onyabe and Conn 1999; Wilkerson et al. 2004). This possibility should be accounted for before rDNA is used for phylogenetic or population studies or as a basis for species-specific PCR primers.

ITS2 Variation

The ITS2 of all 4 RAPD-determined species in the Albitoris Complex were intragenomically and interspecifically variable. Length variation was limited (344-365 bp) and mostly attributable to the 2 variable microsatellite regions. In addition, there were a number of indels and base substitutions accounting for both the length and sequence variabilities (see Results and Table 2). Anopheles albitoris differed from the other species in having intragenomically variable ACC and GC indels (positions 236 and 271) and a variable T/C mutation at position 43. The ACC indel and the T/C mutations were
Figure 2. Predicted secondary structure of rDNA ITS2, including a combined 134 bases from the flanking 5.8S (91) and 28S (43) subunits. The secondary structure was common to clones from all species (A) except for the 2 clones shown in (B) and (C), which were found in 2 individuals of *Anopheles marajoari* from Cojedes, Venezuela.

used by Li and Wilkerson (2005) to design species-specific primers to identify *A. albitorris*, *A. albitorris B* and *A. demeicinum* as a group, and *A. albitorris*, respectively. Even though the above 3 (ACC, GC, and T/C) differences are not fixed in *A. albitorris*, PCR primers designed based on them still amplified as if there were only target sequence present and therefore still functioned to diagnose the species or groups of species.

Clones of *A. albitorris* ITS2 showed greater diversity than the other 3 species. In this case, intragenomic ITS2 variation within *A. albitorris* was greater than that between species in the complex. For example, the genetic distance between A1.1 and A1.3 (*A. albitorris*) was 0.57%, whereas the difference between A1.1 and D1.1 (*A. demeicinum*) was 0.28% (Table 3). This is an apparent example of mutation rates that are higher than homogenization rates. Intragenomic variation at ITS2, and in other parts of the rDNA gene array, is probably very common (Harris and Crandall 2000). In *Anopheles* mosquitoes, intragenomic variation has also been found in a number of other *Anopheles* species (Onyabe and Conn 1999; Wilkerson et al. 2004; Fairley et al. 2005) and in other mosquitoes in subfamily Culicinae (Black et al. 1989; Wesson et al. 1992; Miller et al. 1996; Beebe et al. 2000).

Effect of Microsatellites on Phylogenetic Reconstruction

Highly variable microsatellites may have confounding effects on phylogenetic and population genetics analyses. Harris and Crandall (2000) noted that if the multicopy nature of a marker is not recognized, inconsistent results can occur because alleles will not be distributed in a Mendelian manner. Cloning results verified our hypothesis that microsatellite variation was responsible for ambiguous sequencing results. However, in our case (data not shown), and in that of Vogler and DeSalle (1994), phylogenetic results were not affected by exclusion of microsatellite regions.

Chromosome Location of rDNA Arrays

Figure 1 shows 2 clusters of clones from the same individual of *A. albitorris* that are as different from each other as they are from the other 3 species. This suggests either that there are 2 rDNA loci within *A. albitorris* or that there are semi-independently evolving homologous rDNA loci. This could be caused by inefficient gene conversion and gene recombination. Multiple rDNA locations are not unusual, for example, there are 5 in humans (Gonzalez and Sylvestre 2001) and at least 2 in *Drosophila hydei* (Hennig et al. 1975) and grasshoppers (White et al. 1982). Similar explanations were considered for other *Anopheles* mosquitoes by Onyabe and Conn (1999) and Beebe et al. (2000).

The rDNA arrays are usually on chromosomes associated with sex determination. Kumar and Rai (1990) and Marchi and Pili (1994) mapped dozens of species of mosquitoes and found rDNA loci on the autosomes of culicine mosquitoes and on the X and Y chromosomes of *Anopheles*. In addition, they found loci on heterologous chromosomes in genus *Aedes*, the only confirmed example of loci on different chromosomes found so far in mosquitoes. In the Gambiase Complex (subgenus *Cella*), *Anopheles gambiae* Giles and
Anopheles arabiensis Patton have rDNA only on the X chromosome, whereas in the other species of the complex, it is on the X and Y chromosomes (Collins et al. 1989). In 2 Anopheles subgenus Nyssorhynchus species, Rafael et al. (2003) found rDNA on both the X and Y chromosomes. If rDNA was associated only with the X chromosome, as it is in *A. gambiae*, then males would be expected to have half the number of rDNA cistron copies (Collins et al. 1989) and half the haplotype diversity. If there were a subset of rDNA associated with the Y but not the X chromosome, then only males would be expected to have the Y-associated rDNA. In our sample, we did not see higher haplotype diversity associated with males or females.

Polanco et al. (2000) proposed 2 models to account for apparently correlated sets of rDNA other than loci on separate chromosomes: a haploptic single-lineage model for ITS evolution and a multilinage model for IGS evolution. The X and Y chromosomes in *Anopheles* are only partially homologous, and X chromosome variants do occur (Baimai et al. 1993; Rafael et al. 2003). Such factors may contribute to incomplete homogenization and could explain our finding of partially correlated intragenomic ITS2 haplotypes. As employed in the above studies, physical mapping using in situ hybridization is needed to confirm the location of rDNA loci in the *A. albifasciatus* complex species.

**Anopheles albifasciatus** Species E

Based on complete sequence of the mitochondrial COI, Lehr et al. (2005) proposed a fifth species (*A. albifasciatus* E) for the *albitarsis* complex in northern Brazil and Venezuela. We found no evidence from ITS2 sequence to support their conclusions. Isosequential ITS2 can occur in closely related *Anopheles* species (see above), and additional data are necessary to resolve this question.

**Variant A. marajoara** Clones

*Anopheles marajoara* individuals COJ9 and COJ10 from Cojedes, Venezuela, each had a different highly divergent clone (Table 2). The sequences are similar to the other *A. marajoara* ITS2 but differ from each other by about as much as *A. marajoara* does from the other 3 species. One of the clones (C3.1) has many mutations throughout its length, whereas the other (C4.1) is the same as all the other *A. marajoara* clones up to position 207, after which it mirrors the mutations in the more divergent clone. This "half-variant" could be due to template jumping, which could anomalously combine normal and variant sequence (Thompson et al. 2002). The relatively high sequence variation between these 2 clones suggests that these copies could be from nonfunctioning rDNA (pseudogenes). To test this possibility, we compared estimated minimum free energy levels and looked at the secondary structure predicted by the program MFOLD (Zuker et al. 1999). We found that the folding structures of these 2 clones have statistically significantly lower energies than all other clones (see above) and therefore lower structural stability. In addition, the variant clones lack a stem and loop at the ITS2 excision site present in all other clones (Figure 2). It is possible that this structural variation could affect cleavage efficiency of the precursor RNA, and it leads us to conclude that these copies probably come from nonfunctioning rDNA (pseudogenes). To our knowledge, this is the first of such report in a mosquito, but they have been documented in other organisms (Brownell et al. 1983; Benevolenskaya et al. 1997; Razafimandimbison et al. 2004). Further work is clearly needed to verify this observation.

**Application of ITS2 Intragenomic Variation**

Unambiguous identification of *Anopheles* malaria vector species is essential for the study of an array of factors that affect control and disease transmission. When morphological characters are not available, molecular alternatives must be found. In the case of the *Albitarsis* Complex, we initially looked at sequence of the rDNA ITS2 hoping to find a way to separate the 4 species. Ordinarily, it is possible to directly sequence the ITS2 without ambiguity, but in the Albitarsis Complex, direct sequence results were not clear because of intragenomic variation. Using ITS2 clones, we were able to identify primer locations that were not compromised by intraspecific and intragenomic variability (Li and Wilkerson 2005). Such variability often cannot be seen in direct sequencing and could lead to design of primers that will give erroneous or ambiguous results. For example, at position 236 (Table 2) in *A. albifasciatus*, there are 2 alleles, ACC present and ACC absent. In a consensus sequence, ACC absent copies are preferentially amplified because they are more common. If a primer was designed based on ACC present, then an *A. albifasciatus* sample would be misidentified as *A. albifasciatus* B or *A. dunnorum*. Similar results could occur with primers designed based on positions 43 and 328. With these data, we were able to design primers for the 4 species previously determined using RAPDs and provide an identification tool for an important malaria vector group.

**Acknowledgments.**

We greatly appreciate the support of those who helped collect and rear many of the specimens used in this study. These include Terry Klein and Eric Milsrey who were then members of the US Army Medical Research Unit in Rio de Janeiro. But we especially thank Jose Benito Lima without whose extra efforts many of samples would not have been available to us. We also thank Lee Weig at the Smithsonian Institution, Laboratory of Analytical Biology, for use of laboratory and sequencing facilities; David Erickson, Desmond Foley, and Jan Conn for thoughtful review of a draft of the manuscript; and Carl Schlichting for advice on statistical analysis. This research was performed under a Memorandum of Understanding between the Walter Reed Army Institute of Research and the Smithsonian Institution, with institutional support provided by both organizations. The material published reflects the views of the authors and should not be construed to represent those of the Department of the Army or the Department of Defense. This work was sponsored by NIH AI RO154139 to Jan Conn.

**References**


Beebe NW, Cooper RD, Folsey DH, Ellis JT. 2000. Populations of the south-west Pacific malaria vector *Anopheles farauti* s.s. revealed by ribosomal DNA transcribed spacer polymorphisms. Heredity. 84:244-253.


Rafael MS, Tadil WP, Beccon-Pimentel SM. 2003. Location of ribosomal genes in the chromosomes of *Anopheles darlingi* and *Anopheles nuneztovari* (Diptera: Culicidae) from the Brazilian Amazon. Mem Inst Oswaldo Cruz. 98:629-635.


Received September 19, 2005
Accepted September 15, 2006

Corresponding Editor: Robert Fleischer