Syst. Biol. 50(4):540–556, 2001

Toward Understanding Anophelinae (Diptera, Culicidae) Phylogeny: Insights from Nuclear Single-Copy Genes and the Weight of Evidence

JAROSLAW KRZYWINSKI,[1] RICHARD C. WILKERSON,[2] AND NORA J. BESANSKY[1]

[1]*Department of Biological Sciences, University of Notre Dame, P.O. Box 369, Notre Dame, Indiana 46556-0369, USA; E-mail: jaroslaw.krzywinski.1@nd.edu*

[2]*Walter Reed Biosystematics Unit, Museum Support Center, Smithsonian Institution, 4210 Silver Hill Road, Suitland, Maryland 20746, USA*

Abstract.—A phylogeny of the mosquito subfamily Anophelinae was inferred from fragments of two protein-coding nuclear genes, *G6pd* (462 bp) and *white* (801 bp), and from a combined data set (2,136 bp) that included a portion of the mitochondrial gene *ND5* and the *D2* region of the ribosomal *28S* gene. Sixteen species from all three anopheline genera and six *Anopheles* subgenera were sampled, along with six species of other mosquitoes used as an outgroup. Each of four genes analyzed individually recovered the same well-supported clades; topological incongruence was limited to unsupported or poorly supported nodes. As assessed by the incongruence length difference test, most of the conflicting signal was contributed by third codon positions. Strong structural constraints, as observed in *white* and *G6pd*, apparently had little impact on phylogenetic inference. Compared with the other genes, *white* provided a superior source of phylogenetic information. However, *white* appears to have experienced accelerated rates of evolution in few lineages, the affinities of which are therefore suspect. In combined analyses, most of the inferred relationships were well-supported and in agreement with previous studies: monophyly of Anophelinae, basal position of *Chagasia*, monophyly of *Anopheles* subgenera, and subgenera *Nyssorhynchus* + *Kerteszia* as sister taxa. The results suggested also monophyletic origin of subgenera *Cellia* + *Anopheles*, and the *white* gene analysis supported genus *Bironella* as a sister taxon to *Anopheles*. The present data and other available evidence suggest a South American origin of Anophelinae, probably in the Mesozoic; a rapid diversification of *Bironella* and basal subgeneric lineages of *Anopheles*, potentially associated with the breakup of Gondwanaland; and a relatively recent and rapid dispersion of subgenus *Anopheles*. [*Anopheles*; biogeography; evolution; *G6pd*; mosquitoes; phylogeny; simultaneous analysis; *white*.]

Anopheline mosquitoes (Culicidae, Anophelinae) are of prime medical importance as human malaria vectors, yet their phylogeny is poorly known. Traditionally, the subfamily is subdivided into three genera: *Anopheles*, *Bironella*, and *Chagasia*. *Chagasia*, a Neotropical genus, is regarded as sister to the other genera (Ross, 1951; Harbach and Kitching, 1998). *Anopheles*, with 97% of all anopheline species, is the most diversified genus, with 437 species classified into six subgenera: the cosmopolitan *Anopheles*, Old World *Cellia*, and the Neotropical *Kerteszia*, *Nyssorhynchus*, *Lophopodomyia*, and *Stethomyia*. Previous studies of relationships within Anophelinae have been taxon-limited, but some of them have hinted that the existing classification does not reflect natural groups (Conn, pers. comm.; Foley et al., 1998). Recently, a comprehensive morphology-based analysis of Anophelinae phylogeny was conducted by Sallum et al. (2000), who hypothesized that the subgenus *Anopheles*, as traditionally defined, is paraphyletic. Accordingly, they proposed a change of the existing status of genus *Bironella* and subgenera *Stethomyia* and *Lophopodomyia* into informal groups within the subgenus *Anopheles*.

In contrast to their morphology-based hypothesis regarding the status of subgenus *Anopheles*, molecular evidence tends to support traditional systematics. The analyses of nuclear *white* (Besansky and Fahey, 1997) and mitochondrial *COII* (Foley et al., 1998) genes placed *Bironella* as sister to *Anopheles* lineages. Because those results might have been biased as a result of limited taxon sampling, Krzywinski et al. (2001) addressed the issue of Anophelinae phylogeny by using the mitochondrial *ND5* gene and the *D2* region of ribosomal *28S* nuclear gene sequences from an expanded sample of taxa. Although no decisive support for *Bironella* as a sister taxon to *Anopheles* was shown, their data also favored traditional relationships and rejected the hypothesis of close affinity between *Bironella* and the subgenus *Anopheles*. They hypothesized that lack of resolution of those relationships reflected the rapid radiation of the *Bironella* and *Anopheles* subgeneric lineages. However, poorly supported

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2001		**00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Toward Understanding Anophelinae (Diptera, Culicidae) Phylogeny Insights from Nuclear Single-Copy Genes and the Weight of Evidence	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Walter Reed Biosystematics Unit,Smithsonian Institution,Washington,DC,20560**	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:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**17**	

relationships might have resulted from bias introduced through mutational saturation and unequal evolutionary rates among lineages observed in those genes.

A better picture of anopheline evolution not only would allow us to understand the relationships within the group, but also might help answer more fundamental biological questions concerning causes of the peculiar geographic distribution of Anophelinae. However, this area of research has been virtually neglected. To approach this goal, additional phylogenetic information is needed, preferably from an alternative source. Promising molecular candidates are protein-coding nuclear genes, which are not as strongly biased in nucleotide composition as mitochondrial genes, and are relatively easy to align—in contrast to rDNA (Brower and DeSalle, 1994). Here we analyze sequences of two protein-coding single-copy nuclear genes, *glucose-6-phosphate dehydrogenase* (*G_6pd*) and *white*, from anophelines. Glucose-6-phosphate dehydrogenase (*G_6pd*) plays a key role in regulating carbon flow through the pentose shunt pathway. The enzyme is considered to have an important housekeeping function and for this reason is expected to be relatively conservative in terms of amino acid changes. Soto-Adames et al. (1994) showed that *G_6pd* was informative for insect systematics over a very broad range, from sibling species to the ordinal level. The protein product of the *white* gene belongs to a superfamily of Traffic ATPase membrane transporters and helps transport eye pigment precursors, guanine and tryptophan, into pigment cells (Ewart et al., 1994). The gene was useful for reconstructing higher-level relationships in mosquitoes (Besansky and Fahey, 1997).

To determine the phylogenetic relationships within Anophelinae and to test the hypothesis of rapid radiation of the group, we have performed maximum parsimony and maximum likelihood analyses of the *G_6pd* and *white* gene fragments. Further, we explore the influence of unequal evolutionary rates and structural constraints, two attributes of the sequence data detected in *white* and *G_6pd*, on the inference of Anophelinae phylogeny. Because three of the four loci available for a simultaneous analysis appear incongruent with each other, we have attempted to localize the source of conflict and address the issue of treatment of multiple data sets containing conflicting information. We use the inferred trees to evaluate the phylogenetic hypothesis of Sallum et al. (2000). In addition, we propose a hypothesis for the evolutionary history of Anophelinae in a biogeographic framework.

Materials and Methods

The present data set contains 16 species of Anophelinae representing all anopheline genera and *Anopheles* subgenera and 6 species of other mosquitoes used as an outgroup (Table 1). All species except a representative of the subgenus *Lophopodomyia* were included in our previous analysis of rDNA and mtDNA genes (Krzywinski et al., 2001). Discussion of outgroup sampling is also presented there.

Genomic DNA was extracted following Collins et al. (1987) and resuspended in 100 μl of TE buffer (10 mMTris, 1 mM EDTA), pH 7.4. Sequences of the *white* gene primers WZ2E, WZ4E, and WZ11X used for polymerase chain reaction (PCR) are given in Zwiebel et al. (1995). G6PDF and G6PDR are modified from Soto-Adames et al. (1994), by exclusion of the Kpn I linker. Internal primers were used in conjunction with the flanking primers to facilitate amplification of the *white* and *G_6pd* gene from more difficult templates (Fig. 1, Table 2). PCR amplification conditions for the *white* gene were as described previously (Besansky and Fahey, 1997). *G_6pd* was amplified in 50 μl (total volume) with 2.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 0.001% gelatin, 200 μM each dNTP (GibcoBRL), 50 pmol of each primer, 2.5 U of *Taq* polymerase (GibcoBRL), and 1 μl of template DNA. Amplification was performed in the Perkin-Elmer 9600 thermocycler, with an initial denaturation at 94°C for 3 min, followed by 35 cycles of 94°C for 15 s, 50°C for 15 s, and 72°C for 60 s, followed by the final elongation step at 72°C for 10 min. PCR products were cloned directly into pGEM-T vectors (Promega). Cloned products were PCR-amplified, purified (StrataPrep PCR purification kit, Stratagene), and sequenced by using ABI BigDye terminator chemistry (Perkin-Elmer Applied Biosystems) on an ABI377 sequencer. Sequences of both strands were obtained from single clones. The sequencing error introduced by this method (3×10^{-4}, estimated by Kwiatowski et al., 1991)

TABLE 1. List of taxa examined, with geographical distribution of anophelines.

Subfamily	Genus	Subgenus	Species	Distribution
Anophelinae	*Anopheles*	*Anopheles*	*coustani* Laveran	Afrotropical/Palearctic
			intermedius (Peryassu)	Neotropical
			mattogrossensis Lutz and Neiva	Neotropical
			quadrimaculatus Say	Nearctic
			pseudopunctipennis Theobald	Neotropical
		Cellia	*gambiae* Giles	Afrotropical
			stephensi Liston	Oriental
		Kerteszia	*bellator* Dyar and Knab	Neotropical
			cruzii Dyar and Knab	Neotropical
			neivai Howard, Dyar and Knab	Neotropical
		Lophopodomyia	*squamifemur* Antunes	Neotropical
		Nyssorhynchus	*albimanus*Wiedemann	Nearctic/Neotropical
			albitarsis Lynch Arribalzaga	Neotropical
		Stethomyia	*kompi* Edwards	Neotropical
	Bironella	*Bironella*	*gracilis* Theobald	Australasian
	Chagasia		*bathana* (Dyar)	Neotropical
Culicinae	*Aedeomyia*		*squamipennis* (Lynch Arribalzaga)	
	Armigeres		*subalbatus* (Coquillett)	
	Orthopodomyia		*alba* Baker	
	Toxorhynchites		*amboinensis* (Doleschall)	
			rutilus (Coquillett)	
	Uranotaenia		*sapphirina* (Osten Sacken)	

should have no influence on the results of phylogenetic analyses for this level of divergence. Sequences have been deposited in GenBank, with accession numbers AF317805–AF317824 for G_6pd and AF318192–AF318209 for *white*. The *white* gene sequences of *An. albimanus*, *An. gambiae*, *Bi. gracilis*, and *Toxorhynchites rutilus* were obtained from GenBank (accession numbers U73839, U29486, U73829, and U73836, respectively).

The identity of both sets of sequences was confirmed by comparison of the conceptual translation obtained with TRANSLATE (Genetics Computer Group [GCG], 1997) to the sequences published by Besansky and Fahey (1997) and Soto-Adames et al. (1994). Before amino acid alignment with PILEUP (GCG), sequences corresponding to introns were identified and removed. After visual inspection, slight manual adjustments were performed in the *white* gene alignment. Nucleotide sequences of both genes were aligned according to the resulting amino acid alignments.

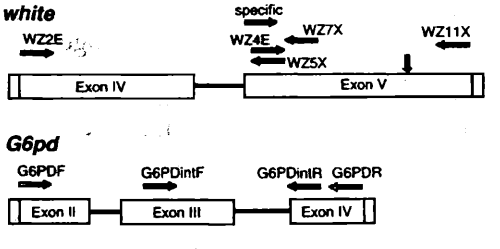


FIGURE 1. Structure of the *white* and G_6pd genes and the strategy for their amplification. The amplified fragments of exons are represented by open boxes; nonamplified regions are hatched. Lines connecting boxes are introns; horizontal arrows are primers. The location of an additional intron in the *white* gene of nonanophelines is indicated by a vertical arrow.

Phylogenetic analyses based on maximum parsimony (MP) and maximum likelihood (ML) were carried out with PAUP*4.d65 (Swofford, 1999), using heuristic searches and TBR branch-swapping. MP analyses were done by stepwise random addition of taxa with 1,000 replications; confidence in the inferred topologies was estimated by bootstrapping (500 bootstrap pseudoreplicates, each with 10 random additions of sequences). Apart from equal weighting, three other weighting schemes were applied to explore the influence of potential multiple substitutions on recovery of basal anopheline relationships: third position transitions given zero weight (nt3Ti = 0), third positions given zero weight (nt3 = 0), and amino acid sequences. The models of DNA substitution used in the ML analyses that best fit each of the data sets were determined by a likelihood ratio test, using MODELTEST 2.0 (Posada and Crandall, 1998). Probabilities of substitution classes and the Γ shape parameter (α) used in the subsequent analyses were estimated iteratively from the data by

TABLE 2. Primer sequences used to generate the *white* and G_6pd fragments.

Primer name	Sequence (5′ to 3′)[a]
WZ5X	XCC(AG)TT(AGT)AT(AG)TTCATIACICC
WZ7X	XTC(AG)AAIAC(AG)TT(TC)TC(AG)AAIGTCAT(AG)TTIGT
G6PDintF	GAA(AG)AAGT(AT)(CT)GA(ACT)GAGTTTTGG
G6PDintR	TTCTC(AC)AC(AG)AT(AGT)AT(AC)C(GT)(AG)TTCCA

[a]X is an Xba I linker (5′—CGCTCTAGA—3′); degeneracy is indicated by parentheses; I is inosine.

using the "tree scores" PAUP* option. Trees obtained from the unweighted MP analysis were used for an initial estimation of the parameters. The parameters were fixed in a ML heuristic search and the resulting tree was used to reoptimize the parameter values. Parameter estimation and tree searching were continued until both parameters and tree likelihood stabilized. For maximum likelihood tree searches and bootstrap analyses we used 100 replications. Potential effects of base frequency differences among taxa on phylogenetic reconstruction were explored by implementing LogDet/paralinear transformation (Lockhart et al., 1994) to calculate evolutionary distances and construct minimum-evolution trees.

Interior-branch tests (Rzhetsky and Nei, 1992) and relative rate tests (Takezaki et al., 1995) implemented in PHYLTEST 2.0 written by S. Kumar (with Kimura 2-parameter + Γ correction for multiple substitutions) were used to test the hypotheses of star phylogeny and rate constancy among lineages, respectively. The hypothesis of long-branch attraction was tested by Monte Carlo simulations according to Huelsenbeck (1997), using the Siminator program. Parametric bootstrapping was performed to assess whether the presence of a given clade in a tree might result from cumulative phylogenetic error in component branches rather than from significant phylogenetic signal (Huelsenbeck and Rannala, 1997).

The simultaneous analysis involved sequences of *white* and G_6pd combined with the sequences of the mitochondrial gene *ND5* and expansion segment *D2* of the nuclear ribosomal *28S* gene (Krzywinski et al., in press; also see Table 3, present paper). Combining the morphological data set of Sallum et al. (2000) with our molecular data was not possible because only four species were common to both studies. Before the combined analysis, the presence of conflict between the data sets was evaluated by using the incongruence length difference (ILD) test (Farris et al., 1995), implemented in PAUP* as described by Cunningham (1997).

RESULTS

Gene Sequences and Alignment

The *white* gene in *An. gambiae* contains five exons (Besansky et al., 1995). The segment used for this study, encompassing most of exon IV, intron 4, and the 5′ half of exon V, was amplified and sequenced from 15 mosquito species. Attempts to amplify the whole fragment from three outgroup species, *Aedeomyia squamipennis*, *Armigeres subalbatus*, and *Uranotaenia sapphirina*, were unsuccessful, yielding in these cases only partial sequences from the 5′-end of the fragment. The gene structure described above was observed in all anophelines except *An. albitarsis*, for which the sequence was intronless. Characteristic of nonanophelines was an additional intron within exon V (Fig. 1), as reported earlier by Besansky and Fahey (1997). The nucleotide alignment of the *white* gene with intron sequences removed was 801 characters long. The coding sequences of the complete fragment varied in length, ranging from 726 bp in *Ch. bathana* and *Orthopodomyia alba* to 780 bp in *An. albitarsis*, because of a highly variable region spanning codons 40–83. Except for *Nyssorhynchus* species, the sequence length was conserved within subgenera of *Anopheles*.

The G_6pd gene structure, as determined in *Drosophila melanogaster*, has four exons (Fouts et al., 1988). The fragment analyzed in this study, including nearly half of exon II, intron 2, exon III, intron 3, and the 5′-end of exon IV (Fig. 1), was obtained from 16 species. Most of this fragment (73% of the 5′-end) was also obtained from *An. intermedius*, *An. kompi*, *An. pseudopunctipennis*, and *Ch. bathana*. The gene could not be amplified from *Tx. rutilus* and *An. squamifemur*, possibly because of DNA degradation, long introns, mismatches between primers and target sequences, or some

TABLE 3. Mean nucleotide composition and numbers of aligned, variable, and informative sites for the *white*, *G_6pd*, *ND5*[a], and *D2*[a] genes.

			nt2			
	Overall	nt1	All sites	Hydrophobic[b]	Hydrophilic[c]	nt3
white						
A	20.4	27.0	20.9	12.3	25.0	13.4
C	27.4	21.1	25.6	31.3	22.9	35.4
G	26.3	29.9	15.0	4.2	20.4	34.1
T	25.8	22.0	38.4	52.2	31.7	17.1
Variable[d]	402 (50.1)	93 (34.8)	64 (24.0)	13 (15.5)	51 (27.9)	245 (91.8)
Informative	336	65	40	5	35	231
Aligned	801	267	267	84	183	267
G_6pd						
A	23.6	23.8	35.3	16.3	46.4	11.8
C	23.7	23.7	16.3	13.5	18.0	30.9
G	29.5	30.6	19.4	20.8	18.6	38.6
T	23.2	21.8	29.0	49.4	29.0	18.7
Variable[d]	247 (53.5)	60 (39.0)	43 (27.9)	13 (21.7)	30 (31.9)	144 (93.5)
Informative	197	40	25	5	20	132
Aligned	462	154	154	60	94	154
ND5						
Variable[d]	309 (58.9)	98 (56.0)	66 (37.7)			147 (84.0)
Informative	233	73	47			115
Aligned	525	175	175			175
D2[e]						
Variable[d]	244 (70.1)					
Informative	180					
Aligned	348					

[a]From Krzywinski et al. (2001).
[b]Refers to transmembrane or buried regions in *white* and *G_6pd*, respectively.
[c]Refers to external or exposed regions in *white* and *G_6pd*, respectively.
[d]In parentheses is given percent of aligned characters.
[e]For *D2*, only numbers of characters in regions included in the analysis are given.

combination of these. Introns 2 and 3 were found in all Anophelinae, whereas in the other mosquitoes intron 3 was absent. The coding sequences were equal in length, except for *Or. alba*, in which the sequence was one codon longer than in the other species. The resulting *G_6pd* nucleotide alignment was 462 characters long.

Nucleotide Composition and Sequence Divergence

Within genes, overall nucleotide frequencies were nearly equal (Table 3). Strong bias was found at the third codon positions (nt3), where G + C accounted for 70% of all bases. Between genes, mean nucleotide composition across species was similar, except for second codon positions (nt2), where *white* was rich in T + C and A + T predominated in *G_6pd*.

Analyses of nt3 sites in each species individually also revealed strong differences in base composition (Fig. 2). This heterogeneity was highly significant for both genes as revealed by a χ^2 test of independence ($P \ll 0.01$). In *G_6pd*, relatively low G + C content in two outgroup species, *Ad. squamipennis* and *Ur. sapphirina*, accounted for most of the heterogeneity, as we found by running the test after sequential exclusion of taxa with extreme base composition. Only exclusion of both species eliminated the heterogeneity in base composition ($P = 0.10$). In *white* the strongest differences were observed within the genus *Anopheles*: Members of *Kerteszia*, as well as *An. intermedius* and *An. mattogrossensis* (belonging to the Arribalzagia Series of the subgenus *Anopheles*), showed no or slight G + C bias, whereas G + C content in *An. gambiae* exceeded 90%. In contrast, the two Arribalzagia species were more biased for G + C at *G_6pd* nt3 sites. The homogeneity of base frequencies in *G_6pd* was not rejected by the χ^2 test when all codon positions combined were analyzed. In contrast, this homogeneity was rejected for the *white* gene ($P \ll 0.01$). This result reflects strong

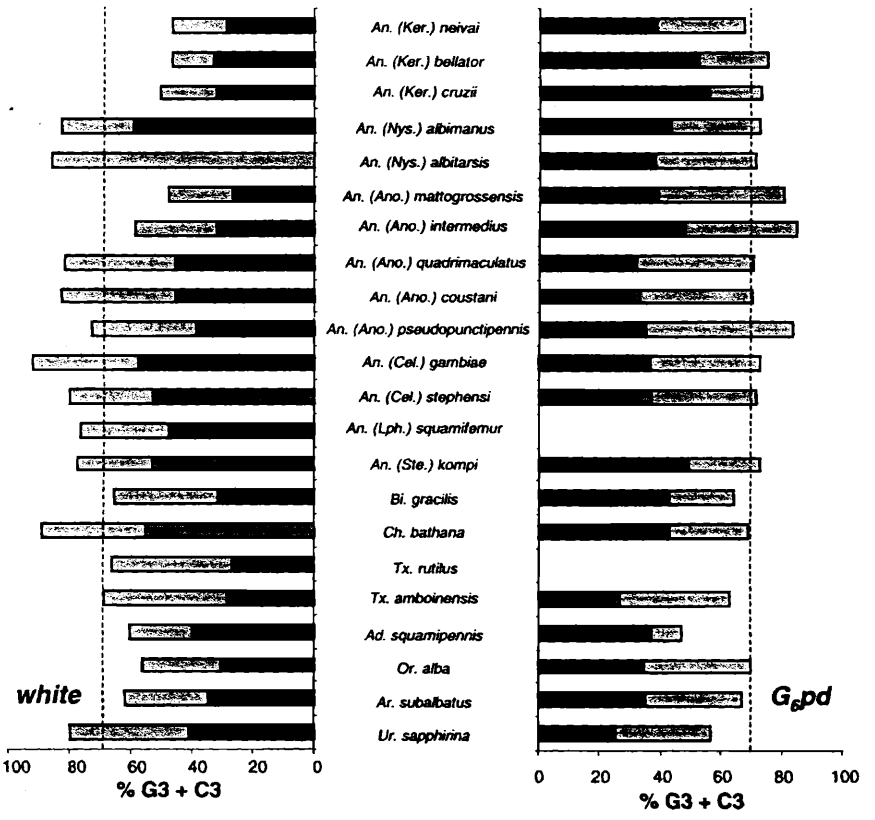


FIGURE 2. G + C content at third codon positions in the *white* and G_6pd genes in the mosquito species studied. Intron G + C percentage (hatched areas) is mapped onto exon G + C content to show the correlation between base composition of coding and noncoding regions of the genes. Dashed lines represent mean G + C content across all species. Note that *An. albitarsis* lacks introns in the *white* gene. G_6pd sequences from *An. squamifemur* and *Tx. rutilus* were not available.

influence of the nt3 sites on overall base frequencies in *white*, because nt1 and nt2 sites are quite homogeneous ($P = 0.999$ in both cases).

Ranges of sequence divergences at increasing taxonomic levels for G_6pd and *white* gene fragments are presented in Table 4.

TABLE 4. Ranges of uncorrected pairwise sequence divergences at increasingly inclusive taxonomic levels for G_6pd and *white* gene fragments. Note that two subgenera of the genus *Anopheles* are included.

	G_6pd		*white*	
	Min	Max	Min	Max
sg. *Kerteszia*	0.051	0.082	0.071	0.109
sg. *Anopheles*	0.101	0.165	0.137	0.238
Anopheles	0.051	0.247	0.071	0.298
Anophelinae	0.051	0.247	0.071	0.317
Outgroup	0.263	0.364	0.123	0.271
Ingroup-outgroup	0.218	0.323	0.178	0.336

Phylogenetic Analysis

G_6pd.—The MP analyses of G_6pd data under different weighting schemes resulted in trees showing little agreement in relationships within Anophelinae. Of the deeper nodes, only the basal position of *Chagasia* was recovered in all trees, whereas the position of other clades depended on the weighting applied. The position of *Bironella* as a sister group to *Anopheles* was recovered in only one of the trees derived under the nt3(Ti) = 0 weighting (and also in ML tree). In all other trees, *Bironella* was associated either with subgenus *Cellia* or as a sister taxon to a clade consisting of *Cellia* and a subset of species from the subgenus *Anopheles*. None of the searches recovered monophyly of the subgenus *Anopheles*. Subgenera *Cellia* and *Kerteszia* were found in interchanging positions either among basal or most-derived clades. Not surprising,

the bootstrap majority-rule consensus trees were very poorly resolved. Of the relationships inferred by using ML, only the clades well-supported in MP received ML bootstrap proportions >50% (Fig. 3).

white.—Phylogenetic analyses of the *white* gene based on both MP and ML consistently recovered deep relationships within the Anophelinae, with *Chagasia* as a basal lineage and *Bironella* as a sister group of *Anopheles* (Fig. 4). All but one analysis indicated monophyly of *Anopheles* subgenera; only equally weighted parsimony inferred that subgenus *Anopheles* was paraphyletic, showing a clustering of *An. pseudopunctipennis* with *Kerteszia* rather than with the remaining subgenus *Anopheles* species. In most trees *An. kompi* (subgenus *Stethomyia*) assumed a basal position relative to all other species of the genus *Anopheles*. Within *Anopheles*, two major monophyletic lineages were suggested but were not supported in MP trees: (*Nyssorhynchus* + *Kerteszia*) and (*Cellia* + subgenus *Anopheles*). *Lophopodomyia* was placed in various locations, depending on the inference method: as a clade branching early from an *Anopheles* stem (after *Stethomyia*), grouped with *Stethomyia* forming a basal lineage within

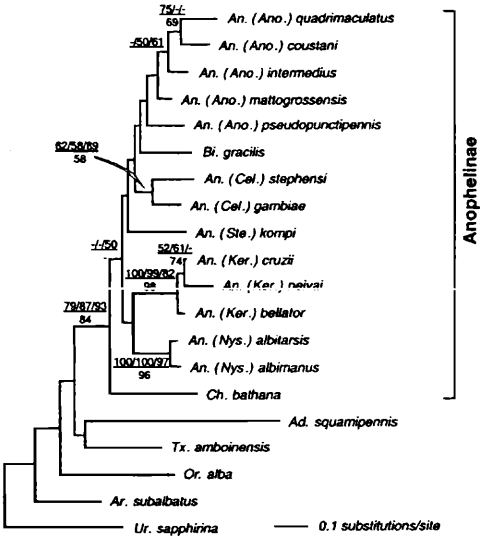
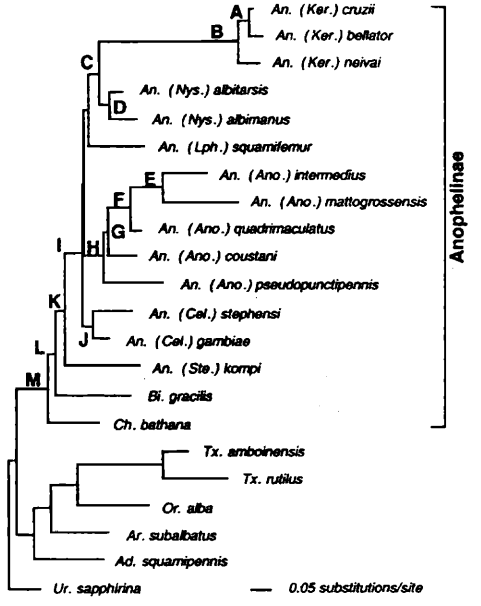


FIGURE 3. Phylogenetic relationships within mosquitoes inferred from the G_6pd gene sequences by maximum likelihood using a submodel (SYM; Zharkikh, 1994) within the general time reversible (GTR) model, assuming equal base frequencies, six substitution rates, and adjustments for among-site rate heterogeneity (SYM + I + Γ). The best-fit tree (−ln L = 4265.82) was obtained using transformation probabilities (A ↔ C = 1.935, A ↔ G = 5.237, A ↔ T = 2.136, C ↔ G = 3.580, C ↔ T = 13.680, and G ↔ T = 1) and gamma shape parameter (α = 1.2053) estimated with PAUP* and an observed proportion of invariable sites (I = 0.4654). Numbers at nodes represent bootstrap support >50% for equally/nt3Ti = 0/nt3 = 0 weighted parsimony above the line, and maximum likelihood analyses below the line.



Node	MP 1[a]	2[a]	3[a]	aa[a]	ML
A	94	74	62	76	78
B	100	100	100	99	100
C	-	-	-	-	-
D	97	88	98	97	93
E	75	67	-	52	95
F	65	65	-	68	91
G	59	-	-	55	52
H	-	86	93	56	75
I	-	53	-	60	70
J	72	95	90	91	89
K	57	-	64	63	62
L	-	89	93	90	62
M	74	76	81	92	92

[a]equal weighting
[b]nt3(Ti) = 0
[c]nt3 = 0
[d]amino acids

FIGURE 4. Phylogenetic relationships within mosquitoes based on the *white* gene sequences. The ML tree (−ln L = 7738.48) was inferred by using the Kimura (1981) three-parameter model with unequal (observed) base frequencies taken into account (K3Puf) and site-specific rate differences accommodated by assuming an observed proportion of sites to be invariable (I = 0.4956) and the remaining sites assumed to follow a discrete approximation of the gamma distribution (α = 1.3644 estimated with PAUP*). The following rates of the substitution were estimated by using PAUP*: A ↔ C and G ↔ T = 1, A ↔ G and C ↔ T = 5.416, A ↔ T and C ↔ G = 1.704. The table shows the bootstrap support for the nodes marked with letters.

Anopheles, or associated with (*Nyssorhynchus* + *Kerteszia*).

Despite strong variability in nt3 base composition among taxa, the tree constructed with the LogDet/paralinear method was similar to those inferred with MP or ML (data not shown). The only difference was paraphyly of *Cellia* (not supported by bootstrap), which was always derived as monophyletic by other inference methods and with other data sets.

Combined analysis.—The ILD test to assess congruence among data sets was applied to the nuclear genes used in this study and to the *ND5* and *D2* genes from Krzywinski et al (2001). (For *ND5* and *D2* description see Table 3.) The results suggested that all genes except ND5 were significantly incongruent (Table 5), despite the fact that the tree topologies derived from each gene separately were congruent for the more strongly supported relationships. Because conducting both separate and combined phylogenetic analyses may lead to better understanding of the data at hand (Sullivan, 1996), we combined all available sequences for a simultaneous analysis.

We conducted the analysis with and without *An. squamifemur*, a representative of the small and rare subgenus *Lophopodomyia*, because sequences of two genes, *ND5* and *G6pd*, were not available from this taxon (for opposing views on the effects of incomplete data matrices in phylogeny reconstruction, see Huelsenbeck, 1991; Wiens and Reeder, 1995).

In the full (22-species) ML tree, *An. squamifemur* was inferred as a sister taxon of *Nyssorhynchus* + *Kerteszia* (Fig. 5). Inclusion of this species had no effect on the position of other clades, although the support for *Anopheles* minus *Stethomyia* and *Cellia* + subgenus *Anopheles* was substantially less than for the 21-species data set. Parsimony analyses of the extended versus 21-species data set led to minor changes in tree topology. However, apart from *Nyssorhynchus* + *Kerteszia*, none of the relationships among *Bironella* and subgeneric clades of *Anopheles* were well-supported.

DISCUSSION

Phylogenetic Utility of the Genes

In their study of insects, Soto-Adames et al. (1994) suggested that the *G6pd* should be useful in phylogenetic reconstruction from generic to ordinal levels. Poor resolution and low support for the inferred clades show that the *G6pd* gene has limited utility as a phylogenetic marker within Anophelinae and perhaps in mosquitoes generally. This example supports the observation of Mardulyn and Whitfield (1999) that good performance of a gene in one taxon is not always easily extrapolated on the performance in other, even closely related, taxa.

In contrast to *G6pd*, the *white* gene appears much more informative in anophelines. However, the *white* gene alignment was nearly twice as long as the *G6pd* data. Because the proportions of variable and informative sites overall and for each codon position were nearly identical in both genes, one might wonder whether the higher phylogenetic information content of *white* is merely a matter of longer sequences analyzed. To answer this, we used a 5′-fragment of *white* (with complete sequences for all taxa) roughly equal in length to the *G6pd* fragment (453 aligned characters containing 206 informative sites) and reanalyzed this fragment by a MP

TABLE 5. Results of the ILD test. *P*-values are given for every gene pair comparison and each gene versus all other genes combined.

	G6pd		*ND5*		*D2*	All genes	
	All sites	1+2	All sites	1+2		All sites	1+2
White							
All sites	**0.001**	0.089	0.177	0.261	**0.001**	0.003	
1+2	0.077	0.041	0.910	0.945	**0.001**	0.563	0.182
G6pd							
All sites			0.557	0.611	0.013	**0.001**	
1+2			0.877	0.579	0.022	0.256	0.040
ND5							
All sites					0.055	0.511	
1+2					**0.005**	0.753	0.621
D2						**0.010**	**0.001**

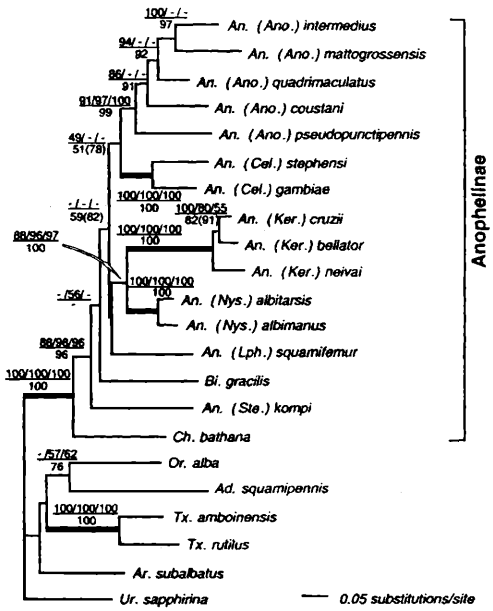


FIGURE 5. Phylogenetic relationships within mosquitoes based on the extended combined data set of *D2*, *ND5*, *G6pd* and *white* genes. The ML tree (−ln L = 21309.77) was inferred using a submodel (TIV; Rodriguez et al., 1990) within the GTR model and assuming unequal (observed) base frequencies, one transition rate (A ↔ G and C ↔ T = 5.406), four transversion rates (A ↔ C = 1.127, A ↔ T = 3.888, C ↔ G = 2.851, G ↔ T = 1), and adjustments for among-site rate heterogeneity (observed value of I = 0.310, α = 1.038). Substitution rates and gamma shape parameter α were estimated by using PAUP*. Numbers at nodes represent bootstrap support >50% for equally/nt3Ti = 0/nt3 = 0-weighted parsimony above the line, and maximum likelihood analyses below the line for 22 species and (in parentheses), for 21 species data sets. Thick lines represent the branches with 100% bootstrap support in all four analyses.

approach. The inferred trees were similar in topology and resolution to the trees inferred from the total *white* gene fragment, although bootstrap support for some clades dropped (e.g., 62% vs. 74% for Anophelinae). This result suggests that *white* is a better source of phylogenetic information in mosquitoes than *G6pd*.

Unequal Evolutionary Rates and Phylogeny Reconstruction

The relationship between *Nyssorhynchus* and *Kerteszia* inferred from different data sets apparently represents an interesting test case of long-branch attraction. The clade formed by these subgenera was recovered with high bootstrap support regardless of the inference method from *D2*, *D2* + *ND5*, all four genes combined (Krzywinski et al., 2001; and present data), and morphological characters (Sallum et al., 2000), strongly suggesting true phylogenetic relationships. Surprisingly, under unweighted MP of the *white* gene, *Kerteszia* was clustered with the subgenus *Anopheles*—the only instance when a clade was not supported by the *white* gene but was strongly supported by other genes or combined data. Felsenstein (1978) pointed out that parsimony converges to an incorrect phylogeny if the evolutionary rates along the lineages are strongly unequal. Relative rate tests indicate that the rates of the *white* gene indeed strongly depart from constancy (Table 6). Increased rates within long-branched lineages apparently led to multiple changes at numerous nt3 sites, obscuring the true phylogenetic signal for *Kerteszia*'s sister taxon relationship in unweighted MP. Elimination from the analysis of third codon transitions, elimination of third codon positions altogether, or implementation of the ML method that corrects for multiple substitutions results in the recovery of *Nyssorhynchus* + *Kerteszia* clade. Although the Monte Carlo simulations did not clearly show that branches are long enough to attract, long-branch attraction should not be excluded here. The HKY85 + Γ model used for data simulations may be too simple for the data at hand, or the test may not perform well when strong base composition differences exist across taxa (see below); these together make the test very conservative (Huelsenbeck, pers. comm.).

Another possible factor underlying incongruence between the unweighted MP analysis of the *white* gene and other analyses may be convergence in nucleotide content between different *Anopheles* lineages. Strong differences in nt3 base composition among species (Fig. 2) and associated differences in synonymous codon usage are observed in the *white* gene (see Besansky and Fahey, 1997). Nonrandom usage of codons is attributable to either mutational bias or selection. Moriyama and Powell (1997) suggested that most codon bias in *Drosophila* results from selection for efficient translation related to the isoaccepting tRNA availability in highly expressed genes. The information concerning levels of *white* expression in mosquitoes is very scarce. However, the evidence available from *An. gambiae*, the species

TABLE 6. Two-cluster relative rate tests for the *white* gene within *Anopheles*. Gapped regions and those with missing data were excluded before making distance calculations. In all comparisons *Chagasia* was used as a reference taxon.

Taxon A	Taxon B	La–Lb	V(La–Lb)	Z
Kerteszia	*Nyssorhynchus*	0.6848	0.2171	3.1548*
sg. *Anopheles*	*Cellia*	0.3409	0.0999	3.4107*
sg. *Anopheles*	*Kerteszia*	−0.3750	0.1882	1.9924*
sg. *Anopheles*	*Lophopodomyia*	0.2586	0.0965	2.6789*
Cellia	*Lophopodomyia*	−0.0823	0.0660	1.2466

*Z-value significant at 0.05.

with the greatest codon bias, suggests that this gene is expressed at very low levels (Besansky et al., 1995). Even if we assume that *white* is actually expressed at very high levels, but in few tissues and in short enough bursts to escape detection, substantial differences in codon usage among species are difficult to explain. Thus, selection seems unlikely to have played an important role in codon bias in *white*. Rather, a clear positive correlation between G + C content of introns and exon nt3 sites (Pearson product-moment correlation coefficient $r = 0.84$, $P < 0.01$; see also Fig. 2) strongly suggests that mutation bias is responsible for the observed patterns of codon usage. Interestingly, no such correlation was evident in the *G6pd* gene ($r = 0.43$, $P > 0.05$).

Sequence Conservation and Protein Structure

Base composition at the nt2 sites suggests that different evolutionary forces act on the *G6pd* and *white* genes, reflecting the structural constraints imposed on their protein products.

G6PD, a cytosolic globular protein, has a highly conserved three-dimensional structure of hydrophilic external parts and a hydrophobic core (Naylor et al., 1996; Notaro et al., 2000). The G6PD fragment under study is located close to the NH_2 terminus of the molecule and encodes portions of both external and core regions. When mosquito *G6pd* sequences are partitioned into nucleotide triplets encoding exposed or buried residues, as predicted on the basis of human G6PD tertiary structure (Notaro et al., 2000), sharply different patterns of nucleotide composition at nt2 are revealed in both groups (Table 3). Buried amino acids, more than half of which are hydrophobic, are strongly biased toward T, in accord with Naylor et al. (1995). In contrast, most of the exposed amino acids are hydrophilic in nature, with a predominance of A or G at nt2 positions. Nonsynonymous substitutions are located mainly in the hydrophilic external parts of the protein, similar to the findings of Notaro et al. (2000). However, most changes are conservative, replacements involving amino acids of similar properties.

The *white* gene encodes a protein belonging to a superfamily of ABC transporters (Higgins, 1992). Characteristic of these proteins are two domains: an ATP-binding domain located at the cytoplasmic face of the membrane, and a transmembrane domain spanning the cellular membrane. The fragment of *white* chosen for this study consists of the carboxy terminus of the cytoplasmic domain and most of the transmembrane domain, which in turn encompasses five putative membrane-spanning α-helices and four intervening loops located outside the membrane (Zwiebel et al., 1995). To preserve the conformational stability of the protein, the membrane-spanning fragments are expected to be rich in hydrophobic residues. Indeed, in comparison with the external (loop) regions, base composition in this gene partition is strongly biased toward C + T in second positions (Table 3).

These examples suggest that the structural constraints limiting character-state space at nt2 may be widespread in nature. In phylogenetic reconstruction, particularly in the case of more distantly related taxa, such constraints are a probable source of homoplasy in characters traditionally treated as most reliable (Naylor et al., 1995). In the present study, these constraints are unlikely to have contributed substantial homoplasy, given the small number of informative sites at nt2 in hydrophobic regions (Table 3); moreover, excluding external nt2 sites from the *G6pd* gene did not improve MP results. However, the

relationship between protein structure and sequence conservation revealed by our data may have some bearing on the improvement of existing models of amino acid sequence evolution (Liò and Goldman, 1999, and references therein).

The Weight of Evidence: Combining Independent Data Sets

The problem of how to analyze independent data sets is a subject of persistent controversy. Some authors (Miyamoto and Fitch, 1995) suggest that data partitions always should be considered separately in a taxonomic congruence framework (Mickevich, 1978). Others (e.g., Kluge, 1989) claim that data always should be combined in a simultaneous analysis because this maximizes the informativeness of the data and yields a strong estimate of phylogeny. Proponents of a third alternative (Bull et al., 1993; de Queiroz, 1993) suggest that the decision to combine the data should depend on the degree of incongruence between separate partitions. Several statistical tests have been used to evaluate incongruence among data partitions (Templeton, 1983; Kishino and Hasegawa, 1989; Rodrigo et al., 1993; Farris et al., 1995; Huelsenbeck and Bull, 1996). However, the existing tests seem too conservative and inadequate to address the issue of when simultaneous analysis should be performed (Sullivan, 1996; Cunningham, 1997; Remsen and DeSalle, 1998). Here we applied the ILD test, which performs better in predicting the compatibility of combined data than goodness-of-fit tests do (Cunningham, 1997) and which is commonly used in phylogenetic studies (Caterino et al., 2000). According to this test, only *ND5* sequences can be combined with any other gene (Table 5). Interestingly, when topology and bootstrap values were examined in separate gene trees, topological incongruence was generally limited to unsupported or poorly supported nodes, whereas highly supported branches were congruent across the trees. When $P = 0.01$ was taken as a significance threshold (Cunningham, 1997), G_6pd was congruent with *D2*, and also with the *white* gene at nt1 + nt2 positions. Congruence was also suggested when nt1 + nt2 positions from one gene were compared with all positions of another protein-coding gene. The discrepant results between all positions and nt1 + nt2 positions cannot be completely accounted for by a lack of resolution, and therefore the *perception* of congruence, in the latter data partition because nt1 + nt2 of *white* produced well-resolved topology (Fig. 4). Even after exclusion of nt3 sites, the ILD test indicated that the *white* gene and also the *ND5* gene were incongruent with *D2*. Such an incongruence may result from extreme difference in the evolutionary rates along some branches. Taken together, these results indicate that weak conflicting signals, probably coming from sites affected by multiple substitutions combined with differing compositional biases, have profound effects on the ILD test results. Moreover, they suggest that improving the phylogenetic reconstruction model by eliminating such sites will improve the congruence between data sets. Despite the professed incongruence, simultaneous analysis did not reduce, and in some cases substantially increased, support for all clades. We agree with the notion that when different partitions yield strongly different and well-supported relationships, simultaneous analysis should not be performed. However, when the topological incongruence is concentrated in unsupported clades, as in the present study, simultaneous analysis appears beneficial. Apparently, when the partitions are combined, phylogenetic signals from separate partitions have additive properties, resulting in stronger support for the inferred clades. Moreover, different partitions resolve different regions of the tree, a property discussed earlier by Pennington (1996). Our analysis suggests that the congruence, or lack thereof, between data sets from real taxa is a complex problem not yet well understood.

Anophelinae Phylogeny

Most of the relationships inferred with the combined data (Fig. 5) are well supported and in agreement with previous morphological and molecular studies. Because the analyzed loci are unlinked and independently give congruent topology for better supported clades, with the areas of conflict limited to unsupported or poorly supported branches, we believe that our phylogenetic hypothesis is a reliable estimate of Anophelinae phylogeny, with two exceptions.

The position of *Bironella* diverging after *Stethomyia* is probably incorrect; instead, we

believe the *white* gene tree, with *Bironella* as a sister taxon to *Anopheles*, reflects a true evolutionary history that is swamped by the noise introduced by other genes. Monophyly of *Anopheles* relative to *Bironella* was suggested by earlier molecular studies based on a more limited sampling of *white* (Besansky and Fahey, 1997) and *COII* (Foley et al., 1998) genes. In contrast, Sallum et al. (2000) stated that *Anopheles*, as traditionally defined, is paraphyletic. According to their newly proposed phylogeny of the subfamily, one of the lineages within the genus *Anopheles* would contain species of *Bironella* as well as the subgenera *Lophopodomyia* and *Stethomyia* arising among species of the subgenus *Anopheles*. Moreover, *An. pseudopunctipennis* of the subgenus *Anopheles* would occupy a basal position within this lineage, before the divergence of *Lophopodomyia* and well before *Bironella* and *Stethomyia* diverge. The present results, however, contradict this hypothesis. In both the *white* gene and combined data tree, *Bironella* is one of the basal branches of Anophelinae, and all species of subgenus *Anopheles*, including *An. pseudopunctipennis*, form a very strongly supported monophyletic clade. According to the Kishino–Hasegawa test, the tree constrained to reflect the hypothesis of Sallum et al. (2000) is significantly less likely ($t = 3.8099$, $P = 0.0001$) than the tree shown on Fig. 5. Similarly, we rejected a close affinity of *Bironella* to the subgenus *Anopheles* in our previous study (Krzywinski et al., 2001). Monophyly of the subgenus *Anopheles* sensu Sallum et al. (2000), that is, including two other *Anopheles* subgenera and *Bironella*, is based on six synapomorphies. However, the authors indicate that those characters are homoplasious and rather inconsistent (bootstrap <50%, Bremer support = 2). Low support from morphology and strong contradictory evidence from molecular data indicate that the hypothesis of Sallum et al. (2000) concerning conflicting clades is based on data compromised by homoplasy. As such, it is difficult to argue that the characters used in their study have enough resolving power for those problematic relationships to serve as a foundation for a substantial change of the established systematics of the group. We conclude that discrepant hypotheses reflect a different interpretation of the results rather than real conflict between morphology and molecular data.

In fact, reliable inference of the relationships among *Bironella* and basal *Anopheles* clades may be problematic. Lack of resolution at these levels, characteristic of all phylogenetic studies to date, can result from (1) combining conflicting signal if different data partitions experienced different evolutionary histories (gene trees vs. species trees), (2) strongly unequal rates of evolution, or (3) nearly contemporaneous radiations. Because the genes used in this and our previous study (Krzywinski et al., 2001), did not yield major phylogenetic conflicts, we discount the first explanation. Separate analyses of the *G6pd* and *white* data, whether partitioned (1) according to nt1 + nt2 versus nt3 positions or (2) as hydrophilic versus hydrophobic regions (data not shown), indicate that rate differences, although contributing to the problem, are unlikely culprits. The most probable explanation, which reconciles morphological and molecular results, is a rapid radiation of *Bironella* and basal clades within *Anopheles*. This scenario, recently suggested on the basis of analysis of the mitochondrial and nuclear ribosomal genes (Krzywinski et al., 2001), is also consistent with the present results, in which the relevant branches are very short (Table 7) and poorly supported.

The position of the subgenus *Lophopodomyia*, represented by a single species *An. squamifemur*, remains unsupported and uncertain. First, character sampling for *An. squamifemur* was sparse, with only *D2* and *white* sequences being obtained. Second, *Lophopodomyia* might have arisen in the process of a rapid radiation of basal *Anopheles* clades, and more characters may not be a remedy for the lack of support. The decreased bootstrap values for the lineages of subgenus *Anopheles* + *Cellia* and the genus *Anopheles* excluding *Stethomyia* in the combined analysis of the extended data set versus the 21-species data set (Fig. 5) suggests that *An. squamifemur* is indeed a problematic taxon. Third, some weak conflict over the position of *Lophopodomyia* seems to exist between the data sets. In contrast to the *white* data tree, the *D2* data tree placed *An. squamifemur* as the most basal clade of Anophelinae (data not shown). However, consideration of this apparent conflict should be tempered by the possibility that the position of *Lophopodomyia* on the *D2* gene tree may result from misalignment of some fragments of *An. squamifemur* sequence relative to other

TABLE 7. Results of the interior-branch test based on selected four-cluster trees with the topology ((A,B),(C,D)). Ch, = *Chagasia*, Bi = *Bironella*, Ano = subgenus *Anopheles*, Cel = *Cellia*, Ker = *Kerteszia*, Nys = *Nyssorhynchus*, Sth = *Stethomyia*. Although addition more data increases branch lengths, some of the branches connecting ingroup taxa are still not significantly different from zero, consistent with a "star phylogeny."

Data set	A	B	C	D	CP[a]
white[b]	Ch	Bi	Ker	Nys	0.949
	Ch	Bi	Lph	Ker + Nys	0.941
	Bi	Sth	Ker	Nys	0.668
all 21[c]	Ch	Sth	Ano + Cel	Ker + Nys	0.907
	Ano	Cel	Sth	Ker + Nys	0.774
	Bir	Ano + Cel	Ker	Nys	0.957*[d]
	Bi	Sth	Ker	Nys	0.965*

[a]Complement of the probability $(1 - \alpha)$ that interior branch in the tree is significantly different from zero.
[b]*white* gene with gaps and all missing sites removed.
[c]The 21-species data set of all four genes combined.
[d]Asterisk denotes significant values.

aligned *D2* sequences. For a discussion of the *D2* alignment, see Krzywinski et al. (2001).

Despite topological incongruence over the position of *Bironella* and subgenera *Lophopodomyia* and *Stethomyia* inferred in the present study and in Sallum et al. (2000), other relationships are fully congruent: monophyly of Anophelinae; basal position of *Chagasia;* monophyly of the *Anopheles* subgenera *Cellia, Kerteszia,* and *Nyssorhynchus;* sister taxon relationship between *Nyssorhynchus* and *Kerteszia;* and monophyly of the Arribalzagia Series within the subgenus *Anopheles.* Present results of the combined analysis are congruent with the hypothesis presented in our previous study of *ND5* and *D2* genes (Krzywinski et al., 2001) and some of these relationships are also suggested in other studies. Monophyly of Anophelinae and the ancient origin of *Chagasia* are congruent with the traditional notion (Ross, 1951) and previous phylogenetic analysis of morphological characters (Harbach and Kitching, 1998). The close relationship of *Anopheles* subgenera *Kerteszia* and *Nyssorhynchus* was suggested by Root (1922) and Zavortink (1973). Edwards (1932), following Christophers (1924), treated *Kerteszia* as a species group within *Nyssorhynchus,* but Peyton et al. (1992) demonstrated the distinctness of both taxa. Finally, monophyly of the Arribalzagia Series was hypothesized by Wilkerson and Peyton (1990).

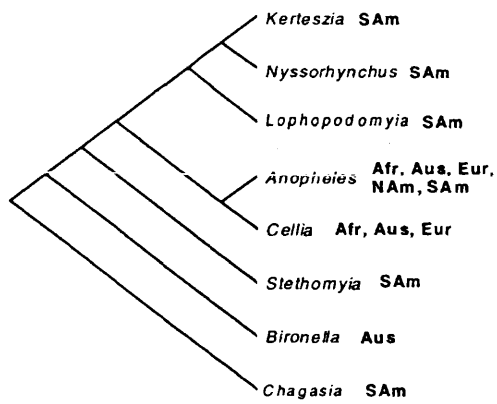
The clade of the subgenera *Anopheles* + *Cellia* probably reflects a historical relationship. Its monophyly, relatively well supported by ML analysis of the 21-species data set (78%) and *D2* + *ND5* genes (70%; Krzywinski et al., 2001), was also recovered by the weighted MP analyses of the *white* gene. Moreover, very high parametric bootstrap support (100%) from the simulated combined data indicates that this grouping is unlikely to be the result of cumulative random error. Foley et al. (1998) argued that the subgenera *Anopheles* and *Cellia* are paraphyletic with regard to each other. In contrast, the present analysis suggests that both are monophyletic. Different taxa used in both studies prevent us from testing the hypothesis of Foley et al. (1998), but very poor support of lineages at these levels suggests that the phylogenetic inference in their *COII* study was strongly influenced by homoplasy. Nevertheless, monophyly of some subgenera, as delineated now, conceivably may not stand when denser taxon sampling is performed.

Biogeography

Strongly supported relationships derived in the present analysis, congruent with other studies, provide an opportunity for interpreting the results in a biogeographic framework with relative confidence. Because Anophelinae and mosquitoes in general have received little phylogenetic attention (Munstermann and Conn, 1997), almost nothing is known about their origins and historical biogeography. Belkin (1962) speculated that "the initial differentiation of the subfamily took place in the American Mediterranean Region". Harbach and Kitching (1998) followed Belkin (1962) and pointed at the New World as a possible center of origin of the subfamily, because Neotropical *Chagasia* took a basal

position within Anophelinae in their analysis. Our results also support Belkin's speculation and additionally strengthen it, based on the reasoning of Bremer (1992), who developed a procedure for estimating ancestral areas of individual groups from topological information in their area cladograms. His method is based on the assumption that areas positionally plesiomorphic (*basal*) and *frequent* in the area cladogram are more likely to be parts of the ancestral range than are positionally apomorphic (placed at the top of the cladogram) and rare areas. According to Bremer's (1992) rationale, basal placement of *Chagasia* relative to other anophelines and the data on the Neotropical distribution of this genus and of four out of six subgenera of *Anopheles* suggest that South America was the center of origin of the subfamily (Fig. 6).

The inference of a South American origin of Anophelinae and the monophyly of subgenera *Anopheles* + *Cellia* and their derived position have important biogeographic implications. The evidence from plate tectonics as well as fossils and the distribution of other organisms (Cox and Moore, 1993; Pitman et al., 1993) allows us to propose the following hypothesis of Anophelinae history. Cosmopolitan distribution of the subgenus *Anopheles* should not be regarded as an ancestral state but rather as a result of relatively recent dispersal. This proposal predicts that the lineage of *Anopheles* + *Cellia*, in which most extant clades are found in the Old World, originated before the breakup of the western Gondwanaland in the Late Cretaceous. The first branching events within the subgenus *Anopheles* would have taken place before the loss of the land connection between Africa and South America, ~95 Mya. Breakup of the continents, leading to effective isolation of faunas, may have resulted in segregating the Neotropical Pseudopunctipennis Group, the basal lineage of the subgenus *Anopheles*, from the other stem lineages in Africa. Creation of the land bridges between Africa and Europe in the Paleocene and the connection from Europe to North America via Greenland, which existed until the end of the Eocene, allowed further dispersal of the subgenus to Eurasia and the Nearctic. The corridor for migration between North and South America, consisting of Aves Ridge and the Greater Antilles, which probably existed until the Late Eocene, ~49 Mya, could provide a route for certain lineages of the subgenus *Anopheles* (Cycloleppteron and Arribalzagia Series) to reenter South America from the north. Alternatively, dispersal from the north could have taken place in the mid-Miocene (~15 Mya) via the Panama island arc. Eastward movement of the Caribbean plate in the Late Eocene and sea level fluctuations might cause some migrating species to reach the islands only to be cut off and become essentially isolated (for example, *An. grabhami* of the Cycloleppteron Series). This scenario is fully congruent with the branching order in the trees inferred in the present study and largely consistent with the hypothesis of Sallum et al. (2000) regarding the basal position of the Pseudopunctipennis Group, the intermediate position of Old World and Nearctic taxa, and the position of the Cycloleppteron and Arribalzagia Series among the most derived lineages. The absence of *Cellia* in the New World suggests that the radiation of this subgenus might not have been triggered until the



	G	L	G/L	AA
South America (SAm)	4	2	2	1
Australia (Aus)	2	3	0.67	0.335
Africa (Afr)	1	4	0.25	0.125
Eurasia (Eur)	1	4	0.25	0.125
North America (NAm)	1	4	0.25	0.125

FIGURE 6. Estimation of ancestral area for Anophelinae based on the method of Bremer (1992). G = number of necessary gains under forward Camin-Sokal parsimony; L = number of necessary losses under reverse Camin-Sokal parsimony; $AA = G/L$ quotients rescaled to a maximum value of 1 by dividing with the largest G/L value.

Late Eocene, when the connection between Europe and North America was lost. The assumed basal position of *Bironella* relative to *Anopheles* clades implies that the ancestors of this lineage migrated to the landmass of Australia well before Australia and Antarctica separated from South America. The timing of separation is uncertain, but they seem to have parted some time in the Early Cenozoic. According to this sequence of events, Australasian *Anopheles* fauna have an Eurasian origin. This accords with Belkin's view (1962) but contrasts with the opinion of Foley et al. (1998), who hinted at a two-way exchange between Australasian and Oriental anopheline fauna rather than immigration.

Our hypothesis suggests that the molecular clock, proposed by Foley et al. (1998), predicting divergence of the lineages leading to *D. yakuba* and *Aedes/Culex* at 106–46 Mya, is seriously underestimated. This is not unexpected for a clock based on a highly saturated *COII* gene. The present data support the opinion of Hennig (1981) that the superfamily Culicoidea might have existed in the Upper Triassic (215 Mya). It is also consistent with Edwards (1932), who suggested an ancient origin of mosquitoes and their existence by the Jurassic. A very limited and relatively young fossil mosquito record contributes little to our understanding of the early evolution of the group. The earliest known fossils, from the Late Eocene, indicate that the main lineages were already well differentiated by about 38 Mya (Poinar, 1992), which are concordant with the notion of a long history of mosquitoes.

The hypotheses presented above are congruent with all the available phylogenetic and biogeographic evidence. Because they are based on analyses of relatively small samples of taxa, further studies of Anophelinae with extended sampling are needed to test them. Careful sampling of representatives of subgenera *Cellia* and *Anopheles* is probably the key to a better understanding of the biogeographic patterns within the genus *Anopheles*.

ACKNOWLEDGMENTS

We gratefully acknowledge the help of R. Cooper, A. Cornell, and D. Roberts in obtaining valuable mosquito specimens from their collections. J. Huelsenbeck provided helpful comments on long-branch attraction tests. Keith Crandall, Richard Olmstead, and two anonymous reviewers provided numerous constructive suggestions that substantially improved our manuscript. N.J.B. has received support from the UNDP/World Bank/WHO Special Programme and Training in Tropical Diseases, from the John D. and Catherine T. MacArthur Foundation, from the National Institutes of Health, and from the University of Notre Dame.

REFERENCES

BELKIN, J. N. 1962. The mosquitoes of the South Pacific (Diptera: Culicidae), volume 1. Univ. of California Press, Berkeley.

BESANSKY, N. J., J. A. BEDELL, M. Q. BENEDICT, O. MUKABAYIRE, D. HILFIKER, AND F. H. COLLINS. 1995. Cloning and characterization of the *white* gene from *Anopheles gambiae*. Insect Mol. Biol. 4:217–231.

BESANSKY, N. J., AND G. T. FAHEY. 1997. Utility of the *white* gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae). Mol. Biol. Evol. 14:442–454.

BREMER, K. 1992. Ancestral areas: A cladistic reinterpretation of the center of origin concept. Syst. Biol. 41:436–445.

BROWER, A. V. Z., AND R. DESALLE. 1994. Practical and theoretical considerations for choice of a DNA sequence region in insect molecular systematics, with a short review of published studies using nuclear gene regions. Ann. Entomol. Soc. Am. 87:702–716.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDEL. 1993. Partitioning and combining characters in phylogenetic analysis. Syst. Biol. 43:278–287.

CATERINO, M. S., S. CHO, AND F. A. H. SPERLING. 2000. The current state of insect molecular systematics: A thriving Tower of Babel. Annu. Rev. Entomol. 45:1–54.

CHRISTOPHERS, S. R. 1924. Provisional list and reference catalogue of the Anophelini. Indian Med. Res. Mem. 3:1–105.

COLLINS, F. H., M. A. MENDEZ, M. O. RASMUSSEN, P. C. MEHAFFEY, N. J. BESANSKY, AND V. FINNERTY. 1987. A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. Am. J. Trop. Med. Hyg. 37:37–41.

COX, C. B., AND P. D. MOORE. 1993. Biogeography: An ecological and evolutionary approach, 5th edition. Blackwell, Oxford.

CUNNINGHAM, C. W. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Syst. Biol. 46: 733–740.

DE QUEIROZ, A. 1993. For consensus (sometimes). Syst. Biol. 42:368–372.

DESALLE, R., AND A. V. Z. BROWER. 1997. Process partitions, congruence, and the independence of characters: Inferring relationships among closely related Hawaiian *Drosophila* from multiple gene regions. Syst. Biol. 46:751–764.

EDWARDS, F. W. 1932. Diptera fam. Culicidae. Fasc. 194 *in* Genera Insectorum (P. Wytsman, ed.). Desmet-Verteneuil, Brussels, 258 pp.

EWART, G. D., D. CANNELL, G. B. COX, AND A. J. HOWELLS. 1994. Mutational analysis of the traffic ATPase (ABC) transporters involved in uptake of eye pigment precursors in *Drosophila melanogaster*. J. Biol. Chem. 269:10370–10377.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1995. Testing significance of incongruence. Cladistics 10:315–319.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

FOLEY, D. H., J. H. BRYAN, D. YEATES, AND A. SAUL. 1998. Evolution and systematics of *Anopheles*: Insights from a molecular phylogeny of Australasian mosquitoes. Mol. Phylogenet. Evol. 9:262–275.

FOUTS, D., R. GANGULY, A. G. GUTIERREZ, J. C. LUCCHESI, AND J. E. MANNING. 1988. Nucleotide sequence of the *Drosophila* glucose-6-phosphate dehydrogenase gene and comparison with the homologous human gene. Gene 63:261–275.

GENETICS COMPUTER GROUP. 1997. Program manual for the GCG package. Version 9.1. Madison, Wisconsin.

HARBACH, R. E., AND I. J. KITCHING. 1998. Phylogeny and classification of the Culicidae (Diptera). Syst. Entomol. 23:327–370.

HENNIG, W. 1981. Insect phylogeny. Wiley, Chichester, U.K.

HIGGINS, C. F. 1992. ABC transporters: From microorganisms to man. Annu. Rev. Cell Biol. 8:67–113.

HUELSENBECK, J. P. 1991. When are fossils better than extant taxa in phylogenetic analysis? Syst. Zool. 40:458–469.

HUELSENBECK, J. P. 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46:69–74.

HUELSENBECK, J. P., AND J. J. BULL. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. Syst. Biol. 45:92–98.

HUELSENBECK, J. P., AND B. RANNALA. 1997. Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. Science 276:227–232.

KIMURA, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. USA 78:454–458.

KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.

KLUGE, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38:7–25.

KRZYWINSKI J., R. C. WILKERSON, AND N. J. BESANSKY. 2001. Evolution of mitochondrial and ribosomal gene sequences in Anophelinae (Diptera: Culicidae): Iimplications for phylogeny reconstruction. Mol. Phylogenet. Evol. 18:479–487.

KWIATOWSKI, J., D. SKARECKY, S. HERNANDEZ, D. PHAM, F. QUIJAS, AND F. J. AYALA. 1991. High fidelity of the polymerase chain reaction. Mol. Biol. Evol. 8:884–887.

LIÒ, P., AND N. GOLDMAN. 1999. Using protein structural information in evolutionary inference: Transmembrane proteins. Mol. Biol. Evol. 16:1696–1710.

LOCKHART, P. J., M. A. STEEL, M. D. HENDY, AND D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11:605–612.

MARDULYN, P., AND J. B. WHITFIELD. 1999. Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera; Braconidae): Evidence of a high diversification rate in this group of parasitoids. Mol. Phylogenet. Evol. 12:282–294.

MICKEVICH, M. F. 1978. Taxonomic congruence. Syst. Zool. 27:143–158.

MIYAMOTO, M. M., AND W. F. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44:64–76.

MORIYAMA, E. N., AND J. R. POWELL. 1997. Codon usage bias and tRNA abundance in *Drosophila*. J. Mol. Evol. 45:514–523.

MUNSTERMANN, L. E., AND J. E. CONN. 1997. Systematics of mosquito disease vectors (Diptera, Culicidae): Impact of molecular biology and cladistic analysis. Annu. Rev. Entomol. 42:351–369.

NAYLOR, C. E., P. ROWLAND, A. K. BASAK, S. GOVER, P. J. MASON, J. M. BAUTISTA, T. J. VULLIAMY, L. LUZZATTO, AND M. J. ADAMS. 1996. Glucose 6-phosphate dehydrogenase mutations causing enzyme deficiency in a model of the tertiary structure of the human enzyme. Blood 87:2974–2982.

NAYLOR, G. J. P., T. M. COLLINS, AND W. M. BROWN. 1995. Hydrophobicity and phylogeny. Nature 373:565–566.

NOTARO, R., A. AFOLAYAN, AND L. LUZZATTO. 2000. Human mutations in glucose-6-phosphate dehydrogenase reflect evolutionary history. FASEB J. 14:485–494.

PENNINGTON, R. T. 1996. Molecular and morphological data provide phylogenetic resolution at different hierarchical levels in *Andira*. Syst. Biol. 45:496–515.

PEYTON, E. L., R. C. WILKERSON, AND R. E. HARBACH. 1992. Comparative analysis of the subgenera *Kerteszia* and *Nyssorhynchus* of *Anopheles* (Diptera: Culicidae). Mosq. Syst. 24:51–69.

PITMAN, W. C., S. CANDE, J. LABRECQUE, AND J. PINDELL. 1993. Fragmentation of Gondwana: The separation of Africa from South America. Pages 15–34 *in* Biological relationships between Africa and South America (P. Goldblatt, ed.). Yale Univ. Press, New Haven, Connecticut.

POINAR, G. O., JR. 1992. Life in amber. Stanford Univ. Press, Stanford, California.

POSADA, D., AND K. A. CRANDALL. 1998. MODEL TEST: Testing the model of DNA substitution. Bioinformatics 14:817–818.

REMSEN, J., AND R. DESALLE. 1998. Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. Mol. Phylogenet. Evol. 9:225–235.

RODRIGO, A. G., M. KELLY-BORGES, P. R. BERGQUIST, AND P. L. BERGQUIST. 1993. A randomization test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. N. Z. J. Bot. 31:257–268.

RODRIGUEZ, F., J. F. OLIVER, A. MARÍN, AND J. R. MEDINA. 1990. The general stochastic model of nucleotide substitutions. J. Theor. Biol. 142:485–501.

ROOT, F. M. 1922. The classification of American *Anopheles* mosquitoes. Am. J. Hyg. 2:321–322.

ROSS, H. H. 1951. Conflict with *Culex*. Mosq. News 11:128–132.

RZHETSKY, A., AND M. NEI. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. 9:945–967.

SALLUM, M. A. M., T. R. SCHULTZ, AND R. C. WILKERSON. 2000. Phylogeny of Anophelinae based on morphological characters. Ann. Entomol. Soc. Am. 93:745–775.

SOTO-ADAMES, F. N., H. M. ROBERTSON, AND S. H. BERLOCHER. 1994. Phylogenetic utility of partial DNA

sequences of *G6pdh* at different taxonomic levels of Hexapoda with emphasis on Diptera. Ann. Entomol. Soc. Am. 87:723–736.

SULLIVAN, J. 1996. Combining data with different distributions of among-site rate variation. Syst. Biol. 45:375–380.

SWOFFORD, D. L. 1999. PAUP*: Phylogenetic analysis using parsimony (and other methods). Version 4.0b2. Sinauer, Sunderland, Massachusetts.

TAKEZAKI, N., A. RZHETSKY, AND M. NEI. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. 12:823–833.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. Evolution 37:221–244.

WIENS, J. J., AND T. W. REEDER. 1995. Combining data sets with different numbers of taxa for phylogenetic analysis. Syst. Biol. 44:548–558.

WILKERSON, R. C., AND E. L. PEYTON. 1990. Standardized nomenclature for the costal wing spots of the genus *Anopheles* and other spotted-wing mosquitoes (Diptera: Culicidae). J. Med. Entomol. 27:207–224.

ZAVORTINK, T. J. 1973. Mosquito studies (Diptera, Culicidae) XXIX. A review of the subgenus *Kerteszia* of *Anopheles*. Contrib. Am. Entomol. Inst. (Ann Arbor) 9:1–54.

ZHARKIKH, A. 1994. Estimation of evolutionary distances between nucleotide sequences. J. Mol. Evol. 39:315–329.

ZWIEBEL, L. J., G. SACCONE, A. ZACHAROPOULOU, N. J. BESANSKY, G. FAVIA, F. H. COLLINS, C. LOUIS, AND F. C. KAFATOS. 1995. The *white* gene of *Ceratitis capitata*: A phenotypic marker for germline transformation. Science 270:2005–2008.

Received 6 October 2000; accepted 5 November 2000
Associate Editor: K. Crandall