Utility of the white Gene in Estimating Phylogenetic Relationships Among Mosquitoes (Diptera: Culicidae)

Nora J. Besansky and G. Thomas Fahey

Division of Parasitic Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia; and Department of Biology, Emory University

The utility of a nuclear protein-coding gene for reconstructing phylogenetic relationships within the family Culicidae was explored. Relationships among 13 species representing three subfamilies and nine genera of Culicidae were analyzed using a 762-bp fragment of coding sequence from the eye color gene, white. Outgroups for the study were two species from the sister group Chaoboridae. Sequences were determined from cloned PCR products amplified from genomic DNA, and aligned following conceptual intron splicing and amino acid translation. Third codon positions were characterized by high levels of divergence and biased nucleotide composition, the intensity and direction of which varied among taxa. Equal weighting of all characters resulted in parsimony and neighbor-joining trees at odds with the generally accepted phylogenetic hypothesis based on morphology and rDNA sequences. The application of differential weighting schemes recovered the traditional hypothesis, in which the subfamily Anopheleinae formed the basal clade. The subfamily Toxorhynchitinae occupied an intermediate position, and was a sister group to the subfamily Culicinae. Within Culicinae, the genera Sabethes and Tripteroides formed an ancestral clade, while the Culex-Deinocerites and Aedes-Huemgogus clades occupied increasingly derived positions in the molecular phylogeny. An intron present in the Culicinae-Toxorhynchitinae lineage and one outgroup taxon was absent in the basal Anopheleinae lineage and the second outgroup taxon, suggesting that intron insertions or deletions may not always be reliable systematic characters.

Introduction

The family Culicidae contains about 3,500 species of mosquitoes, which have been classified into three subfamilies: Anopheleinae, Toxorhynchitinae, and Culicinae (Knight and Stone 1977). Over 80% of these species are Culicines, making this the most diversified subfamily. Culicidae is considered to be monophyletic and a sister group of the family Chaoboridae (Wood and Borkent 1989). The fossil record indicates that the main mosquito lineages were well differentiated by the late Eocene, about 38 MYA (Bates 1949; Poinar 1992). The oldest mosquito fossil was found in Canadian amber (Bates 1949; Poinar 1992), but since all the main Dipteran lineages were present in the Upper Triassic 210-220 MYA (Krzeminski 1992 as cited in Oosawa 1976; Poinar 1992), the mosquito stem lineage probably existed by the Jurassic (Edwards 1932; Belkin 1962).

The Culicidae as a group have received relatively little attention from phylogenetic systematists, in spite of their spectacular diversity and worldwide distribution. A partial explanation must lie in the economic or medical importance of a handful of species and the consequent focus on identification and incrimination of these vectors of malaria, filariasis, and arboviruses. A further historical explanation is provided by the extreme rarity of fossils, which made it difficult to deduce the evolutionary sequence of morphological characters. Nevertheless, the Culicidae present an interesting set of evolutionary questions that might fruitfully be explored in light of accurately reconstructed phylogenetic relationships. For example, what features might have been responsible for the increased diversification seen in the Culicine clade? Referring to the superfamily containing both mosquitoes and chaoborids, Rohdendorf (1974, p. 51) described it as relatively monolithic and monoto nous. In mosquitoes, chromosome complement (2n = 6) is also invariant, with a single known exception in the Anopheleinae genus Chagasia (Chagasia balthana, 2n = 8) (White 1980; Rao and Rai 1987a). However, mosquitoes have diversified into species capable of breeding in almost any conceivable body of fresh, polluted, or brackish water, from bromeliads, tree holes, and crab holes to man-made containers and warm mineral water springs (e.g., Bates 1949). Radiation at both microevolutionary (sibling speciation) and macroevolutionary levels has been accompanied by chromosomal repatter ning (e.g., Green 1982; Coluzzi, Petrarca, and Di Deco 1985; Matthews and Munstermann 1994) and changes in genome size and organization (e.g., McLain, Rai, and Fraser 1987, Rao and Rai 1987b, 1990, Black and Rai 1988; Kumar and Rai 1990).

The vast majority of molecular phylogenies at all taxonomic levels have relied on either mitochondrial DNA (mtDNA) or nuclear ribosomal DNA (rDNA) (Hillis and Dixon 1991; Simon et al. 1994). Both share the advantages of technical ease, availability of conserved PCR primers useful in a wide range of taxa, and a large database from preceding studies. However, mtDNA may be less useful at higher taxonomic levels because of bias in base composition and multiple substitutions at silent sites. Regions of rDNA can be difficult to align, and nucleotide sites subject to saturation are more difficult to predict. Nuclear protein-coding genes offer an attractive alternative, combining the advantages of relative ease of alignment and unambiguous identification of rapidly evolving sites with the flexibil-
ity of choosing genes that evolve at appropriate rates for a given systematic question (Brower and DeSalle 1994; Soto-Adames, Robertson, and Berlocher 1994). Some recent examples have been the glucose-6-phosphate dehydrogenase gene in insects (Soto-Adames, Robertson, and Berlocher 1994), the alcohol dehydrogenase gene in *Drosophila* and related genera (Russo, Takezaki, and Nei 1995), and the phosphoenolpyruvate carboxykinase gene (Friedlander et al. 1996) and elongation factor-1α gene (Cho et al. 1995) in Lepidoptera. Several laboratories have been working with the single-copy nuclear gene *white*, a homolog of the *Drosophila melanogaster* eye color gene (Besansky et al. 1995; Zwiebel et al. 1995). Although a primary motivation has been the development of phenotypic markers for germline transformation, we have been exploring its utility in Dipteran systematics.

Decades before the wide acceptance of molecular systematics, Bates (1949) wrote, "The phylogeny of mosquitoes, like that of other animal groups, offers a fascinating field for speculation, but speculation it must remain because of the completely fragmentary character of fossil material in the group." Just 2 years later, Ross (1951) published his speculation, an evolutionary taxonomic classification of Culicidae based on morphological characters (fig. 1), unfortunately without providing character state information. Although it seems likely that the relationships were interpreted based on a subjective assessment of similarities, the placement of Anophelinae at the base of his tree reflected, and continues to reflect, the prevailing opinion of most mosquito taxonomists (Harrison and Scanlon 1975, p. 24), and agrees with preliminary results based on 18S and 28S rDNA sequences (Miller, Crabtree, and Savage 1996a; Pawlowski et al. 1996). However, the rDNA-based studies were not focused on the Culicidae per se, and included only four taxa from this family. No comprehensive, testable phylogenetic hypothesis of mosquito phylogeny has been published to date. In the present study, we reconstruct the evolutionary history of Culicidae as told by the *white* gene, using sequences from 13 species, representing three subfamilies and nine genera, and 2 species from the sister group Chaoboridae. The *white* gene is potentially informative for this purpose, but the information is not distributed uniformly across all characters, nor is base composition uniform across all taxa. Congruence between the *white* gene tree and either morphology-based or rDNA-based classifications depends upon the weighting scheme adopted.

**Materials and Methods**

Molecular Methods

Mosquitoes were field collected or obtained from colonies (table 1). Genomic DNA was extracted from individual specimens (Collins et al. 1987). Degenerate oligonucleotide primers for PCR were designed based on an amino acid alignment of *D. melanogaster* and *Anopheles gambiae* white genes (Besansky et al. 1995). These were WZ2E, 5′-AA[T,C]T[A,T]C[A,T]AA(T,C)-CCIGCIGA(T,C)TT(T,C)TA-3′, and WZ11X, 5′-TTIA{G,A}AA{G,A}AA{G,A}AAICCCIC{G,A}AA-3′, where brackets enclose degenerate positions and I = inosine. These primers correspond to positions 12483–

![Fig. 1.—Evolutionary taxonomic tree of the Culicidae, pruned from that of Ross (1951). At that time, *Chaoborus* was placed within Culicidae, and the designation "Megarhinus" (no longer in use) contained *Toxorhynchites*.](image-url)
12505 and 13314–13333, respectively, in the published An. gambiae white sequence (Besansky et al. 1995; GenBank accession number U29485). Each 50-μl PCR reaction contained 1.5 mM MgCl₂; 50 mM KCl; 10 mM Tris-HCl, pH 8.3; 0.001% gelatin, 200 μM each dNTP (Perkin-Elmer), 2.5 U Taq polymerase (Perkin Elmer, Boehringer Mannheim, or GibcoBRL), 100 pmol each primer, and 1 μl template DNA (one 100th of the DNA extracted from a single mosquito, except Bironella gracilis, one 2,000th). After 60 s at 93°C, 35 cycles of 20 s at 93°C, 60 s at 45°C, and 2 min at 72°C were performed. PCR products had restriction sites incorporated into both primers. Initially, PCR products were cloned into pBluescript II SK⁺ (Stratagene) following digestion with appropriate enzymes (EcoRI and Xba I). Subsequently, we found it more efficient to clone directly into pCRII using the TA Cloning Kit (Invitrogen). Southern analysis from a subset of taxa (An. gambiae, An. albimanus, Aedes aegypti) suggested that white was present in a single copy; we presume this is the case for the other taxa.

Sequences were determined manually from double-stranded plasmid template using the Sequenase Version 2 Kit (Amersham Life Science) with 35SdATP, or on an ABI 377 automated sequencer using the Dye-deoxy terminator kit (Perkin-Elmer/ABI). Universal and reverse primers as well as sequence-specific internal primers were used. Comparison of cloned sequences obtained from different individuals, or multiple clones from the same individuals, showed mismatches only in introns or, rarely, in third codon positions. The misincorporation rate of Taq polymerase during PCR was not quantified for these conditions, but based on rates of 0.22%–0.25% estimated for similar conditions (Saiki et al. 1988; Eckbush and Eickbush 1995), this low amount of error should not have significantly affected the results. Sequences have been deposited in GenBank (accession numbers U73826–U73839).

Sequence and Phylogenetic Analysis

Sequences were analyzed with the GCG Sequence Analysis Package (Genetics Computer Group 1994) as follows. Conceptual translations were handled with the MAP program to confirm their identity as white gene sequences. Guided by canonical intron splicing signals and comparisons to translations of available white gene cDNAs (D. melanogaster, Pepling and Mount 1990; An. gambiae, Besansky et al. 1995; Ceratitis capitata, Zwiebel et al. 1995), introns were identified and “spliced out” using the program SEQUED. (This was necessary because the introns could not be aligned reliably; their presence or absence at a particular location, however, is considered below). After eliminating the primer sequences, the remaining coding sequences were translated into amino acids (TRANSLATE) and aligned (PILEUP), with the default settings of gap weight = 3, gap length weight = 0.1. Visual inspection of the alignment revealed no obvious problems. Manual adjustment was therefore avoided to remove the possibility of subjective bias. However, the stability of this alignment under gap weights ranging from 0.66 to 100 was explored (Gatesy, DeSalle, and Wheeler 1993), and will be presented in the Results section. The amino acid alignment was used to guide the nucleotide alignment (fig. 3), upon which all subsequent phylogenetic analyses were based.

Basic sequence statistics and distance estimations were calculated, and the neighbor-joining method of phylogenetic analysis implemented, using MEGA (Kumar, Tamura, and Nei 1993). Codon usage bias was calculated using CODONS (Lloyd and Sharp 1992). Cladistic analysis was carried out with PAUP 3.1.1 (Swofford 1993), using the branch-and-bound algorithm unless otherwise noted. For both types of phylogenetic analyses, two chaoborid midges, Chaoborus asictopus and Eucorethra underwoodi, were used as outgroups. Alignment gaps were eliminated from all sequences in distance computations; in PAUP, gaps were treated as missing data and were also coded as two-state characters at the end of the data set. Successive-approximations character weighting, when applied, was performed with the reweight characters option of PAUP. Bootstrapping (Felsenstein 1985) based on 500 replications, each with 10 random additions of taxa and tree bisection-reconnection branch swapping, was performed on all analyses with the heuristic search option of PAUP 3.1.1.

Results

Sequence Variation

The structure of the white gene in An. gambiae (Besansky et al. 1995) is given in figure 2. The region selected for this study represents approximately one-third of the coding portion of the gene, and spans most of exon IV and part of exon V. Intron 4 was inserted between codons 107 and 108 (as defined in fig. 3) in all mosquitoes and both chaoborid midges examined. An additional intron interrupted codon 228 in all mosquitoes except anophelines, and in one of the midges, C. asictopus. This intron was not present in the white genes of D. melanogaster (Pepling and Mount 1990) or C. capitata (L. Zwiebel, personal communication). Because both introns were short in mosquitoes, ranging from 56 to 181 bp and from 53 to 469 bp, respectively, they did not pose a problem for PCR amplification. Thus, we were able to use genomic DNA as template rather than cDNA, an advantage given the relative instability of RNA and the difficulty in obtaining live or properly preserved specimens of certain taxa.

Alignment of the inferred amino acid sequences was not problematic, except for an uncertain area spanning amino acids 40–63 in which evolutionary constraints on sequence and structure may be relaxed. The
### FIG. 3.

-Nucleotide alignment of white coding sequences from the sampled taxa. These sequences have been assigned GenBank accession numbers. U match the representative *An. albimanus* sequence, dashes represent gaps introduced to improve the alignment, arrows represent intron positions. Above the letter code are the predicted amino acids occurring at each codon position.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Accession Numbers</th>
<th>Nucleotide Alignment</th>
<th>Amino Acid Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>An. albimanus</em></td>
<td>U73826473839</td>
<td>GGA ACG UCC TCG CGA GGG GCC GGT GGT ATT GAG TTG ACG CGC ACC</td>
<td>E. together</td>
</tr>
<tr>
<td><em>An. gambiae</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>An. freebordi</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bi. gracilis</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Sa. cyaneus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Tp. bambosa</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>An. triseriatus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hg. equinus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>As. aegypti</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>As. albopictus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cq. quinquefasciatus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>De. canescens</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Tx. rutilis</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cq. astictopus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. underwoodi</em></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The amino acid sequences are aligned based on the nucleotide sequences with gaps and inframeshifts introduced to maximize the alignment.
<table>
<thead>
<tr>
<th>SLA</th>
<th>VQK</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAT</td>
<td>NVS</td>
</tr>
</tbody>
</table>

An. albimanus

- TAT
  - G
  - C
  - T

An. gambiae

- TAT
  - G
  - C
  - T

An. freeborni

- TAT
  - G
  - C
  - T

Bi. gracilis

- TAT
  - G
  - C
  - T

Sa. cyaneus

- TAT
  - G
  - C
  - T

T. bubus

- TAT
  - G
  - C
  - T

Ae. triseriatus

- TAT
  - G
  - C
  - T

Hg. equinus

- TAT
  - G
  - C
  - T

Ae. aegypti

- TAT
  - G
  - C
  - T

Hg. gambiense

- TAT
  - G
  - C
  - T

Bi. gracilis

- TAT
  - G
  - C
  - T

Sa. cyaneus

- TAT
  - G
  - C
  - T

T. bubus

- TAT
  - G
  - C
  - T

Ae. triseriatus

- TAT
  - G
  - C
  - T

Hg. equinus

- TAT
  - G
  - C
  - T

Ae. aegypti

- TAT
  - G
  - C
  - T

Hg. gambiense

- TAT
  - G
  - C
  - T

Fig. 3 (Continued)
<table>
<thead>
<tr>
<th>Species</th>
<th>Accession Numbers</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>An. albimanus</td>
<td>CTC ACC ACG CTC TTC ACG CTC GCC ACG TTC ACC TTC GGC TAC CTC TTC TCC GCC ACG TTC TCC GCC ACG TCC TCC AXC TCC AXC GCC TCC TCC ACG TCC TCC GCC TCC</td>
<td></td>
</tr>
</tbody>
</table>
| An. gambiae      | G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ......
The data set was partitioned by three codon positions, and some first codon positions, to high, not much less than that between mosquitoes and species within Culicidae (~31%).

The degeneracy of the genetic code should allow third codon positions, and some first codon positions, to evolve at a faster rate. Indeed, table 2 shows that nonsynonymous changes represent only about one third of overall substitutions. The relative uncorrected frequencies of substitutions at first (nt1), second (nt2), and third (nt3) codon positions were found to be 18.2%, 10.3%, and 71.5%, respectively. The data set was partitioned by codon position, and the average pairwise divergence at each position within and among selected clades was calculated. From the results presented in tables 2 and 3, two important points deserve emphasis. First, overall divergence among the mosquito subgenera is relatively high, not much less than that between mosquitoes and chaoborids. This is consistent with an ancient radiation. Second, the nt3 positions, with average pairwise divergences above 0.50, are very likely to be saturated.

Saturation of third codon positions may obscure the historical phylogenetic signal with random noise. To explore this possibility, the skewness test statistic, g1, was calculated based on tree length distributions for 10,000 random samples of trees using the random-trees option of PAUP, for all positions, nt1 + nt2, and nt3 positions. The premise of the test is that the shape of the length distribution of all possible trees, or a sufficiently large random subset of them, is a sensitive indicator of phylogenetic signal (Hillis and Huelsenbeck 1992). Left-skewed distributions (g1 < 0) are expected for signal-containing data sets versus symmetrical (g1 = 0) or right-skewed (g1 > 0) distributions for random ones. The skewness statistics indicated significant structure (P << 0.01) for the complete data set, the nt1 + nt2, and the nt3 partitions (g1 = -0.63, g1 = -0.72, and g1 = -0.62, respectively), providing no justification for the a priori exclusion of the nt3 data.

Saturation at nt3 can also overwhelm the historical signal by one based on convergence if the pattern of nucleotide substitution is biased. One factor that could contribute to nonrandom change is constraint on base composition, reflecting either mutational bias or selection among synonymous codons. Equality of base composition was therefore examined by codon position within and between each species (table 4). For within-species comparisons, this was done indirectly by calculating an estimator of codon bias, the effective number of codons (ENc), developed by Wright (1990). The ENc statistic can vary from 20 to 61, reflecting use of only one codon per amino acid at one extreme through use of all possible codons at the other extreme. The value of ENc was correlated with G+C-richness at nt3 and introns, but not with base composition at nt1 or nt2. Among the most strongly biased in terms of codon usage and G+C composition at nt3 were the Anopheles and Culex species. Between-species comparisons of the equality of base composition showed that composition did not differ significantly at nt1 or nt2 (χ², P ≫ 0.05 in each case), but differed quite significantly at nt3 (χ², P < 0.01). Contributing most strongly to departure from equality were, again, the Anopheles and Culex species. Inequality of base composition can mislead methods of distance estimation (Gillespie 1986) and tree-building (Lockhart et al. 1994 and references therein). That it is the nt3 position that was biased in composition, yet this

Table 3

<table>
<thead>
<tr>
<th>Taxa</th>
<th>nt1</th>
<th>nt2</th>
<th>nt3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anopheles</td>
<td>0.071</td>
<td>0.053</td>
<td>0.439</td>
</tr>
<tr>
<td>Aedini</td>
<td>0.052</td>
<td>0.017</td>
<td>0.392</td>
</tr>
<tr>
<td>Anopheles-Culicinae</td>
<td>0.124</td>
<td>0.069</td>
<td>0.549</td>
</tr>
<tr>
<td>Anopheles-Toxorhynchitinae</td>
<td>0.153</td>
<td>0.131</td>
<td>0.544</td>
</tr>
<tr>
<td>Culicinae-Toxorhynchitinae</td>
<td>0.148</td>
<td>0.114</td>
<td>0.551</td>
</tr>
<tr>
<td>Culicidae-Chaoboridae</td>
<td>0.217</td>
<td>0.117</td>
<td>0.620</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Taxa</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>An. albimanus</td>
<td>0.14</td>
<td>0.16</td>
<td>0.19</td>
<td>0.29</td>
<td>0.25</td>
<td>0.76</td>
<td>0.75</td>
<td>0.24</td>
<td>0.19</td>
<td>0.22</td>
<td>0.24</td>
<td>0.29</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>An. gambiae</td>
<td>0.04</td>
<td>0.15</td>
<td>0.18</td>
<td>0.28</td>
<td>0.24</td>
<td>0.24</td>
<td>0.24</td>
<td>0.22</td>
<td>0.18</td>
<td>0.22</td>
<td>0.24</td>
<td>0.30</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>An. freeborni</td>
<td>0.03</td>
<td>0.02</td>
<td>0.20</td>
<td>0.29</td>
<td>0.26</td>
<td>0.26</td>
<td>0.25</td>
<td>0.24</td>
<td>0.20</td>
<td>0.22</td>
<td>0.27</td>
<td>0.32</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bi. gracilis</td>
<td>0.04</td>
<td>0.04</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.23</td>
<td>0.25</td>
<td>0.24</td>
<td>0.20</td>
<td>0.22</td>
<td>0.27</td>
<td>0.32</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sa. cyaneus</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.08</td>
<td>0.21</td>
<td>0.23</td>
<td>0.24</td>
<td>0.27</td>
<td>0.28</td>
<td>0.27</td>
<td>0.26</td>
<td>0.33</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tp. bambusa</td>
<td>0.10</td>
<td>0.09</td>
<td>0.08</td>
<td>0.04</td>
<td>0.23</td>
<td>0.22</td>
<td>0.23</td>
<td>0.21</td>
<td>0.22</td>
<td>0.23</td>
<td>0.25</td>
<td>0.30</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ae. triseriatus</td>
<td>0.08</td>
<td>0.08</td>
<td>0.09</td>
<td>0.07</td>
<td>0.07</td>
<td>0.10</td>
<td>0.18</td>
<td>0.16</td>
<td>0.22</td>
<td>0.23</td>
<td>0.26</td>
<td>0.30</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hq. equinus</td>
<td>0.08</td>
<td>0.08</td>
<td>0.09</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.02</td>
<td>0.18</td>
<td>0.18</td>
<td>0.23</td>
<td>0.24</td>
<td>0.27</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ae. aegypti</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.02</td>
<td>0.13</td>
<td>0.22</td>
<td>0.24</td>
<td>0.26</td>
<td>0.32</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ae. albopictus</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>0.20</td>
<td>0.22</td>
<td>0.25</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cx. quinquefasciatus</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.07</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.25</td>
<td>0.31</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>De. cancer</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.08</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
<td>0.05</td>
<td>0.25</td>
<td>0.31</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tx. rutillus</td>
<td>0.11</td>
<td>0.10</td>
<td>0.11</td>
<td>0.11</td>
<td>0.09</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.11</td>
<td>0.32</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. aegypti</td>
<td>0.14</td>
<td>0.15</td>
<td>0.15</td>
<td>0.14</td>
<td>0.16</td>
<td>0.15</td>
<td>0.15</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE.—Alignment gaps were excluded.
same position that contributed two thirds of the characters informative for parsimony, raises the possibility that conclusions drawn from these data may be misleading.

Phylogenetic Analysis

Of 762 aligned positions, 425 were variable and 341 were informative for parsimony. In the parsimony analysis, alignment gaps occurring between positions 117 and 189 were treated as missing data but coded as binary characters at the end of the data set. Heuristic searching with 1,000 replicates of random addition sequence, furthest addition option, tree-bisection-reconnection branch-swapping, and all characters weighted equally produced a single most parsimonious tree (length = 1,534 steps, consistency index excluding uninformative characters [cix] = 0.470, retention index [ri] = 0.405; fig. 4). There was strong support for the monophyletic origin of mosquitoes, of Anophelinae, and of Aedini (Aedes-Hemagogus) clades, and moderate support for a Sabethini (Sabethes-Tripteroides) clade, in agreement with prior morphological classifications. However, deeper relationships within Culicidae were not robust, and, contrary to expectation, Anophelinae occupied the most derived position of the tree, with Culex as sister group. Traditionally, Anophelinae has been considered basal within Culicidae on the basis of morphologic, cytogenetic, and other available molecular data (Ross 1951; Harrison and Scanlon 1975; Besansky, Finnerty, and Collins 1992; Miller, Crabtree, and Savage 1996a; Pawlowski et al. 1996; but see Belkin 1962).

Interestingly, neighbor-joining (NJ) analysis (Saitou and Nei 1987) gave similar results, using pairwise distances for all codon positions, corrected by the Tajima-Nei model (Tajima and Nei 1984; fig. 4).

Although alignment gaps were omitted prior to the NJ analysis, these factored into the parsimony analysis and could have misled it if the alignment in this region was unreliable. To assess the stability of the alignment, the amino acid sequences were realigned 14 times, using gap costs of 0.67, 1–10, 20, 50, and 100 following Gatesy, DeSalle, and Wheeler (1993). The alignment used above, resulting from a gap cost of 3, was identical to that produced with a gap cost of 4. However, gap costs between 5 and 20 produced a single alignment that differed slightly from the first, and each gap cost higher

Table 4

<table>
<thead>
<tr>
<th></th>
<th>NT1</th>
<th></th>
<th></th>
<th>NT2</th>
<th></th>
<th></th>
<th>NT3</th>
<th></th>
<th></th>
<th>INTRON(S)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>An. albimanus</td>
<td>28.1</td>
<td>22.1</td>
<td>21.3</td>
<td>28.5</td>
<td>20.9</td>
<td>38.3</td>
<td>26.1</td>
<td>14.6</td>
<td>6.3</td>
<td>10.7</td>
<td>45.5</td>
<td>37.5</td>
<td>17.1</td>
</tr>
<tr>
<td>An. gambiae</td>
<td>26.3</td>
<td>20.2</td>
<td>23.5</td>
<td>30.0</td>
<td>19.8</td>
<td>39.5</td>
<td>25.5</td>
<td>15.2</td>
<td>2.1</td>
<td>5.3</td>
<td>46.5</td>
<td>46.1</td>
<td>12.8</td>
</tr>
<tr>
<td>An. freeborni</td>
<td>26.9</td>
<td>22.9</td>
<td>20.8</td>
<td>29.4</td>
<td>19.6</td>
<td>38.4</td>
<td>26.5</td>
<td>15.5</td>
<td>9.0</td>
<td>11.4</td>
<td>41.2</td>
<td>38.4</td>
<td>21.8</td>
</tr>
<tr>
<td>Br. graciulus</td>
<td>27.7</td>
<td>20.1</td>
<td>23.7</td>
<td>28.5</td>
<td>21.7</td>
<td>37.8</td>
<td>25.7</td>
<td>14.9</td>
<td>18.1</td>
<td>15.7</td>
<td>32.9</td>
<td>33.3</td>
<td>37.2</td>
</tr>
<tr>
<td>Sa. cyaneus</td>
<td>28.2</td>
<td>24.5</td>
<td>17.4</td>
<td>25.9</td>
<td>21.6</td>
<td>40.2</td>
<td>24.9</td>
<td>13.5</td>
<td>27.0</td>
<td>32.4</td>
<td>16.6</td>
<td>24.1</td>
<td>35.9</td>
</tr>
<tr>
<td>Tp. bambusa</td>
<td>27.2</td>
<td>23.5</td>
<td>20.2</td>
<td>29.2</td>
<td>21.8</td>
<td>39.1</td>
<td>24.7</td>
<td>14.4</td>
<td>14.4</td>
<td>25.5</td>
<td>27.2</td>
<td>32.9</td>
<td>25.0</td>
</tr>
<tr>
<td>Ae. triseriatus</td>
<td>26.9</td>
<td>19.8</td>
<td>22.3</td>
<td>31.0</td>
<td>21.5</td>
<td>40.5</td>
<td>23.6</td>
<td>14.5</td>
<td>20.2</td>
<td>23.1</td>
<td>24.0</td>
<td>32.6</td>
<td>33.8</td>
</tr>
<tr>
<td>Hg. equinus</td>
<td>28.1</td>
<td>19.4</td>
<td>21.9</td>
<td>30.6</td>
<td>21.9</td>
<td>39.7</td>
<td>24.8</td>
<td>13.6</td>
<td>24.4</td>
<td>18.2</td>
<td>27.7</td>
<td>39.8</td>
<td>30.7</td>
</tr>
<tr>
<td>Ae. aegypti</td>
<td>27.7</td>
<td>21.9</td>
<td>20.2</td>
<td>30.2</td>
<td>22.3</td>
<td>38.8</td>
<td>25.2</td>
<td>13.6</td>
<td>18.6</td>
<td>19.4</td>
<td>26.9</td>
<td>35.1</td>
<td>27.2</td>
</tr>
<tr>
<td>Ae. albopictus</td>
<td>26.4</td>
<td>22.7</td>
<td>21.1</td>
<td>29.8</td>
<td>21.9</td>
<td>38.8</td>
<td>25.2</td>
<td>14.0</td>
<td>14.0</td>
<td>14.0</td>
<td>31.0</td>
<td>40.9</td>
<td>30.8</td>
</tr>
<tr>
<td>Cx. quinquefasciatus</td>
<td>26.0</td>
<td>21.9</td>
<td>21.5</td>
<td>30.6</td>
<td>21.9</td>
<td>38.8</td>
<td>26.0</td>
<td>13.2</td>
<td>5.4</td>
<td>7.4</td>
<td>45.0</td>
<td>42.1</td>
<td>21.9</td>
</tr>
<tr>
<td>De. cancer</td>
<td>28.5</td>
<td>21.5</td>
<td>21.1</td>
<td>28.9</td>
<td>21.9</td>
<td>39.3</td>
<td>24.8</td>
<td>14.0</td>
<td>13.2</td>
<td>17.4</td>
<td>32.2</td>
<td>37.2</td>
<td>31.2</td>
</tr>
<tr>
<td>Tx. rutilus</td>
<td>28.5</td>
<td>23.3</td>
<td>17.3</td>
<td>30.9</td>
<td>20.1</td>
<td>38.2</td>
<td>24.9</td>
<td>16.9</td>
<td>14.1</td>
<td>19.3</td>
<td>30.9</td>
<td>35.7</td>
<td>28.3</td>
</tr>
<tr>
<td>C. astictopus</td>
<td>27.0</td>
<td>23.7</td>
<td>19.3</td>
<td>29.9</td>
<td>22.0</td>
<td>41.5</td>
<td>21.6</td>
<td>14.9</td>
<td>22.4</td>
<td>19.5</td>
<td>31.1</td>
<td>27.0</td>
<td>35.8</td>
</tr>
<tr>
<td>E. underwoodi</td>
<td>24.9</td>
<td>27.8</td>
<td>16.6</td>
<td>30.7</td>
<td>21.2</td>
<td>39.8</td>
<td>24.5</td>
<td>14.5</td>
<td>24.9</td>
<td>24.1</td>
<td>23.2</td>
<td>27.8</td>
<td>25.5</td>
</tr>
<tr>
<td>Average</td>
<td>27.2</td>
<td>22.3</td>
<td>20.6</td>
<td>29.9</td>
<td>21.3</td>
<td>39.2</td>
<td>24.9</td>
<td>14.5</td>
<td>15.6</td>
<td>17.5</td>
<td>32.2</td>
<td>34.7</td>
<td>29.3</td>
</tr>
</tbody>
</table>

* Effective number of codons, Wright (1990).

Fig. 4.—White phylogeny of Culicidae, using both chaoborid midges as outgroups and weighting all positions equally. Branch lengths are proportional to the number of steps. Numbers above and below branches are bootstrap values (>50) for parsimony and neighbor-joining analyses, respectively. Brackets enclose subfamilies: A, Anophelinae. B, Toxorhynchitinae. C, Culicinae. Tribes: D, Culicini F, Aedini. F, Sabethini according to the classification of Knight and Stone (1977).
than 20 or lower than 3 produced additional alignments. Ultimately, there were no alignment-invariant positions found in the gap-filled region between positions 117 and 189 of figure 3, although outside of this region the alignment was constant. Accordingly, heuristic searching was repeated after exclusion of these alignment-ambiguous positions. This time, two most-parsimonious trees were found (length = 1,356 steps, cix = 0.458, ri = 0.402), one of which was identical to that in figure 4. The second differed only in the clustering relationships among the three \textit{Anopheles} species. Although the initial alignment (i.e., gap weight = 3) may not have been optimal, it had little apparent effect on tree topology and therefore was retained during further analyses.

Farris (1969) introduced an \textit{a posteriori} approach of successive approximations weighting to identify and differentially weight unreliable characters. As implemented in this study (using PAUP), after the initial search for the most parsimonious tree(s), characters were given a baseline weight of 1,000 and then reweighted according to the highest observed value of the rescaled consistency index across all trees. This process was repeated until the character weights and tree topologies stabilized. Although not normally implemented on a single parsimony tree, successive weighting was applied in this case to assess the stability of the hypothesis in figure 4, under the assumption that many highly variable nt3 positions would be downweighted or given zero weight. In the resulting tree (cix = 0.778, ri = 0.855) the relationships among almost all taxa were fully resolved (fig. 6). Like the analysis based on all

\textit{Culex} in particular, the elimination of nt3 positions should avoid the problem, since compositional bias was only detected at nt3. The exclusion of nt3 positions reduced to 115 the number of characters informative for parsimony. A branch-and-bound search resulted in five most-parsimonious trees (length = 394 steps, cix = 0.545, ri = 0.553). The strict consensus of these trees is shown in figure 5. Culicidae was monophyletic and the Anophelinae, Aedini, and Culicini (\textit{Culex-Deinocerites}) clades were recovered, but interrelationships among these clades were unresolved.

To explore the effect of differentially weighting characters based on their performance across the five equally parsimonious trees, successive-approximations weighting was implemented as described above. After the first iteration, a single tree was recovered whose topology was congruent with one of the original five trees. By the second iteration, the tree topology was unchanged and the character weights had stabilized (as determined by a further iteration). An additional 34 characters had been assigned zero weight; these were positioned either within six-fold degenerate codons or within highly variable regions of the protein (i.e., codons 26-70 and 204-218). In the resulting tree (cix = 0.778, ri = 0.855) the relationships among almost all taxa were fully resolved (fig. 6). Like the analysis based on all
positions, this weighted analysis recovered the Anopheles, Sabethini, and Aedini clades, and the topological relationships among their component taxa were congruent. However, in the total evidence tree, Toxorhynchitinae was basal, with Anopheles in the most derived position and clustering with Culicini. By contrast, the tree in figure 6 shows the Anopheles occupying the basal position, with Toxorhynchitinae intermediate and Culicini clustering with Aedini. This basic topology is congruent with the 18S+5.8S rDNA parsimony tree of Miller, Crabtree, and Savage (1996a). Within the subfamily Anopheles, Bironella was basal and the Anopheles subgenus Nyssorhynchus (represented by An. albimanus) was sister to the clade formed by the other two subgenera, Anopheles (An. freeborni) + Cellia (An. gambiae). Within the subfamily Culicinæ, Sabethini was basal, Culicini intermediate, and Aedini in a derived position. The NJ method produced nearly concordant results, when only nonsynonymous pairwise distances corrected by the Jukes-Cantor model (Jukes and Cantor 1969) were analyzed. The only difference between the parsimony and NJ trees was the arrangement of taxa within Aedini. Whereas the parsimony tree implied (albeit with very low bootstrap support) that the genus Aedes is paraphyletic (Ae. triseriatus clusters with Hg. equinus instead of the other Aedes species), the NJ tree preserved the monophyly of Aedes.

The inferred amino acid sequences were also subjected to parsimony analysis. Under the assumption of uniform weights, branch-and-bound searching recovered 104 equally parsimonious trees of 288 steps each. After two iterations of successive-approximations weighting as described above, character weights stabilized and three trees were recovered (not shown). Their topologies differed only in the relationships among three Culicinæ clades: the Sabethini, Aedini, and Culicini. Other than this, they were identical to one another and to the tree shown in figure 6. The poorer resolution afforded by the amino acid analysis is likely due to the reduction of informative characters. Even after elimination of third codon positions and successive reweighting, the nucleotide data set retained 81 informative characters, versus 67 for the amino acid data set.

Are the nt3 versus nt1+nt2 partitions really in conflict? To answer this question, parsimony analysis was performed using only nt3 positions. Heuristic searching, as described above, found a single tree (length = 1,143, cix = 0.455, ri = 0.375) in which the Anopheles clade was not only in a derived position, but now included Culex (not shown). Deinocerites, traditionally placed with Culex in the Culicini, was now positioned near the base of the tree, along with Toxorhynchitinae. The Aedini and Sabethini clades, however, were recovered as before. Disturbingly, bootstrapping did not support the monophyly of mosquitoes, an issue that has been beyond debate. If the nt1+nt2 data were forced into this nt3 topology, the most parsimonious arrangement of characters required a tree length of 421, an increase of 27 steps (~7%). On the other hand, forcing the nt3 data into the nt1+nt2 topology required 1,171 steps, an increase of 28 steps (~2%). A statistical approach, the Wilcoxon signed-rank test, was applied to determine whether the data partitions were significantly in conflict or actually estimating the same topology with error (Templeton 1983; Larson 1994). This test, applied separately to each data set, evaluates characters that change a different number of times on alternative topologies and asks whether these changes are significantly more parsimonious for one of the topologies. For the nt1+nt2 data set, the nt1+nt2 tree was significantly more parsimonious than the nt3 tree (n = 29, Ts = 55, two-tailed P < 0.01). However, for the nt3 data set, the nt3 tree was not significantly more parsimonious than the nt1+nt2 tree (n = 79, Ts = 1,191, t[4] = 1.9, two-tailed P > 0.05). If the validity of this statistical approach is accepted, these results suggest that the nt3 topology was a suboptimal estimation of the nt1+nt2 topology and no significant conflict exists.

Discussion

The protein product of the white gene belongs to a superfamily of Traffic ATPase (ABC) membrane transporters (Higgins 1992). Representatives from this superfamily are present in both prokaryotes and eukaryotes, where they perform a variety of import or export functions. Based on what is known from D. melanogaster, white helps transport eye pigment precursors, guanine and tryptophan, into pigment cells (Ewart et al. 1994). Characteristic of this superfamily are two domains, an ATP-binding domain that is highly conserved evolutionarily, and a transmembrane domain, which bears no sequence similarity between different types of transporters (Higgins 1992). The carboxy terminal region of white selected for this study encompasses five of six putative α-helices that make up the hydrophobic transmembrane domain (fig. 1 of Zwiebel et al. 1995). Within Culicinæ, pairwise sequence divergence in this region is below 30% at the nucleotide level, low enough to permit amplification from a broad range of specimens but high enough to contain phylogenetic information. Intron length is short enough to permit PCR amplification of the target fragment directly from genomic DNA, a practical advantage given the lability of mRNA.

The best estimate of mosquito phylogeny based on the white gene indicates that Anopheles forms the basal clade, Toxorhynchitinae occupies an intermediate position, and Culicinae occupies a derived one. Within Culicinæ, the tribe Sabethini forms an ancestral clade, while the Culicini and Aedini tribes occupy increasingly derived positions in the molecular phylogeny. However, these conclusions are premised on the exclusion of nt3 data and the a posteriori weighting of nt1+nt2, weighting schemes that we feel are justified for this group. Although the true phylogenetic history of a group of wild organisms can never be known, it is known that different genes and different regions within a gene can evolve at different rates, and recent analyses have shown that this may mislead attempts to estimate phylogenetic history (e.g., Yang 1996 and references therein). It is also known that heterogeneities in base composition can confound tree-building methods (e.g., Lockhart et al. 1997).
and (2) that mosquito radiation did not occur in sudden bursts of speciation, which would restrict the number of phylogenetically informative characters no matter how much time elapsed. Insufficient data exist to address either objection, so that the white gene phylogeny presented here should be interpreted cautiously, as a guideline and hopefully a stimulus for future work with this and other molecules, as well as morphology. Future studies could improve on this one by including representatives from other lines of culicids and by increasing the sampling density to include at least two species from within a given genus. Increased resolution and confidence in the phylogenetic hypothesis would also be expected by increasing the number of loci analyzed, rather than by increasing the length of the white gene analyzed.

Although much emphasis has been placed on the (visually) discordant placement of the subfamily Anophelinae in trees from different data partitions, it should be stressed that most shallower relationships recovered by the complete and partitioned (weighted nt1 + nt2) white gene data are not only congruent but also in accord with morphological and other molecular evidence. These clades are the subfamily Anophelinae and genus Anopheles within Anophelinae, the tribe Aedini and subgenus Stegomyia within Aedini (Ae. aegypti + Ae. albopictus), and the tribe Sabethini. In addition to the absence of a well-developed siphon in larvae, the subfamily Anophelinae has been united by the shape of the pupal respiratory trumpets. The genera Anopheles and Bironella have been distinguished by the length of the maxillary palps, which in adult female Anopheles are approximately as long as the proboscis (Komp 1942) but are shorter in Bironella. The tribe Sabethini is marked by the characteristic development of the ventral brush and siphon in larvae, and by the position or absence of specific hairs in larvae and pupae (Belkin 1962). The morphological classification of Aedini has been called into question (Belkin 1962; Wesson, Porter, and Collins 1992). Indeed, the white gene data agree with the rDNA ITS2 data of Wesson, Porter, and Collins (1992) in the separation of Ae. aegypti + Ae. albopictus (Stegomyia) from taxa of New World origin, Ae. triseriatus and Haemagogus. In fact, Wesson, Porter and Collins (1992) questioned the elevation of the Haemagogus lineage alone to generic status, and proposed instead the splitting of Aedes into two genera, one of which would presumably contain Haemagogus as well as Ae. triseriatus and allies. While premature, this proposal receives some support from the present study.

A number of recent reports have hinted at the phylogenetic utility of intron presence or absence (e.g., Simon et al. 1994). The rationale behind this approach is the reasonable assumption that intron insertion or de-
letion is a very rare event compared to the rate of nucleotide substitution. Data from the white gene indicate that both chaoborids and mosquitoes are polymorphic for the presence of an intron that, because it interrupts the same codon in the same place, is presumably the "same" intron. This intron is present in one of the chaoborid species, C. astictopus, but is absent from E. underwoodi. It is absent in the Anophelesinae, but present in all other subfamilies (see fig. 6). This same intron was absent in the white genes of the higher flies, D. melanogaster (Pepling and Mount 1990) and C. capitata (L. Zwiebel, personal communication). If chaoborids and mosquitoes have a monophyletic origin, and the Anophelesinae are accepted as basal, these data imply at least two independent gains (one in the lineage leading to C. astictopus, one in the lineage leading to Toxorhynchitinae and Culicinae) or three independent losses (one in the lineage leading to E. underwoodi, one in the lineage leading to Anophelesinae, and one in the lineage leading to Drosophila and Ceratitis). Interestingly, this intron has not been found in any other higher fly examined to date, but is present in certain other lower flies (unpublished data). This implies an even greater number of independent losses, gains, or both, and suggests that in spite of the fact that these insertion or deletion events are rare relative to base substitution, we may not understand the rules governing them well enough to rely on intron presence-absence for phylogeny reconstruction at deep levels of divergence.

Acknowledgments

Thanks especially to D. M. Wesson, who generously provided some of the DNA for this study, to R. Cooper for kindly providing Bironella specimens, to G. Fritz for the chaoborid specimens, and to F. Collins, A. Cornel, and C. Porter for additional specimens and T. Lehmann for helpful discussion. The constructive criticisms of R. DeSalle, J. Powell, F. Collins, and an anonymous reviewer were invaluable. We thank B. Holloway and the staff of the CDC Biotechnology Core Facility for synthesizing oligonucleotide primers. This investigation received financial support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. T. Fahey performed work contained herein in partial fulfillment of the requirements for his bachelor's degree at the Georgia Institute of Technology.

LITERATURE CITED


Jeffrey R. Powell, reviewing editor

Accepted January 13, 1997