Genetic Differentiation of *Anopheles gambiae* s.s. Populations in Mali, West Africa, Using Microsatellite Loci


*Anopheles gambiae* sensu stricto is a principal vector of malaria through much of sub-Saharan Africa, where this disease is a major cause of morbidity and mortality in human populations. Accordingly, population sizes and gene flow in this species have received special attention, as these parameters are important in attempts to control malaria by impacting its mosquito vector. Past measures of genetic differentiation have sometimes yielded conflicting results, in some cases suggesting that gene flow is extensive over vast distances (6000 km) and is disrupted only by major geological disturbances and/or barriers. Using microsatellite DNA loci from populations in Mali, West Africa, we measured genetic differentiation over uniform habitats favorable to the species across distances ranging from 62 to 536 km. Gene flow was strongly correlated with distance ($r^2 = 0.77$), with no major differences among chromosomes. We conclude that in this part of Africa, at least, genetic differentiation for microsatellite DNA loci is consistent with traditional models of isolation by distance.

*Anopheles gambiae* sensu stricto, a principal vector of malaria, is widely distributed in Africa south of the Sahara Desert to northern South Africa. Throughout the region it is highly polymorphic for gene arrangements on the second chromosome (Bryan et al. 1982; Coluzzi et al. 1979). The genetic structure of this species is complex, with several recognized “chromosomal forms” that differ ecologically. These chromosomal forms are of uncertain taxonomic status; for the purposes of this article we will simply treat them as distinct units that potentially interbreed within themselves. In Mali, the site of our study, three chromosomal forms are widely distributed: Bamako, Savanna, and Mopti. These forms commonly occur in sympatry and their relative abundance varies greatly from location to location and season to season. Correlation studies have established that the composition by chromosomal form varies with climate (Touré et al. 1998b), the $2R^b$ arrangement is closely tied to the amount of vegetation (Thomson et al. 1997), and the canonical correlation between all gene arrangement frequencies and climate, especially dry-cold season temperature and precipitation, is very high ($r > 0.88$) (Massumi et al. 2000). Geographic variation for other genetic markers has been less clear. In particular, for mitochondrial DNA (mtDNA) Besansky et al. (1997) found little or no population subdivision across considerable distances in East Africa and reported that there was no evidence for isolation by distance there. Likewise, Lehmann et al. (1996, 1997) found no differences in microsatellite DNA frequencies between populations of the Savanna form on the east and west coasts of Africa, thousands of kilometers apart. In that study, the one barrier to gene flow they did find separated populations on opposite sides of the Rift Valley complex in East Africa, a pattern also observed using mtDNA markers (Lehmann et al. 2000). At the same time Lanzaro et al. (1998) reported that populations in Mali separated by uniformly favorable habitat showed differences at microsatellite DNA loci comparable to those reported by Lehmann et al. (2000) for the Rift Valley.

Extensive karyotype data from Mali, West Africa, has shown populations to be geographically structured for frequencies of certain chromosomal arrangements (Touré et al. 1998b). While these data do not directly show populations of the same form to be isolated by distance, they do support the notion that such arrangements vary both seasonally and geographically. Unfortunately these chromosomal inversions are known to be nonneutral
Figure 1. Map of Mali, showing locations of sample locations and land cover vegetation. Land cover vegetation was determined by Hansen et al. (2000) from a variety of satellite imaging methods to a resolution of 1 km, which is less than the area over which An. gambiae is expected to move during its lifetime. The species is found more or less continuously throughout the vegetated area included in this study, though not over the bare ground of the Sahara Desert, to the north of Tombouctou.

and therefore may not be adequate markers to test whether gene flow is restricted across geographic distance.

The purpose of this study was to inquire further into the question of geographic population genetic structure for microsatellite DNA loci. We looked at populations of An. gambiae s.s. in Mali, West Africa, separated by a range of distances of continuously suitable habitat, and examined whether these distances were correlated with differences in microsatellite allele frequencies.

Methods

An. arabiensis and An. gambiae female mosquitoes were collected from human habitations in six villages in Mali, West Africa: Banambani (12°48'N, 08°03’W), Mopti (14°31’N, 04°16’W), Pimperena (11°28’N, 05°42’W), Selinkenyi (11°42’N, 08°17’W), Soulouba (11°47’N, 07°43’W), and Tene (13°24’N, 04°35’W). Their locations are shown in Figure 1. The six villages are located along the Niger River valley and the geographic distance between them ranges from 62 to 536 km. Descriptions of the plant communities of this region may be found in White (1983) and Touré et al. (1998b). The vegetation throughout this region is quite uniform, as shown by the land cover map derived from remotely sensed data by the University of Maryland Global Land Cover Facility (Figure 1). It is based on NASA/NOAA Pathfinder land datasets (AVHRR sensor) with a spatial resolution of 1 km, collected from 1981 to 1994. Land cover was inferred using a supervised classification procedure from the Landsat Multispectral Scanner System, Landsat Thematic Mapper, and Linear Imaging Self-Scanning Sensors (Hansen et al. 2000). The area included in our samples was almost continuously covered by grassland, shrubland, or woodland. These are all known to support high densities of An. gambiae (Touré et al. 1998b). The dispersal distance of this species is several hundred meters to several kilometers per day (Touré et al. 1998a), so any significant barriers to dispersal should be recognizable from the 1 km resolution of this figure. Because no such barriers are evident from the figure, we infer that there were no significant impediments to dispersal in these populations.

Female mosquitoes were captured daily in the morning using mouth aspirators. They were temporarily stored in small, netted cardboard containers that were held in the shade and kept humidified until they were transported to the laboratory at the Malaria Research and Training Center in Bamako.

An. gambiae s.s. populations at Tene and Mopti are known to contain only the Mopti form (Touré et al. 1998b), whereas populations at the other sites are typically mixed, and therefore each specimen collected from these sites required cytological examination to determine the form to which it belonged. Chromosome preparations were made by extracting ovaries from the abdomen of each sample, follow-
ing established protocols (Coluzzi 1968; Hunt 1973). Individuals from Mopti and Tene were identified to species by diagnostic polymerase chain reaction (PCR) (Scott et al. 1993). For the remaining locations, species identification and karyotype scoring was done using a phase-contrast microscope and classified according to Coluzzi et al. (1979). The rare individuals that were classified as hybrids by karyotype were excluded from this analysis. Following cytological examination, the carcass of each specimen was placed in a correspondingly numbered 1.5 ml tube containing silica gel for DNA extraction. Individual female mosquitoes were removed from the tubes in which they had been stored and DNA was extracted from the carcass using a standard extraction protocol (Post et al. 1993). The microsatellite primer sequences and PCR conditions employed were as described by Zheng et al. (2000). Pairwise genetic differences between populations were determined from weighted estimates of $F_{ST}$ and $R_{ST}$ over all loci, using the number of different alleles or the sum of squared size differences, respectively. As expected, $F_{ST}$ had smaller confidence limits than $R_{ST}$, so we used the former as the measure of differentiation. Values of $N_{m}$ derived from $F_{ST}$ were used to correlate with distance, where $N_{m}$ is the effective population size and $m$ is the fraction of the population replaced by immigrants. In some models of gene flow, $N_{m}$ can be interpreted as an unbiased estimate of the number of individuals moving between populations, and when correlated with geographic distance, can be used to test isolation by distance among populations. The confidence limits of $F_{ST}$ can be used to predict gene flow. So long as the populations are more or less in equilibrium and that mutation rate is negligible compared to migration, thus calculating the number of migrants ($N_{m}$) from estimates of $F_{ST}$ using the relation $F_{ST} = 1/(N_{m} + 1)$ (Crow and Kimura 1970).

Strict applicability of this estimate of $N_{m}$ depends on many assumptions, especially that the populations are at equilibrium with respect to drift and migration. Additional assumptions or shortcomings have been discussed by Goldstein and Pollack (1997), Whitlock and McCauley (1999), and in this particular context by Taylor et al. (2001). While this estimate for $N_{m}$ derived from $F_{ST}$ may not be strictly interpretable as the number of migrants, there is a monotonic relation between the two measures, so it still offers a heuristically useful way to compare levels of gene flow. So long as the assumptions are borne in mind, $N_{m}$ derived from $F_{ST}$ should still be appropriate for relative comparison of genetic differentiation as it is employed in this study.
Results

The loci in this study were highly polymorphic, with an average of 15.5 alleles per locus and an average gene diversity of 0.74. Significance tests of departures from Hardy–Weinberg expectations were performed for each locus using the method described by Guo and Thomson (1992). This method is equivalent to Fisher’s exact test using a contingency table of arbitrary size and computing the probability of observing a table with the same marginal counts under the null hypothesis of no association. There were few significant deviations \((P > .001)\) from Hardy–Weinberg expectations at each of the loci: 5 of the 13 populations sampled had no significant deviations at any of the loci, and all but 1 of the remaining 8 populations had no more than two loci that deviated. We attribute this to the presence of null alleles, as described in Lanzaro et al. (1995, 1998). The one exception, Pimperena-Savanna, had a total of nine loci that showed a significant departure from Hardy–Weinberg. The reason for this exception is unclear.

The majority of the loci away from Hardy–Weinberg equilibrium were located on chromosome 2, which is known to be under strong selection (Touré et al. 1998b), but some loci on both chromosomes 3 and X were also affected. Some of the populations, such as Soulouba-Bamako, had small sample sizes. They were included in this study only if they were offset by a large number of loci sampled (more than 20). Conversely, samples with small numbers of usable loci were used only if they were offset by a larger number of individuals.

The pairwise estimates of \(F_{ST}\) and \(N_m\) are shown with the geographic distance between each location in Table 1. Estimates of \(\log(N_m)\) are plotted against geographic distance for all loci together and for the loci on each chromosome separately (Figure 2). The correlation between genetic and geographic distance was determined using least-squares regression. The significance test here is one-tailed because we do not expect a positive correlation between \(N_m\) and geographic distance. For all loci, there was a significant \((P < .005)\) inverse correlation with \(r^2 = 0.77\) (Figure 2A). Parallel analyses of loci from each of the three individual chromosomes gave similar results. Correlations on both the third \((r^2 = 0.55)\) and X \((r^2 = 0.68)\) chromosomes were statistically significant \((P < .05)\); correlations on the second chromosome were not statistically significant \((P > .2)\), but were of similar magnitude to the others.

Discussion

These results demonstrate that different levels of genetic differentiation exist between populations of \textit{An. gambiae} in West Africa. A strong correlation exists between genetic and geographic distance. This correlation is significant for all microsatellite loci across chromosomes and individually on two out of three chromosomes. Nonsignificant correlation for loci on the second chromosome may be due to the small number of comparisons or because of selection for inversions, documented by Touré et al. (1998b). All else being equal, selection would be confounded with the effects of isolation by distance, giving an overall lower correlation between differentiation and distance. Multilocus values of \(N_m\) ranged between 1.21 and 64.43, with the number of migrants between populations closely associated with their distance apart.

Recent mark-release-recapture (MRR) experiments have been used to estimate gene flow among populations in Mali by determining the amount of intervillage movement (Taylor et al. 2001). For the period from 1993 to 1998, movement between Banambani and its nearest neighboring village, Donégoubougou, was 0.039. The distance between the two neighboring villages is about 7 km. If we assume that the intervillage rate is analogous to the interpopulation rate, then the estimates for \(N_m\) were 35.9, 59.4, 75.0 for the Bamako, Savanna, and Mopti forms, respectively, quite in line with what was observed in this study.

These results contrast with several recent studies in both East and West Africa. Lehmann et al. (1996) found only small differences in microsatellite allele frequencies between populations of \textit{An. gambiae} in Kenya and Senegal, a distance of approximately 6000 km. The average value of \(N_m\) was 2.5 using \(R_{st}\) and 7.7 using \(F_{ST}\) for microsatellite loci. These results differ markedly from the results of this study, though the distances in our study were not so large. One possible reason the authors suggest for this low differentiation is range constraints on the evolution of the loci. If the number of repeats is bounded then, according to the stepwise mutation model, estimates of genetic differentiation using \(F_{ST}\) and \(R_{st}\) in this way may be problematic for populations separated for long periods of time (Feldman et al. 1997; Goldstein and Pollack 1997; Nauta and Weissing 1996). So one possible explanation for the discrepancy is that the populations in Mali have not been separated by a large enough distance or long enough time for the range constraints to have an effect.

In East Africa, only the Rift Valley complex appears to be a barrier to gene flow.
for both microsatellite and mtDNA loci. Lehmann et al. (1997, 2000) found differences across the Rift Valley complex, while differences among populations on either side were much lower. Microsatellite values for $F_{ST}$ and $R_{ST}$ across the Rift Valley complex were 0.104 and 0.032, respectively, with an average distance between sites of approximately 700 km. On either side of the Rift Valley complex, both values of $F_{ST}$ and $R_{ST}$ were less than 0.01 with geographic distances up to 220 km. In our study, similar values (0.167, 0.140, and 0.034) of $F_{ST}$ to those found across the Rift Valley complex were found for populations separated by 200–500 km with no obvious barriers. Our results do not discount the notion that barriers exist. However, given that we observed such a strong correlation between geographic and genetic distance across uniformly favorable habitat, it is evident that barriers are not the only cause for a reduction in gene flow in this species. The situation in Mali, West Africa, is perhaps markedly different than that in East Africa.

References


Received November 11, 2000
Accepted November 26, 2001

Corresponding Editor: Ross MacIntyre