Gene Flow Among Populations of the Malaria Vector, *Anopheles gambiae*, in Mali, West Africa

Charles Taylor,* Yeya T. Touré,† John Carnahan,* Douglas E. Norris,‡,[1] Guimogo Dolo,† Sekou F. Traoré,† Frances E. Edillo* and Gregory C. Lanzaro‡

**Department of Organismic Biology, Ecology and Evolution, University of California, Los Angeles, California 90095-1606,* †*Malaria Research and Training Center, Faculté de Médecine de Pharmacie et d'Odonto-Stomatologie, Bamako B.P. 1805, Mali and* ‡*Department of Pathology, University of Texas Medical Branch, Galveston, Texas 77555-0609*

Manuscript received March 19, 2000
Accepted for publication November 7, 2000

ABSTRACT

The population structure of the *Anopheles gambiae* complex is unusual, with several sibling species often occupying a single area and, in one of these species, *An. gambiae sensu stricto*, as many as three "chromosomal forms" occurring together. The chromosomal forms are thought to be intermediate between populations and species, distinguishable by patterns of chromosome gene arrangements. The extent of reproductive isolation among these forms has been debated. To better characterize this structure we measured effective population size, N_e, and migration rates, m, or their product by both direct and indirect means. Gene flow among villages within each chromosomal form was found to be large ($N_em > 40$), was intermediate between chromosomal forms ($N_em \approx 3$–30), and was low between species ($N_em \approx 0.17$–1.3). A recently developed means for distinguishing among certain of the forms using PCR indicated rates of gene flow consistent with those observed using the other genetic markers.

THE *Anopheles gambiae* complex in Africa consists of six species, distinguishable by their gene arrangements. They are morphologically similar to one another, in some cases indistinguishable, so are collectively termed *An. gambiae sensu lato*. Two members of this complex, *An. gambiae s.s.* and *An. arabiensis*, are very common at our focal study site, Banambani village near Bamako, Mali, where we have been studying local populations of these species for almost two decades. One of these species, *An. gambiae s.s.*, is itself quite polymorphic for gene arrangements on the right arm of chromosome 2 (Touré *et al.* 1998a). These gene arrangements are in marked departure from Hardy-Weinberg equilibrium, prompting the suggestion to further divide *An. gambiae s.s.* into several "chromosomal forms," the most common of which in Mali are "Bamako," "Mopti," and "Savanna" (Figure 1). This unusual pattern of population structure probably contributes to the great ecological flexibility of *An. gambiae* (Bryan *et al.* 1982; Coluzzi *et al.* 1979, 1985; Touré *et al.* 1994), and underlies their importance as vectors of malaria, responsible for hundreds of millions of malaria cases per year (Petrarca and Beier 1992). An understanding of this structure assumes special importance for attempts to control malaria by genetic manipulation of the vector species (Collins 1994).

While gene flow across the recognized species boundaries is known to occur at low levels, the extent of reproductive isolation among the forms has been debated. On the basis of attempts to fit gene arrangement frequencies to Hardy-Weinberg equilibria, it would appear that the Bamako and Mopti forms rarely mate in nature, though both forms seem to hybridize with the Savanna form. Estimates for percentage of hybrids in the population range from 0 to 11%, depending on location, time of year, and the form in question (see Tables 7–10 in Touré *et al.* 1998a).

As a result, the taxonomic status of these forms is unclear. On the one hand, they mate freely in the laboratory and hybrid offspring are just as viable, or more so, than the parental forms (Di Deco *et al.* 1980; Persiani *et al.* 1986). This would suggest they are not reproductively isolated. On the other hand, Favia *et al.* (1997) have suggested that the forms are almost completely reproductively isolated, at least in Mali. There are ecological differences among forms (Touré *et al.* 1998a), a moderate difference of allele frequencies for microsatellite DNA loci on the 2R chromosome (Lanzaro *et al.* 1998), and some differences on the X chromosome (Favia *et al.* 1997). These support the interpretation that the forms should be regarded as "good" species. The core issue here relates to the extent of gene flow among several subpopulations.

A variety of methods for estimating gene flow now exist. Direct, or ecological, methods attempt to measure

Corresponding author: Charles Taylor, Department of Organismic Biology, Ecology and Evolution, University of California, 621 Charles Young Dr. S., Box 951606, Los Angeles, CA 90095-1606.
E-mail: taylor@biology.ucla.edu

[1] *Present address:* School of Hygiene and Public Health, Johns Hopkins University, Baltimore, MD 21205.

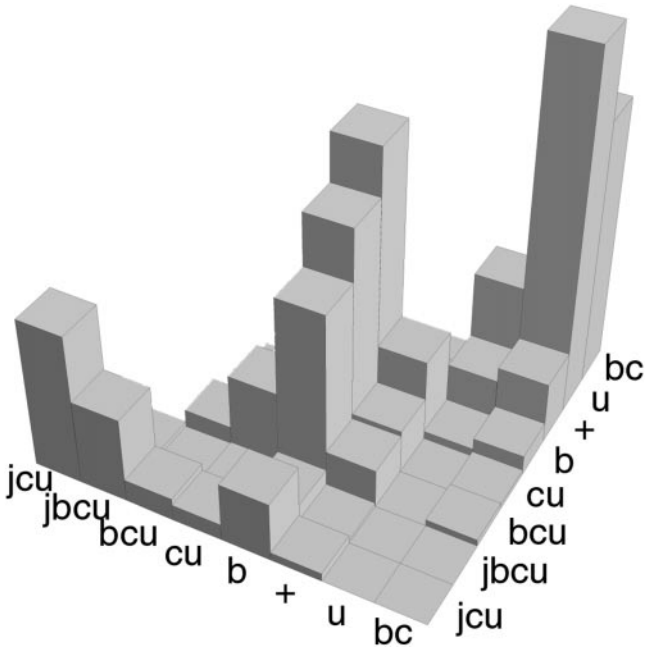

FIGURE 1.—Gene arrangements of the 2R chromosome in *Anopheles gambiae s.s.* Height of column indicates representation in a sample of 1422 females from Banambani, Mali reported by TOURÉ *et al.* (1998a, Table 7). The maximum height corresponds to 219 individuals. The clusters of genotypes along the diagonal correspond to the chromosomal forms, distinguished by gene arrangements. There is a clear deficiency of hybrid forms from overall Hardy-Weinberg equilibrium that would be expected if the populations were panmictic. But there is a large excess of hybrids, 8.1%, if the populations were assumed to be reproductively isolated. Such hybrids are most evident for the Bamako × Savanna crosses, *jcu* with *bcu*, *cu*, and *b* heterozygotes.

population size and migration rates with direct ecological and behavioral methods like mark-release-recapture (MRR). Direct methods have the advantage of unambiguously distinguishing contributing behaviors, but are difficult or inaccurate when population densities are low and are likely to overlook rare but important events, like population crashes or occasional migrants. Indirect, or genetic, methods examine patterns of allele and genotype frequencies and infer what population sizes, migration rates, etc., would have had to be responsible. Such methods for estimating population structure assume the populations are in equilibrium; if that is not the case, for example, if they are recovering from a crash or colonization event, then the estimates of population parameters can be quite erroneous. Further, it may occur that several explanations could produce the same patterns or observations. Both methods have their advantages and shortcomings. A good understanding can come only from examination by both approaches (TAYLOR and POWELL 1983; SLATKIN 1985); we have attempted to do so for the relevant populations, forms, and species of *An. gambiae s.l.* found at Banambani.

MATERIALS AND METHODS

Microsatellite DNA: Gene flow among forms, species, and locations was estimated by indirect means using 21 microsatellite DNA loci. The chromosomal locations, primers, and methods for collecting and processing the DNA were identical to those used in our earlier study (LANZARO *et al.* 1998). Allele frequencies for *An. arabiensis* and the Mopti and Bamako forms of *An. gambiae s.s.* were taken from that study. Markers were chosen to achieve a representative sample of the entire genome ($2n = 6$), as follows: *AGXH38*, *AGXH25*, *AGXH8*, *AGXH7*, and *AGXH293* on the X chromosome; *AG2H175*, *AG2H156*, *AG2H79*, *AG2H26*, *AG2H125*, *AG2H95*, *AG2H135*, *AG2H637*, *AG2H603*, and *AG2H143* on chromosome 2; and *AG3H128*, *AG3H119*, *AG3H83*, *AG3H158*, *AG3H88*, and *AG3H577* on chromosome 3. The numbers of Savanna forms from those collections at the principal collection site, Banambani, were small, and there were no Savanna forms from the second site, Selenkenyi. For this study these were supplemented by a collection in 1998 from another village, Pimperena ($N = 44$), ~380 km from Banambani where the Savanna form was more abundant. Details of allelic frequencies are available from the senior author or online at the Anopheles Database site, AnoDB, at http://konops.imbb.forth.gr/AnoDB/. Differences in allele frequencies were computed with the ARLEQUIN software package (SCHNEIDER *et al.* 1996) and F_{ST} and R_{ST} were measured. N_em was estimated from those values using the relationship F_{ST} (or R_{ST}) $= (1 + 4N_em)^{-1}$ (SLATKIN 1995). Estimates of gene flow, N_em, depend somewhat on which model of mutation is used, determining the choice of F_{ST} or R_{ST} (COLSON and GOLDSTEIN 1999). The two measures gave concordant results. For conciseness the values presented here were based on F_{ST}, because the estimates of error using resampling methods in the ARLEQUIN software package were smaller.

Mark-release-recapture experiments: Direct, or ecological, measures of gene flow were based largely on MRR experiments at Banambani village for 1993–1994 and 1996–1998. Details about methods and analysis for the 1993–1994 experiments have been published (TOURÉ *et al.* 1998b); the 1996–1998 experiments were performed in a similar fashion. The MRR studies involved capturing large numbers of blood-fed females from the walls of huts, transporting them to a central location where they were counted, dusted with fluorescent powder, and then released. On subsequent days blood-fed females were again collected and checked to see how many of them were marked. Because this method involves only females, the total population size of both sexes requires that the numbers estimated with the Lincoln index or Fisher-Ford methods be doubled. The numbers in Table 2 are the summed *An. arabiensis* and *An. gambiae s.s.*

Gene flow among villages was estimated by looking at intervillage movement during MRR experiments. A marked individual was recorded from a nearby village, Donéguébougou, ~7 km from Banambani, in 1994. But the number of those unmarked was unfortunately not recorded at that time. To study this further we captured mosquitoes during the 1996–1998 MRR studies from four nearby villages (Donéguébougou, Sirakoro, N'Toubana, and Seguintambougou), all separated by a few kilometers of normal northern Sudan-savannah habitat. The numbers of both marked and unmarked mosquitoes captured in the days following release were recorded.

Chromosome studies: Details of chromosome preparation and identification have been described in TOURÉ *et al.* (1998a), where the observations for 1982–1994 were also reported; those for 1996–1998 are new. Resting female mosquitoes were captured from the walls of huts and taken to the laboratory, where half-gravid females were preserved in Carnoy's solution

until further processing. They were later dissected, their ovaries removed, and chromosome squashes prepared using the method of Hunt (1973). The use of karyotype information to classify females to species and form and to identify hybrids has been described in Touré *et al.* (1998a).

PCR characterization: Larvae were sampled from all of the known breeding sites in Banambini and its immediate surrounds (Edillo *et al.* 1999). They were dried and stored individually at −80° until prepared for DNA extraction. Extraction of DNA was performed on individual larvae following the protocol of Post *et al.* (1993). Species were diagnosed by PCR, according to Scott *et al.* (1993). Further aliquots of those larvae identified as *An. gambiae s.s.* were amplified according to the method of Favia *et al.* (1997), using the modified probes of Lanfrancotti *et al.* (1998). Aliquots of the PCR products thus generated were digested, individually, with the restriction enzymes *Tru91* and *HhaI* and electrophoresed on 2% agarose gels to generate the diagnostic banding phenotypes as described by Favia *et al.* (1997). This method is reported to be able to distinguish the Mopti form from members of the Savanna and Bamako forms, but is not capable of distinguishing Savanna from Bamako. We evaluated this technique, using a sample of 140 karyotyped individuals collected from Banambani village in 1993. Our results confirmed the accuracy of this approach, yielding complete concordance between the two methods with no anomalous banding patterns.

RESULTS

Indirect measures of gene flow: Values of N_em calculated from F_{ST} among the forms of *An. gambiae s.s.* and *An. arabiensis* at the Banambani site are shown in Table 1. The rates for 7 loci on chromosome 2, linked to the distinguishing chromosome arrangements, are above the diagonal and the averages of all 21 loci are below the diagonal. In all cases the values in this table were significantly different from 0 at the $P \leq 0.01$ level.

Values of gene flow across the good species boundary, *An. arabiensis vs.* Bamako, Mopti, or Savanna form of *An. gambiae s.s.*, were consistently lower, with N_em in the range of 1.1–1.8, than they were across forms. Estimated values of N_em from Table 1 are, in all cases of interform gene flow, somewhat lower for loci on 2R than for the 21-loci average, suggesting less gene flow for loci on this chromosome than for the genome as a whole. This might be a result of selection for the gene arrangements. The pattern is reversed across the border between *An. arabiensis* and *An. gambiae s.s.*, with somewhat more gene flow occuring across species boundaries for genes on the 2R chromosome. This is expected in view of results from caged populations of the two species, where genes were exchanged across species boundaries more freely for loci on 2R than on the other chromosomes examined (della Torre *et al.* 1997).

Gene flow of microsatellite DNA alleles across locations seems to show no simple pattern where it has been studied in East Africa (Kamau *et al.* 1998; Simard *et al.* 1999). The pattern in West Africa has received less analysis. Our earlier study of the Bamako and Mopti forms of *An. gambiae s.s.* between Banambani and Selenkenyi, ~170 km distant, gave no significant differences of gene frequencies between locations. When *F*-statistics were nonetheless computed, they gave averaged values of $F_{ST} = 0.00217$ for Bamako and 0.00566 for Mopti (R_{ST} values were −0.01690 and 0.02312). The corresponding values of N_em from these values ranged from 10.6 to infinity; the median value was 79.4. In this study the Savanna forms from Bamako and Pimperena, ~380 km distant, gave $F_{ST} = 0.034$ and $N_em = 7.1$ for the average across all loci; this is statistically significant from 0 at the 0.01 level, and conforms to the expectation of less gene flow across greater distances.

Direct measures of gene flow: Table 2 summarizes the MRR studies that were conducted at Banambani, Mali from 1993 to 1998. Population sizes were calculated with the Fisher-Ford method (Service 1993) using estimates of daily survival for 1993–1994 described in Touré *et al.* (1998b) and obtained from parity rates for 1996–1998. The numbers include both species and all three forms of *An. gambiae s.l.* during the peak of the rainy season. They range from 20,000 to 79,000. It is desirable to distinguish among the various forms and groups as much as possible. However, while some individuals were typed in these samples, the numbers were typically small and unevenly distributed over the years. In addition, the relative abundance of the different forms and species varies with the time of year. Accordingly, we used the long-term data for estimating N_e. At Banambani the overall proportions observed for 5414 individuals of the two species over 7 years (Table 2 in Touré *et al.* 1998a) were 61.5% *An. gambiae s.s.* and 38.5% *An. arabiensis.* Of *An. gambiae s.s.*, the forms and hybrids that could be

TABLE 1

N_em of *An. gambiae s.l.* estimated from F_{ST}

OTU	Mopti	Savanna	Bamako	*arabiensis*
Mopti		3.83	2.22	1.74
Savanna	7.97		1.86	1.58
Bamako	3.89	3.17		1.22
arabiensis	1.32	1.14	1.11	

Operational taxonomic units (OTU) are the chromosomal forms of *An. gambiae s.s.* and *An. arabiensis.*

TABLE 2

Population sizes from mark-release-recapture experiments

Year	Reps	Released	Recaptured	Daily survival	Estimated *N*
1993	3	938	83	0.80	20,178
1994	4	1,913	57	0.80	64,002
1996	4	1,421	44	0.92	63,006
1997	2	1,002	24	0.97	53,400
1998	4	1,205	21	0.97	79,280

Reps refers to the number of replicate releases. Other values refer to data summed across all replicates for the year.

TABLE 3

Composition determined from chromosome arrangements

Year	Month	*N*	Bam	Mopti	Sav	Bam × Mopti	Bam × Sav	Mopti × Sav
1982	6, 8–10	227	45.81	17.62	29.50	0.00	3.52	0.88
1983	6–9	1526	10.16	34.21	41.28	0.00	5.96	1.57
1983	11	58	5.17	62.07	18.97	0.00	1.72	3.45
1984	3–5	165	0.61	96.97	0.61	0.00	0.00	1.21
1984	6–10	499	19.04	44.49	28.66	0.00	2.81	1.20
1985	7–9	370	12.16	26.49	48.38	0.00	2.16	1.62
1986	6–9	203	31.53	42.36	17.73	0.00	2.96	0.49
1989	7–10	222	40.99	22.97	23.87	0.00	1.80	2.25
1992	9	89	28.09	10.11	44.94	0.00	2.25	1.12
1993	7–8	243	14.40	62.14	12.76	0.00	1.23	4.12
1994	8	215	20.47	26.05	38.60	0.00	3.26	1.86
1996	7–8	16	12.50	37.50	43.75	0.00	6.25	0.00
1997	8	14	21.43	21.43	57.14	0.00	0.00	0.00
1998	8	22	0.00	72.73	27.27	0.00	0.00	0.00
Ave.	(Total)	(3869)	18.74	39.08	30.96	0.00	2.42	1.41

Bam, Bamako; Sav, Savanna; ave., average.

identified in collections from Banambani are shown in Table 3. The karyotypes that could not be classified were rare and are not shown in the table. Data for 1982–1994 are from Touré *et al.* (1998a), and those for 1996–1998 are new. The values in the bottom line are the unweighted averages over all collections. It should be recognized that there are significant changes in composition that occur within years and also signficant year-to-year fluctuations. The unweighted mean composition for collections over the 16-year period was 19% Bamako, 31% Savanna, 39% Mopti, and the remainder hybrids or otherwise unclassifiable.

All of our MRR studies reported here were conducted during the wet season, June–September. Calculation of N_e requires that we also know population sizes during other times in the year. If we assume that the dry-season population densities are $\sim 1/10$ of the wet-season high, that the wet-season population size is the mean value calculated from the MRR studies listed in Table 2, and that the species composition remains approximately as above, then N_e can be calculated from the harmonic mean, $1/N_e = (1/t)(1/N_1 + \ldots + 1/N_t)$, of the population sizes (Hartl and Clark 1997). This is only approximate and is subject to error because the harmonic mean is especially sensitive to the lower population sizes during the dry season, which are currently not well known (Taylor *et al.* 1993). The values of N_e calculated in this way were 922 for Bamako, 1523 for Savanna, 1922 for Mopti, and 3079 for *An. arabiensis*.

Of 1836 individuals captured in neighboring villages, 1 was marked, again from Donéguébougou. This is compared with captures within the village where the releases were made, of which 89 marked and 6416 unmarked were captured. The relative recapture rate of nearby villages to those from the release village gives an intervillage movement rate of $m = 0.039$. [Costantini *et al.* (1996) also recorded considerable intervillage movement in Burkina Faso; $m = 0.08$ and $m = 0.24$ of released mosquitoes were estimated to move to neighboring villages. These values are higher than at Banambani, but the villages in Burkina Faso were closer to one another than were the villages around Banambani, and so are consistent with what we observed.] Assuming that the intervillage rate of movement is approximately the same for each form near Banambani, then we obtain estimates of $N_e m = 35.9$, 59.4, and 75.0 for the Bamako, Savanna, and Mopti forms, respectively (Table 4).

Gene flow across species boundaries can be estimated by the number of hybrids identified by chromosome studies. *An. arabiensis* and *An. gambiae s.s.* are fixed for different arrangements on the X chromosome (Coluzzi *et al.* 1979) so identification of hybrids is unambiguous. Touré *et al.* (1998a) observed three hybrids in their survey of 17,705 polytene chromosomes throughout Mali, the remainder clearly belonging to one species or the other. In this case then, $m = 0.00017/2$. (The factor of 2 enters in because the hybrids observed might have come from crosses in either direction.) If we can assume that this occurs equally among the various chromosomal forms and that the approximate value of N_e is the average value of the two populations involved, then we obtain $N_e m = 0.17$ for Bamako/*arabiensis*, 0.19 for Savanna/*arabiensis*, and 0.21 for Mopti/*arabiensis* (Table 4).

Gene flow across form boundaries is more ambiguous than across species boundaries because the forms share some gene arrangements in common (*e.g.*, the + arrangement is shared by both the Mopti and Savanna forms), because of ambiguity as to the separate gene arrangements that might give some karyotype configurations and because some apparently hybrid gene arrangements can be generated by recombination within

TABLE 4

Summary of gene flow estimates among *An. gambiae s.l.*

	Locations		*gambiae s.s.*		Species	
Method	Form[a]	N_em	Forms	N_em	Species[b]	N_em
Indirect						
	B	115.0	B × S	3.2	B × A	1.1
	S	7.1	S × M	8.0	S × A	1.1
	M	43.9	B × M	3.9	M × A	1.3
Direct						
	B	35.9	B × S	28.4	B × A	0.17
	S	59.4	S × M	17.0	S × A	0.19
	M	75.0	B × M	0.0	M × A	0.21
PCR-based			S × M	9.6		

[a] B, M, and S refer to the Bamako, Mopti, and Savanna chromosomal forms, respectively, of *An. gambiae s.s.*
[b] A, *An. arabiensis.*

forms (*e.g.*, through recombination a bc/+ female of Mopti might give the b arrangement, thought to be characteristic of Savanna). These complications recognized, the long-term averages of the collections at Banambani summarized in Table 3 give 2.4% of *An. gambiae s.s.* to be Bamako/Savanna hybrids and 1.4% to be Savanna/Mopti hybrids. Bamako/Mopti hybrids seem to be nearly absent. Assuming that introgression is equal in both directions, then m should equal half this value, divided by the overall proportions of the hybriding forms, giving $N_em = 28.4$ for the Bamako/Savanna crosses and 17.0 for Savanna/Mopti (Table 4).

Because identification of forms by karyotype is limited to half-gravid females and even then admits some ambiguity, it would be desirable to have good molecular markers that are able to distinguish the forms. It is for this reason that the PCR-based method of distinguishing between *An. gambiae s.s.* and *An. arabiensis* developed by Scott *et al.* (1993) has proven so valuable. Favia *et al.* (1997) described a similar PCR-based method that is able to distinguish the Mopti form from the Bamako and Savanna forms, though not yet the Bamako and Savanna from one another. The hybrids between laboratory colonies of Mopti with Bamako/Savanna are also readily identified. We used this probe on collections of larvae from breeding sites in 1997, 1998, and 1999. These included puddles, swampy areas, and rock pools, as illustrated in Touré *et al.* (1998a). All collections were made during the end of July to mid-August. Multiple dips were taken from the breeding sites, as prescribed by WHO (1975).

The Favia probes of these larvae indicated 152 Bamako/Savanna, 60 Mopti, and 1 hybrid in 1997, 164 Bamako/Savanna, 18 Mopti, and 0 hybrids in 1998, and 273 Bamako/Savanna, 34 Mopti, and 4 hybrids in 1999. Taken together the frequency of Mopti with Bamako/Savanna hybrids was 0.71%. The meaning ascribed to this number is sensitive to the relative proportions of the Bamako and Savanna forms. If we assume that all the hybrids were Mopti with Savanna; *i.e.*, there were no Bamako/Mopti in accord with their absence in Table 3, then the numbers of Bamako + Savanna, Mopti, and hybrids were approximately the same as those identified chromosomally ($\chi^2 = 4.04$, 1 d.f., $P = 0.044$). Further, assuming the proportions of the various forms in the larval sample were similar to the long-term weighted average, then we obtain $N_em = 8.4$, between the corresponding values for indirect measures ($N_em = 8.0$) and for the direct measures ($N_em = 17.0$). On the basis of these results there is little reason to believe that the PCR-based method of Favia *et al.* (1997) indicates a different pattern of population structure at Banambani than do the chromosome arrangements or microsatellite DNA.

DISCUSSION

Favia *et al.* (1997) observed that crosses between laboratory strains of Mopti and Bamako or Savanna forms gave clearly identifiable F_1 banding patterns. But no such hybrid patterns were observed in their samples from nature, prompting them to infer that such hybrids do not occur. In our survey we observed a frequency of hybrids of 0.86%. These findings are not necessarily in conflict; expected numbers cannot be calculated from their study because their typed individuals were not selected at random, nor were the frequencies of all three forms reported, and the expectations would depend on that. Further, we observed large year-to-year fluctuations in the numbers of hybrids identified using the PCR-based method. This emphasizes the differences between direct and indirect measurements of gene flow, and illustrates that what is found in one year or location might well be different from another. At the same time, the sample of Favia *et al.* (1997) contained a number of individuals that would have been classified as putative hybrids or recombinants by chromosomal means, and

none of these was classified as hybrid by the PCR-based method. Further study is clearly warranted.

It should be recognized that the estimates of gene flow in Table 4 are all approximate, if for no reason other than that they rely on assumptions that may not be true. The indirect estimates could be affected by several sources of error. Mutation rates for microsatellite DNA loci are sometimes sufficiently high and bounded to sufficiently narrow ranges that N_em is not well estimated from F_{ST} or R_{ST} (Nauta and Weissing 1996). While this cannot be ruled out entirely, there is no evidence that mutation rates for Anopheles microsatellite DNA loci are higher than those of *Drosophila melanogaster* (L. Zheng, personal communication), which are small enough that this is unlikely to substantially affect such estimates (Schug *et al.* 1998). More important is the assumption that populations are at equilibrium for drift and gene flow for N_em to be estimated from F_{ST} in the way used here. Coluzzi (1999) and Powell *et al.* (1999) make a strong argument that there have been recent and dramatic changes in the structure of these populations, due to human activities making new habitats available and selection on the mosquitoes to occupy them. The importance of this is difficult to evaluate, underlining the importance of using independent estimates of gene flow. For direct measures we assumed that population sizes contract during the dry season to $\frac{1}{10}$ their rainy season high and then expand again from local individuals when the rains resume. The best available evidence indicates this is appropriate (Taylor *et al.* 1993), but it is far from firmly established. So while the patterns of gene flow do seem consistent and do appear to be robust, they also remain just approximations.

The most glaring discrepancy in Table 4 is the absence of B × M hybrids in direct measures compared to an expected N_em of 3.9 from the indirect methods. It should first be recognized that B × M may still occur, as one such probable individual was reported from nearby Bancumana, reported in Touré *et al.* (1998a), though it is certainly quite rare. It may also occur that there is some gene flow between these forms, with the Savanna acting as an intermediary, though this too should be only a trickle if it occurs in Banambani, based on our observations. But other locations, especially where the forest form is present, might well be places where genes flow across the form boundaries, then pass to Banambani by the fairly large rates of gene flow that occur within forms. More likely, however, the pattern of crossing observed today might be very different from what occurred in the past, and the degree of separation measured by F_{ST} does not reflect the current situation. In the absence of good information about mutation rates it is difficult to calculate how long it will take for F_{ST} to reach equilibrium, but surely the changes in human populations during the past 200 years are sufficiently great to cast doubt on any assumption of equilibrium.

It would appear from Figure 1 that the Savanna form might be a conduit for gene flow between the Bamako and Mopti forms. The indirect estimates in Table 1 show gene flow equal or greater between Bamako and Mopti as between Bamako and Savanna, indicating that Savanna is not simply an intermediate. It is apparent from these measurements that Savanna is much more similar to the Mopti than the Bamako form, albeit it is distinct in its own right. The greater similarity to Mopti takes on special interest in view of the finding by Favia *et al.* (1997) that PCR primers associated with X-linked ribosomal DNA place Savanna closer to the Bamako, rather than Mopti, chromosomal form. This underscores that gene flow varies from chromosome to chromosome and shows that if the Savanna form is in some way a transitional one, then it is still subject to its own independent selection and drift.

It is well established that rates of gene flow with $N_em < 1$ are required for the dispersive effects of genetic drift to outweigh the homogenizing effects of gene flow. How, then, could the rates of gene flow that we observed among chromosomal forms ($3 < N_em < 30$) be consistent with their continued coexistence? The most obvious explanation is that the chromosomal forms are in a transient phase, perhaps moving toward complete reproductive isolation, perhaps just drifting in a long-term disequilibrium. Recent studies on sympatric speciation (Dieckmann and Doebeli 1999; Higashi *et al.* 1999; Kondrashov and Kondrashov 1999; Tregenza and Butlin 1999) clearly show that disruptive selection and sexual selection, both of which are likely to occur in these populations, may lead to eventual sympatric speciation. This process can take a very long time, and if the forces of selection are subject to change, a virtual "permanent transience" can result. Clearly this invites further investigation.

In summary, the picture of population structure that emerges from both the direct and indirect studies of gene flow is one of local populations concentrated within human villages. According to this view there would be four populations of *An. gambiae s.l.* in each village—one each of *An. arabiensis* and the Bamako, Savanna, and Mopti forms of *An. gambiae s.s.* The populations of each chromosomal form exchange a few migrants with one another and with like populations from neighboring villages. The sizes of these populations vary through the year, peaking during the wet season of June–September, at ~15,000 individuals per form/species. The numbers during the dry season are much less, so that the effective population size per form/species is only in the range of N_e = 1000–3000 individuals. The amount of gene flow from village to village depends on how close the villages are to one another; in this study they were ~5–7 km apart, so that N_em equaled 30–70 individuals per generation. This would result from "active dispersal," as opposed to "passive transport" (Dobzhansky 1973), with the total rates of gene flow being

much greater and not confined to immediate neighbors. Based on the 21 microsatellite loci used in this study, the total rate of gene flow among nearby locations is estimated to be $N_em > 40$ individuals per generation.

It appears that the amount of gene flow across the species boundaries is low, $N_em \approx 0.15$–1.3, consistent with independent genetic drift within the two species. And while it seems that the chromosomal forms that together comprise *An. gambiae s.s.* exchange genes at lower rates than if they were in neighboring villages even several kilometers apart, they do so at significantly higher rates than if they were distinct species. The rates of gene exchange among forms depend on chromosome and on which forms are considered, but appear to be in the range of $N_em = 3$–30. It is hard to imagine what could be keeping the forms distinct in the face of so much gene flow, unless the chromosomal forms of *An. gambiae s.s.* represent populations that are in transient stages of sympatric speciation. In view of the importance this species has for malaria control, it is evident that the phenomenon of chromosomal forms merits further study, and that gene flow among the forms, or the lack of it, is likely to play a very significant role in any attempts to reduce malaria by genetic control of the vector populations.

We are grateful to Bakary Sissoko, Fah Niaré, and the village of Banambani, who have been very helpful to us for undertaking the MRR experiments and for the regular collections of mosquitoes for microsatellite studies. The studies benefited from the technical support of Dr. Richard Sakai, Dr. Robert W. Gwadz, and the Laboratory of Parasitic Diseases at the National Institutes of Health. This research was supported by grants from the National Institutes of Health, from the John Sealy Memorial Endowment Fund, and from a World Health Organization partnership grant.

LITERATURE CITED

Bryan, J. H., J. H. Di Deco, V. Petrarca and M. Coluzzi, 1982 Inversion polymorphism and incipient speciation in *Anopheles gambiae s. str.* in The Gambia, West Africa. Genetica **59:** 167–176.

Collins, F. H., 1994 Prospects for malaria control through the genetic manipulation of its vectors. Parasitol. Today **10:** 370–371.

Colson, I., and D. B. Goldstein, 1999 Evidence for complex mutations at microsatellite loci in Drosophila. Genetics **152:** 617–627.

Coluzzi, M., 1999 The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of *Plasmodium falciparum.* Parassitologia **41:** 277–283.

Coluzzi, M., A. Sabatini, V. Petrarca and M. A. Di Deco, 1979 Chromosomal differentiation and adaptation to human environments 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.

Costantini, C., S.-G. Li, A. della Torre, N'F. Sagnon, M. Coluzzi *et al.*, 1996 Density, survival and dispersal of *Anopheles gambiae* complex mosquitoes in a West African Sudan Savanna village. Med. Vet. Entomol. **10:** 203–219.

della Torre, A., L. Merzagora, J. R. Powell and M. Coluzzi, 1997 Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. Genetics **146:** 239–244.

Di Deco, M. A., V. Petrarca, F. Villani and M. Coluzzi, 1980 Polimorfismo cromosomico d inversioni paracentriche ed eccesso degli eterocariotipi in ceppi di *Anopheles* allevati in laboratorio. Parassitologia **22:** 304–307.

Dieckmann, U., and M. Doebeli, 1999 On the origin of species by sympatric speciation. Nature **400:** 354–357.

Dobzhansky, Th., 1973 Active dispersal and passive transport in *Drosophila.* Evolution **27:** 565–575.

Edillo, F. E., G. C. Lanzaro, Y. T. Touré, G. Dolo and C. E. Taylor, 1999 Do *Anopheles gambiae s.s.* and *Anopheles arabiensis* have a habitat choice in Banambani Village, Mali, West Africa? Am. J. Trop. Med. Hyg. **61:** 365–366.

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.

Hartl, D. L., and A. G. Clark, 1997 *Principles of Population Genetics,* Ed. 3. Sinauer Associates, Sunderland, MA.

Higashi, M., G. Takimoto and N. Yamamura, 1999 Sympatric speciation by sexual selection. Nature **402:** 1523–1525.

Hunt, R. H., 1973 A cytological technique for the study of *Anopheles gambiae.* Parassitologia **15:** 137–139.

Kamau, L., T. Lehmann, W. A. Hawley, A. S. Orago and F. H. Collins, 1998 Microgeographic genetic differentiation of *Anopheles gambiae* mosquitoes from Asembo Bay, Western Kenya: a comparison with Kilifi in coastal Kenya. Am. J. Trop. Med. Hyg. **58:** 64–69.

Kondrashov, A. S., and F. A. Kondrashov, 1999 Interactions among quantitative traits in the course of sympatric evolution. Nature **400:** 351–354.

Lanfrancotti, A., A. della Torre and G. Favia, 1998 Improvement of a PCR-based assay for the identification of sympatric forms of *Anopheles gambiae s.s.* Parassitologia **40** (Suppl. 1): 87.

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.

Nauta, M. J., and F. J. Weissing, 1996 Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics **143:** 1021–1032.

Persiani, A., M. A. Di Deco and G. Petrangeli, 1986 Osservazioni di laboratorio su polimorfismi da inversione originata da incroci tra popolazioni diverse de *Anopheles gambiae s.s.* Ann. Ist. Super. Sanita **22:** 221–224.

Petrarca, V., and J. C. Beier, 1992 Intraspecific chromosomal polymorphism in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. Am. J. Trop. Med. Hyg. **46:** 229–237.

Post, R. J., P. K. Flook and A. L. Millest, 1993 Methods for the preservation of insects for DNA studies. Biochem. Syst. Ecol. **21:** 85–92.

Powell, J. R., V. Petrarca, A. della Torre, A. Caccone and M. Coluzzi, 1999 Population structure, speciation and introgression in the *Anopheles gambiae* complex. Parassitologia **41:** 101–114.

Schneider, S., J.-M. Kueffer, D. Roessli and L. Excoffier, 1996 *ARLEQUIN, A Software Package for Population Genetics.* Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva.

Schug, M. D., C. M. Hutter, M. A. Noor and C. F. Aquadro, 1998 Mutation and evolution of microsatellites in *Drosophila melanogaster.* Genetica **102–103:** 359–367.

Scott, J. A., W. G. Brogdon and F. H. Collins, 1993 Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. Am. J. Trop. Med. Hyg. **49:** 520–529.

Service, M. W., 1993 *Mosquito Ecology: Field Sampling Methods,* Ed. 2. Elsevier, London.

Simard, F., D. Fontenille, T. Lehmann, R. Girod, L. Brutus *et al.*, 1999 High amounts of genetic differentiation between populations of the malaria vector *Anopheles arabiensis* from West Africa and eastern outer islands. Am. J. Trop. Med. Hyg. **60:** 1000–1009.

Slatkin, M., 1985 Gene flow in natural populations. Annu. Rev. Ecol. Syst. **16:** 393–430.

Slatkin, M., 1995 A measure of population subdivision based on microsatellite allele frequencies. Genetics **139:** 457–462.

Taylor, C. E., and J. R. Powell, 1983 Population structure of Drosophila: genetics and ecology, pp. 29–60 in *The Genetics and Biology of Drosophila,* Vol. 3, edited by M. Ashburner, H. L. Carson and J. R. Thompson. Academic Press, London.

Taylor, C. E., Y. T. Touré, M. Coluzzi and V. Petrarca, 1993 Ef-

fective population size and persistence of *Anopheles arabiensis* during the dry season of West Africa. Med. Vet. Entomol. **7:** 351–357.

Touré, Y. T., V. Petrarca, S. F. Traoré, A. Coulibaly, H. M. Maïga *et al.*, 1994 Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae s. str.* in Mali, West Africa. Genetica **94:** 213–223.

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

Touré, Y. T., G. Dolo, V. Petrarca, S. F. Traoré, M. Bouaré *et al.*, 1998b Mark-release-recapture experiments with *Anopheles gambiae s.l.* in Banambani Village, Mali, to determine population size and structure. Med. Vet. Entomol. **12:** 74–83.

Tregenza, T., and R. K. Butlin, 1999 Speciation without isolation. Nature **400:** 311–312.

WHO, 1975 *Manual on Practical Entomology in Malaria.* World Health Organization, Geneva.

Communicating editor: W. F. Eanes