Copyright © 2002 by the Genetics Society of America

Genetic Differentiation in the African Malaria Vector, *Anopheles gambiae* s.s., and the Problem of Taxonomic Status

Gabriele Gentile,*,† Alessandra della Torre,‡ Bertha Maegga,§ Jeffrey R. Powell,* and Adalgisa Caccone*,,[1]**

**Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520-8106, †Dipartimento di Biologia, Università di Roma "Tor Vergata," 00133 Rome, Italy, ‡Istituto di Parassitologia Università di Roma "La Sapienza," 00185 Rome, Italy, §National Institute for Medical Research, Tukuyu Research Station, Tukuyu, Tanzania and **Yale Institute for Biospherics Studies, Yale University, New Haven, Connecticut 06520*

Manuscript received March 18, 2002
Accepted for publication May 22, 2002

ABSTRACT

Of the seven recognized species of the *Anopheles gambiae* complex, *A. gambiae* s.s. is the most widespread and most important vector of malaria. It is becoming clear that, in parts of West Africa, this nominal species is not a single panmictic unit. We found that the internal transcribed spacer (ITS) of the X-linked rDNA has two distinct sequences with three fixed nucleotide differences; we detected no heterozygotes at these three sites, even in areas of sympatry of the two ITS types. The intergenic spacer (IGS) of this region also displays two distinct sequences that are in almost complete linkage disequilibrium with the distinct ITS alleles. We have designated these two types as S/type I and M/type II. These rDNA types correspond at least partly to the previously recognized chromosomal forms. Here we expand the geographic range of sampling to 251 individuals from 38 populations. Outside of West Africa, a single rDNA type, S/type I, corresponds to the Savanna chromosomal form. In West Africa, both types are often found in a single local sample. To understand if these findings might be due to unusual behavior of the rDNA region, we sequenced the same region for 46 *A. arabiensis*, a sympatric sibling species. No such distinct discontinuity was observed for this species. Autosomal inversions in one chromosome arm (2R), an insecticide resistance gene on 2L, and this single X-linked region indicate at least two genetically differentiated subpopulations of *A. gambiae*. Yet, rather extensive studies of other regions of the genome have failed to reveal genetic discontinuity. Evidently, incomplete genetic isolation exists within this single nominal species.

SINCE the introduction of explicitly genetic concepts in defining species (Dobzhansky 1935), genetic analysis of natural populations has been important in distinguishing species that otherwise would not be recognized on the basis of classical features such as morphology. The demonstration of genetic distinctness, or genetic differentiation, between samples from natural populations, especially in areas of sympatry, is often used as a criterion to erect new species. At least for sexually reproducing organisms, genetic differentiation between groups can come about only if the two groups are not interbreeding. This led to the so-called biological species concept (BSC) that defines species as entities that are incapable of exchanging genes via interbreeding. As the sophistication of methodologies to genetically analyze populations has increased dramatically, the ability to detect more subtle genetic distinctness among natural units has increased. This has sometimes led to unexpected findings. For example, it was initially assumed that "reproductive isolation" was a genome-wide phenomenon; *i.e.*, once two groups ceased to interbreed or ceased to exchange genes due to sterility of hybrids, all genes in the two groups would evolve independently. Contrary to this classical assumption, there is a growing number of examples of cases in which two species (or named taxa of some sort) have genomes that vary in the level of isolation; *i.e.*, parts of the genome can still be exchanged while other parts cannot. This has variously been termed semipermeable barriers to gene exchange (Harrison 1990), selective introgression (della Torre *et al.* 1998), or gene-by-gene isolation (Barton and Hewitt 1981; Wu 2001). Plant biologists have long been comfortable with the notion that "good species" can still exchange some genes via introgressive hybridization (Anderson 1949; Arnold 1997) as have biologists concerned with hybrid zones (Barton and Hewitt 1981; Harrison 1990). Recently, empirical evidence from Drosophila has indicated that in the historical process of speciation, isolation does not occur simultaneously for the whole genome (Machado *et al.* 2002). Noor *et al.* (2001) demonstrated that the least likely parts to exchange are associated with hybrid sterility or mate choice. The subject of this article, the mosquito *Anopheles gambiae*, has genetic substructur-

[1] *Corresponding author:* Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520-8106.

ing that is very likely the result of partial isolation; *i.e.*, parts of the genome display strong differentiation while most of the genome remains nearly completely undifferentiated, implying ongoing gene flow for these regions.

The *A. gambiae* complex is composed of at least seven morphologically indistinguishable species (reviewed in White 1974; Coluzzi *et al.* 1979). Members of this complex exist throughout sub-Saharan Africa, including neighboring islands. Genetic distinctness was first detected during crosses intended to reveal the genetic basis of insecticide resistance (Davidson 1964; Davidson and Hunt 1973); F_1 male sterility was sometimes observed between crosses of what were considered the same species. Eventually six species were described on the basis of reproductive compatibility; the seventh was only recently detected (Hunt *et al.* 1998). Chromosomal inversions were then shown to also distinguish the species (Coluzzi and Sabatini 1967, 1968, 1969). Now, DNA-based diagnostic procedures are available (Collins *et al.* 1987). Thus, there is little doubt that the seven recognized species of this complex meet the criteria of species recognition defined by the BSC.

One of the seven species is *A. gambiae sensu stricto*. This is a widespread species, closely associated with human habitats, and a very efficient vector of malaria. Evidence has been accumulating that this species is composed of subunits that are genetically distinct. The first evidence for this came from extensive studies of frequencies of naturally occurring chromosomal inversions by Coluzzi and colleagues (Coluzzi *et al.* 1985; Touré *et al.* 1998). In some West African populations the karyotype frequencies are often far from Hardy-Weinberg expectations with a notable excess of homokaryotypes for inversions in the right arm of the second chromosome, 2R. When samples are subdivided into certain sets of inversions, the subsets of karyotypes are in Hardy-Weinberg proportions. On the basis of these observations, chromosomal "forms" have been proposed (Coluzzi *et al.* 1985) consisting almost exclusively of different sets of 2R gene arrangements. The designation forms (*sensu* Mayr *et al.* 1953) is intentionally ambiguous and is used to denote units of uncertain taxonomic status. The names Bamako, Mopti, Savanna, Forest, and Bissau have been given to these chromosomal forms. Unlike the presently recognized species of the complex, these chromosomal forms appear to display no postmating reproductive isolation, with hybrids between forms actually having higher fecundity under laboratory conditions (Di Deco *et al.* 1980). This implies that the lack of conformity to Hardy-Weinberg expectations is due to nonrandom mating. However, it needs to be emphasized that the fitness of hybrids in nature has not been studied and, as noted in the discussion, there is indirect evidence of lower viability of hybrids in natural populations.

Attempts have been made to determine the magnitude and consistency of the genetic differentiation among the chromosomal forms of *A. gambiae* s.s., especially for genetic markers not associated with chromosome 2R by which they were originally defined. The reasoning is that if the substructuring detected for the 2R chromosome is due to complete reproductive isolation, then genetic differentiation should extend to the whole genome. Favia *et al.* (1994) found distinctive RAPD (randomly amplified polymorphic DNA) markers among forms, but this was not confirmed by Mukabayire *et al.* (2001). Favia *et al.* (1997) found that Mopti is consistently different from Savanna and Bamako at nucleotide sites in the intergenic spacer of rDNA (IGS), an X-linked region (Collins *et al.* 1987; see Figure 1). A diagnostic procedure was developed on the basis of PCR amplification followed by a restriction enzyme digest that reliably distinguished Mopti from Savanna and Bamako in samples from Mali and Burkina Faso (Favia *et al.* 1997). More recently Favia *et al.* (2001) developed a primer-specific PCR approach for the IGS that allows the same distinction in samples from Mali. The use of this PCR-restriction fragment length polymorphism analysis has been extended to a number of populations in West Africa and to a limited extent elsewhere in Africa; virtually no mixed molecular patterns (indicative of hybrids) have been detected (della Torre *et al.* 2001); the terms M and S were proposed to indicate the two IGS types. Outside of Mali and Burkina Faso, M and S do not always correspond to the previously described chromosomal forms (della Torre *et al.* 2001; Gentile *et al.* 2001). A study of microsatellites throughout the genome failed to find significant and consistent differences among the chromosomal forms, except for loci on 2R (Lanzaro *et al.* 1998). Later studies using microsatellites detected reduced gene exchange between forms compared to within forms, although not as great a reduction as found between the recognized species *A. gambiae* and *A. arabiensis* (Taylor *et al.* 2001). In a similar microsatellite study, Wang *et al.* (2001) confirmed little differentiation between M and S molecular forms except for a microsatellite tightly linked to, but distinct from, the rDNA region of the X. Mukabayire *et al.* (2001) failed to find diagnostic differences at two nuclear loci for samples from Mali, but did find a fixed difference between Mopti and Savanna/Bamako in the internal transcribed spacer (ITS) of the rDNA (Figure 1); Gentile *et al.* (2001) confirmed and extended these ITS results to populations outside Mali. To distinguish the chromosomal forms from the distinct rDNA sequences, we have used the term "types" in referring to the rDNA differentiation (Gentile *et al.* 2001).

There is evidence for further genetic differentiation of M and S rDNA types for genes not on the X chromosome. This is for an insecticide resistance gene, *knockdown resistance* (*Kdr*; Chandre *et al.* 1999; Weill *et al.* 2000; della Torre *et al.* 2001), located on the left arm of the second chromosome, 2L, location 20C (Ranson *et al.* 2000). In the case of *Kdr*, the S type has a much

higher frequency of the resistance allele compared to M types captured at the same time. So this is not a case of fixed differences; rather, there are extreme frequency differences between sympatric S and M.

One problem has been that most of the data pertaining to these chromosomal forms and molecular types have come from a restricted region of West Africa, although some data exist for samples taken outside West Africa (della Torre *et al.* 2001; Gentile *et al.* 2001). It is not clear how widespread the correspondence is between the chromosomally defined forms and molecular types. Especially for the ITS and DNA sequence data of the IGS, the data presented here expand the geographic range over which the rDNA types of *A. gambiae* have been determined. Second, we also attempted to address the question of whether rDNA differentiation of the sort we have documented between forms of *A. gambiae* could occur in the absence of any reproductive isolation, perhaps due to some unusual evolutionary dynamics (*e.g.*, unusually rapid concerted evolution). In the closely related *A. arabiensis* there is no indication of any cryptic reproductive isolation. Because *A. arabiensis* is so closely related to *A. gambiae* and they are largely sympatric (*i.e.*, both co-occur in human disturbed habitats), it is likely that the molecular evolutionary dynamics of the rDNA region would be similar. To compare to *A. gambiae,* we sequenced the same rDNA region in several specimens of *A. arabiensis* collected at the same localities.

MATERIALS AND METHODS

A total of 120 individual mosquitoes were sampled in this study. By pooling data from this study with those from Gentile *et al.* (2001) we obtained a data set of 251 specimens. Samples are from 38 and 11 populations of *A. gambiae* and *A. arabiensis*, respectively. A list of localities is in Table 1 with their locations on the map in Figure 2. Table 1 also reports the number of individuals studied for each population for the ITS and IGS regions together with information on the chromosomal form and molecular type of each specimen. Mosquitoes and DNA preparations were kindly provided by Mario Coluzzi's laboratory in Rome and Tovi Lehmann at the Centers for Disease Control and Prevention (CDC) in Atlanta. DNA extractions provided by the Coluzzi laboratory were from Carnoy preserved specimens as described in Favia *et al.* (1994). DNA extractions from the CDC laboratory were performed as described in Collins *et al.* (1987). DNA extractions performed in our laboratory were done using the protocol for hair samples in the Easy-DNA kit (Invitrogen, Carlsbad, CA).

Primers and PCR conditions to amplify an 890-bp fragment of rDNA, which includes the internal transcribed spacers (ITS1 and ITS2 illustrated in Figure 1) and the intervening 5.8S rDNA, are as in Gentile *et al.* (2001). A 1500-bp fragment of the rDNA IGS was amplified using the following two primers: RT9FOR (5′-ATTCCTGTGCAGTACGAGAGG-3′) and RT8REV (5′-GTGTTTGCAGACTTTYGCCAAGC-3′). They correspond to positions 3–23 (RT9FOR) and 1518–1540 (RT8REV) of the published *A. gambiae* IGS (GenBank accession no. U10135). Double-stranded amplifications were performed in 50 μl of a solution containing 10 mm Tris-HCl (pH 8.8), 50 mm KCl, 1.5 mm $MgCl_2$, each dNTP at 2.5 mm, 200 ng for each primer, 10–100 ng of genomic DNA, and 1 unit of Amplitaq (Perkin Elmer-Cetus, Norwalk, CT). After a 2-min denaturation step (94°), each cycle of the polymerase chain reaction consisted of denaturation for 1 min at 94°, annealing for 1.5 min at 50°, and extension for 2 min at 72°. This cycle was repeated 35 times and followed by a 5-min incubation step at 72°. PCR fragments were gel purified by glass milk extraction (GeneClean III kit; BIO 101, Vista, CA). Sequences were determined with an automated sequencer (377; Applied Biosystems, Foster City, CA) following the manufacturer's protocols. Both strands were sequenced in both directions for each individual. Sequences and locations of the internal primers used for sequencing are available from the authors. GenBank accession numbers for all the new DNA fragments studied are AF470117–AF470230 (ITS) and AF470093–AF470116 (IGS).

Sequences were analyzed using the program Sequencer 3.1.1 (Gene Codes Corporation, Ann Arbor, MI). Alignments were carried out by eye and by using CLUSTAL W (Thompson *et al.* 1994); they are available from the authors upon request. Values of π (observed number of nucleotide differences per site) and θ ($4N\mu$, predicted number of nucleotide differences per site at equilibrium for neutral mutations; Watterson 1975) were calculated using the MEA 2000 computer program (Molecular Evolutionary Analysis, year 2000 version; developed by Etsuko Moriyama, Center for Genetic Research, University of Nebraska).

To obtain a genealogical relationship among different ITS haplotypes we constructed a parsimony network, using the TCS software (Clement *et al.* 2000), which implements the method of statistical parsimony developed by Templeton *et al.* (1992). This method is particularly appropriate for population level analysis as it does not involve many of the assumptions of phylogenetic reconstruction methods. For example, it does not assume that the ancestral sequence is missing and does not require bifurcating relationships. The method uses parsimony (as defined in Templeton *et al.* 1992) to construct pairwise distances (number of mutational steps) between all haplotypes until the probability exceeds 95%. The matrix just above this cutoff point represents the maximum number of mutational steps justified by the 95% parsimony criteria. The TCS program then connects the haplotypes based on the criteria into a network with the number of mutational steps indicated on the lines connecting haplotypes. On the basis of coalescent theory, this program also identifies the most probable ancestral haplotype among the collection of samples (Donnelly and Tavaré 1986; Castelloe and Templeton 1994).

RESULTS

Table 2 shows the variable sites and the different haplotypes found in the 890-bp-long ITS/5.8S fragment in *A. arabiensis* and *A. gambiae.* Of all the changes found at the variable sites, 22 (47.8%) were transitions, 17 (37%) were transversions, and 7 (15.2%) were insertions/deletions (indels). We found species-specific differences at 28 sites. Of these, 19 were found in the ITS1, 2 in the 5.8S region, and 8 in the ITS2. The number of variable sites and the overall number of heterozygotes observed were 14 and 16 in *A. gambiae* and 2 and 7 in *A. arabiensis*, respectively. *A. gambiae* showed a number of singletons higher than the number observed in *A. arabiensis* (four and one, respectively). Percentages of sites that are polymorphic are 1.69% in *A. gambiae*

TABLE 1

Sample localities

Localities	Chromosomal form	IGS	Sample size (ITS)	Sample size (IGS)	ITS
		A. gambiae			
Senegal					
Ndialahar (1)	Savanna	M	5	—	gII
The Gambia					
Kaur (3)	Savanna	M	4	—	gII
Bassè (4)	Savanna	M	4	—	gII
	Savanna	S	1	—	gI
Ballingho (5)	Sav/Bissau	M	6	—	gII
Bambali (6)	Savanna	M	1	—	gII
Kani Kunda (7)	Bissau	M	1	—	gII
Guinea Conakry					
Sombili (8)	Savanna	S	4	—	gI
Timbi Madina (9)	Savanna	S	4	—	gI
Mali					
Banambani (10)	Mopti	M	5	—	gII
	Sav/Mopti	M	1	—	gII
	Savanna	S	4	—	gI
	Bamako	S	2	—	gI
	Sav/Mopti	S	2	—	gI
Moribabougou (11)	Mopti	M	14	—	gII
	Savanna	S	2	—	gI
	Bamako	S	3	—	gI
Pimperena (12)	Savanna	S	5	—	gI
Ivory Coast					
Danta (14)	Forest	M	6	—	gII
Ziglo (15)	Forest	S	3	—	gI
Yaokiffikro (16)	Savanna	S	4	—	gI
Toliakowandrokro (17)	Savanna	S	5	—	gI
M-bè (18)	Mopti	M	1	—	gII
	Savanna	M	2	—	gII
	Savanna	S	2	—	gI
Burkina Faso					
Goudry (21)	Mopti	M	5	—	gII
	Savanna	S	1	—	gI
Benin					
Bohicon (22)	NA	M	2	1	gII
	NA	S	5	1	gI
Lemà (23)	Savanna	S	5	1	gI
Bamè (24)	Savanna	M	6	2	gII
Agbabilame (25)	For/Sav	M	6	—	gII
Ghana					
Kassena-Nankana district (26)	NA	M	8	—	gII
Cameroon					
Deschang (27)	NA	S	2	—	gI
Nkoteng (28)	NA	S	1	—	gI
Obala (29)	NA	S	1	—	gI
Simbock (30)	NA	M	2	—	gII
São Tomé					
São Tomé (31)	Forest	M	11	—	gIII

(*continued*)

and 0.23% in *A. arabiensis* although it is important to note this parameter is dependent on sample size, which is smaller for the latter species. These levels of intraspecific variation were higher than those estimated by Paskewitz *et al.* (1993). Of 499 ITS/5.8S haplotypes scored, we found 15 different haplotypes in *A. gambiae* and 3 in *A. arabiensis*. It needs to be emphasized that in calculating the numbers of haplotypes analyzed we made some assumptions. Because the rDNA region is X linked, we assumed we were sampling 2 haplotypes

TABLE 1

(Continued)

Localities	Chromosomal form	IGS	Sample size (ITS)	Sample size (IGS)	ITS
Malawi					
Mangochi (32)	NA	S	9	—	gI
Tanzania					
Nyakariro (33)	Savanna	S	8	—	gI
Kyela district (34)	NA	S	9	—	gI
Kenya					
Asembo (35)	NA	S	10	—	gI
Kisian (36)	NA	S	9	—	gI
Malindi (37)	NA	S	9	—	gI
Madagascar					
Beforona (38)	Savanna	S	5	—	gI
		A. arabiensis			
Senegal					
Guia (2)	—	—	7	—	aI
The Gambia					
Bassè (4)	—	—	4	—	aI
Ballingho (5)	—	—	2	—	aI
Mali					
Moribabougou (11)	—	—		3	—
N'Gabacoro Droit (13)	—	—	6	—	aI
Burkina Faso					
Goden (19)	—	—	7	—	aI
Zaghtouli (20)	—	—		3	—
Benin					
Bohicon (22)	—	—	11	—	aI
Madagascar					
Ankazomorona (39)	—	—	1	4	aI
Morafeno (40)	—	—	3	3	aI
Tsarara (41)	—	—	5	3	aI

NA, not available.

for each female and 1 for each male. When we detected heterozygosity at a site for females, we assumed we sampled 2 different haplotypes. This may not actually be the case because of the repetitive nature of rDNA. The "heterozygosity" may be due to different copies of the repeat unit on the same chromosome being different. However, because of ample evidence that rDNA copies on the same chromosome undergo concerted evolution, scoring frequencies in this manner is not unreasonable. The fact that we never observed a heterozygote for the males in our sample also supports these assumptions. In our previous study (Gentile *et al.* 2001), we sometimes cloned these fragments and sequenced three to five independent copies; we never detected any differences between clones from males or females scored as homozygotes.

Table 3 shows the geographic distribution of the different haplotypes in the populations of *A. gambiae* and *A. arabiensis*. For all but one sample, complete linkage disequilibrium was observed at the three positions 59, 100, and 834 in the alignment used in this article. [These sites correspond to positions 88, 127, and 859 in Gentile *et al.* (2001). We have changed the numbering of sites due to several small insertion/deletions in the alignment to the *A. arabiensis* sequence.] São Tomé remained the only locality where the linkage disequilibrium between the three sites was not complete. Our data confirm the existence of three major types of rDNA: I, II, and III as defined by Gentile *et al.* (2001). Types I and II correspond, respectively, to the molecular forms S and M defined by Favia *et al.* (1994). Type I is distributed over the whole of sub-Saharan Africa, while type II seems to be localized to West Africa. Type III has been found only on the isolated island of São Tomé. Type I showed the highest number of sites at which heterozygotes were observed (4) and the highest overall number of observed heterozygotes (13). Percentages of polymorphism were equal to 0.8, 0.5, and 0.1% in types I, II, and III, respectively. *A. arabiensis* shared one fixed difference with type I (cytosine at site 59) and one fixed difference with type II (deletion at site 834). The geographic distribution of the three types is shown in Figure 2.

Table 4 reports the variable sites and the haplotypes

TABLE 2

Internal transcribed spacer (ITS) variable sites and haplotypes in *A. gambiae* and *A. arabiensis*

Haplotypes	No. of haplotypes	Variable sites																								
		12	16	18	21	22	26	39	40	**59**	64	65	92	93	**100**	101	103	109	110	114	125	139	141	153	208	254
		A. gambiae																								
gIA	177	A	T	G	:	:	T	C	G	C	:	:	A	C	A	C	C	G	T	T	T	C	G	A	T	C
gIB	16	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gIC	10	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gID	17	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gIE	1	—	—	—	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—
gIF	2	—	—	—	—	—	—	—	—	—	—	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—
gIG	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gIH	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gIK	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gIIA	147	—	—	—	—	—	—	—	—	A	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—
gIIB	8	—	—	—	—	—	—	—	—	A	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—
gIIC	1	—	—	—	—	—	—	A	—	A	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—
gIID	2	—	—	—	G	T	—	—	—	A	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—
gIIIA	14	—	—	—	—	—	—	—	—	A	—	—	—	—	T	—	—	—	—	—	—	—	—	—	G	—
gIIIB	8	—	—	—	—	—	—	—	—	A	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—
		A. arabiensis																								
aIA	81	G	A	A	—	—	—	—	A	—	T	G	G	—	G	G	A	T	C	C	C	A	—	G	—	T
aIB	2	G	A	A	—	—	C	—	A	—	T	G	G	—	G	G	A	T	C	C	C	A	—	G	—	T
aIC	9	G	A	A	—	—	—	—	A	—	T	G	G	—	G	G	A	T	C	C	C	A	—	G	—	T

(*continued*)

TABLE 2

(Continued)

Haplotypes	No. of haplotypes	Variable sites																
		290	302	320	347	348	562	588	599	605	631	798	833	**834**	847	881	883	887
A. gambiae																		
gIA	177	A	A	T	C	T	C	T	A	G	C	T	T	G	G	T	C	C
gIB	16	—	—	—	—	—	—	—	—	C	—	—	—	—	—	—	—	—
gIC	10	—	—	—	—	—	—	—	—	—	—	—	—	—	:	—	—	—
gID	17	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	—	—
gIE	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gIF	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gIG	2	G	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gIH	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	G	—
gIK	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	G	—
gIIA	147	—	—	—	—	—	—	—	—	—	—	—	—	:	—	—	—	—
gIIB	8	—	—	—	—	—	—	—	—	—	—	—	—	:	—	—	—	—
gIIC	1	—	—	—	—	—	—	—	—	—	—	—	—	:	—	—	—	—
gIID	2	—	—	—	—	—	—	—	—	—	—	—	—	:	—	—	—	—
gIIIA	14	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gIIIB	8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
A. arabiensis																		
aIA	81	—	G	C	A	C	T	C	G	—	—	G	A	:	—	—	—	T
aIB	2	—	G	C	A	C	T	C	G	—	—	G	A	:	—	—	—	T
aIC	9	—	G	C	A	C	T	C	G	—	A	G	A	:	—	—	—	T

Diagnostic sites for the types I, II, and III are in boldface.

TABLE 3

Distribution of the ITS haplotypes in *A. gambiae* and *A. arabiensis*

Localities	Haplotypes																	
	gIA	gIB	gIC	gID	gIE	gIF	gIG	gIH	gIK	gIIA	gIIB	gIIC	gIID	gIIIA	gIIIB	aIA	aIB	aIC
										A. gambiae								
Senegal																		
Ndialahar (1)	—	—	—	—	—	—	—	—	—	10	—	—	—	—	—	—	—	—
The Gambia																		
Kaur (3)	—	—	—	—	—	—	—	—	—	—	8	—	—	—	—	—	—	—
Bassè (4)	2	—	—	—	—	—	—	—	—	6	—	—	2	—	—	—	—	—
Ballingho (5)	—	—	—	—	—	—	—	—	—	11	—	1	—	—	—	—	—	—
Bambali (6)	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—
Kani Kunda (7)	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—
Guinea Conakry																		
Sombili (8)	8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Timbi Madina (9)	6	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Mali																		
Banambani (10)	16	—	—	—	—	—	—	—	—	12	—	—	—	—	—	—	—	—
Moribabougou (11)	10	—	—	—	—	—	—	—	—	28	—	—	—	—	—	—	—	—
Pimperena (12)	8	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—
Ivory Coast																		
Danta (14)	—	—	—	—	—	—	—	—	—	12	—	—	—	—	—	—	—	—
Ziglo (15)	5	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Yaokiffkro (16)	2	—	—	6	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Toliakowandrokro (17)	6	—	—	2	—	—	—	1	1	—	—	—	—	—	—	—	—	—
M-bè (18)	4	—	—	—	—	—	—	—	—	6	—	—	—	—	—	—	—	—
Burkina Faso																		
Goudry (21)	2	—	—	—	—	—	—	—	—	10	—	—	—	—	—	—	—	—
Benin																		
Bohicon (22)	10	—	—	—	—	—	—	—	—	4	—	—	—	—	—	—	—	—
Lemà (23)	7	—	—	1	—	2	—	—	—	—	—	—	—	—	—	—	—	—
Bamè (24)	—	—	—	—	—	—	—	—	—	12	—	—	—	—	—	—	—	—
Aghabilame (25)	—	—	—	—	—	—	—	—	—	12	—	—	—	—	—	—	—	—
Ghana																		
Kasseni-Nankana district (26)	—	—	—	—	—	—	—	—	—	16	—	—	—	—	—	—	—	—
Cameroon																		
Deschang (27)	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nkoteng (28)	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Obala (29)	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Simbock (30)	—	—	—	—	—	—	—	—	—	4	—	—	—	—	—	—	—	—
São Tomé																		
São Tomé (31)	—	—	—	—	—	—	—	—	—	—	—	—	—	14	8	—	—	—
Malawi																		
Mangochi (32)	18	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

(*continued*)

TABLE 3

(Continued)

Localities	Haplotypes																	
	gIA	gIB	gIC	gID	gIE	gIF	gIG	gIH	gIK	gIIA	gIIB	gIIC	gIID	gIIIA	gIIIB	aIA	aIB	aIC
Tanzania																		
Nyakariro (33)	—	16	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Kyela district (34)	14	—	—	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—
West Kenya																		
Asembo (35)	16	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Kisian (36)	17	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—
East Kenya																		
Malindi (37)	18	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Madagascar																		
Beforona (38)	—	—	10	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
										A. arabiensis								
Senegal																		
Guia (2)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14	—	—
The Gambia																		
Bassè (4)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8	—	—
Ballingho (5)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4	—	—
Mali																		
N'Gabacoro Droit (13)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	—	—
Burkina Faso																		
Goden (19)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	2	—
Benin																		
Bohicon (22)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	13	—	9
Madagascar																		
Ankazomorona (39)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—
Morafeno (40)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6	—	—
Tsarara (41)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10	—	—

The sample is composed of females, with the exception of three males from Asembo.

TABLE 4

Haplotypes in the intergenic spacer (IGS)

	Haplotypes																				
	A. gambiae							*A. arabiensis*													
Variable sites	M1	M2	M3	M4	S1	S2	S3	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13
82	G	—	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
213	C	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
311	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	:	:	:
408	T	G	—	—	—	—	—	—	—	—	—	—	—	G	G	G	G	—	—	—
418	:	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	T	T
433	A	—	—	—	—	—	—	T	T	—	—	—	—	—	—	—	—	—	—	—
470	:	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
476	T	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
488	A	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
498	T	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
518	T	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
536	A	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
541	G	—	A	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
542	T	—	—	—	C	C	C	—	—	—	—	—	—	—	—	—	—	—	—	—
546	G	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
573	A	—	—	—	G	G	G	—	—	—	—	—	—	—	—	—	—	—	—	—
604	T	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
630	T	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
646	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	G
651	A	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
653	C	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
655	C	—	—	—	—	—	—	T	T	—	—	—	—	—	—	—	—	—	—	—
699	C	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
675	A	—	—	—	—	—	—	—	—	T	—	—	—	T	T	T	T	—	—	—
703	C	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
704	G	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
706	T	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
713	C	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
714	G	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
716	A	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
721	G	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
730	G	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
751	G	—	—	—	—	—	—	C	C	—	—	—	—	—	—	—	—	—	—	—
756	C	G	G	G	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
779	A	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
781	A	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
787	C	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
793	A	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C

(*continued*)

TABLE 4

(Continued)

Variable sites	Haplotypes																			
	A. gambiae							*A. arabiensis*												
	M1	M2	M3	M4	S1	S2	S3	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13
811	C	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
812	A	—	—	C	C	C	C	—	—	:	:	:	:	:	:	:	:	:	:	:
813	T	—	—	—	—	—	—	—	—	:	:	:	:	:	:	:	:	:	:	:
837	G	—	—	—	A	A	A	—	—	—	—	—	—	—	—	—	—	—	—	—
839	T	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
844	T	—	—	—	—	—	—	—	G	—	—	—	—	—	—	—	—	—	—	—
848	G	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
859	T	—	—	—	—	—	—	:	:	—	—	—	—	—	—	—	—	—	—	—
875	G	—	—	—	—	—	—	:	:	:	:	:	:	:	:	:	:	:	:	:
876	T	—	—	—	—	—	—	:	:	:	:	:	:	:	:	:	:	:	:	:
877	A	—	—	—	T	T	T	:	:	:	:	:	:	:	:	:	:	:	:	:
893	:	—	—	—	—	—	—	C	C	—	—	—	—	—	—	—	—	—	—	—
894	:	—	—	—	—	—	—	A	A	—	—	—	—	—	—	—	—	—	—	—
898	C	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
901	C	T	T	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
916	T	—	—	—	C	C	C	—	—	—	—	—	—	—	—	—	—	—	—	—
971	A	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
1007	C	—	—	—	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—	—
1012	G	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
1023	G	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
1025	C	—	—	—	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—	—
1056	A	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
1066	G	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
1077	C	—	—	—	—	—	—	—	—	—	—	T	T	—	—	—	—	—	T	—
1089	A	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
1124	A	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
1126	A	—	—	—	—	—	—	G	G	—	G	G	G	—	—	—	G	G	G	G
1132	T	—	—	—	—	—	—	—	—	C	—	—	—	C	C	C	C	—	—	—
1167	C	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
1172	A	—	—	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—	—	—
1192	T	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
1204	G	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
1207	G	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
1213	T	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
1214	G	—	—	—	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
1217	G	—	—	—	—	—	—	A	A	—	—	—	—	—	—	—	—	—	—	—
1223	C	—	—	—	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
1228	G	—	—	—	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T

(*continued*)

TABLE 4

(Continued)

	Haplotypes																			
	A. gambiae							*A. arabiensis*												
Variable sites	M1	M2	M3	M4	S1	S2	S3	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13
1245	A	—	—	—	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
1258	C	—	—	?	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
1261	A	—	—	?	T	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1264	T	—	—	?	C	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1266	A	—	—	?	—	—	—	G	G	G	G	G	G	G	G	G	G	G	G	G
1275	C	—	—	?	—	—	—	—	—	—	T	T	T	—	—	—	T	T	T	T
1287	T	—	—	?	A	A	—	A	A	A	A	A	A	A	A	A	A	A	A	A
1289	T	—	—	?	—	—	—	—	—	—	—	—	—	A	A	—	—	—	—	—
1296	T	C	C	?	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1303	T	—	—	?	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
1313	G	—	—	?	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
1334	A	—	—	?	—	—	—	—	—	—	T	T	T	—	—	—	T	T	T	T
1339	A	—	—	?	—	—	—	C	C	C	C	C	C	C	C	C	C	C	C	C
1347	G	—	—	?	—	—	—	A	A	A	A	A	A	A	A	A	A	A	A	A
1352	T	G	—	?	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1361	C	—	—	?	G	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1376	G	—	—	?	—	—	—	T	T	T	T	T	T	T	T	T	T	T	T	T
1398	?	T	T	?	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
	Localities																			
	22	24	24	[a]	22[b]	23	[c]	11	11	20	39, 40	39, 40	39	22	22	22	22	41	41	41

No haplotype is shared by more than two samples. Diagnostic sites for M and S types in *A. gambiae* are in boldface. At Bohicon (22) M and S types and *A. arabiensis* occur in sympatry. Names of localities are given in Table 1.

[a] Favia *et al.* (1997), MaliM.

[b] Favia *et al.* (1997), MaliS.

[c] U10135 (G3 S) The Gambia.

TABLE 5

Summary of polymorphism data for the ITS and IGS DNA regions in *A. arabiensis* and *A. gambiae*

	$\pi \pm$ SE	$\theta \pm$ SE
	ITS	
A. arabiensis	0.0001 ± 0.0005	0.0005 ± 0.0004
A. gambiae	0.0013 ± 0.0005	0.0013 ± 0.0013
M form	0.0001 ± 0.0003	0.0002 ± 0.0002
S form	0.0003 ± 0.0007	0.0009 ± 0.0004
	IGS	
A. arabiensis	0.0033 ± 0.0020	0.0034 ± 0.0008
A. gambiae	0.0066 ± 0.0034	0.0058 ± 0.0013
M form	0.0037 ± 0.0028	0.0039 ± 0.0012
S form	0.0010 ± 0.0013	0.0016 ± 0.0008

For *A. gambiae* polymorphism data are also provided separately for the M and S molecular form. The observed average proportion of nucleotide differences between alleles sequenced (π) and the mean number of nucleotides segregating per site (θ) are given together with their standard errors (SE).

found in both *A. gambiae* and *A. arabiensis* in 1399 bp of the IGS region. Among the 94 variable sites we observed 45 (46.9%) transitions, 40 transversions (41.7%), and 11 (11.5%) indels. Species-specific differences were found at 53 positions. Twenty-one variable sites were found in *A. gambiae*, with only a single heterozygote scored. *A. arabiensis* has 24 variable sites (including five singletons) with three heterozygotes scored at 3 different sites. Percentages of polymorphism were 1.7% in *A. arabiensis* and 1.5% in *A. gambiae*. Linkage disequilibrium that characterized types I and II in the ITS extended also to the IGS. Favia *et al.* (2001) found 10 sites in the IGS in complete linkage disequilibrium. By increasing the sample size, especially including a few specimens from Benin, the number of sites from the IGS in complete disequilibrium was reduced to 7. We will refer to the composite rDNA types as S/type I and M/type II to encompass both the IGS and ITS variants.

Levels of genetic variation in the rDNA region of *A. gambiae* and *A. arabiensis* are summarized in Table 5. Values of π and θ were estimated for various taxonomic units. If one considers *A. gambiae* as a single taxonomic unit, it has considerably more variation than *A. arabiensis*. However, if *A. gambiae* is considered two units (S/type I and M/type II) the level of genic diversity is virtually the same as in *A. arabiensis*. A similar trend was observed for IGS, even though π and θ values were on the whole higher by one order of magnitude with respect to those estimated from ITS (Table 5). This is consistent with previous observations on these species (Paskewitz *et al.* 1993). The relative conservation of ITS sequences is thought to be due to constraints on secondary structure required for processing (Thweatt and Lee 1990; Yeh and Lee 1990). It is somewhat remarkable that the relative level of variation in the two regions appears different in the two types; *i.e.*, the ITS is more variable in type I whereas the IGS is more variable in type II. However, given the error estimates in Table 5, it is difficult to draw any firm conclusions in this regard.

It should also be noted that the polymorphic sites in the ITS (not including the diagnostic sites) also segregate by molecular type (Table 2). That is, the polymorphic sites in the major type I and II are nonoverlapping. This is further evidence of little or no gene exchange in this region. Also, these polymorphic sites indicate possible differences in levels of gene flow among populations within types. However, given the sample sizes, any conclusions concerning population differentiation within types needs to be tentative at this point.

The network shown in Figure 3 was obtained by applying the method of statistical parsimony (Templeton *et al.* 1992) to the ITS haplotypes using the TCS program package (Crandall *et al.* 2000). The figure shows the maximum number of steps parsimoniously connecting two haplotypes. Connections between conspecific haplotypes are established only if the probability of parsimony for mutational steps exceeds 95%. The size of the circles or box is proportional to the number of individuals sharing the same haplotype. Only a limited amount of reticulation is present (the two loops in Figure 3), indicating either low amounts of homoplasy in the data set or rare hybridization events. The two major types within *A. gambiae*, gIA and gIIA, are separated by 3 steps and are connected through type III. Both *A. gambiae* types gIA and gIIA are separated from the commonest *A. arabiensis* type, aIA, by 28 steps. It is important to note that in this case both links were below the 95% cutoff point.

DISCUSSION

It is clear that within *A. gambiae* s.s. two distinct types of rDNA exist, which we have designated S/type I and M/type II to indicate the IGS/ITS sequences. In West Africa, these two types may be found sympatrically and synchronously. Including the data presented here and those of della Torre *et al.* (2001), Chandre *et al.* (1999), and Tripet *et al.* (2001), the total number of populations sampled in which both types were present (*i.e.*, sympatric) is 24 from five countries with a total of 1086 adult female mosquitoes analyzed. Two hybrid IGS molecular patterns were detected by della Torre *et al.* (2001) and one by Tripet *et al.* (2001). In none of the 251 adults we sampled were any ITS hybrids observed. Taylor *et al.* (2001) and Edillo *et al.* (2002) sampled larvae at a single site where both rDNA types were present and observed four hybrids in an estimated sample of 350. [Taylor *et al.* (2001) reported five hybrids in the original publication; however, one was subsequently found to be in error as reported in Edillo *et al.* (2002). Because hybrids for this X-linked marker can be de-

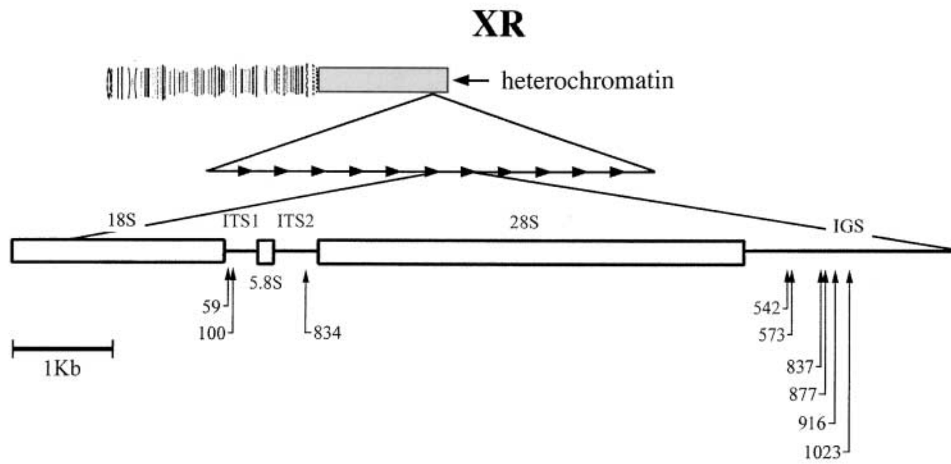

FIGURE 1.—Location and organization of the rDNA in *A. gambiae*. The rDNA locus consists of one very long array of repeat units (500–700 repeats per genome) located in the heterochomatic region of the X chromosome (COLLINS *et al.* 1989). Each repeat is ~9 kb long and consists of the genes for the 18S, 5.8S, and 28S rDNA (open rectangles). These genes are separated by spacers (solid lines), the internal transcribed spacers (ITS1 and ITS2), and the intergenic spacer (IGS). Arrows point to the three diagnostic sites in the ITS (type I and type II) and the six diagnostic sites in the IGS (M and S). The ITS and IGS sites are ~5 kb apart.

tected only in females, the sample size was estimated assuming a 1:1 sex ratio of larvae.] TRIPET *et al.* (2001) examined IGS genotypes (M *vs.* S) of females and the genotypes of the sperm they carried in their spermatheca; they found three females with the "wrong" sperm in a sample of 251 females examined. While it is difficult to precisely calculate the expected numbers (due to varying frequencies of the chromosomal forms present in different populations and at different collecting times), clearly the observed numbers of hybrid genotypes are a small fraction of those expected in a freely interbreeding panmictic unit. If we simply calculate the frequencies as observed, the rate of hybrid matings (*i.e.*, M × S) observed is 1.2%, and the percentage of hybrid larvae and adults is 1.1 and 0.3%, respectively. This may reflect reduced viability of hybrids from the larval to adult stages. No systematic study has been done under field conditions to test for viability. It is also of note that the frequency of adult hybrids between the established species *A. gambiae* and *A. arabiensis* is on the order of 0.05% (8 hybrids out of 15,745 individuals sampled in areas where both species are at 10% or greater; WHITE 1970; PETRARCA *et al.* 1991; TOURÉ *et al.* 1998). Thus, at this point, the observed frequency of adult hybrid rDNA molecular types is about the same as the frequency of detected hybrids between established species in this group.

One could question whether the methods used are sensitive enough to detect "mixtures" of the types in a DNA preparation, especially since the rDNA is a repeat unit of at least 100 copies per chromosome. While detection of an F_1 hybrid predicted to have equal numbers of each rDNA type would be easy, backcrosses could produce individuals with varying numbers of copies of each type. FAVIA *et al.* (2001) addressed this issue by mixing DNA preparations from the two types in varying ratios; even at a mixture of 1:100, a "hybrid" molecular pattern was detected. Similarly, our method of DNA sequencing is capable of detecting heterozygosity at single sites. We observed 23 heterozygotes in the ITS confirmed by sequencing both strands. None of these heterozygotes was at the sites used to define the major rDNA types.

The geographic distributions of the two types of rDNA are only partly known; the data set presented here is, to date, the most geographically widespread for both the IGS and ITS variants. So far M/type II has been found only in West Africa north of the equator, where S/type I also exists (Figure 1). All samples from East Africa, including Madagascar, are S/type I. It is not clear how far M/type II might extend westward through central Africa nor how far southward; these regions have simply not been sampled. [Preliminary data indicate that the M/type II form exists in the Congo (T. LEHMANN, our unpublished data).]

If new taxa are to be defined on the basis of studies of the sort reported here, it is important to explore whether the diagnostic characteristics used extend to populations throughout the range. Initially, on the basis of samples from Mali and Burkina Faso, the IGS M form was thought to correspond to the chromosomally defined Mopti form while S could be either Savanna or Bamako (FAVIA *et al.* 1997). With further geographic sampling reported here and in DELLA TORRE *et al.* (2001) and GENTILE *et al.* (2001) it is now clear that the chromosomally defined forms and the rDNA types do not have a one-to-one correspondence throughout the range. S/type I can be found in populations that, chromosomally, would be characterized as Bamako, Savanna, or Forest; M/type II can be Mopti, Savanna, Forest, or Bissau (Table 1; see also DELLA TORRE *et al.* 2001 and GENTILE *et al.* 2001). Furthermore, it is important to determine whether all diagnostic characters (in this case, nucleotides) extend to different geographic regions. So far, with the exception of the single São Tomé population, all three diagnostic sites in the

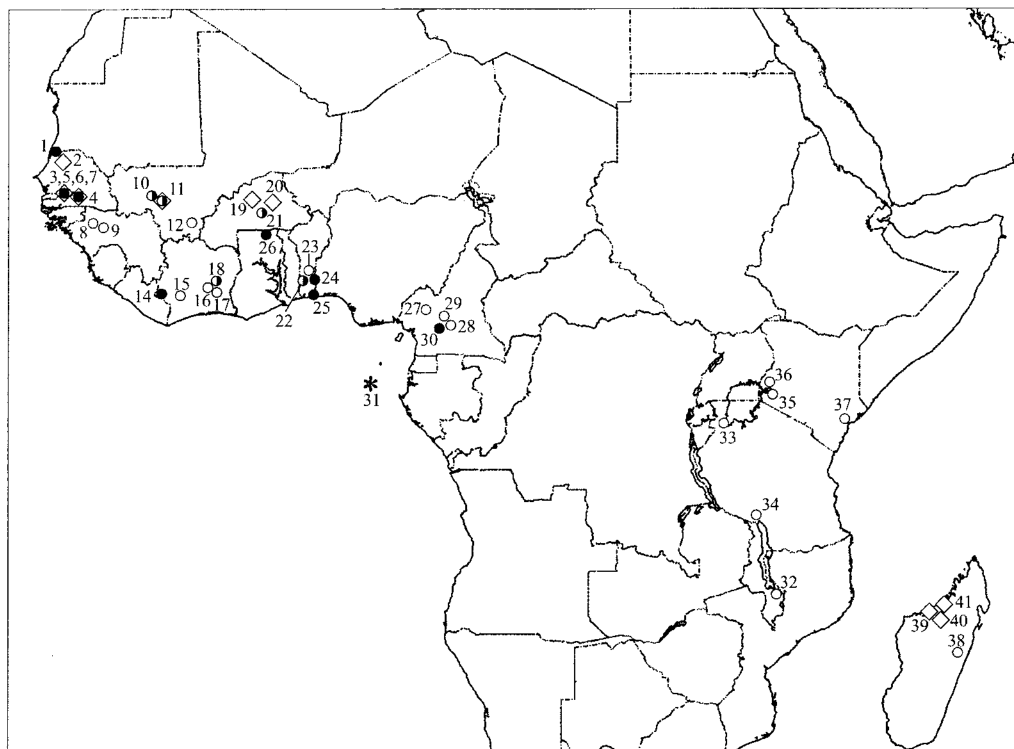

FIGURE 2.—Location of populations sampled. The names of the numbered localities are in Table 1. The map also reports the distribution of the three major types found for the ITS region in *A. gambiae.* Open circles correspond to type I (S form) and solid circles correspond to type II (M form). Localities where the two types are sympatric are indicated by solid and open semicircles. Type III (on Saõ Tomé) is represented by an asterisk. Locations where *A. arabiensis* populations have been sampled are indicated by open diamonds. Symbols overlap where sampled populations are sympatric or geographically proximate.

ITS remain diagnostic throughout the range sampled (Figure 2). Initially the M and S IGS forms were thought to have fixed differences at 10 nucleotide sites (FAVIA *et al.* 1997) on the basis of samples from Mali and Burkina Faso. The routine manner in which M and S are distinguished, either a restriction site difference or a primer sequence difference, assays only one or two of the fixed differences. More thorough direct DNA sequencing of the IGS in a limited number of our samples revealed that adding just five individuals from Benin reduced the number of IGS diagnostic sites to six (Table 4; Figure 1).

The problem of geographic variation for diagnostic sites is highlighted by the samples from São Tomé, an island about 300 km off the coast of West Africa (Figure 2). This is the only sample in which the complete linkage disequilibrium between the defining sites in the ITS breaks down. In a network analysis such as Figure 3, the São Tomé haplotype (gIII) connects the major rDNA types within *A. gambiae.* It is not possible to assess whether this represents further significant genetic substructuring or is simply a case of one nucleotide substitution fixed in a geographically isolated population.

If indeed the types defined by the rDNA patterns represent reproductively isolated taxa, then we need to confront the conundrum of why have not other parts of the genome diverged? GENTILE *et al.* (2001) and MUKABAYIRE *et al.* (2001) found no fixed differences between the rDNA types for regions of the mtDNA as well as for introns of nuclear genes studied on each chromosomal arm. As noted in the Introduction, the only nuclear gene other than rDNA that has been found significantly differentiated between S/type I and M/type II is *Kdr*, an insecticide resistance locus. Even here, the data are not clear-cut. In the Ivory Coast and Burkina Faso, M/type II does not have the resistant allele while sympatric S/type I has it in frequencies ranging from 17 to 90% (CHANDRE *et al.* 1999; DELLA TORRE *et al.* 2001). However in Benin, M/type II has the resistant allele at a frequency of ~50% (DELLA TORRE *et al.* 2001); evidently it is the same allele that has been passed between S/type I and M/type II via hybridization (WEILL *et al.* 2000). So while *Kdr* displays distinct frequency differences between sympatric forms of the rDNA types in some regions, this is not true throughout the range of the species. Similarly, microsatellite studies have revealed virtually no differentiation between chromosome arms except for 2R (LANZARO *et al.* 1998; TAYLOR *et al.* 2001) and a single microsatellite closely linked to the rDNA region (WANG *et al.* 2001).

Given evidence (so far) that the rest of the genomes of these two rDNA types are not differentiated, we need to ask why is it that the rDNA region is? For example, is the concerted evolutionary dynamics of the rDNA in these mosquitoes of a sort to cause the patterns we are seeing in the absence of any reproductive isolation? In the extreme, concerted evolution might occur in every individual every generation so no heterozygotes would ever be observed. If so, then we might expect a similar pattern in the close sibling species of *A. gambiae,* especially in *A. arabiensis,* which is very similar to *A. gambiae* in ecology, geographic distribution, population sizes,

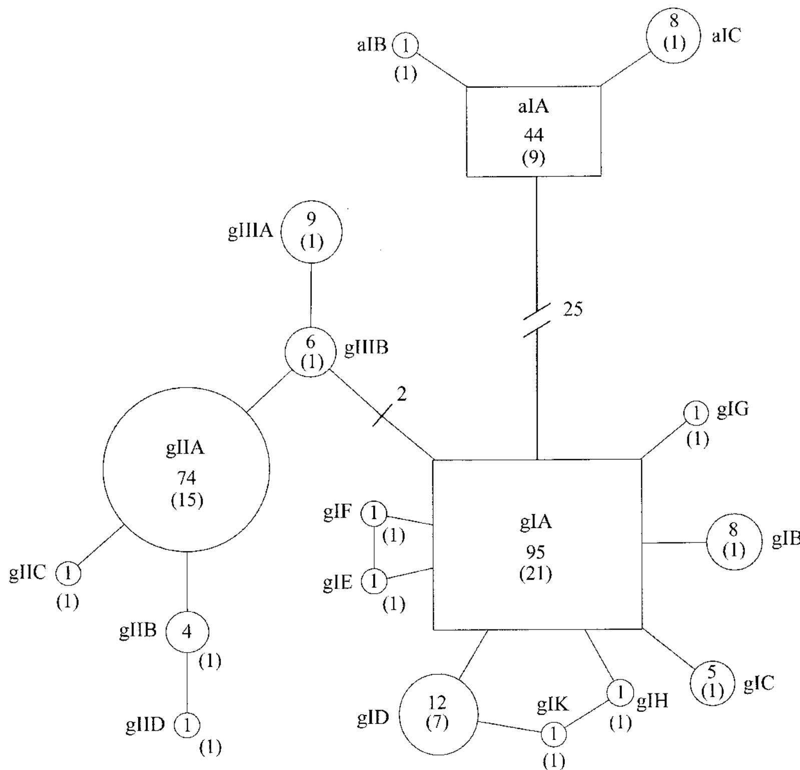

Figure 3.—Network analysis based on statistical parsimony (Templeton *et al.* 1992) showing the genealogical relationships of the ITS haplotypes in *A. gambiae* and *A. arabiensis*. The size of the circle or rectangle corresponds to the haplotype frequency. Rectangles identify the most abundant haplotypes; by the criteria of Donnelly and Tavaré (1986) and Castelloe and Templeton (1994) these would be considered the probable ancestral types. Designations beginning with "g" are from *A. gambiae* and those beginning with "a" are from *A. arabiensis*. Roman numerals indicate major types followed by letters indicating minor variants, usually a single nucleotide difference. Numbers within each circle or rectangle refer to the individuals carrying that haplotype and to the number of populations where each haplotype is found (within parentheses). All the haplotypes are linked by 1 step except for the two major haplotypes in *A. gambiae* (gIA and gIIA; 2 steps). Connections are established on the basis of a set of plausible solutions with a 95% of parsimony probability. The haplotypes gIA, gIIA (*A. gambiae*), and aIA (*A. arabiensis*) differ by 28 steps.

and migration rates. Furthermore, we expect the process of genomic evolution, including concerted evolution, to be very similar, if not identical, in the two species. Thus we sequenced a small sample of *A. arabiensis* for the rDNA regions from some of the same localities as we had for *A. gambiae*. There is no evidence for a similar pattern in this species, best seen in Figure 3. While *A. gambiae* breaks into two major groups in the network, *A. arabiensis* is a single major haplotype with some minor variants. It should be noted that while small, this sample of *A. arabiensis* represents a wide geographic distribution. Similarly, it is instructive to examine the overall genetic variation in these species. If we consider all *A. gambiae* as a single taxon, then the nucleotide heterozygosity of the rDNA region in this species is much greater than that in *A. arabiensis* (Table 5); however, if *A. gambiae* samples are subdivided into S/type I and M/type II, the genetic variation is virtually identical to that of *A. arabiensis* (Table 5). The overall heterozygosity is inflated when considering *A. gambiae* a single unit due to matching sequences between major rDNA types S/I and M/II. So if one assumes that the expected or "typical" heterozygosity for a species for the rDNA of these mosquitoes is like *A. arabiensis*, which all evidence indicates is a single taxon, then each of the rDNA types within *A. gambiae* has the variation typical of a species.

The incipient reproductive isolation among units within what was considered a single species, *A. gambiae* s.s., was first detected by examining chromosomal inversion frequencies (Coluzzi *et al.* 1985). While in some regions the rDNA types defined here also correspond to the chromosomal forms, the relationship is complex and varies among geographic locations. The rDNA types appear to be more reliable and consistent indicators of reproductive isolation compared to the 2R inversion frequencies. Should, then, new taxa be raised on the basis of these observations? There can be little doubt that, as presently defined, *A. gambiae* s.s. is not a single panmictic unit and some taxonomic subdivision is warranted; perhaps even new species should be recognized. The lack of detectable molecular differentiation elsewhere in the genome (Lanzaro *et al.* 1998; Gentile *et al.* 2001; Mukabayire *et al.* 2001; Wang *et al.* 2001) implies that these different molecularly characterized types may have mosaic genomes consisting of parts completely differentiated between which gene flow is barred, whereas other parts of the genome are free to pass between types. There is good evidence that a similar situation exists for *A. gambiae* and *A. arabiensis*; *i.e.*, there is selective introgression between these species (Coluzzi *et al.* 1979, 1985; Besansky *et al.* 1994, 1997; Caccone *et al.* 1996; della Torre *et al.* 1998).

To summarize what is now known of genetic subdivision within *A. gambiae* s.s. in West Africa:

1. There is sympatric differentiation for inversions in the right arm of the second chromosome (Coluzzi *et al.* 1985; Touré *et al.* 1998).
2. Sympatric differentiation at the rDNA region near the centromere of the X chromosome region (this report and references in the Introduction) also includes a microsatellite that is tightly linked to, but is

not within, the rDNA repeat unit (Wang *et al.* 2001). At least in some geographic localities, the rDNA region differentiation corresponds largely to the 2R inversion differentiation.

3. Where the insecticide resistance gene *Kdr* is found, it is often in very different frequencies in the subunits defined by the rDNA and 2R inversions (Chandre *et al.* 1999; della Torre *et al.* 2001). This gene is closely linked to the centromere on 2L. Given that the inversions defining the chromosomal differentiation are on the distal part of 2R, it is unlikely that the *Kdr* differentiation is due to close linkage (hitchhiking) to selectively affected inversions.
4. Despite considerable efforts, no significant differentiation has been noted in other parts of the genome and all indications are that, except for the three regions just noted, gene flow is occurring in a manner nearly indistinguishable from panmixia.

As pointed out in the Introduction, there is a growing recognition and acceptance that reproductive isolation may not be a genome-wide phenomenon especially at the earliest stages of the speciation process. It should be emphasized that this perception is based on empirical evidence and, among animals, comes primarily from insects (Barton and Hewitt 1981; Harrison 1990; Noor *et al.* 2001; Wu 2001; Machado *et al.* 2002). If new taxa are to be erected for these "molelcular types" and/or "chromosomal forms" of *A. gambiae*, the crux of the problem is to provide an accurate and consistent definition for new taxa. There are no detectable postmating barriers, morphological distinctions, or uniform chromosomal differences between the rDNA types. Would then differences at a few nucleotide sites be sufficient "description" to erect new taxa?

It needs to be emphasized that the recognition of new taxa in this group of mosquitoes is not simply an academic exercise. Members of this complex are the most important insects in the world from a medical standpoint. An estimated 2.7 million people die of malaria each year; a majority have had the disease transmitted to them by members of the *A. gambiae* complex. Control measures need to take into consideration the number of taxonomic or genetic units that are being targeted for control, especially considering the evidence that the units may be ecologically distinct. For example, in Mali, where rDNA M/type II corresponds to the Mopti chromosomal form, it is clear that Mopti is a much more dry-adapted mosquito that can breed yearround (Touré *et al.* 1998); this allows continuous malaria transmission that would otherwise be interrupted by the dry season.

We thank the following who collected the samples of mosquitoes used in this study: A. Bosman, J. H. Bryan, C. Carrara, D. Charlwood, C. Costantini, C. F. Curtis, V. do Rosario, D. Fontenille, Bill Hawley, M. Jawara, J. H. Kolaczinski, M. Pinder, O. Ralisoa, R. Rakotoson, G. Sabatinelli, N'F. Sagnon, G. A. Targett, and Y. T. Touré. We also thank M. Coluzzi and T. Lehmann for providing the DNA samples. We thank N. Besansky, G. Lanzaro, C. Taylor, and F. Tripet for comments on an earlier version of this article; two reviewers and especially the communicating editor, Richard Harrison, were also very helpful in stimulating us to clarify the presentation of a complex situation. This investigation received financial assistance from National Institutes of Health grant RO1-AI46018 to J.R.P., the INCO-DC program of the European Union (contract no. IC18CT970244), and the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) to A.C. and to A.d.T.

LITERATURE CITED

Anderson, E., 1949 *Introgressive Hybridization.* John Wiley & Sons, New York.

Arnold, M. L., 1997 *Natural Hybridization and Evolution.* Oxford University Press, New York.

Barton, N. H., and G. M. Hewitt, 1981 Hybrid zones and speciation, pp. 109–145 in *Evolution and Speciation*, edited by W. Atchley and D. Woodruff. Cambridge University Press, Cambridge, UK.

Besansky, N. J., J. R. Powell, A. Caccone, D. M. Hamm, J. A. Scott *et al.*, 1994 Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. Proc. Natl. Acad. Sci. USA **91:** 6885–6888.

Besansky, N. J., T. Lehmann, T. G. Fahey, D. Fontenille, L. Braack Lawrence *et al.*, 1997 Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *An. arabiensis*, suggest extensive gene flow. Genetics **147:** 1817–1828.

Caccone, A., B. A. García and J. R. Powell, 1996 Evolution of the mitochondrial DNA control region in the *Anopheles gambiae* complex. Insect Mol. Biol. **5:** 51–59.

Castelloe, J., and A. R. Templeton, 1994 Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. **3:** 102–113.

Chandre, F., A. Manguin, C. Brengues, Y. Dossou, F. Darriet *et al.*, 1999 Current distribution of pyrethroid resistance gene (*kdr*) in *Anopheles gambiae* complex from West Africa and further evidence for reproductive isolation of the Mopti form. Parassitologia **41:** 319–322.

Clement, M., D. Posada and K. A. Crandall, 2000 TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9:** 1657–1659.

Collins, F. H., M. A. Mendez, M. O. Rasmussen, P. C. Mehaffey, N. J. Besansky *et al.*, 1987 A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. Am. J. Trop. Med. Hyg. **37:** 37–41.

Collins, F. H., S. M. Paskewitz and V. Finnerty, 1989 Ribosomal RNA genes of the *Anopheles gambiae* species complex. Adv. Dis. Vector Res. **6:** 1–28.

Coluzzi, M., and A. Sabatini, 1967 Cytogenetic observations on species A and B of the *Anopheles gambiae* complex. Parassitologia **9:** 73–88.

Coluzzi, M., and A. Sabatini, 1968 Cytogenetic observations on species C of the *Anopheles gambiae* complex. Parassitologia **10:** 155–166.

Coluzzi, M., and A. Sabatini, 1969 Cytogenetic observations on the salt water species, *Anopheles merus* and *Anopheles melas* of the *gambiae* complex. Parassitologia **11:** 177–187.

Coluzzi, M., A. Sabatini, V. Petrarca and M. A. Di Deco, 1979 Chromosomal differentiation and adaptation to human environment in the *Anopheles gambiae* complex. Trans. R. Soc. Trop. Med. Hyg. **73:** 483–497.

Coluzzi, M., V. Petrarca and M. A. Di Deco, 1985 Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae.* Boll. Zool. **52:** 45–63.

Davidson, G., 1964 The five mating types of the *Anopheles gambiae* complex. Rivista di Malariologia **13:** 167–183.

Davidson, G., and R. H. Hunt, 1973 The crossing and chromosome characteristics of a new sixth species in the *Anopheles gambiae* complex. Parassitologia **15:** 121–128.

della Torre, A., L. Merzagora, J. R. Powell and M. Coluzzi, 1998 Selective introgression of paracentric inversions between two sib-

ling species of the *Anopheles gambiae* complex. Genetics **146:** 239–244.

della Torre, A., C. Fanello, M. Akogbeto, J. Dossou-yovo, G. Favia *et al.*, 2001 Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. Insect Mol. Biol. **10:** 9–18.

Di Deco, M. A., V. Petrarca, F. Villani and M. Coluzzi, 1980 Polimorfismo cromosomico da inversioni paracentriche ed eccesso degli eterocariotypi in ceppi di *Anopheles* allevati in laboratorio. Parassitologia **22:** 304–306.

Dobzhansky, T., 1935 A critique of the species concept in biology. Philos. Sci. **2:** 344–355.

Donnelly, P., and S. Tavaré, 1986 The ages of alleles and a coalescent. Adv. Appl. Probab. **18:** 1–19.

Edillo, F. E., Y. T. Touré, G. C. Lanzaro, G. Dolo and C. E. Taylor, 2002 Spatial and habitat distribution of *Anopheles gambiae* and *Anopheles arabiensis* (Diptera: Culicidae) in Banambani Village, Mali. J. Med. Entomol. **39:** 70–77.

Favia, G., G. Dimopoulos, A. della Torre, Y. T. Touré, M. Coluzzi *et al.*, 1994 Polymorphisms detected by random PCR distinguishes between different chromosomal forms of *Anopheles gambiae.* Proc. Natl. Acad. Sci. USA **91:** 10315–10319.

Favia, G., A. della Torre, M. Bagayoko, A. Lanfrancotti, N'F. Sagnon *et al.*, 1997 Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. Insect Mol. Biol. **6:** 377–383.

Favia, G., A. Lanfrancotti, L. Spanos, I. Sidén-Kiamos and C. Louis, 2001 Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s.s. Insect Mol. Biol. **10:** 19–24.

Gentile, G., M. Slotman, V. Ketmaier, J. R. Powell and A. Caccone, 2001 Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae* s.s. Insect Mol. Biol. **10:** 25–32.

Harrison, R. G., 1990 Hybrid zones: windows on evolutionary process, pp. 69–128 in *Oxford Surveys in Evolutionary Biology*, Vol. 7, edited by D. Futuyma and J. Antonovics. Oxford University Press, Oxford.

Hunt, R. H., M. Coetzee and M. Fettene, 1998 The *Anopheles gambiae* complex: a new species from Ethiopia. Trans. R. Soc. Trop. Med. Hyg. **92:** 231–235.

Lanzaro, G. C., Y. T. Touré, J. Carnahan, L. Zheng, G. Dolo *et al.*, 1998 Complexities in the genetic structure of *Anopheles gambiae* populations in West Africa as revealed by microsatellite DNA analysis. Proc. Natl. Acad. Sci. USA **95:** 14260–14265.

Machado, C. A., R. M. Kliman, J. A. Markert and J. Hey, 2002 Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and its close relatives. Mol. Biol. Evol. **19:** 472–488.

Mayr, E., E. G. Linsley and R. L. Usinger, 1953 *Methods and Principles of Systematic Zoology.* McGraw-Hill, New York.

Mukabayire, O., J. Caridi, X. Wang, Y. T. Touré, M. Coluzzi *et al.*, 2001 Patterns of DNA sequence variation in chromosomally recognized taxa of *Anopheles gambiae*: evidence from rDNA and single-copy loci. Insect Mol. Biol. **10:** 33–46.

Noor, M. A. F., K. L. Grams, L. A. Bertucci and J. Reiland, 2001 Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA **98:** 12084–12088.

Paskewitz, S. M., D. M. Wesson and F. H. Collins, 1993 The internal transcribed spacers of ribosomal DNA in five members of the *Anopheles gambiae* complex. Insect Mol. Biol. **2:** 247–257.

Petrarca, V., J. C. Beier, F. Onyango, J. Koros, C. Asiago *et al.*, 1991 Species composition of the *Anopheles gambiae* complex (Diptera:Culicidae) at two sites in Western Kenya. J. Med. Entomol. **28:** 307–313.

Ranson, H., B. Jensen, J. M. Vulule, X. Wang, J. Hemingway *et al.*, 2000 Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. Insect Mol. Biol. **9:** 491–498.

Taylor, C. E., Y. T. Touré, J. Carnahan, D. E. Norris, G. Dolo *et al.*, 2001 Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. Genetics **157:** 743–750.

Templeton, A. R., K. A. Crandall and C. F. Sing, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Thweatt, R., and J. C. C. Lee, 1990 Yeast precursor ribosomal RNA molecular cloning and probing the higher-order structure of the internal transcribed spacer I by kethoxal and dimethylsulfate modification. J. Mol. Biol. **211:** 305–320.

Touré, Y. T., V. Petrarca, S. F. Traoré, A. Coulibaly, H. M. Maiga *et al.*, 1998 Distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. Parassitologia **40:** 477–511.

Tripet, F., Y. T. Touré, C. E. Taylor, D. E. Norris, G. Dolo *et al.*, 2001 DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae.* Mol. Ecol. **10:** 1725–1732.

Wang, R., L. Zheng, Y. T. Touré, T. Dandekar and F. Kafatos, 2001 When genetic distance matters: measuring genetic differentiation at microsatellite loci in whole-genome scans of recent and incipient mosquito species. Proc. Natl. Acad. Sci. USA **98:** 10769–10774.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Weill, M., F. Chandre, C. Berngues, S. Manguin, M. Akogbeto *et al.*, 2000 The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae* s.s. through introgression. Insect Mol. Biol. **9:** 451–455.

White, G. B., 1970 Chromosomal evidence for natural interspecific hybridization by mosquitoes of the *Anopheles gambiae* complex. Nature **231:** 184–185.

White, G. B., 1974 *Anopheles gambiae* complex and disease transmission in Africa. Trans. R. Soc. Trop. Med. Hyg. **68:** 278–301.

Wu, C.-I, 2001 The genic view of the process of speciation. J. Evol. Biol. **14:** 851–865.

Yeh, L. C. C., and J. Lee, 1990 Structural analysis of the internal transcribed spacer 2 in the precursor ribosomal RNA from *Saccharomyces cerevisiae.* J. Mol. Biol. **211:** 699–712.

Communicating editor: R. Harrison