

Genetic differentiation of the malaria vector *Anopheles gambiae* across Nigeria suggests that selection limits gene flow

DAVID Y. ONYABE & JAN E. CONN*

Department of Biology, The University of Vermont, 321 Marsh Life Science Building, Burlington, VT 05405, U.S.A.

Gene flow was investigated in *Anopheles gambiae* from eight localities that span the ecological zones of Nigeria (arid savanna zones in the north gradually turn into humid forest zones in the south). Genetic differentiation was measured over 10 microsatellite loci and, to determine any effects of selection, five loci were located within chromosome inversions and the other five were outside inversions. Over all loci, the largest estimates of differentiation were in comparisons between localities in the savanna vs. forest zones (range F_{ST} 0.024–0.087, Nm 2.6–10.1; R_{ST} 0.014–0.100, Nm 2.2–16.4). However, three loci located within inversions on chromosome II, whose frequencies varied clinally from north to south, were responsible for virtually all of the differentiation. When the three loci were removed, genetic distances across the remaining seven loci were markedly reduced even between localities in the forest and savanna zones (range F_{ST} 0.001–0.019, Nm 12.7–226.1) or no longer significant ($P > 0.05$) in the case of R_{ST} . Although tests of isolation by distance gave seemingly equivocal results, geographical distance does not appear to limit gene flow. These observations suggest that gene flow is extensive across the country but that selection on genes located within some inversions on chromosome II counters the homogenizing effects of gene flow.

Keywords: *Anopheles gambiae*, ecological zones, gene flow, microsatellite DNA, Nigeria.

Introduction

Due largely to its highly anthropophilic nature, *Anopheles gambiae* is the most important vector of malaria in sub-Saharan Africa. An understanding of the genetic structure of vector populations is required for the proposed release of transgenics (mosquitoes genetically altered with genes that confer malaria refractoriness) (Collins & Besansky, 1994; Curtis, 1994; Ashburner *et al.*, 1998). Moreover, gene flow is a useful indicator of direction and rates of dispersal among populations (Slatkin, 1987) and may be useful in designing and evaluating mosquito control strategies (reviewed in Coluzzi *et al.*, 1979; Tabachnick & Black, 1995).

Studies of the genetic structure of *A. gambiae* have reached conflicting conclusions. Analyses of mitochondrial DNA (mtDNA) sequences (Besansky *et al.*, 1997), allozymes (Lehmann *et al.*, 1996) and microsatellite markers (Lehmann *et al.*, 1996; Kamau *et al.*, 1999) suggest extensive gene flow across Africa in *A. gambiae* collected at least 5000 km apart. In contrast, a complex differentiation of paracentric inversions on chromosome

II exists in West Africa, such that five chromosomal forms are recognized: Savanna, Forest, Bamako, Mopti and Bissau (Bryan *et al.*, 1982; Coluzzi *et al.*, 1985; Toure *et al.*, 1998). The Savanna form occurs throughout much of sub-Saharan Africa and is therefore the most widespread form. The Forest form refers to forest-breeding *A. gambiae*, whereas Bissau occurs only in The Gambia and Senegal. Bamako occurs in Mali and northern Guinea, and Mopti is found in Mali, Ivory Coast, Guinea and Burkina Faso. In parts of Mali where the Mopti, Savanna and Bamako forms exist in sympatry, hybrids are found at low frequencies (Bamako \times Savanna, Mopti \times Savanna) or not at all (Mopti \times Bamako) (Bryan *et al.*, 1982; Coluzzi *et al.*, 1985; Toure *et al.*, 1998). Recent studies (Black & Lanzaro, 2001; della Torre *et al.*, 2001; Favia *et al.*, 2001; Gentile *et al.*, 2001; Mukabayire *et al.*, 2001) showed that ribosomal DNA spacers (intergenic spacer and internal transcribed spacers) distinguish two major groups of *A. gambiae* across Africa, the S and M molecular forms. These studies suggest at least partial barriers to gene flow between the two forms in the Ivory Coast and other West African countries to the north and west, whereas introgression occurs between them in the Benin

*Correspondence. E-mail: jconn@zoo.uvm.edu

Republic and countries to the east. Nigeria is located to the east of the Benin Republic and therefore introgression is very likely to occur between our samples.

There are few studies of the population genetics of *A. gambiae* in Nigeria. Coluzzi *et al.* (1979, 1985) demonstrated a north to south cline in the frequency of some inversions on chromosome II in *A. gambiae* across the ecological zones of Nigeria (arid savanna zones in the north gradually turn into humid forest zones in the south. To avoid confusion with the chromosomal forms above, we use the terms 'savanna zones' and 'forest zones' to refer to ecological zones). The most significant differences in the frequencies of inversions were between the northern and southern extremes of the country. Thus, *A. gambiae* in Nigeria consists of two chromosomal forms – the Savanna form in the north and the Forest form in the south – between which there appears to be normal intergradation (Coluzzi *et al.*, 1979, 1985).

The magnitude of gene flow across ecological zones is unclear from Coluzzi *et al.* (1979). Chromosome inversions are probably poor indicators of gene flow because they are not selectively neutral. The distribution of inversions across Nigeria suggests that gene flow is restricted by geographical distance, that is, isolation by distance, because the largest disparities in inversion frequencies were between the extremes of the country (Coluzzi *et al.*, 1979). Alternatively,

assuming the island model (Wright, 1931), i.e. each geographical locale exchanges migrants with each of the other localities at an equal rate, the correlation of some inversions, but not others, with ecological zones (Coluzzi *et al.*, 1979, 1985) suggests that selection counters extensive gene flow across the country. Thus, parts of the genome that are located within inversions, especially on chromosome II, might be expected to measure higher levels of differentiation than those that are located outside inversions (Lanzaro *et al.*, 1998; Black & Lanzaro, 2001). If both isolation by distance and selection restrict gene flow, loci within inversions should detect higher levels of differentiation than those outside inversions, but loci located outside inversions should measure significant differentiation that increases with geographical distance.

The objective of this study was to determine the extent of gene flow across the ecological zones of Nigeria in *A. gambiae* sampled from eight localities (Fig. 1) spanning a distance of 90–833 km. Genetic differentiation was estimated by employing 10 micro-satellite loci, short blocks of tandemly repeated simple DNA sequences of length 1–5 bp (reviewed in Bruford & Wayne, 1993). To measure the effect of selection on genetic structure, five of the loci employed are located within inversions, whereas the other five loci occur outside inversions.

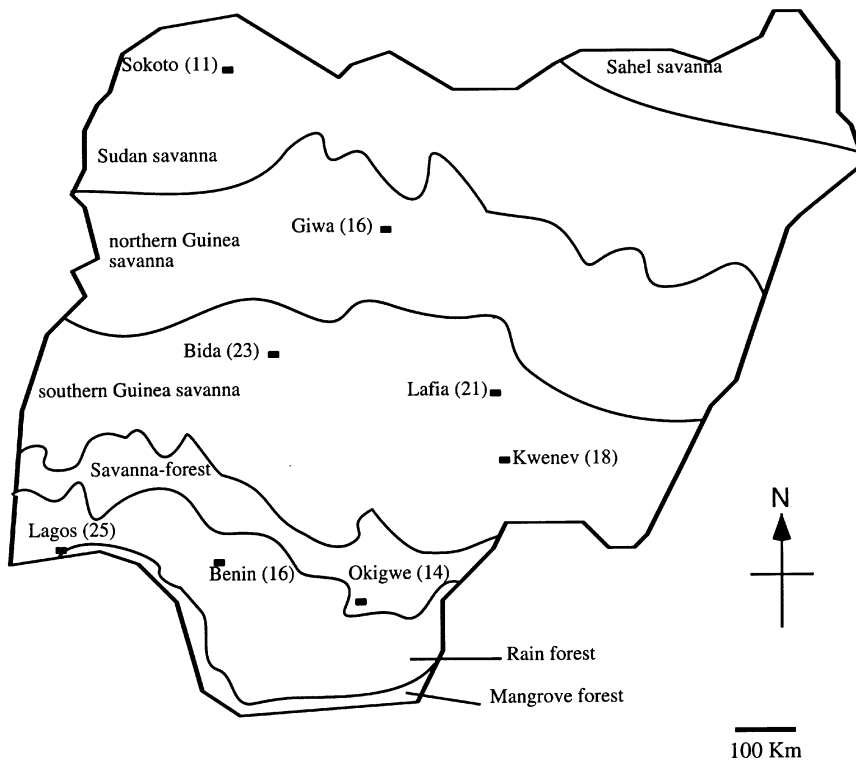


Fig. 1 Sampling sites of *Anopheles gambiae* across Nigeria. Numbers in parentheses indicate number of pools sampled per locality.

Materials and methods

Study area

As one travels from north to south across Nigeria, mean annual rainfall increases, the number of dry season months decreases, and the vegetation becomes taller and more dense (Davies, 1977). The country is therefore divided into seven major ecological zones (Fig. 1) (Coluzzi *et al.*, 1979), such that arid savanna zones in northern Nigeria gradually turn into humid forest zones in the south. Mean annual rainfall ranges from less than 500 mm in the Sudan and Sahel savanna zones to more than 2500 mm in the forests, so that mosquito breeding is seasonal in the north, but is subcontinuous in the south (Coluzzi *et al.*, 1979).

Mosquito collection

Mosquitoes were collected from eight localities across Nigeria (Fig. 1) during the rainy season, between June and August 1997 and between May and June 1999. Samples from Kwenev, Okigwe and Lagos were collected in 1999, whereas the other localities (Fig. 1) were sampled in 1997. Giwa, Lafia, Bida and Benin were each sampled over 2 or more days, whereas all other localities were each sampled in 1 day.

In each locality, larvae of all available instars, or pupae, or both were collected from shallow, temporary, sunlit pools within a radius of approximately 1 km. Samples were collected from at least 11 pools per locality (Fig. 1), but sample size per pool was not determined and collections from all pools in each locality were mixed. At least 200 larvae were collected in each locality and larvae were reared on location in paper cups to the adult stage because they were more easily distinguished as *A. gambiae sensu lato* by the keys of Gillies & Coetzee (1987). Rearing success was greater than 95%, and emerging adults and occasional dead larvae or pupae were fixed in 95% ethanol. Specimens were then hand-carried to the University of Vermont. Due to logistic difficulties, samples were not collected from the Sahel savanna and Mangrove forest (Fig. 1).

Species identification

Samples for identification to species were selected by random draw and DNA was then extracted from whole specimens according to Collins *et al.* (1987). Because *A. gambiae* and *A. arabiensis* occur in sympatry in many localities across Nigeria (Coluzzi *et al.*, 1979; D. Y. Onyabe & J. E. Conn, unpublished data), each

specimen was identified to species by the polymerase chain reaction (PCR) method of Scott *et al.* (1993). Only samples identified as *A. gambiae* were included in this study.

Microsatellite genotyping

Ten microsatellite loci (Table 1) were selected from the published genomic map of *A. gambiae* (Zheng *et al.*, 1996). Location of loci in the genome (Table 1) was determined from genomic (Zheng *et al.*, 1996) and cytogenetic maps (Coluzzi & Sabatini, 1967). At least 39 specimens per locality (Table 2) were genotyped at each locus using the primers of Zheng *et al.* (1996) by PCR amplification in a PTC-100 thermal cycler (MJ Research, Waltham, MA). The forward primer of each pair was labelled with a fluorescent dye (6FAM, HEX, or TET, Applied Biosystems, Foster City, CA). The reaction mixture contained approximately 1/100th of the genomic DNA of each specimen, 10× PCR buffer, 0.2 mM of each dNTP, 10 pMols per primer, 1.5 mM MgCl₂, and 0.5 units DNA polymerase (Amersham Pharmacia Biotech, Piscataway, NJ). DNA amplification was conducted through 30 cycles as follows: initial denaturation for 5 min at 94°C, primer annealing for 30 s at 55°C, and extension for 20 s at 72°C. In subsequent cycles, denaturation was for 20 s at 94°C, whereas the final extension step was for 5 min. Each PCR reaction (1.5 µL aliquot) was prepared for electrophoresis by mixing with 2.4 µL formamide, 0.5 µL blue dextran and 0.6 µL GeneScan-350 (TAMRA) size standard (Applied Biosystems, Foster City, CA). Electrophoresis was performed using 1.3 µL of the final mixture in a 4.5% acrylamide gel on an ABI Prism 377

Table 1 Microsatellite loci studied in *Anopheles gambiae*†

Locus	Cytological location‡	Inversion§	Repeat motif
AG2H175	2R	OI¶	CA
AG2H637	2L	2La	CA
AG2H26	2R:12	2Rb	GT
AG2H523	2L	2La	GT
AG2H79	2R	2Ra	GT
AG3H249	3R	OI	GT
AG3H88	3L	OI	GT
AG3H577	3L:42A	OI	GT
AG3H128	3R	OI	GT
AGXH99	XR:2C	Xag	GT

†Data from Zheng *et al.* (1996).

‡R refers to right arm of chromosome and L to left arm. Subdivision, if known, is given.

§From Coluzzi & Sabatini (1967).

¶Outside inversion.

Table 2 Genetic variability in *Anopheles gambiae* from eight localities across Nigeria (a) Sokoto, Giwa, Lafia and Kwenev

		Sokoto (2n = 90)	Giwa (2n = 90)	Lafia (2n = 90)	Kwenev (2n = 78)
AG2H175	No. alleles	9	8	9	6
	H_O †	0.84	0.80	0.71	0.77
	H_E ‡	0.87	0.76	0.79	0.76
AG2H26	No. alleles	10	16	12	11
	H_O	0.58	0.62	0.76	0.54
	H_E	0.71	0.69	0.77	0.74
AG2H523	No. alleles	19	22	18	19
	H_O	0.80	0.73	0.64*	0.74
	H_E	0.79	0.87	0.91	0.74
	$D\$/r$ ¶	—	—	0.29/0.17	—
AG2H79	No. alleles	5	6	7	9
	H_O	0.58	0.62	0.53	0.64
	H_E	0.67	0.62	0.66	0.78
AG2H637	No. alleles	7	10	12	11
	H_O	0.30*	0.76	0.68	0.71
	H_E	0.59	0.80	0.79	0.77
	D/r	0.49/0.32	—	—	—
AG3H249	No. alleles	11	12	10	11
	H_O	0.84	0.87	0.90	0.82
	H_E	0.84	0.84	0.84	0.88
AG3H88	No. alleles	13	13	14	13
	H_O	0.44*	0.62*	0.58*	0.68
	H_E	0.86	0.84	0.88	0.87
	D/r	0.48/0.32	0.26/0.15	0.34/0.20	—
AG3H577	No. alleles	9	10	9	8
	H_O	0.53	0.60	0.67	0.49
	H_E	0.58	0.65	0.65	0.50
AG3H128	No. alleles	21	26	22	18
	H_O	0.93	0.91	0.84	0.94
	H_E	0.94	0.91	0.92	0.90
AGXH99	$2n$ ††	70	75	73	59
	No. alleles	6	6	5	4
	H_O	0.60	0.57*	0.85	0.55
	H_E	0.73	0.73	0.73	0.71
	D/r	—	0.21/0.12	—	—
Mean		11.0 (1.6)	12.60 (2.0)	12.1 (1.6)	11.0 (1.5)
no. alleles (\pm SE)					
Mean H_E (\pm SE)		0.76 (0.03)	0.77 (0.03)	0.79 (0.03)	0.76 (0.03)

DNA sequencer (Applied Biosystems). Gels were analysed and project and sample files were generated using ABI GENESCAN software, version 3.1. (Applied Biosystems). Subsequently, allele sizes were scored using ABI GENOTYPER software version 2.0 (Applied Biosystems).

Statistical analyses

All analyses, except where noted, were performed using ARLEQUIN software version 2.000 (Schneider *et al.*, 2000). The genetic diversity indices, observed and expected heterozygosity (H_O and H_E , respect-

ively), and number and frequency of alleles were calculated per locus for each locality. The H_E data were first tested for normality by the method of Shapiro & Wilk (1965) and non-normally distributed data were then square-root arcsine-transformed to better approximate a normal distribution (Zar, 1996). The significance of differences among localities in number of alleles per locus and in H_E was tested in separate nested analyses of variance (ANOVA) performed in JMP software, version 3.1 (SAS Institute Inc., Cary, NC). The factors in the ANOVA were: (1) ecological zone; (2) localities nested within ecological zones; (3) location of loci with respect to inversions;

Table 2 (Continued) (b) Bida, Okigwe, Benin and Lagos

		Bida (2n = 90)	Okigwe (2n = 90)	Benin (2n = 92)	Lagos (2n = 90)
AG2H175	No. alleles	9	6	8	8
	H_O	0.73	0.64	0.89	0.73
	H_E	0.79	0.69	0.84	0.81
AG2H26	No. alleles	9	10	8	9
	H_O	0.71	0.71	0.59	0.76
	H_E	0.77	0.79	0.80	0.77
AG2H523	No. alleles	19	20	17	17
	H_O	0.45*	0.34*	0.43*	0.56*
	H_E	0.86	0.87	0.89	0.83
	D/r	0.47/0.31	0.60/0.43	0.51/0.34	0.32/0.19
AG2H79	No. alleles	9	6	18	15
	H_O	0.70	0.58	0.83	0.73
	H_E	0.70	0.59	0.83	0.65
AG2H637	No. alleles	9	7	9	7
	H_O	0.53	0.60	0.43	0.67
	H_E	0.65	0.56	0.49	0.75
AG3H249	No. alleles	13	16	15	13
	H_O	0.89	0.87	0.85	0.91
	H_E	0.84	0.88	0.86	0.84
AG3H88	No. alleles	13	8	11	11
	H_O	0.69	0.41*	0.65	0.31*
	H_E	0.84	0.82	0.84	0.80
	D/r	—	0.5/0.33	—	0.61/0.44
AG3H577	No. alleles	11	10	13	12
	H_O	0.69	0.67	0.64	0.67
	H_E	0.68	0.62	0.63	0.55
AG3H128	No. alleles	25	22	23	21
	H_O	0.91	0.91	0.87	0.87
	H_E	0.92	0.93	0.93	0.91
AGXH99	2n	72	63	69	65
	No. alleles	6	4	5	6
	H_O	0.41*	0.44	0.82	0.55
	H_E	0.75	0.65	0.68	0.74
	D/r	0.45/0.29	—	—	—
Mean		12.3 (1.8)	10.9 (1.9)	12.7 (1.7)	11.9 (1.5)
no. alleles (\pm SE)					
Mean H_E (\pm SE)		0.78 (0.02)	0.74 (0.04)	0.78 (0.04)	0.76 (0.03)

†Observed heterozygosity.

‡Unbiased heterozygosity.

§Proportion of observed heterozygote deficiency. A '—' indicates no significant heterozygote deficiency.

¶Estimated frequency of null alleles.

††Number of chromosomes scored differed for the X chromosome marker and may be an odd number due to inclusion of males. For samples from Benin, 2n at this locus was approximated by assuming a 1:1 ratio between males and females.

* $P < 0.05$ after sequential Bonferroni correction.

(4) loci nested within or outside inversions; and (5) interaction between ecological zones and location of loci with respect to inversions.

Each locus was tested separately for departures from Hardy–Weinberg equilibrium (HWE) using the Markov chain algorithm of Guo & Thompson (1992). The number of steps in the Markov chain was 100 000 and the number of dememorization steps

was 10 000. The proportion of observed heterozygote deficiencies (D) and the frequencies of null alleles (r) that caused the deficiencies were estimated by the method of Chakraborty *et al.* (1992). The following expressions were used: $D = H_E - H_O/H_E$ and $r = (H_E - H_O)/(H_E + H_O)$.

The significance of pairwise linkage disequilibrium was determined by the exact test, which also employed

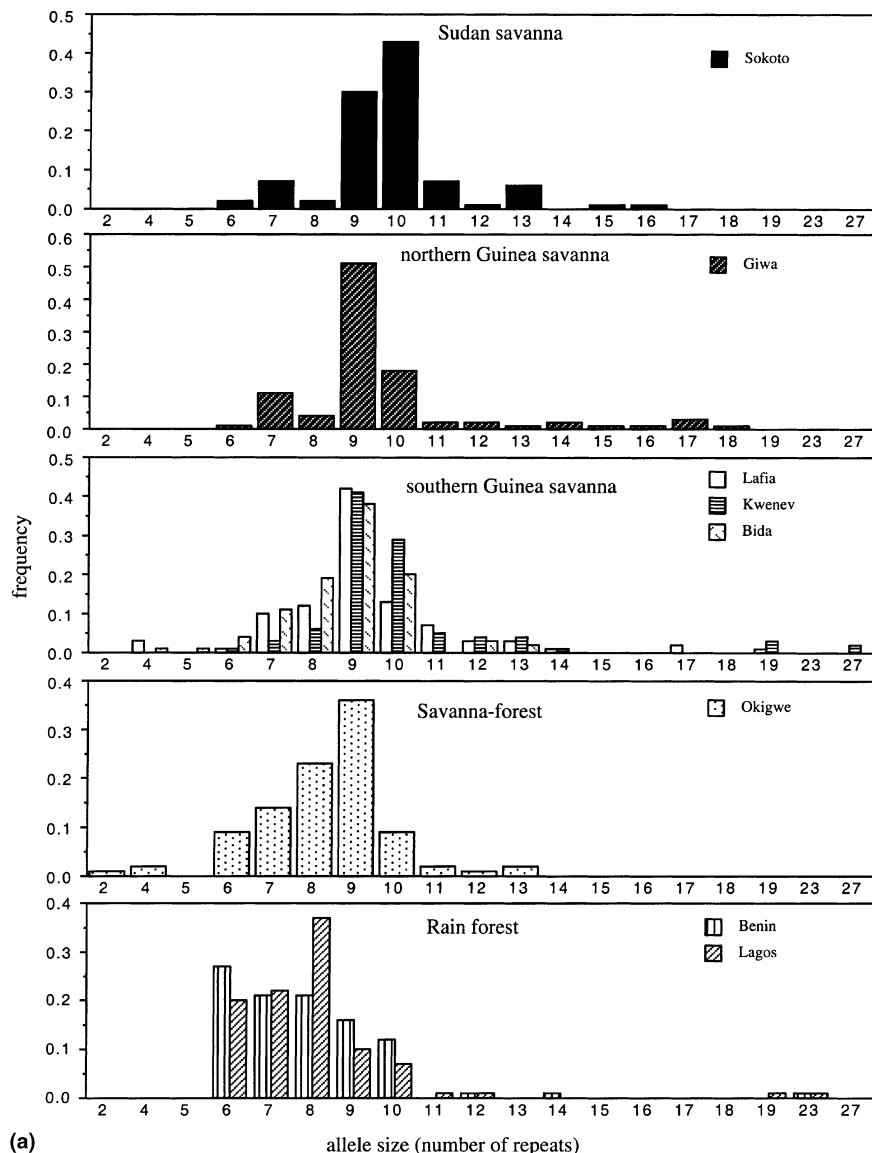


Fig. 2 Distribution and frequencies of alleles at three microsatellite loci in *Anopheles gambiae* across Nigeria. Notice the different scales on the y axis between zones in each figure. Fig. 2(a), AG2H26; (b), AG2H79; (c), AG2H637.

the Markov chain described above. Significance levels were adjusted using the sequential Bonferroni method to account for multiple comparisons (Holm, 1979) in tests of HWE and linkage disequilibrium.

Two indices of genetic differentiation were estimated between the eight localities as follows: (1) F_{ST} (Weir & Cockerham, 1984); and (2) R_{ST} (Slatkin, 1995). F_{ST} was calculated based on the absolute frequencies of alleles, whereas R_{ST} was estimated from the sum of squared number of repeat differences (Slatkin, 1995). (The number of repeats for alleles at each locus was determined by comparison with the sequence of one allele from each locus or with published data (Zheng *et al.*, 1996)). The two indices assume different mutation models: F_{ST} assumes the infinite alleles model (IAM), whereas R_{ST} assumes that alleles evolve

according to the stepwise mutation model (SMM), that there are no constraints on allele size and that the mutation rate is similar across alleles (Slatkin, 1995). The significance of F_{ST} and R_{ST} was determined by permuting genotypes between localities. The number of migrants per population per generation (Nm) between localities was estimated from pairwise F_{ST} and R_{ST} (Slatkin, 1995).

Isolation by distance, the correlation between genetic and geographical distance, was tested by the regression of $F_{ST}/1 - F_{ST}$ on the natural logarithm (\ln) of straight-line geographical distance (Rousset, 1997). A similar regression was performed using R_{ST} instead of F_{ST} . Straight-line geographical distances were determined by using Microsoft Encarta 97 World Atlas.

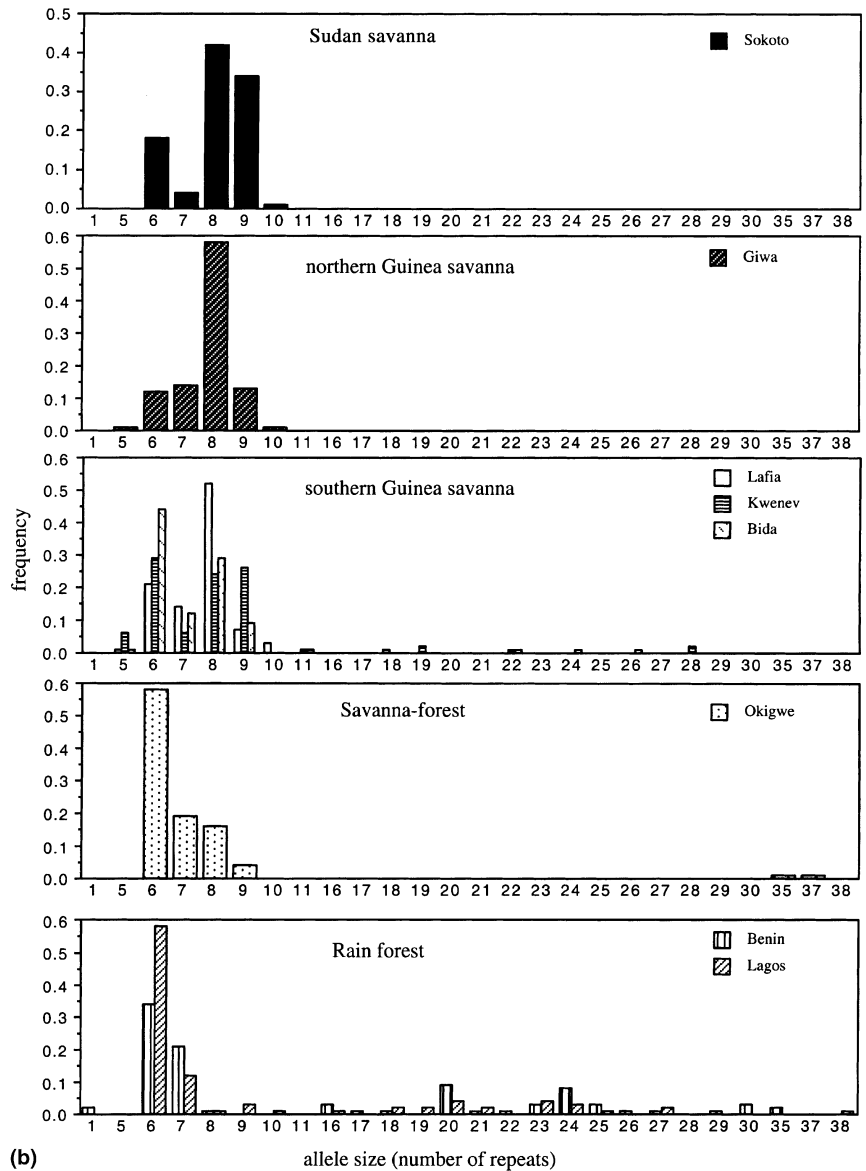


Fig. 2 (Continued).

Results

Distribution of alleles

North to south clines in allele frequencies occurred at three loci, *AG2H26*, *AG2H79* and *AG2H637* (Fig. 2a,b,c), all of which are located within inversions on chromosome II. At the other seven loci, the modal allele was the same (*AG2H175*, *AG2H523*, *AG3H88* and *AG3H577*) or nearly so (*AG3H249*, *AG3H128* and *AGXH99*) across localities (data not shown).

The range of mean number of alleles per locus per locality was 10.9–12.7 (Table 2). There was no significant difference in number of alleles per locus among localities within ecological zones ($F_{3,59} = 0.791$,

$P = 0.503$) or among ecological zones ($F_{4,59} = 1.012$, $P = 0.408$).

Mean H_E per locus per locality ranged from 0.74 to 0.79 (Table 2) and it did not differ significantly among localities within ecological zones ($F_{3,59} = 0.728$, $P = 0.538$) or among ecological zones ($F_{4,59} = 0.705$, $P = 0.591$).

Hardy-Weinberg equilibrium and linkage disequilibrium

A comparison of H_E and H_O is shown in Table 2. Thirteen of 70 tests showed significant departures ($P < 0.05$) from HWE after sequential Bonferroni correction. Two loci, *AG2H523* and *AG3H88*, accounted for 10 of the 13 significant deficits of heterozygotes.

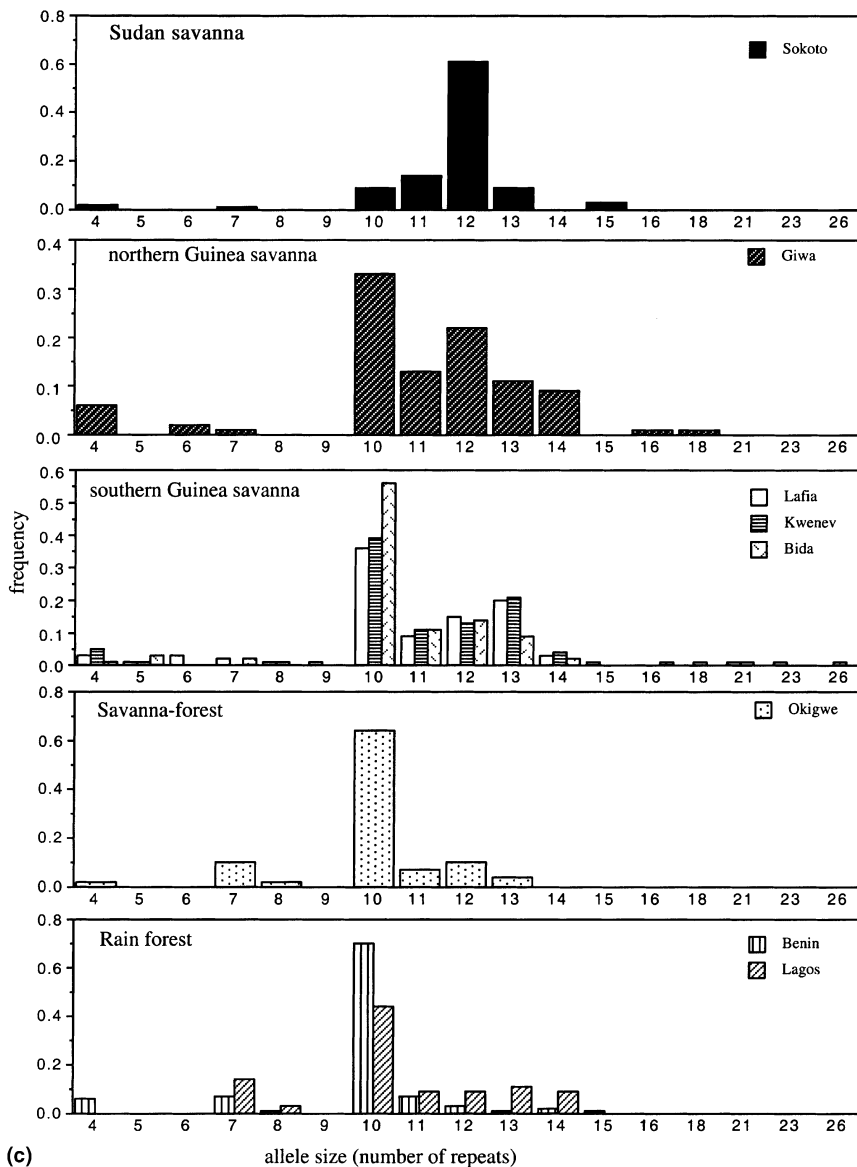


Fig. 2 (Continued).

Heterozygote deficiency at microsatellite loci is usually attributed to null alleles, which are caused by mutations in the primer-binding site. The estimated frequency of null alleles (r) is shown in Table 2. At least one specimen per locality failed to yield PCR products at *AG2H523* and *AG3H88* after two or three attempts. Yet, the same specimens were successfully amplified at other loci. This strongly suggests the existence of null alleles at the two loci above.

Eight of 360 tests for linkage disequilibrium were significant ($P < 0.05$) after sequential Bonferroni correction. None of the tests were significant in Giwa and Bida, whereas all other localities had at least one significant pair. None of the eight significant pairs occurred in more than one locality; physical linkage is

thus an unlikely explanation for the significant results, but the Wahlund effect or assortative mating cannot be ruled out.

Genetic differentiation

Over all loci, genetic differentiation was significant ($P < 0.05$), especially in comparisons between localities in the savanna vs. forest zones – range of F_{ST} 0.028–0.087 and R_{ST} 0.014–0.100 (Tables 3 and 4). In contrast, estimates of genetic differentiation were usually lower within both forest and savanna zones (Tables 3 and 4). In comparisons between localities in the savanna zones, F_{ST} values ranged from 0.000 to 0.048 and none of the R_{ST} values was significant (Tables 3 and 4).

Table 3 Genetic distance, F_{ST} , and Nm (in parentheses) between *Anopheles gambiae* across Nigeria. Values below the diagonal were calculated over 10 loci and those above the diagonal were calculated over seven loci

	Lagos	Benin	Okigwe	Kwenev	Lafia	Bida	Giwa	Sokoto
Lagos	—	0.000 (530.6)	0.017* (14.4)	0.011* (21.2)	0.008* (29.3)	0.007 (31.1)	0.008* (30.3)	0.007 (34.8)
Benin	0.012* (20.3)	—	0.015* (15.5)	0.019* (12.7)	0.001 (226.1)	0.010* (22.8)	0.008* (30.8)	0.009* (26.1)
Okigwe	0.022* (10.8)	0.022* (10.8)	—	0.013* (18.4)	0.001 (127.8)	0.008* (28.7)	0.004 (55.6)	0.024* (9.8)
Kwenev	0.041* (5.7)	0.045* (5.1)	0.031* (7.7)	—	0.010* (24.1)	0.012* (19.2)	0.016* (14.6)	0.011* (20.9)
Lafia	0.045* (5.1)	0.039* (6.1)	0.027* (8.9)	0.015* (16.1)	—	−0.007 (Inf†)	−0.003 (Inf)	0.006 (35.7)
Bida	0.023* (10.5)	0.024* (10.1)	0.009* (27.4)	0.015* (16.2)	0.008* (29.6)	—	0.001 (127.1)	0.009* (26.0)
Giwa	0.062* (3.7)	0.056* (4.1)	0.042* (5.6)	0.023* (10.3)	0.000 (815.3)	0.018* (13.5)	—	0.011* (21.3)
Sokoto	0.079* (2.8)	0.087* (2.6)	0.086* (2.6)	0.033* (7.2)	0.035* (6.7)	0.048* (4.9)	0.032* (7.4)	—

* $P < 0.05$.

†Inf, infinity.

Table 4 Genetic distance, R_{ST} , and Nm (in parentheses) between *Anopheles gambiae* across Nigeria. Values below the diagonal were calculated over 10 loci and those above the diagonal were calculated over seven loci

	Lagos	Benin	Okigwe	Kwenev	Lafia	Bida	Giwa	Sokoto
Lagos	—	−0.013 (Inf†)	0.009 (25.5)	−0.014 (Inf)	−0.005 (Inf)	−0.007 (Inf)	−0.004 (Inf)	−0.009 (Inf)
Benin	0.018* (13.5)	—	0.001 (208.5)	−0.031 (Inf)	−0.008 (Inf)	−0.001 (Inf)	−0.011 (Inf)	−0.019 (Inf)
Okigwe	0.031* (7.6)	0.099* (2.2)	—	0.001 (226.0)	0.002 (117.2)	0.012 (20.2)	−0.003 (Inf)	0.000 (373.5)
Kwenev	0.014* (16.4)	0.069* (3.3)	0.020* (11.9)	—	−0.006 (Inf)	−0.017 (Inf)	−0.004 (Inf)	−0.005 (Inf)
Lafia	0.022* (11.0)	0.100* (2.2)	0.007 (35.3)	−0.001 (Inf)	—	−0.007 (Inf)	−0.010 (Inf)	−0.008 (Inf)
Bida	0.017* (14.3)	0.100* (2.2)	0.011 (22.2)	−0.003 (Inf)	−0.006 (Inf)	—	−0.010 (Inf)	−0.010 (Inf)
Giwa	0.023* (10.3)	0.095* (2.3)	0.006 (41.0)	−0.002 (Inf)	−0.009 (Inf)	−0.007 (Inf)	—	−0.005 (Inf)
Sokoto	0.021* (11.6)	0.092* (2.4)	0.017* (14.1)	−0.004 (Inf)	−0.005 (Inf)	−0.001 (Inf)	−0.004 (Inf)	—

* $P < 0.05$.

†Inf, Infinity.

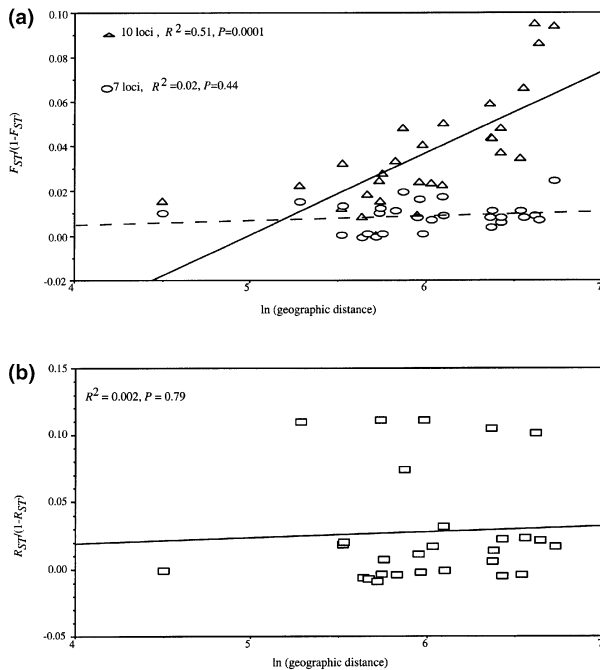


Fig. 3 Isolation by distance in *Anopheles gambiae* across Nigeria based on (a) F_{ST} calculated over 10 loci (bold line) and seven loci (dotted line); and (b) R_{ST} calculated over 10 loci.

Single-locus comparisons between all localities revealed that genetic distances were generally highest and Nm lowest at three loci: *AG2H26*, *AG2H79* and *AG2H637* (data not shown). When these three loci were removed from the data-set, genetic distances calculated over the seven remaining loci were reduced to very low or insignificant levels even between the localities in forest and savanna zones (Tables 3 and 4). The range of F_{ST} values was 0.001–0.024 and all R_{ST} values were no longer significant ($P > 0.05$) in any comparison. These observations indicate that the three loci above were responsible for nearly all of the genetic differentiation.

Isolation by distance

Tests of isolation by distance gave seemingly equivocal results. The test based on $F_{ST}/(1-F_{ST})$ was highly significant ($R^2 = 0.51, P = 0.0001$) (Fig. 3A), whereas the R_{ST} -based test was not significant ($R^2 = 0.002, P = 0.798$) (Fig. 3B). Following Bossart & Prowell (1998), the F_{ST} -based test was further analysed to determine if the significant result was due to the influence of one or a few localities. The analysis was performed by removing, in a stepwise procedure, all the data associated with each locality or ecological zone and then repeating the test. The test was no longer significant when both Benin and Lagos were removed from the data-set ($R^2 = 0.29, P = 0.11$). Moreover, the analysis

performed on data derived from F_{ST} calculated over seven loci (excluding *AG2H26*, *AG2H79* and *AG2H637*) was not significant ($R^2 = 0.02, P = 0.440$) (Fig. 3A).

Discussion

Three loci, *AG2H26*, *AG2H79* and *AG2H637* were responsible for nearly all of the genetic differentiation in this study. *AG2H26* and *AG2H637* are located within inversions *2Rb* and *2La*, respectively; like the microsatellite loci, the frequencies of these inversions vary clinally from north to south in Nigeria (Coluzzi *et al.*, 1979, 1985). *AG2H79* is located within inversion *2Ra* (Coluzzi & Sabatini, 1967), the frequency of which varies clinally across Nigeria in *A. arabiensis*, but not in *A. gambiae* (Coluzzi *et al.*, 1979). Removal of the three loci from the data-set resulted in low or insignificant estimates of differentiation even between localities in savanna and forest zones. The preceding observation suggests that gene flow is extensive across the country but that selection on genes located within some inversions on chromosome II counters the homogenizing effects of gene flow. It is likely that the three microsatellite loci above merely hitch-hike on nearby genes that are under selective pressure. It is unlikely that entire inversions are the units of selection because *AG2H637* and *AG2H523* are both located within inversion *2La*, yet the modal allele at *AG2H523* was the same across localities. Local selection, which probably results in adaptation to the ecological zones (Coluzzi *et al.*, 1979), can result in differentiation by reducing survival and fecundity in immigrants. If, for example, an immigrant does not carry a particular inversion, it may experience reduced survival and reproduction. The extensive genetic exchange measured by parts of the genome that are located outside inversions suggests that migrants survive and reproduce. Furthermore, there is probably recombination among regions outside inversions, such that inversion heterozygous offspring give rise to a mixture of gametes. However, only zygotes that possess the inversion survive and reproduce. The present study is in agreement with Lanzaro *et al.* (1998), who concluded that selection on genes located on chromosome II, but not on the other chromosomes, is responsible for genetic differentiation between the Bamako and Mopti chromosomal forms in Mali.

Removal of *AG2H26*, *AG2H79* and *AG2H637* from the data-set significantly reduced the genetic distances in some comparisons within ecological zones (Benin \times Lagos in the rain forest and Bida \times Lafia in the southern Guinea savanna). These observations suggest that selection is not simply zone-dependent. The correlation of some inversions with indoor vs. outdoor biting and resting behaviour (Coluzzi *et al.*, 1979) and with higher

infection rates with malaria parasites (Petrarca & Beier, 1992) suggests multiple roles for selection.

Over all loci, genetic differentiation was low among localities within each of the savanna and forest zones, although further sampling within the forest zone is required. The largest genetic distances were measured between localities in the savanna and forest zones, so that Nm ranged from 2.2 to 16.4 (from both F_{ST} and R_{ST}). This level of gene flow exceeds the threshold ($Nm < 1$) at which substantial differentiation by genetic drift may accrue (Slatkin, 1987). The findings above are consistent with chromosome inversion data from Nigeria (Coluzzi *et al.*, 1979, 1985): *A. gambiae* samples from forest zones (Forest chromosomal form) are virtually uniform for the standard arrangement on chromosome II (but see della Torre *et al.*, 2001), whereas samples from savanna zones (Savanna chromosomal form) show a diversity of floating inversions on the same chromosome. However, forest and savanna samples do intergrade. The high levels of gene flow over all loci in this study among savanna localities (Nm at least 4.9) is also consistent with the observation of extensive gene flow across Africa within the Savanna chromosomal form (Lehmann *et al.*, 1996).

Although tests of isolation by distance gave seemingly equivocal results, geographical distance does not appear to limit gene flow in *A. gambiae* across Nigeria (Fig. 3). If, in addition to selection, isolation by distance limits gene flow, parts of the genome that are not under selective pressure would be expected to measure significant levels of differentiation that increase with geographical distance. The data show instead that differentiation estimated over seven of 10 loci was low or insignificant and was not correlated with geographical distance.

It is not clear what strategy will be employed for releasing transgenic mosquitoes. Assuming a transposable element is found that is capable of germ line transformation (reviewed in Curtis, 1994; Ashburner *et al.*, 1998), the current study suggests that its spread throughout the *A. gambiae* genome will be rapid as long as the insertion event is not biased towards parts of the genome that are located within inversions. Similarly, if a stable transformed mosquito (reviewed in Curtis, 1994; Ashburner *et al.*, 1998) is released, for example, the present study suggests that the spread of the transgene will be rapid, provided that it is located outside an inversion.

Acknowledgements

We acknowledge Bill Kilpatrick, Alison Brody and Dawn Wesson for discussions and comments on the manuscript. We thank Tovi Lehmann and colleagues

and members of the Conn Laboratory for many useful suggestions. Nick Gotelli and Carl Schlichting advised us on statistics and Tessa Robinson, Tim Hunter and Jan Nicklaus provided technical assistance in the laboratory. Charles Vajime, Ishaya Nock, Iliya Ndams, Alex Akpa and Anyanwu Alaribe assisted in mosquito collection. We thank UNDP/WORLD BANK/WHO for supporting field work. This work was supported by National Institutes of Health Grant AI 40116 to JEC and by the University of Vermont.

References

- ASHBURNER, M., HOY, M. A. AND PELOQUIN, J. J. 1998. Prospects for the genetic transformation of arthropods. *Insect Mol. Biol.*, **7**, 210–213.
- BESANSKY, N. J., LEHMANN, T., FAHEY, G. T., FONTENILLE, D., BRAACK, L. E. O., HAWLEY, W. A. *ET AL.* 1997. Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *A. arabiensis*, suggest extensive gene flow. *Genetics*, **147**, 1817–1828.
- BLACK IV, W. C. AND LANZARO, G. C. 2001. Distribution of genetic variation among chromosomal forms of *Anopheles gambiae* s.s.: introgressive hybridization, adaptive inversions, or recent reproductive isolation? *Insect Mol. Biol.*, **10**, 3–7.
- BOSSART, J. L. AND PROWELL, D. P. 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends Ecol. Evol.*, **13**, 202–206.
- BRUFORD, M. W. AND WAYNE, R. K. 1993. Microsatellites and their application to population genetic studies. *Curr. Opin. Genet. Devel.*, **3**, 939–943.
- BRYAN, J. H., DI DECO, M. A., PETRARCA, V. AND COLUZZI, M. 1982. Inversion polymorphism and incipient speciation in *Anopheles gambiae* s.str. in Gambia, West Africa. *Genetica*, **59**, 167–176.
- CHAKRABORTY, R., DE ANDRADE, M., DAIGER, S. P. AND BUDOWLE, B. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Ann. Hum. Genet.*, **56**, 45–57.
- 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., BESANSKY, N. J. AND FINNERTY, V. 1987. A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *Am. J. Trop. Med. Hyg.*, **37**, 37–41.
- COLUZZI, M., PETRARCA, V. AND DI DECO, M. A. 1985. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Boll. Zool.*, **52**, 45–63.
- COLUZZI, M. AND SABATINI, A. 1967. Cytogenetic observations on species A and B of the *Anopheles gambiae* complex. *Parassitologia*, **9**, 73–88.
- COLUZZI, M., SABATINI, A., PETRARCA, V. AND DI DECO, M. A. 1979. Chromosomal differentiation and adaptation to human environment in the *Anopheles gambiae* complex. *Trans. Roy. Soc. Trop. Med. Hyg.*, **73**, 483–497.

- CURTIS, C. F. 1994. The case for malaria control by genetic manipulation of its vectors. *Parasitol. Today*, **10**, 371–374.
- DAVIES, H. 1977. *Tsetse Flies in Nigeria: A Handbook for Junior Control Staff*. Oxford University Press, Ibadan.
- DELLA TORRE, A., FANELLO, C., AKOGBETO, M., DOSSOU-YOVO, J., FAVIA, G., PETRARCA, V. ET AL. 2001. Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Mol. Biol.*, **10**, 9–18.
- FAVIA, G., LANFRANCOTTI, A., SPANOS, L., SIDEN-KIAMOS, I. AND LOUIS, C. 2001. Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s.s. *Insect Mol. Biol.*, **10**, 19–23.
- GENTILE, G., SLOTMAN, M., KETMAIER, V., POWELL, J. R. AND CACCONE, A. 2001. Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae* s.s. *Insect Mol. Biol.*, **10**, 25–32.
- GILLIES, M. T. AND COETZEE, M. 1987. *A Supplement to the Anophelinae of Africa South of the Sahara*. South African Institute for Medical Research, Johannesburg.
- GUO, S. W. AND THOMPSON, E. A. 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.
- HOLM, S. 1979. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, **6**, 65–70.
- KAMAU, L., MUKABANA, W. R., HAWLEY, W. A., LEHMANN, T., IRUNGU, L. W., ORAGO, A. A. S. ET AL. 1999. Analysis of genetic variability in *Anopheles arabiensis* and *Anopheles gambiae* using microsatellite loci. *Insect Mol. Biol.*, **8**, 287–297.
- LANZARO, G. C., TOURE, Y. T., CARNAHAN, J., ZHENG, L., DOLO, G., TRAORE, S. ET AL. 1998. Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proc. Nat. Acad. Sci. U.S.A.*, **95**, 14260–14265.
- LEHMANN, T., HAWLEY, W. A., KAMAU, L., FONTENILLE, D., SIMARD, F. AND COLLINS, F. H. 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. *Heredity*, **77**, 192–208.
- MUKABAYIRE, O., CARIDI, J., WANG, X., TOURE, Y. T., COLUZZI, M. AND BESANSKY, N. J. 2001. Patterns of DNA sequence variation in chromosomally recognized taxa of *Anopheles gambiae*: evidence from rDNA and single-copy loci. *Insect Mol. Biol.*, **10**, 33–46.
- PETRARCA, V. AND BEIER, J. C. 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.*, **49**, 520–529.
- ROUSSET, F. 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
- SCHNEIDER, S., ROESSLI, D. AND EXCOFFIER, L. 2000. ARLEQUIN ver 2.000: *A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- 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.
- SHAPIRO, S. S. AND WILK, M. B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.
- SLATKIN, M. 1987. Gene flow and the geographic 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.
- TABACHNICK, W. J. AND BLACK IV, W. C. 1995. Making a case for molecular population genetic studies of arthropod vectors. *Parasitol. Today*, **11**, 27–30.
- TOURE, Y. T., PETRARCA, V., TRAORE, S. F., COULIBALY, A., MAIGA, H. M., SANKARE, O. ET AL. 1998. The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia*, **40**, 477–511.
- WEIR, B. S. AND COCKERHAM, C. C. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- WRIGHT, S. 1931. Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- ZAR, J. H. 1996. *Biostatistical Analysis*. Prentice Hall, Upper Saddle River, NJ.
- ZHENG, L., BENEDICT, M. Q., CORNEL, A. J., COLLINS, F. H. AND KAFATOS, F. C. 1996. An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. *Genetics*, **143**, 941–952.