Heredity 77 (1996) 192–208 Received 8 November 1995

Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci

TOVI LEHMANN*†, WILLIAM A. HAWLEY†‡, LUNA KAMAU‡, DIDIER FONTENILLE§, FREDERIC SIMARD§ & FRANK H. COLLINS†¶

†*Division of Parasitic Diseases, Centers for Disease Control and Prevention, MS F22, 4770 Buford Highway, Chamblee, GA 30341, U.S.A.,* ‡*Kenya Medical Research Institute, Clinical Research Centre, Nairobi, Kenya,* §*Laboratoire ORSTOM de Zoologie Medicale, Institut Pasteur, BP 220, Dakar, Senegal and* ¶*Department of Biology, Emory University, Atlanta, GA 30322, U.S.A.*

Genetic variation of *Anopheles gambiae* was analysed to assess interpopulation divergence over a 6000 km distance using short tandem repeat (microsatellite) loci and allozyme loci. Differentiation of populations from Kenya and Senegal measured by allele length variation at five microsatellite loci was compared with estimates calculated from published data on six allozyme loci (Miles, 1978). The average Wright's F_{ST} of microsatellite loci (0.016) was lower than that of allozymes (0.036). Slatkin's R_{ST} values for microsatellite loci were generally higher than their F_{ST} values, but the average R_{ST} value was virtually identical (0.036) to the average allozyme F_{ST}. These low estimates of differentiation correspond to an effective migration index (Nm) larger than 3, suggesting that gene flow across the continent is only weakly restricted. Polymorphism of microsatellite loci was significantly higher than that of allozymes, probably because the former experience considerably higher mutation rates. That microsatellite loci did not measure greater interpopulation divergence than allozyme loci suggested constraints on microsatellite evolution. Alternatively, extensive mosquito dispersal, aided by human transportation during the last century, better explains the low differentiation and the similarity of estimates derived from both types of genetic markers.

Keywords: allozymes, *Anopheles gambiae,* gene flow, microsatellites, population genetic structure, population genetics.

Introduction

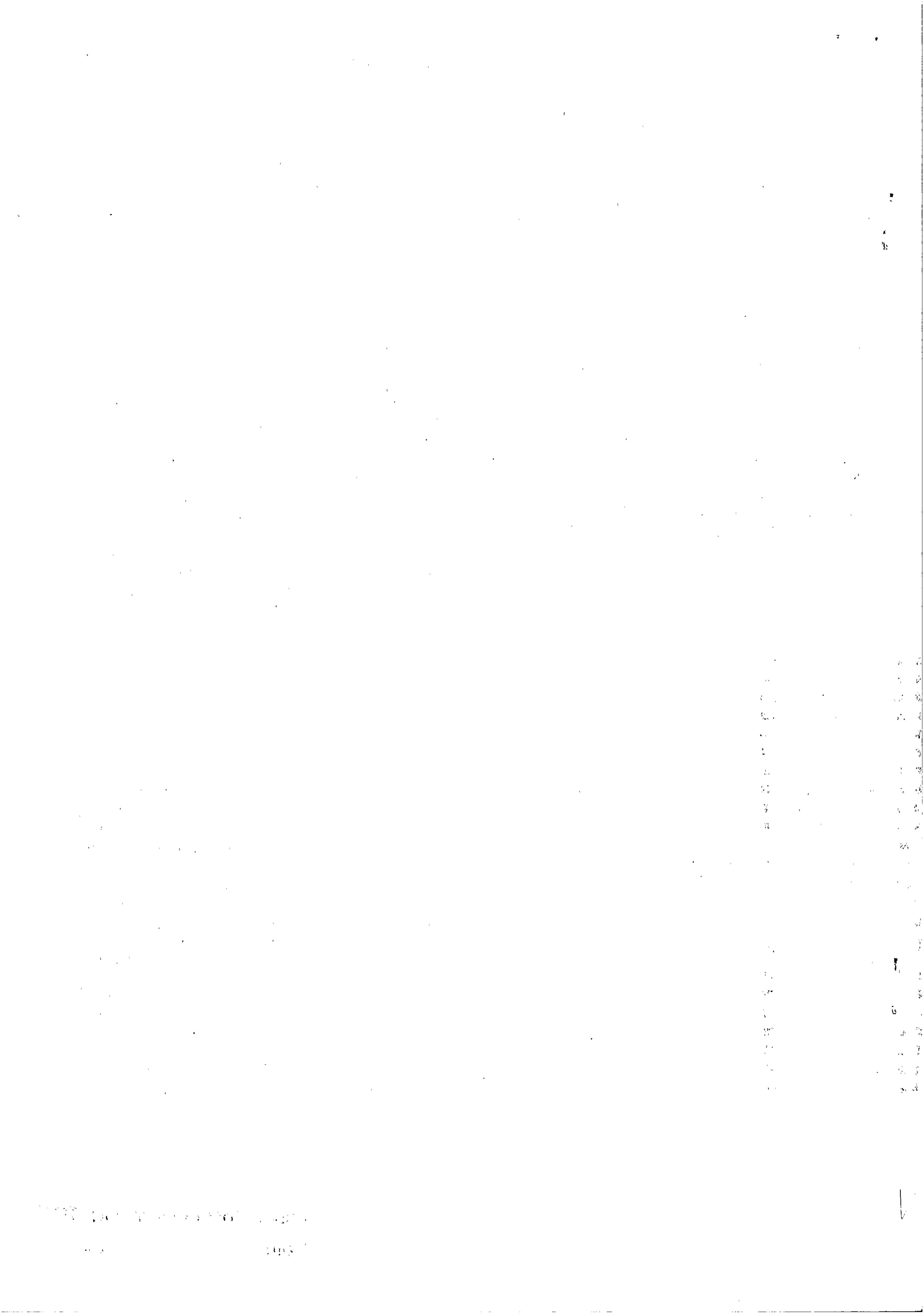
In sub-Saharan Africa *Anopheles gambiae* is the principal vector of human malaria, a disease which continues to inflict immense misery despite substantial efforts to bring it under control. An understanding of the genetic structure of *A. gambiae* populations is critical in evaluating the possibility of genetic manipulation of this species to block malaria transmission (e.g. Collins & Besansky, 1994; Crampton *et al.,* 1994). Moreover, such understanding could also aid control based on currently available technology, such as in the management of insecticide resistance.

*Correspondence.

Microsatellite loci have been described as powerful markers for measuring intraspecies differentiation because of their high polymorphism, codominance, abundance throughout the genome, and relative ease of scoring (Bowcock *et al.*, 1994; Estoup *et al.*, 1995). A microsatellite survey in *A. gambiae* (Lanzaro *et al.*, 1995) demonstrated the above features and concluded that microsatellite loci are superior to allozymes for studies of population structure. Currently several groups are using microsatellite loci to assess gene flow and related phenomena in the *A. gambiae* complex of species. The forces that shape allele composition at such loci are poorly understood, however (Edwards *et al.*, 1992; Di Rienzo *et al.*, 1994; FitzSimmons *et al.*, 1995; Garza *et al.*, 1995), and some evidence suggests that these forces include biased mutation rates (Garza *et*

©1996 The Genetical Society of Great Britain.

Fonds Documentaire ORSTOM
Cote : B✻15465 Ex : 1



al., 1995) and/or selection acting on allele size (Epplen *et al.*, 1993). If such forces strongly influence allele composition at these loci, estimates of differentiation between populations and rates of gene flow will be misleading, because they are derived on the basis of genetic drift, migration, and random mutation as the main forces in operation.

Using five polymorphic microsatellite loci, we measured differentiation between populations of *A. gambiae* from Kenya and Senegal that represent two geographical extremes (6000 km) in the range of this species. Assuming that gene flow is restricted by distance, genetic differentiation between these populations would be expected to be near the maximum possible for the species. Additionally, we compared differentiation based on microsatellite loci, using both Wright's F_{ST} and Slatkin's R_{ST}, to differentiation based on the allozyme data (Miles, 1978) using Wright's F_{ST} derived from populations in Kenya and The Gambia. The rationale for this comparison was to evaluate the possibility that constraints on microsatellite loci such as biased mutation rates or selection on allele size could yield lower estimates of population differentiation than those based on allozyme data.

Materials and methods

Study sites

The Asembo Bay area in western Kenya is located on the northern shores of Lake Victoria. It is a relatively flat, densely populated landscape, traversed by semipermanent streams. During the major rainy season (April to July), many mosquito breeding sites are available. Mosquito populations are much reduced during the dry season, when larval breeding sites are scarce. The village of Barkedji in north Senegal is located in the Sahelian region, and *A. gambiae* can be found only during the June to December rainy season. In both sites, mosquitoes were collected from houses less than 2 km apart to minimize the possibility of sampling members of different demes.

Mosquito collection

In Kenya, mosquitoes were aspirated at dawn from bed nets hung the previous evening over the beds of sleeping volunteers. The nets were hung in a manner to leave a space for the mosquitoes to enter. Thus, samples consisted of blood-fed and blood-seeking females. In Senegal, mosquitoes were aspirated after landing on human volunteers or collected from the floor after spraying the interior of houses with pyrethrum insecticide early in the morning. Collections were carried out in Kenya between 28 June and 6 July 1994 and in Senegal on 5–6 October 1994. Only the savanna cytotype of *A. gambiae* (Coluzzi *et al.*, 1985; Fontenille, unpublished data) and *A. arabiensis* of the *A. gambiae* complex, were present in both study sites. Only *A. gambiae* were included in the analysis, however, after species identification (Scott *et al.*, 1993).

DNA extraction and genotype scoring

DNA from individual specimens (or parts of a specimen) was extracted as described by Collins *et al.* (1987) and resuspended in water or TE buffer (Sambrook *et al.*, 1989). Loci *33C1*, *29C1*, *1D1* and *2A1* were identified in cloned *A. gambiae* genes (Table 1). Locus *33C1* is from the *dopa decarboxylase* (*Ddc*) gene (P. Romans, unpublished data), *29C1* is from the *xanthine dehydrogenase* gene (F. Collins, unpublished data), *1D1* is from the *actin1D* gene (Salazar *et al.*, 1993), and *2A1* is from the *white* gene (Besansky *et al.*, 1995). Locus *AG2H46* was isolated from an *A. gambiae* chromosome division-specific library (Zheng *et al.*, 1991) by probing with a labelled GT-repeat oligonucleotide. Microsatellite alleles were PCR amplified and viewed by autoradiography using incorporation of alpha-^{33}P dATP (Amersham) into the PCR product or using a primer end-labelled with gamma-^{33}P dATP (Dupont). Both techniques provided identical results. Standard PCR in 20 μL reaction volume was run in a Perkin-Elmer 9600 thermal cycler. For incorporation of radiolabelled dATP, a mixture containing 0.2 mM each of dGTP, dCTP, dTTP; 0.05 mM of dATP (Perkin-Elmer) and 0.4 μL of alpha-^{33}P dATP at 1000–3000 Ci/mmol; 5 ng/μL of each primer (approx. 15 pmol); 1 × reaction buffer (Boehringer Mannheim); and 0.035 units of Taq polymerase (Boehringer Mannheim) was used. For PCR reactions with one end-labelled primer, all dNTP concentrations were 0.2 mM, 2 pmol radiolabelled primer with 8 pmol of the same primer, which was not radiolabelled, 10 pmol of the complementary primer and the other components unchanged. An equivalent of 1/100 or less of genomic DNA extracted from a whole mosquito was used. PCR conditions were: denaturation at 94°C for 5 min, followed by 30 cycles consisting of 94°C for 25 s, 55°C for 28 s, and 72°C for 30 s. The last elongation step was at 72°C for 5 min. The PCR product was mixed (3:2) with formamide stop solution (Amersham), denatured at 94°C for 5 min before loading

Table 1 Microsatellite loci in *Anopheles gambiae*: cytological location, repeat sequence and primer sequences

Locus	Cytol. Location*	Repeat	Primers†
AG2H46	IIR:7A	GT	CGC CCA TAG ACA ACG AAA GG
			TGT ACA GCT GCA GAA CGA GC
			CAA AGA AAG CGC CCA TAG AC
			CGC TGT GTT TTC GTC TTG TA
33C1	IIIR:33C	AGC	TTG CGC AAC AAA AGC CCA CG
			ATG AAA CAC CAC GCT CTC GG
29C1	IIIR:29C	TGA	ATG TTC CAG AGA CGA CCC AT
			TGT TGC CGG TTT GTT GCT GA
1D1	X:1D	CCA	TAA TGG TCC CAA ATC GTT GC
			GTT ATC CAC TGC GCA TCA TG
2A1	X:2A	10/6 bases	GAA TTC GTT TAG AGT CTT TC
			GTA TAC AGG CCT TTG TTT CC

*The cytological position of each locus was determined by polytene chromosome *in situ* hybridization (Kumar & Collins, 1993).
†For locus *AG2H46* the upper pair refers to the original primers and the lower pair to alternative primers (see Materials and methods).

3 μL of the mixture onto a 6 per cent acrylamide, 7 M urea sequencing gel (Life Technologies) in parallel to a 2-lane-ladder standard, which was loaded every 10–20 lanes. The standard, constructed by sequencing an AT-rich region of *A. gambiae* mtDNA with the dideoxy terminators ddA and ddT (Beard *et al.*, 1993), allowed exact determination of allele size. Autoradiographs of gels, developed after overnight exposure, were visually inspected and allele size was determined. Alleles were distinguished from occasional artefacts by intensity and size. These criteria were tested by scoring the alleles at loci *1D1* and *AG2H46* in laboratory-reared progeny of parent mosquitoes with known genotypes. All scored genotypes were in complete concordance with the genotypes expected from the test crosses.

At locus *AG2H46*, no PCR product was visible in repeated reactions for approx. 5 per cent of the specimens that were scored successfully for other loci. Using alternative primers, which flank the sites of the original primers (Table 1), PCR products were obtained. Subsequent sequencing revealed a mutation in one of the original primer annealing regions. For this locus 2/3 of the homozygotes scored with the original primers were found to be heterozygotes when PCR amplification was carried out with the alternative primers. No 'hidden' heterozygotes were found in the other microsatellite loci using alternative primers.

Allozyme data

To identify allozyme alleles that could be used to identify the different cryptic species in the *A. gambiae* complex, Miles (1978) examined variation at 18 loci from *A. gambiae* complex populations from different parts of Africa. Allele frequency data based on sample sizes larger than six mosquitoes were summarized only for the loci *alpha-naphthyl acetate esterases* (*EST-1, EST-2, EST-3*), *octanol dehydrogenase* (*ODH*) and *phosphoglucomutases* (*PGM-1, PGM-2*). These data for the populations from Kenya (Chulaimbo) and Gambia (Mandinari, having the largest sample size) were used. Additional analyses including samples from East, West, and Central Africa were carried out to evaluate the consistency of these results (see Discussion). We have assumed that the population structure of *A. gambiae* across the continent has not changed significantly between the time of Miles's study and the present.

Data analysis

Goodness of fit tests of genotype distributions with Hardy–Weinberg expectations in each population were performed for each microsatellite locus, after pooling rare alleles to achieve expected values per cell higher than two. Because genotype data were not available for the published allozyme data, analy-

sis was carried out on allele frequencies, assuming random mating in each population. *F*-statistics were calculated based on Wright (1978) for microsatellite and allozyme data using BIOSYS (Swofford & Selander, 1989). This method adjusts for sampling variation and does not require genotype frequencies. R_{ST} (Slatkin, 1995), a statistic related to F_{ST} developed specifically for microsatellite loci, accounts for sampling variation and especially for the different mutation process thought to occur in microsatellite loci (high mutation rate and partial dependence of the mutant allele size on the original allele size). Slatkin's method relies on the assumptions of no constraints on allele size and that the mutation process is similar across all allele sizes. Calculation of the repeat number for each allele was based on a known sequence, where the length of the regions flanking the repeated motif was subtracted from the total allele length, and the result was divided by the length of the repeat unit. The sizes of a few alleles at locus *33C1* differed by an amount that was not equal to the size of a repeat (Fig. 1) and R_{ST} calculation proceeded based on the following assumptions, namely that these alleles were created by an insertion/deletion of one nucleotide that occurred outside the repeat region, allowing rounding of the repeat number to the nearest integer. These noncanonical alleles had low frequency, comprising together eight per cent in both populations, thus even if these assumptions were wrong, the effect would be small. Locus *2A1* had a complex series of allele sizes (Fig. 1) resulting from several repeat motifs of different sizes. Calculation of R_{ST} for locus *2A1* was based on the assumption of a closer relationship between alleles of similar size, and thus rounding was performed based on the smallest motif size (6 bases). However, as the effect of this procedure was not known, average R_{ST} values for the set of microsatellite loci were calculated with and without this locus (Table 4).

Significance of the F_{ST} was evaluated based on a chi-square test of the contingency table of allele frequencies by populations, after pooling rare alleles such that expected cell counts would be higher than 2, and no more than 20 per cent of the cell counts would be lower than 5. Significance of the R_{ST} (and R_{IS}) was evaluated based on an *F*-test in a nested ANOVA on the repeat number in a model including the individual and the population as factors (Slatkin, 1995). The average R_{ST} was calculated from the averages of the within-population and total variance components across loci. Estimates of *Nm* were derived from F_{ST} for two populations according to Slatkin (1995): $Nm = 1/4\ (1/F_{ST} - 1)$. *Nm* was derived from R_{ST} by substituting R_{ST} for F_{ST}. Calculations not available in BIOSYS were carried out by programs written in the SAS language (SAS Institute, 1990).

Results

Microsatellite loci were highly polymorphic in both populations, with an average of 7.8 alleles (ranges of 2–13) per locus per population. The average unbiased heterozygosity was 0.63 (range of

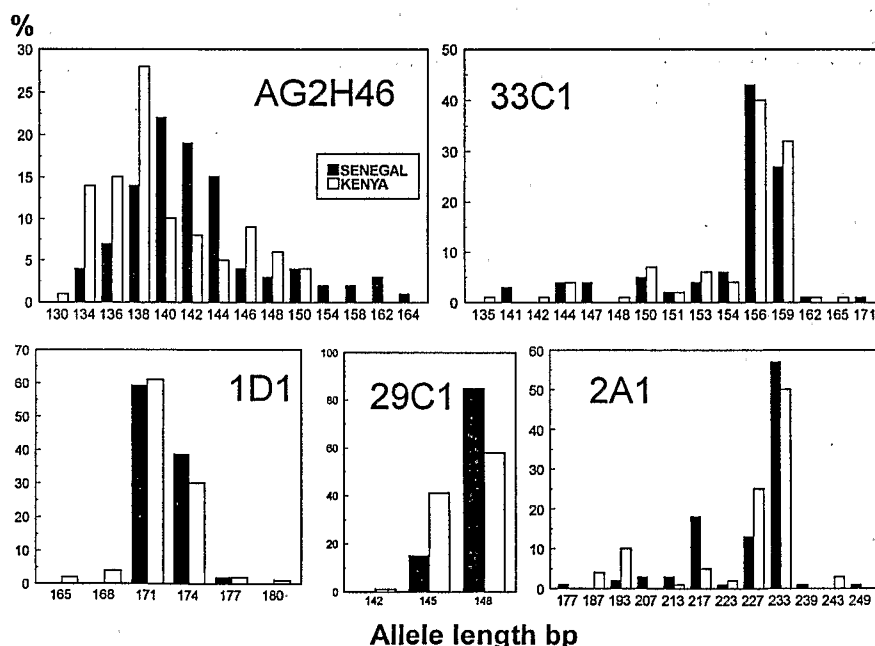


Fig. 1 Allele composition at microsatellite loci in East and West African *Anopheles gambiae*.

0.26–0.87, Table 2 and Fig. 1). Allozyme loci were moderately polymorphic with an average of 3.2 alleles (between 1 and 6) per locus, and the average unbiased heterozygosity was 0.38 (range of 0.0–0.76) (Table 2 and Fig. 2). Although the average unbiased heterozygosities of both microsatellite and allozyme loci were slightly higher in East Africa, the differences were not significant (signed rank paired test by locus = 15, d.f. = 10, $P > 0.21$ and Table 3), suggesting no large difference in effective population sizes

Table 2 Polymorphism of microsatellite and allozyme loci in populations of *Anopheles gambiae* from Kenya and Senegambia

	Kenya*					Senegambia*				
Locus	N	No. of alleles	Common allele %	H_E†	H_O‡	N	No. of alleles	Common allele %	H_E†	H_O‡
Microsatellites										
AG2H46	50	10	28	0.856	0.760	50	13	22	0.870	0.860
33C1	50	12	40	0.732	0.580	50	11	43	0.737	0.720
29C1	50	3	58	0.500	0.480	50	2	85	0.258	0.260
1D1	50	6	61	0.541	0.500	49	3	59	0.504	0.449
2A1	50	8	50	0.679	0.640	50	10	57	0.629	0.640
Average	50	7.8	47.4	0.662	0.592	49.8	7.8	53.2	0.600	0.586
Allozymes										
EST-1	24	3	75	0.415	—	22	3	48	0.591	—
EST-2	24	3	50	0.550	—	22	4	68	0.481	—
EST-3	24	6	42	0.755	—	22	6	52	0.671	—
ODH	22	1	100	0.000	—	22	2	98	0.045	—
PGM1	22	3	75	0.409	—	22	3	77	0.385	—
PGM2	22	2	89	0.206	—	22	2	96	0.089	—
Average	23	3.0	71.1	0.389	—	22	3.3	87.8	0.377	—

*For microsatellites, the Kenyan population was obtained from Asembo Bay and the Senegalian population from Barkedji; for allozymes (Miles, 1978), the Kenyan population was obtained from Chulaimbo and the Gambian population from Mandinari.

†Unbiased heterozygosity (Nei, 1978).

‡Observed heterozygosity based on direct count of heterozygotes.

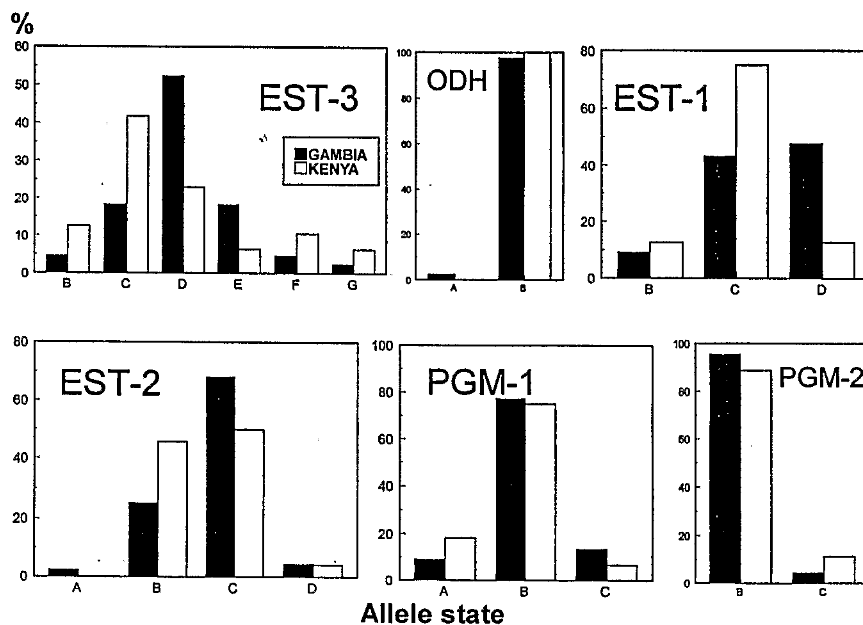


Fig. 2 Allele composition at allozyme loci in East and West African *Anopheles gambiae*.

between these populations from East and West Africa. The heterozygosity of microsatellite loci, however, was significantly higher than that of allozymes (Table 3). This difference was expected, based on known higher mutation rates reported for microsatellite loci (e.g. Dallas, 1992; Weber & Wong, 1993).

Deviations from Hardy–Weinberg expectations were not significant ($P > 0.05$) for all microsatellite loci in both populations (data not shown). Similarly, the within-individual variance component of the number of repeats tested by nested ANOVA (Slatkin, 1995) was not significant for all loci (data not shown). No evidence was obtained for nonrandom mating in the populations.

The allele compositions of both microsatellite and allozyme loci were very similar between West and East Africa (Figs 1 and 2). Every allele with a frequency higher than five per cent in a given population was also found in the other population, and the most common allele was the same in both populations for four of the five microsatellite loci (Fig. 1) and four of the six allozyme loci (Fig. 2). Differentiations of the populations, as measured by F_{ST} and R_{ST} were accordingly low (0.0–0.09, Table 4) and all *Nm* values were larger than 1. The R_{ST} values of microsatellite loci were higher than their corresponding F_{ST} (except for locus *2A1*, see Materials and methods), as expected of loci with high mutation rates, and with high likelihood that the same allele could be produced by independent mutations (Slatkin, 1995). Nevertheless, the R_{ST} (and F_{ST}) values of *AG2H46* and *29C1* were significant, whereas only the F_{ST} but not the R_{ST} of locus *2A1* was significant (Table 4). Likewise, the F_{ST} values of two of the six allozyme loci were significant.

Table 3 The differences in locus unbiased heterozygosity (H_E) between genetic markers and locales (as indicators of N_e): ANOVA results

Source	d.f.	Mean square	*P*
Model*	2	0.170	0.060
Error	19	0.052	—
Genetic marker	1	0.334	0.020
Population	1	0.007	0.724

Model $R^2 = 26\%$.
See Table 2 for group means.
*Interaction term was removed after it was found to be not significant when included ($P > 0.8$).

Discussion

A remarkable similarity in allele profile of *A. gambiae* populations 6000 km apart was evident at microsatellite and allozyme loci. The significance of the divergence indices in two (or three if F_{ST} is used) of the five microsatellite loci and in two of the six allozyme loci implies that a degree of separation between the gene pools does exist. However, low estimates of interpopulation differentiation were measured by Wright's F_{ST} and Slatkin's R_{ST}, corresponding to high estimates of the average migration index ($Nm > 3$) across this enormous distance. Genetic differentiation at 'neutral' loci because of genetic drift is expected if $Nm < 1$ but not if $Nm > 1$ (Slatkin, 1987). The consistency of the differentiation indices across loci, i.e. all 11 estimates fall in a narrow range of 0–0.087, implies that heterogeneity among loci in each marker group was not large. Accordingly, considering the three loci with the

Table 4 Differences between *Anopheles gambiae* populations from Kenya and Senegambia based on microsatellite and allozyme loci

Microsatellites			Allozymes	
Locus (N_k,N_s)†	Wright's F_{ST} (*Nm*)	Slatkin's R_{ST} (*Nm*)	Locus (N_g,N_k)	Wright's F_{ST} (*Nm*)
AG2H46 (50,50)	0.013** (9.5)	0.0624*** (1.9)	*EST-1* (22,24)	0.087*** (1.3)
33C1 (50,50)	0.0 NS (≫1)	0.0000 NS (≫1)	*EST-2* (22,24)	0.018 NS (6.8)
29C1 (50,50)	0.077*** (1.5)	0.08620*** (1.3)	*EST-3* (22,24)	0.037** (3.3)
1D1 (50,49)	0.0 NS (≫1)	0.0097 NS (12.7)	*ODH* (22,22)	0.0 NS (≫1)
2A1 (50,50)	0.01** (12.4)	0.0029 NS (42.5)	*PGM-1* (22,22)	0.0 NS (≫1)
			PGM-2 (22,22)	0.0 NS (≫1)
Average 5 loci	0.016 (7.7)	0.0358 (3.4)	Average	0.036 (3.3)
4 loci		0.0469 (2.5)		

†Sample size (mosquitoes): N_k, Kenya; N_s, Senegal; and N_g, Gambia.
NS $P > 0.05$, ** $P < 0.01$, *** $P < 0.001$.

highest differentiation estimates from each group increased the average R_{ST} of microsatellite loci to 0.061 (corresponding to an $Nm = 1.9$) and the average divergence of allozymes to only 0.047 (corresponding to an $Nm = 2.5$). Thus, even these estimates imply extensive gene flow across the continent. Furthermore, regional analyses of Miles's (1978) six-loci data including either 10 or six localities divided into West (Gambia), East (Kenya and Tanzania), and Central Africa (Nigeria and Cameroon) with samples larger than 12 or 20 mosquitoes, respectively, provided slightly lower estimates of between-region F_{ST} (0.015 and 0.021, data not shown).

For comparison, microsatellite loci in the honey bee, *Apis mellifera*, measured far greater divergence between populations within lineages, with an average F_{ST} value of 0.34 (Estoup *et al.*, 1995), whereas allozyme variation hardly exists. Pairwise average F_{ST} values for populations within the least differentiated honey bee subspecies, separated by less than 2500 and 2000 km (two from France and one from Sweden), were 0.083 and 0.042, respectively, vs. 0.016 in the populations of the present study (Table 4). The mosquito *Aedes aegypti* showed a complex structure with substantial differentiation on a worldwide scale (Powell *et al.*, 1980), although extensive gene flow ($Nm > 9$) was observed across approx. 150 km in Puerto Rico (Apostol *et al.*, 1995). Even greater geographical differentiation was observed in *Aedes albopictus* in Malaysia and Borneo (Black *et al.*, 1988).

Because mark-release-recapture experiments have suggested that active dispersal in *A. gambiae* is restricted to a few kilometers (Gillies & De Meillon, 1968), it is unlikely that active dispersal accounts for the low differentiation. An alternative explanation is that *A. gambiae*, a highly domestic mosquito, has expanded to its current range in recent evolutionary time in response to the major expansion in human population density that followed the development of agriculture in West Africa about 2000–5000 years ago (Coluzzi *et al.*, 1985). It is conceivable that with population expansion occurring within the past few thousand years and with large effective population size, the divergence of populations across the African continent could remain low today.

The similarity of estimates of divergence measured by marker systems so different in their time-scale resolution is surprising. Microsatellite loci evolve considerably faster than allozyme loci, and have mutation rates of 10^{-3}–10^{-4} per locus per gamete per generation (Dallas, 1992; Weber & Wong, 1993) vs. 10^{-6}–10^{-9} for allozyme loci (Ayala, 1976). Thus they are expected to detect differentiation between populations that occurred over shorter periods. The lower polymorphism of the allozymes as compared with the microsatellite loci (Tables 2 and 3) is consistent with the mutation rates of the microsatellite loci being higher than those of the allozyme loci. In spite of mutation rates at least one scale order higher than that of allozymes, the R_{ST} estimate for the microsatellite loci was not higher than the F_{ST} for the allozymes. The failure of microsatellite loci to detect greater divergence than allozyme loci suggests the existence of constraints on their evolution, such as biased mutation rates (Garza *et al.*, 1995) and/or selection for certain allele sizes (Epplen *et al.*, 1993).

Alternatively, dispersal via both active and passive transportation of *A. gambiae* during the past century might be much higher than generally appreciated. Surprisingly high numbers of *A. gambiae* have been captured in trains and boats in Africa (Holstein, 1954). The rapid geographical spread of this mosquito after its introduction into Brazil in 1930 also indicates considerably higher rates of dispersal than commonly appreciated. The spread of *A. gambiae* in Brazil from its point of introduction in 1930 to the extreme extent of its range in 1938–40 was at an average minimal rate of 50 km y^{-1}, based on Soper & Wilson (1943). This is probably an underestimate because (1) it is based on the detection of new, fully established populations rather than small numbers of migrants, (2) the estimate assumes a constant rate of movement over the entire 10-year period, and (3) populations could successfully spread only in the north-westerly direction. If migration at such a rate is also occurring in Africa, it may be sufficient to explain the low levels of genetic divergence between populations in East and West Africa. Furthermore, such high rates of migration would also explain our finding of similar estimates of divergence obtained by the two different marker systems, without the need to invoke constraints on their evolution.

The low levels of genetic differentiation between populations from East and West Africa would seem to be incompatible with the very high levels of chromosomal-inversion-based genetic population structure recorded in *A. gambiae* populations, particularly from West Africa (Coluzzi *et al.*, 1985; Lanzaro *et al.*, 1995). Several distinct chromosomal forms and a series of climato-geographically stable clines in the frequencies of certain chromosomal inversions are characteristic of many populations in West Africa, whereas they have not been found in East Africa despite extensive karyotype data from

Kenya and Tanzania. These reports of highly structured West African populations might suggest restricted gene flow in sympatric populations across dimensions other than geographical distance — the only dimension explored in this study. It is important to note that the Savanna chromosomal form is the only form known to occur in Kenya (Coluzzi *et al.*, 1985) and Barkedji (Fontenille, unpublished data). Additionally, the strong selection operating on many of these chromosomal inversions that is evident by regular seasonal or spatial-climatological change in karyotype frequencies may render them unsuitable for gene flow evaluations.

At this point there is insufficient information to explain the low differentiation of *A. gambiae* populations across Africa by ruling out either the hypothesis of recent expansion, which is coupled with constraints on microsatellite evolution, or the hypothesis of migration augmented by transportation. Information based on additional genetic loci, different genetic markers, and populations on micro- and macrogeographical scales is necessary before final conclusions can be reached.

Acknowledgements

We are grateful to William C. Black IV, Nora J. Besansky, and Michel Tibayrenc for valuable suggestions throughout this study and useful comments on earlier versions of this manuscript. We thank Brian Holloway and staff of the NCID Biotechnology Core Facility for synthesizing the oligonucleotide primers and Diane M. Hamm for valuable advice on the laboratory work. T.L. was supported by an American Society for Microbiology Fellowship. Additional financial support was provided by the John D. and Catherine T. MacArthur Foundation and by the UNDP/WORLD BANK/WHO Special Programme for Research and Training in Tropical Diseases.

References

APOSTOL, B. L., BLACK, W. C., IV., REITER, P. AND MILLER, B. R. 1995. Population genetics with RAPD-PCR markers: the breeding structure of *Aedes aegypti* in Puerto Rico. *J. Med. Entomol.*, (in press).

AYALA, F. J. (ed.) 1976. *Molecular Evolution.* Sinauer Associates, Sunderland, MA.

BEARD, C. B., HAMM, D. M. AND COLLINS, F. H. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.*, **2**, 103–124.

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

BLACK, W. C., IV, HAWLEY, W. A., RAI, K. S. AND CRAIG, G. B., Jr. 1988. Breeding structure of a colonizing species: *Aedes albopictus* (Skuse) in peninsular Malaysia and Borneo. *Heredity*, **61**, 439–446.

BOWCOCK, A. M., RUIZ-LINARES, A., TOMFOHRDE, J., MINCH, E., KIDD, J. R. AND CAVALLI-SFORZA, L. L. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455–457.

COLLINS, F. H. AND BESANSKY, N. J. 1994. Vector biology and control of malaria in Africa. *Science*, **264**, 1874–1875.

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

COLUZZI, M., PETRARCA, V. AND DI DECO, M. 1985. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Boll. Zool.*, **52**, 45–63.

CRAMPTON, J. M., WARREN, A., LYCETT, G. J., HUGHES, M. A., COMLEY, I. P. AND EGGLESTON, P. 1994. Genetic manipulation of insect vectors as a strategy for the control of vector-borne disease. *Ann. Trop. Med. Parasitol.*, **88**, 3–12.

DALLAS, J. F. 1992. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mamm. Genome*, **3**, 452–456.

DI RIENZO, A., PETERSON, A. C., GARZA, J. C., VALDES, A. M., SLATKIN M. AND FREIMER, N. B. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 3166–3170.

EDWARDS, A., HAMMOND, H. A., JIN, L., CASKEY, C. T. AND CHAKRABORTY, R. 1992. Genetic variation of five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, **12**, 241–253.

EPPLEN, C., MELMER, G., SIEDLACZCK, I., SCHWAIGER, F.-W., MÄUELER, W. AND EPPLEN, J. T. 1993. On the essence of "meaningless" simple repetitive DNA in eukaryote genomes. In: Pena, S. D. J., Chakraborty, R., Epplen, J. T., and Jeffreys, A. J. (eds) *DNA Fingerprinting: State of the Science*, pp. 29–45. Birkhauser Verlag, Basel.

ESTOUP, A., GARNERY, L. SOLIGNAC, M. AND CORNUET, J.-M. 1995. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, **140**, 679–695.

FITZSIMMONS, N. N., MORITZ, C. AND MOORE, S. S. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol. Biol. Evol.*, **12**, 432–440.

GARZA, J. C., SLATKIN, M. AND FREIMER, N. B. 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.*, **12**, 594–603.

GILLIES, M. T. AND DE MEILLON, B. 1968. *The Anophelinae*

of Africa South of the Sahara, 2nd edn. Publications of the South African Institute for Medical Research no. 54. Johannesburg, South Africa.

HOLSTEIN, M. H. 1954. *Biology of* Anopheles gambiae, *Research in French West Africa*. WHO, Geneva.

KUMAR, V. K. AND COLLINS, F. H. 1993. A technique for nucleic acid *in situ* hybridization to polytene chromosomes of mosquitoes in the *Anopheles gambiae* complex. *Insect Mol. Biol.*, **3**, 41–47.

LANZARO, G. C., TOURE, Y. T., ZHENG, L., KAFATOS, F. C. AND VERNICK, K. D. 1995. Microsatellite DNA and isozyme variability in a West African population of *Anopheles gambiae*. *Insect Mol. Biol.*, **4**, 105–112.

MILES, S. J. 1978. Enzyme variation in the *Anopheles gambiae* Giles group of species (Diptera: Culicidae). *Bull. ent. Res.*, **68**, 85–96.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

POWELL, J. R., TABACHNICK, W. J. AND ARNOLD, J. 1980. Genetics and the origin of a vector population: *Aedes aegypti*, a case study. *Science*, **208**, 1385–1387.

SALAZAR, C. E., HAMM, D. M., WESSON, D. M., KUMAR, V. K., BEARD, C. B. AND COLLINS, F. H. 1993. A cytoskeletal actin gene in the malaria vector *Anopheles gambiae*. *Insect Mol. Biol.*, **3**, 1–14.

SAMBROOK, J., FRITSCH, E. F. AND MANIATIS, T. 1989. *Molecular Cloning, a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, NY.

SAS INSTITUTE INC. 1990. *SAS language: references.* Version 6, 1st edn. Cary, NC.

SCOTT, J. A., BROGDON, W. G. AND COLLINS, F. H. 1993. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg.*, **49**, 520–529.

SLATKIN, M. 1987. Gene flow and the geographical structure of natural populations. *Science*, **236**, 787–792.

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

SOPER, F. L. AND WILSON, D. B. 1943. Anopheles gambiae *in Brazil 1930–1940*. Rockefeller Foundation, New York.

SWOFFORD, D. L. AND SELANDER, R. B. 1989. BIOSYS-1. *A Computer Program for the Analysis of Allelic Variation in Population Genetics and Biochemical Systematics.* Release 1.7. University of Illinois, Urbana, IL.

WEBER, J. L. AND WONG, C. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.*, **2**, 1123–1128.

WRIGHT, S. 1978. *Evolution and the Genetics of Populations,* vol. 4, *Variability Within and Among Natural Populations.* University of Chicago Press, Chicago.

ZHENG, L., SOUNDERS, R. D. C., FORTINI, D., TORRE, A. D., COLUZZI, M., GLOBER, D. M. AND KAFATOS, F. C. 1991. Low-resolution genome map of the malaria mosquito *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 11187–11191.

© The Genetical Society of Great Britain, *Heredity*, **77**, 192–208.