Insect Molecular Biology (2000) **9**(5), 467–479

Persistence of *Anopheles arabiensis* during the severe dry season conditions in Senegal: an indirect approach using microsatellite loci

Frédéric Simard,*† Tovi Lehmann,† Jean-Jacques Lemasson,* Mathurin Diatta* and Didier Fontenille‡

**Laboratoire de Zoologie Médicale, French Institut de Recherche pour le Développement (IRD), Institut Pasteur, Dakar, Sénégal,* †*Division of Parasitic Diseases, Centers for Disease Control and Prevention, Chamblee, GA, USA,* ‡*Laboratoire de l'Institut de Recherche pour le Développement (IRD), Organisation de lutte Contre les grandes Endémies en Afrique Centrale (OCEAC), Yaoundé, Cameroun*

Abstract

Variation at nine microsatellite loci was investigated to understand how *Anopheles arabiensis* populations survive the dry season in the sahelian region of Senegal. Low estimates of genetic differentiation (F_{ST} = 0.012, R_{ST} = 0.009) between two populations, 250 km apart, suggested extensive gene flow across this distance. Despite extreme seasonal fluctuation in abundance with dry season minima in which mosquitoes virtually disappeared, allele frequencies remained stable over time in the village of Barkedji from August 1994 to December 1997 (including four rainy seasons and three dry seasons). The effective population size (Ne) was estimated to be 601 with 95% CI (281, 1592), providing strong evidence against annual bottlenecks. Differences in measures of genetic diversity and linkage disequilibrium between the dry and the rainy seasons were not detected. These results suggest that despite extreme minima in local density, *An. arabiensis* maintains large permanent deme spread out over large area.

Keywords: *Anopheles arabiensis*, malaria, dry season, microsatellites, effective population size.

Received 23 February 2000; accepted following revision 8 June 2000. Correspondence: Frédéric Simard, Division of Parasitic Diseases, Centers for Diseases Control and Prevention, MS F-22, 4770 Buford Highway, Chamblee, GA 30341, USA. Tel.: (770) 488 4634; fax: (770) 488 4258; e-mail: fes3@cdc.gov

Introduction

Malaria transmission in sub-Saharan Africa is mainly due to mosquito vectors of the *Anopheles gambiae* complex. Among the seven species formally described (Gillies & Coetzee, 1987; Hunt *et al*., 1998), *An. gambiae* and *An. arabiensis* are the most efficient malaria vectors. In the dry savannas of Africa, vector populations typically display strong seasonal fluctuations in abundance, being present in large numbers during the rainy season and dropping to very low levels when breeding sites dry up (Taylor *et al*., 1993; Charlwood *et al*., 1995; Lemasson *et al*., 1997). In some areas, no specimens can be found during the dry season. The low tolerance of eggs to desiccation (Beier *et al*., 1990) implies that only adults can survive the dry season, possibly via estivating females that do not reproduce (Omer & Cloudsley-Thompson, 1968, 1970). Nevertheless, soon after the onset of rains, when breeding sites fill with water, mosquito numbers increase rapidly. It is still unknown how these populations pass the dry season in such arid areas.

Scenarios depicting the possible dry season dynamics include the following. (i) Local populations go through dry season bottlenecks, being sustained by few survivors. (ii) Populations become locally extinct during the dry season and re-colonized at the beginning of the rainy season by few migrants from adjacent areas where permanent breeding is allowed (i.e. stable source populations). (iii) Populations become locally extinct (or nearly so) during the dry season and re-colonized by 'mass migration' or expansion of population(s) inhabiting 'stable' areas. (iv) Large populations do survive locally, while hidden with respect to sampling. Finally (v) extensive mobility of individuals can maintain large population even if local densities are very low. Such a 'diffused deme', which is spread over large area was proposed for *An. gambiae* in Kenya (Lehmann *et al*., 1998). These scenarios have different expected outcomes in terms of the genetic structure of populations and the temporal stability of the genetic composition within populations.

Anopheles arabiensis is the species that inhabits the driest sahelian savannas occupied by the *An. gambiae*

complex. It virtually disappears during the long and severe dry season in these areas, and thus it best represents the enigmatic dry season biology of several African vector species. Knowledge of the dry season biology of this mosquito has implications for understanding malaria epidemiology and for malaria control. For example, if the target population undergoes recurrent events of extinction and colonization from one (or several) source population(s), control efforts should concentrate on the source populations or prevent the establishment and proliferation of the migrants, but if the population is locally stable and geographically isolated, control efforts should be confined to its own limits. Additionally, such knowledge is relevant to understanding parasite transmission dynamics and its population structure (Gupta & Day, 1994), and to the development of genetic strategies of vector control using genetically engineered mosquitoes with reduced vectorial competence (e.g. Collins & Besansky, 1994; Crampton *et al*., 1994).

The direct (ecological) approach to study the dry season dynamics of mosquitoes using mark–release–recapture experiments has little use in the dry season, when no mosquitoes (or only few) are found. One indirect approach relies on the relationship between temporal variation in allele frequencies (genetic drift) and population size, such that large fluctuations are expected in small populations, while minor changes would occur in large populations (Waples, 1991; Taylor *et al*., 1993). The effective population size, denoted Ne, summarizes the effect of random genetic drift on populations. Ne is defined as the size of an ideal population (i.e. a hypothetical panmictic population with nonoverlapping generations, in which the sex ratio is 1 and each individual has the same reproductive potential) that experiences drift at the same rate as the natural population under consideration (Wright, 1931, 1938; Avise, 1994). When the population size varies over generations, Ne approximates the harmonic mean of the effective population sizes in all individual generations, and hence is dominated by the smallest value (Nei & Tajima, 1981; Pollak, 1983). In other words, Ne is sensitive to bottlenecks (Hartl & Clark, 1997).

We focused on *An. arabiensis* populations sampled in Barkedji, a village in the Sahelian region of Senegal (West Africa), in which the severe dry season lasts more than 7 months (Lemasson *et al*., 1997) and in Dielmo, a village located 250 km southwards where breeding sites are available all year round (Fontenille *et al*., 1997). Spatial variation in allele frequencies at nine microsatellite loci was investigated between these two populations to assess the level of differentiation and migration on this geographical scale. Temporal stability of allelic frequencies was monitored during 4 years in Barkedji, where the dry season is extremely severe. Estimates of Ne and genetic diversity in this population were used to evaluate the likelihood of the different scenarios described above.

Results

Genotypes of 49–54 females were scored at nine microsatellite loci in each sample. *Anopheles arabiensis* mosquitoes from the villages of Barkedji and Dielmo (Fig. 1) showed high polymorphism in all loci except locus 49 (that was nearly fixed for one allele with frequency of the most common allele > 0.95 and was therefore discarded from further analysis). The number of alleles per locus varied among loci (from three at locus 29C to sixteen at locus 26) but the average values were similar in all samples (Tables 1 and 2). Allele frequencies were also similar between locations and across time in Barkedji.

Genetic variability within populations

Both Barkedji and Dielmo were sampled in August 1996, therefore conformance with Hardy–Weinberg expectations was first evaluated in these samples as a prerequisite for the spatial comparison. Most loci conformed to Hardy–Weinberg expectations in both populations, but significant deviations were observed (after the sequential Bonferroni procedure was applied to take into account the number of tests within each population) at locus 45C in the Barkedji population and at loci 24D and 141 in the Dielmo population (Table 1). Heterozygote excess resulted in departure from Hardy–Weinberg in locus 45C, while homozygote excess resulted in departures at loci 24D and 141. No significant association between loci was detected by linkage disequilibrium tests in Barkedji or in Dielmo, suggesting random mating within population and independence of the loci.

Across eight time points in Barkedji, twenty single-locus tests were significant out of 64, indicating departures from Hardy–Weinberg equilibrium (Table 2; $P < 10^{-4}$, binomial test with 0.05 'success' rate). Multiple tests on the same loci using samples separated by few generations from the same population are not independent and may reflect the same deviation(s) in a particular locus (or loci), thus inflating significance. For example, deviations from equilibrium due to null alleles (i.e. non amplifying alleles, see Callen *et al*., 1993) would be carried from generation to generation resulting in a high number of significant tests that are locus specific. To assess the heterogeneity between loci, we estimated the 95% confidence interval (CI) around each locus by bootstrapping the F_{IS} values of each locus across time points. Positive mean F_{IS} values were found (i.e. 95% CI above 0, given 10 000 replicates) in all five loci on chromosome II (24D, 147, 141, 26 and 803), negative value was found in locus 45C, and the values of loci 7 and 29C were insignificant (Table 2). These deviations remained significant at the multitest level, indicating that heterozygote deficiency was true for all loci on chromosome II, but not necessarily a genome-wide pattern. No difference was found among time points (ANOVA, $F = 0.33$,

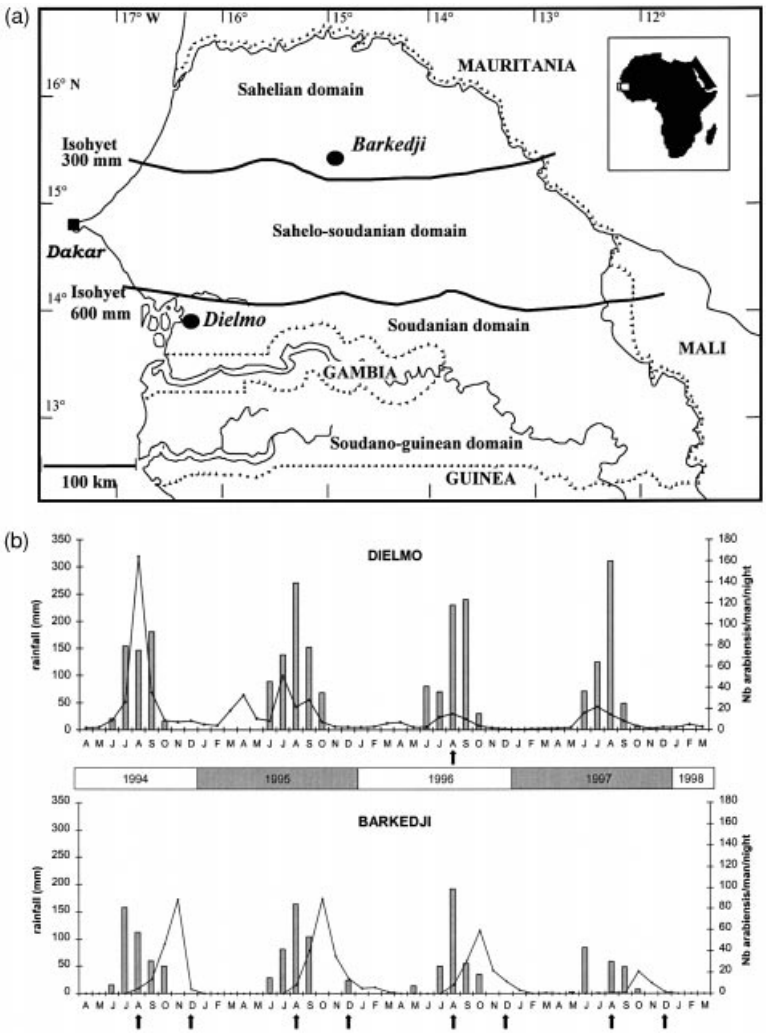

Figure 1. (a) Geographic location of the study sites in Senegal, the villages of Barkedji and Dielmo. (b) Monthly total rainfall (bars) and abundance of *An. arabiensis* (solid line: monthly mean number of bites per man per night based on indoor and outdoor human-bated collections, see Lemasson *et al.*, 1997 and Fontenille *et al.*, 1997) in the villages of Dielmo (above) and Barkedji (below) from April 1994 to March 1998. Arrows indicate sampling events.

Table 1. Genetic variability and significance level for goodness of fit tests to Hardy–Weinberg equilibrium within *An. arabiensis* populations from Barkedji and Dielmo

		Locus								Mean
Populations		7	24D	147	141	26	803	45C	29C	over loci
Barkedji	Nall	8	5	6	8	14	7	6	2	7
2n = 100	H_{obs}	0.620	0.360	0.660	0.520	0.813	0.600	0.460	0.140	0.522
	H_{exp}	0.587	0.489	0.719	0.636	0.884	0.621	0.419	0.165	0.565
	F_{IS}	−0.057	**±0.265**	+0.083	+0.184	+0.081	**+0.035**	**−0.098**	+0.155	+0.078
	r	–	0.087	0.034	0.071	0.038	0.013	–	0.021	nc
Dielmo	Nall	8	7	6	6	15	8	6	2	7.25
2n = 98	H_{obs}	0.653	0.388	0.563	0.469	0.783	0.673	0.429	0.163	0.515
	H_{exp}	0.692	0.587	0.739	0.646	0.856	0.774	0.419	0.151	0.608
	F_{is}	+0.057	**+0.342**	**±0.241**	**+0.276**	+0.087	+0.131	−0.022	−0.079	+0.155
	r	0.023	0.125	0.101	0.108	0.039	0.057	–	–	nc
Pooled	Nall	10	8	6	8	16	8	7	2	8.125
populations	H_{obs}	0.636	0.374	0.612	0.495	0.797	0.636	0.444	0.152	0.517
2n = 198	H_{exp}	0.640	0.546	0.727	0.648	0.872	0.705	0.421	0.158	0.590
	F_{IS}	+0.006	**+0.316**	**±0.158**	**+0.236**	+0.085	+0.098	−0.055	+0.040	+0.121
	r	0.002	0.111	0.067	0.093	0.040	0.040	–	0.005	nc

Samples were collected in both locations in August 1996. *2n*, number of chromosomes scored; Nall, number of alleles per locus; H_{obs}, observed heterozygosity; H_{exp}: expected heterozygosity (Nei, 1978). *r,* null alleles frequency estimated as in Brookfield 1996 (–: irrelevant because of heterozygote excess; nc, not calculated for the mean over loci). F_{IS} is calculated according to Weir & Cockerham (1984). Note that when populations are pooled, $F_{IS} = F_{IT}$. Bolded values: $P < 0.05$; bolded underlined values: $P < 0.05$ after application of the sequential Bonferroni procedure.

Table 2. Genetic variability within *An. arabiensis* in the village of Barkedji.

Sampling date		Locus 7	24D	147	141	26	803	45C	29C	Mean over loci
Aug 94	Nall	9	5	7	6	15	7	7	2	7.25
2n = 108	H_{obs}	0.604	0.481	0.667	0.519	0.833	0.660	0.537	0.167	0.558
	H_{exp}	0.749	0.648	0.773	0.594	0.915	0.714	0.539	0.154	0.636
	F_{IS}	**+0.195**	**+0.259**	+0.138	+0.128	+0.090	+0.076	+0.004	–0.082	+0.122
	r	0.083	0.101	0.060	0.047	0.043	0.032	0.001	–	nc
Dec 94	Nall	8	8	6	7	13	7	7	2	7.25
2n = 100	H_{obs}	0.600	0.380	0.680	0.460	0.826	0.680	0.460	0.260	0.543
	H_{exp}	0.722	0.577	0.768	0.583	0.861	0.736	0.435	0.228	0.614
	F_{IS}	**+0.170**	**+0.343**	+0.115	**+0.213**	**+0.041**	+0.077	–0.057	–0.140	+0.117
	r	0.071	0.125	0.050	0.078	0.019	0.032	–	–	nc
Aug 95	Nall	7	6	7	7	13	7	6	2	6.87
2n = 100	H_{obs}	0.680	0.560	0.520	0.420	0.841	0.660	0.520	0.300	0.563
	H_{exp}	0.685	0.561	0.757	0.535	0.876	0.674	0.477	0.258	0.603
	F_{IS}	+0.007	+0.001	**+0.315**	**+0.216**	**+0.041**	+0.021	–0.090	–0.167	+0.068
	r	0.003	0.001	0.135	0.075	0.019	0.008	–	–	nc
Dec 95	Nall	9	6	6	8	11	7	6	3	7
2n = 100	H_{obs}	0.640	0.460	0.580	0.540	0.694	0.600	0.520	0.200	0.529
	H_{exp}	0.654	0.531	0.712	0.687	0.846	0.625	0.511	0.216	0.598
	F_{IS}	+0.021	+0.134	+0.186	+0.216	**+0.181**	+0.041	–0.018	+0.073	+0.115
	r	0.008	0.046	0.077	0.087	0.082	0.015	–	0.013	nc
Aug 96	Nall	8	5	6	8	14	7	6	2	7
2n = 100	H_{obs}	0.620	0.360	0.660	0.520	0.813	0.600	0.460	0.140	0.522
	H_{exp}	0.587	0.489	0.719	0.636	0.884	0.621	0.419	0.165	0.565
	F_{IS}	–0.057	**+0.265**	+0.083	+0.184	+0.081	**+0.035**	**–0.098**	+0.155	+0.078
	r	–	0.087	0.034	0.071	0.038	0.013	–	0.021	nc
Dec 96	Nall	5	6	5	8	15	7	6	2	6.75
2n = 100	H_{obs}	0.640	0.360	0.680	0.500	0.800	0.640	0.460	0.280	0.545
	H_{exp}	0.668	0.464	0.705	0.619	0.874	0.709	0.451	0.243	0.592
	F_{IS}	+0.042	+0.226	+0.036	**+0.193**	+0.085	+0.098	–0.021	–0.153	+0.079
	r	0.017	0.071	0.015	0.074	0.039	0.040	–	–	nc
Aug 97	Nall	8	5	5	10	14	6	5	2	6.87
2n = 100	H_{obs}	0.776	0.400	0.540	0.551	0.837	0.640	0.340	0.180	0.533
	H_{exp}	0.738	0.413	0.724	0.689	0.899	0.660	0.337	0.165	0.578
	F_{IS}	–0.051	+0.033	**+0.256**	**+0.202**	**+0.070**	+0.031	–0.010	–0.089	+0.079
	r	–	0.009	0.107	0.082	0.033	0.012	–	–	nc
Dec 97	Nall	9	6	6	6	13	6	7	2	6.87
2n = 100	H_{obs}	0.500	0.300	0.700	0.520	0.735	0.600	0.420	0.160	0.492
	H_{exp}	0.573	0.485	0.754	0.630	0.890	0.624	0.423	0.149	0.566
	F_{IS}	+0.128	**+0.384**	**+0.072**	+0.176	**+0.176**	+0.039	+0.008	–0.077	+0.132
	r	0.046	0.125	0.031	0.067	0.082	0.015	0.002	–	nc
Mean over sampling dates	Nall	7.87	5.87	6	7.5	13.5	6.75	6.25	2.12	7
	H_{obs}	0.645	0.423	0.636	0.490	0.803	0.624	0.438	0.210	0.533
	H_{exp}	0.669	0.545	0.731	0.631	0.887	0.661	0.426	0.196	0.593
	F_{IS} †	+0.057	**+0.206**	**+0.155**	**+0.191**	**+0.096**	**+0.052**	**–0.035**	–0.060	+0.099
	r	0.014	0.079	0.055	0.086	0.045	0.022	–	–	nc

2n, number of chromosomes scored; Nall, number of alleles per locus; H_{obs}, observed heterozygosity; H_{exp}, expected heterozygosity (Nei, 1978). *r*, null alleles frequency estimated as in Brookfield 1996 (–: irrelevant because of heterozygote excess; nc, not calculated for the mean over loci). F_{IS} is calculated according to Weir & Cockerham (1984). †Statistical significance of the mean F_{IS} was assessed by bootstraping across time points (10 000 replicates). Bolded values: $P < 0.05$; bolded underlined values: $P < 0.05$ after correction for multiple tests.

df = 7,56 $P > 0.93$), even after all F_{IS} values were normalized with the mean and standard deviation within locus (ANOVA, $F = 0.82$, df = 7,56 $P > 0.57$).

If heterozygote deficits were caused by population substructure (i.e. the Wahlund effect) or inbreeding, it is unlikely that all eight samples consisted of the same composition of the subpopulations. Moreover, we would expect deviations to be found throughout the genome. The clustering of loci with deviations only on chromosome II is therefore inconsistent with such explanations. Likewise, substructure and inbreeding would be detected by linkage disequilibrium because members of the different subpopulations would have different probabilities to carry certain combinations of alleles. Only thirteen out of 224 linkage disequilibrium tests were significant (5.8%), which is insignificant (binomial test, $P > 0.33$). Interestingly, loci mapped on chromosome II that deviated from Hardy–Weinberg were in a state of linkage equilibrium. Altogether, these results are inconsistent with a Wahlund effect or inbreeding. We believe therefore that the Barkedji population is homogenous and that locus specific constraints were the cause of the deviations observed.

Null alleles have been reported in every paper about *An. gambiae* and *An. arabiensis* using microsatellite loci (e.g. Lanzaro *et al.*, 1995; Lehmann *et al.*, 1996a,b; Kamau *et al.*, 1998; Lanzaro *et al.*, 1998; Walton *et al.*, 1998; Donnelly *et al.*, 1999; Simard *et al.*, 1999). They produce heterozygote deficiency because an individual heterozygote for a null allele will be scored as homozygote for the other (amplified) allele. Individuals that carry null alleles on both chromosomes will fail to amplify PCR products at this locus. Indeed, we were unable to obtain a PCR product for a few individuals at a particular locus, despite successful amplification at other loci. Consistent with our results, linkage disequilibrium is not expected as all individuals are equally likely to carry a null allele. Following Brookfield (1996), we estimated null alleles frequency for each locus in each sample (*r*, Tables 1 and 2). Frequencies up to 13.5% were estimated. Why were heterozygote deficits clustered in chromosome II? This could merely be a chance event and we will carefully evaluate the influence of null alleles and heterogeneity among loci in all further analyses. Alternatively, polymorphic chromosomal inversions under selection could result in nonrandom genotype distribution. This hypothesis will be extensively tested below.

Differentiation between populations

Single locus F_{ST} and R_{ST} between Barkedji and Dielmo were low and insignificant, leading to high estimates of gene exchange (*Nm*, see Table 3). Values of differentiation indices varied little between loci ($F_{ST} \leq 0.029$ and $R_{ST} \leq 0.056$) indicating that it was a genome-wide pattern. Average F_{ST} was 0.012 ($P = 0.013$, $Nm = 21.4$) and average R_{ST} was 0.009 ($P = 0.20$, $Nm = 28.1$), suggesting high gene flow on that geographical scale. Nevertheless, the significantly positive mean F_{ST} indicates that despite high gene flow, these two populations do not belong to the same panmictic unit.

Table 3. Genetic differentiation between *An. arabiensis* populations from Barkedji and Dielmo.

Locus	F_{ST}	*P*	*Nm*	R_{ST}	*P*	*Nm*
7[a]	0.005	0.47	68	−0.005	0.61	∞
24D	0.029	0.14	8.4	−0.004	0.59	∞
147	−0.007	0.93	∞	−0.010	0.99	∞
141	0.018	0.18	13.5	0.056	0.054	4.2
26	0.002	0.70	127	0.004	0.38	58
803	0.022	0.16	11.2	0.007	0.30	35.8
45C	0.009	0.31	26.1	0.013	0.27	19.5
29C	−0.001	0.87	∞	−0.010	0.85	∞
All loci	**0.012**	0.013	21.4	0.009	0.20	28.1

P refers to the statistical significance of F_{ST} and R_{ST}, based on 1000 permutations procedure (Schneider *et al.*, 1997). Bolded values: $P < 0.05$.
[a]*Nm* values for locus 7 (on the X chromosome) are adjusted assuming that *Ne* for the X chromosome is three-quarters that for the autosomes.

Temporal stability and effective population size in Barkedji

Allele frequencies distributions appeared stable across time (Fig. 2). Differences in allele frequencies were insignificant across all time points in all loci (Fisher's exact test: $P > 0.16$) except in loci 24D, 26 and 7 ($P < 0.001$), particularly when the sample of August 1994 was included in pair-wise comparisons. In contrast to expectations based on the extreme population minima during the dry season, moderate to high estimates of Ne were obtained for all time intervals as well as for the total period (Table 4). As commonly observed in large populations, the upper 95% confidence limit was infinity for almost all single-locus estimates (63/64) and for all seven seasonal multilocus estimates except one. However, when estimated across the entire time interval (i.e. August 1994 to December 1997), estimates of Ne were more consistent and their 95% CIs were well defined.

The effect of null alleles (above) on our Ne estimates was evaluated. Mutations would not change the frequency of null alleles during the time frame of our study. Changes in null allele frequencies therefore would result from drift (assuming that our microsatellite loci are neutral). Since we could not score null alleles (which did not amplify), the number of chromosomes that were effectively scored was reduced by the fraction of nulls. Ignoring any allele class (including null alleles) would result in a reduced statistical power but would not lead to a biased estimate of Ne. However, to consider that the number of chromosomes scored was twice the number of individuals in our sample would inflate the actual sample size. Thus,

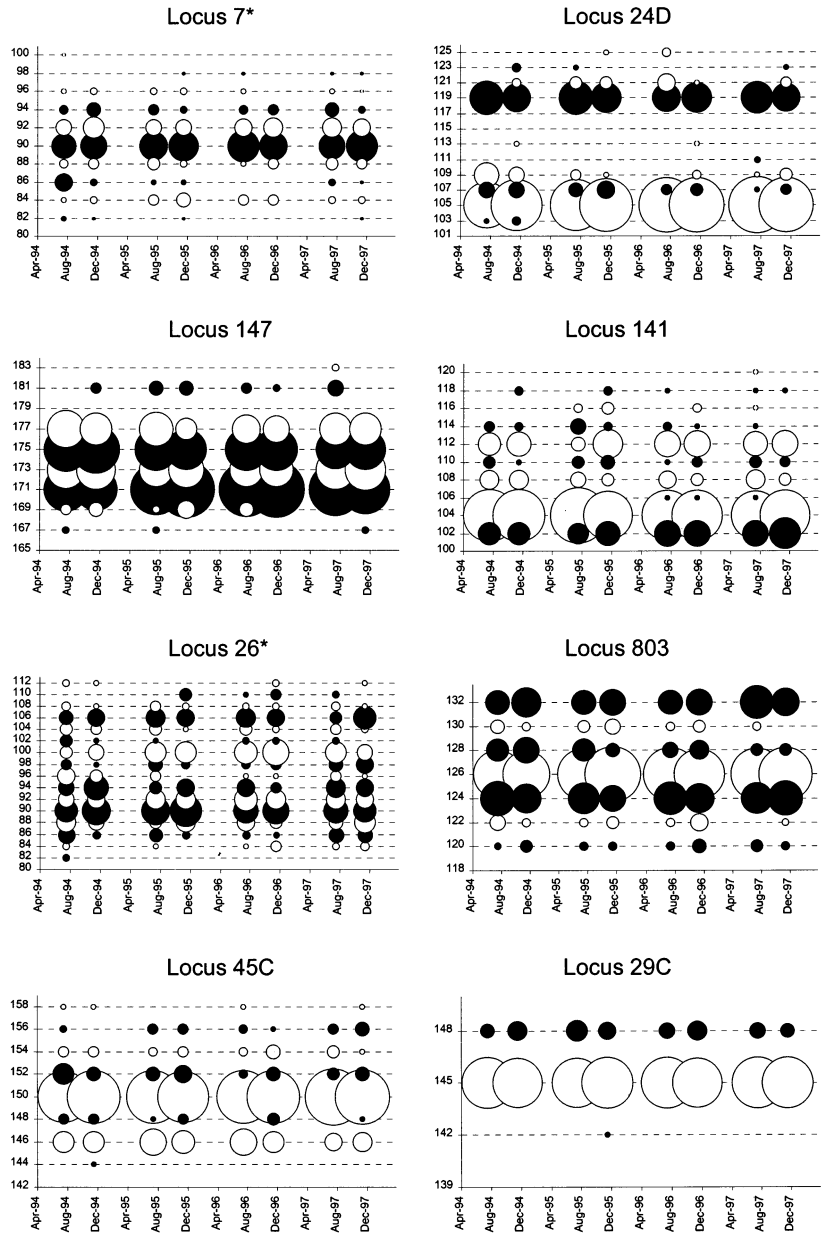

Figure 2. Schematic representation of relative allelic frequencies observed at eight microsatellite loci in the *An. arabiensis* population from Barkedji, at eight consecutive sampling dates from August 1994 to December 1997 (x-axis). 'April' months stand as dry season indicators. Each chart represents data from one locus, named in the upper part. Alleles are denoted by their total size in base pairs on the y-axis. The surface area of each circle is directly proportional to the frequency of the allele in the sample at the sampling date considered. *Surface area of circles has been halved when too many alleles were present at one locus to optimize clarity.

we adjusted our Ne estimates using a corrected sample size, i.e. by removing a fraction *r* (equal to the estimated null alleles frequency, see Table 2) of the number of chromosomes scored in each locus at each time point (*2n*, see Table 2), and presented the adjusted Ne estimates in Table 4.

To consider the possibility of estivation (no oviposition) during the dry season (December–August), resulting in only two generations during this period instead of eight, Ne was also computed with the number of generations adjusted accordingly. The corresponding Ne estimated over all loci (601) was slightly lower than our previous estimate assuming twelve generations a year (Ne = 1046), and their 95% CIs substantially overlapped (Table 4). Thus, the lowest and highest bounds of the 95% CIs of Ne obtained for the total time length were 281 and 2768, respectively.

Chromosome II carries several polymorphic inversions in *An. gambiae* and *An. arabiensis*. Alternative arrangements of *An. gambiae* were found to be under strong selection by environmental conditions, showing geographical clines associated with aridity and cyclic change in frequency between the rainy and dry season (Coluzzi *et al*., 1979; Touré *et al*., 1994, 1998a). Locus 24D is located within inversion *2La*, which is fixed (monomorphic) in *An. arabiensis*. Loci 803 and 147 are outside known inversions (at least two subdivisions away from the major breakpoints,

Table 4. Effective population size (Ne) estimates based on F_k computations (Pollak, 1983)

	Time Interval							Total Time Period	
Locus	Aug 94–Dec 94	Dec 94–Aug 95	Aug 95–Dec 95	Dec 95–Aug 96	Aug 96–Dec96	Dec 96–Aug 97	Aug 97–Dec 97	Aug 94–Dec 97	Aug 94–Dec 97[†]
7	63 (11–449)	871 (56–∞)	1222 (34–∞)	∞ (134–∞)	187 (18–∞)	463 (43–∞)	236 (26–∞)	435 (83–1845)	250 (48–1061)
24D	176 (17–∞)	∞ (77–∞)	∞ (55–∞)	∞ (37–∞)	134 (13–∞)	∞ (40–∞)	697 (19–∞)	569 (62–45 284)	327 (36–26 038)
147	∞ (58–∞)	∞ (61–∞)	247 (16–∞)	∞ (110–∞)	∞ (20–∞)	510 (24–∞)	∞ (51–∞)	∞ (436–∞)	∞ (251–∞)
141	∞ (49–∞)	303 (33–∞)	84 (12–∞)	1716 (60–∞)	∞ (62–∞)	∞ (92–∞)	1108 (20–∞)	2139 (161–∞)	1230 (92–∞)
26	123 (25–∞)	406 (60–∞)	171 (29–∞)	2367 (97–∞)	∞ (74–∞)	623 (78–∞)	299 (47–∞)	1238 (268–∞)	712 (154–∞)
803	∞ (30–∞)	∞ (131–∞)	∞ (38–∞)	∞ (117–∞)	∞ (36–∞)	227 (23–∞)	∞ (52–∞)	954 (113–∞)	549 (65–∞)
45C	∞ (24–∞)	993 (41–∞)	∞ (87–∞)	441 (33–∞)	172 (15–∞)	609 (26–∞)	1566 (23–∞)	4744 (211–∞)	2728 (121–∞)
29C	556 (1–∞)	∞ (1–∞)	∞ (3–∞)	∞ (8–∞)	460 (1–∞)	767 (1–∞)	∞ (2–∞)	∞ (152–∞)	∞ (87–∞)
All loci	229 (82–41 119)	1442 (258–∞)	460 (111–∞)	∞ (450–∞)	753 (131–∞)	762 (187–∞)	809 (151–∞)	**1046 (490–2768)**	**601 (281–1592)**
Chromosome II[§]	466 (95–∞)	1698 (220–∞)	268 (74–∞)	∞ (360–∞)	∞ (156–∞)	1056 (172–∞)	1464 (140–∞)	**1343 (503–7210)**	**772 (289–4146)**
Chromosome X + III	95 (23–1187)	1131 (103–∞)	∞ (99–∞)	∞ (147–∞)	169 (32–∞)	456 (67–∞)	390 (51–∞)	**680 (195–3022)**	**391 (112–1738)**
Within Inversions[¥]	248 (48–∞)	363 (82–∞)	125 (34–8699)	2102 (146–∞)	∞ (145–∞)	3711 (140–∞)	374 (65–∞)	1426 (376–∞)	820 (216–∞)
Outside Inversions	219 (64–∞)	∞ (328–∞)	∞ (189–∞)	∞ (449–∞)	279 (69–∞)	500 (118–∞)	4503 (142–∞)	899 (360–2884)	517 (207–1658)

Twelve generations a year was assumed. August to December (Aug–Dec) time intervals encompass the rainy season and December to August (Dec–Aug) intervals encompass the dry season. 95% confidence intervals are shown in parenthesis. ∞, infinity. [†]Ne has been estimated assuming two generations during the December to August dry seasons time intervals.
[§]Loci on the second chromosome are loci 24D, 147, 141, 26 and 803. [¥]Loci within polymorphic paracentric inversions are loci 26 and 141.

see Zheng *et al*., 1996). Only locus 26 is mapped within the polymorphic inversion 2R*b*, whereas locus 141 may be linked to inversion 2R*d* although its exact cytological location is not clear (see Zheng *et al*., 1996). Unfortunately, no karyotypic information was available for the specimens included in our study. Previous cytological studies detected no departure from panmixia for *An. arabiensis* in this area (Petrarca *et al*., 1987). Variation in microsatellite loci within and adjacent to polymorphic inversions may be confounded by selection on other genes within the inversion due to reduced recombination, a process called 'hitchhiking' (Slatkin, 1995a; see Lanzaro *et al*., 1995, 1998). If such effects encompass the whole chromosome or at least a region that includes several of our loci, we would expect to find linkage disequilibrium between those loci, as was reported by Lanzaro *et al*. (1998) in *An. gambiae*. This pattern was not observed even when we pooled all eight samples from Barkedji. If selection favours certain chromosomal arrangements during the rainy season and alternative arrangements during the dry season, then allele frequencies may change more than expected by drift alone. However, no seasonal (cyclical) pattern in allelic composition and allele frequencies distribution was observed: samples collected in the rainy season (August) were not more similar to each other than to samples collected in the dry season (December, Fig. 1b). Values of Ne estimated from 'August to August' or 'December to December' time intervals were not higher than estimates obtained from 'August to December' and 'December to August' intervals (not shown). Ne estimates based on loci on chromosome II were slightly higher than estimates obtained from the three informative loci mapped on chromosome X (locus 7) and chromosome III (loci 45C and 29C) but their 95% CIs substantially overlapped. The same pattern was observed when comparing loci *within* polymorphic inversions in *An. arabiensis* (e.g. locus 26 and 141) to loci *outside* inversions (Table 4). This trend is inconsistent with rapid cyclical changes in inversions frequencies mediated by selection. This evidence suggests that the influence of inversions on Ne estimated from our loci was negligible. However, even if we exclude all loci on chromosome II, the estimated Ne is inconsistent with dry season bottlenecks (Ne = 391, with 95% CI [112–1738], Table 4).

To evaluate whether mutation, which is expected to be high in microsatellite loci, contributed substantially to the variation in allele frequencies, values of F_{ST} between time points (sensitive especially to drift) were compared with corresponding R_{ST} values (sensitive especially to mutation). Waples (1989a) pointed out that exact tests of homogeneity of allele (or genotype) frequencies and F_{ST} (and R_{ST}) are biased when temporally spaced samples of the same population are compared, but we calculated these statistics solely to assess the role of mutation. Comparing each time point to the others, F_{ST} ranged

	Aug 94	Dec 94	Aug 95	Dec 95	Aug 96	Dec 96	Aug 97	Dec 97
Aug 94	–	−0.006	−0.009	0.001	0.001	0.012	0.001	−0.004
Dec 94	0.005	–	−0.001	−0.001	−0.003	0.007	−0.009	−0.010
Aug 95	0.005	0.002	–	−0.001	−0.008	−0.001	0.001	−0.007
Dec 95	**0.017**	**0.009**	0.002	–	−0.001	0.008	−0.003	0.002
Aug 96	**0.019**	**0.009**	0.001	−0.003	–	−0.002	−0.002	−0.003
Dec 96	**0.015**	0.005	−0.001	−0.003	−0.003	–	0.007	0.003
Aug 97	**0.014**	0.003	0.004	0.005	0.003	−0.002	–	−0.006
Dec 97	**0.014**	**0.008**	0.003	0.003	−0.002	0.002	0.002	–

Bolded values: $P < 0.05$.

Table 5. Pair-wise F_{ST} (below diagonal) and R_{ST} (above diagonal) among temporal samples of the *An. arabiensis* population from Barkedji

between −0.003 and 0.019 with eight significant values out of twenty-eight, while R_{ST} ranged between −0.010 and 0.012 and none was significant (Table 5). This pattern suggests that mutation has not contributed much to the changes in allele frequencies and that these changes were caused by genetic drift (aside from sampling error).

Further insight into the dry season population dynamics may be provided by changes in indices of genetic diversity within and between years. Scenarios involving migration into a small resident population would be accompanied by an increase in genetic diversity during the rainy season because migrants from a distinct gene pool would carry some unique or rare alleles with respect to the small resident population. We would expect this trend to be enhanced as a result of genetic drift during the dry season (population size reduction). The changes in the average observed number of alleles per locus (Nall), average observed heterozygosity (Hobs) and average expected heterozygosity (Hexp) over all loci are shown in Fig. 3. While all three indices slightly declined over the 4-year period, there was no obvious seasonal pattern. Paired-samples Wilcoxon signed-ranks tests (between two consecutive time points for each locus) were insignificant for each of the variable tested (df = 7, $P > 0.07$ for both the observed and expected heterozygosities and $P > 0.27$ for the number of alleles).

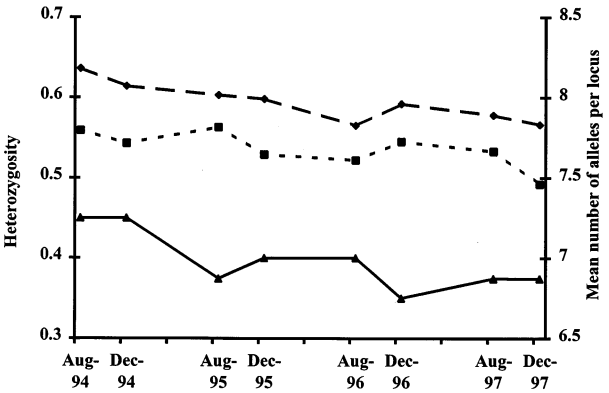

Figure 3. Temporal evolution of the mean observed heterozygosity (---■---), mean expected heterozygosity under Hardy–Weinberg equilibrium (—◆—) and mean number of alleles per locus (—▲—) averaged over all eight polymorphic loci, in the *An. arabiensis* population from Barkedji.

Discussion

We investigated spatial and temporal components of genetic variation of *An. arabiensis* populations to understand how they survive the long and severe dry season in the Sahel where their numbers virtually drop to zero. Despite being separated by 250 km, differentiation between Barkedji and Dielmo was very low. Allele frequencies and indices of genetic diversity remained stable during the 4-year study period, suggesting that *An. arabiensis* was continuously maintained in large numbers as the effective population size in this dry savanna area was in the hundreds or even thousands. As both, spatial and temporal variation in (neutral) allele frequency is caused by genetic drift, these findings agree with each other. These results allow us to reconsider the five scenarios proposed in the introduction to explain how populations pass the dry season.

Scenarios involving dry season bottleneck or strong founder effect (i and ii, see Introduction) are clearly inconsistent with Ne in the range of our estimates and therefore are ruled out. The ecological data show a gradual increase in abundance over 4–8 weeks after the onset of rains rather than a rapid rise associated with the arrival of a large number of mosquitoes within a short time period (Fig. 1b). However, this evidence is insufficient to rule out the hypothesis of mass migration (scenario iii). Indeed, if mosquitoes collected in August are migrants from a large, permanent population, the genetic variability within these samples would reflect Ne in the permanent (source) population and we would not expect any bottleneck effect to be detected. Thus, the mass migration scenario is compatible with our Ne estimates but would require extensive migration of many individuals across tens or hundreds of kilometers. Direct estimates of dispersal by mark–release–recapture experiments suggested that active life-time dispersal is limited to a few kilometers at most,

thus clearly incompatible with such long distance migration (Thomson *et al.*, 1995; Costantini *et al.*, 1996; Touré *et al.*, 1998b). Our genetic data, however, cannot rule out that large populations survive hidden from sampling (scenario iv). During the dry season in Sudan, adults females were found in rodent burrows, disused or ruined houses, and wells (Omer & Cloudsley-Thompson, 1968, 1970), but despite intense sampling efforts in such places and extensive pyrethrum spraying in human dwellings, we failed to find any mosquitoes there (but see below). We propose that either mosquitoes are so well hidden that we failed to find them or that eggs can survive the dry season despite evidence to the contrary (Beier *et al.*, 1990). This scenario cannot be ruled out based on the temporal analysis of genetic data, but it fits poorly with the ecological observations and our findings of low levels of genetic differentiation over 250 km in this area. Alternatively, local mosquito density is indeed below the efficiency threshold of our sampling methods. For example, if there are only twenty to fifty (inseminated) females in a whole village, it is conceivable that we will sample none. If the geographical area associated with a deme includes ten or more villages, then their numbers will add up to 400–1000 in accordance with our Ne estimates. This 'diffused deme' scenario (v) fits well with the genetic results of moderate Ne, and with the ecological records of very low local density. The low differentiation measured between populations 250 km apart also supports this scenario. In this region, we estimated that the diameter of a circle containing ten villages is 25 km.

The 'diffused deme' scenario requires mobility of mosquitoes in the range of 5–10 km. Using mark–release–recapture (MRR) experiments, Costantini *et al.* (1996) reported daily dispersal distance of 350–650 m and noted that one *An. gambiae s.l.* was recaptured 6 km from the release point. In all MRR studies, there was no focused effort to recapture mosquitoes in distances greater than 2–3 km from the release point and in all studies some mosquitoes were recaptured at the most distant sites (Thomson *et al.*, 1995; Costantini *et al.*, 1996; Touré *et al.*, 1998b). Thus, dispersal may have been underestimated. Additionally, all MRR studies were conducted during peak abundance when breeding sites are virtually everywhere, whereas larger distances would be required to reach breeding sites after most of them had dried out (Costantini *et al.*, 1996; Touré *et al.*, 1998b). In summary, the 'diffused deme' scenario refers to permanent populations that pass the dry season in small numbers in most villages, such that their local number is below the sampling 'threshold', but their total number (summed over all villages) reaches hundreds or thousands. This scenario fits best with available data.

Moderate Ne in the face of extreme fluctuations in density (Lemasson *et al.*, 1997) does not mean that variation in abundance had no effect on the genetic composition of the population, but that changes in genetic composition occur rapidly only when Ne drops to very low levels or when changes are monitored over long periods (Waples, 1989b, 1991; Waples & Teel, 1990; Richards & Leberg, 1996; Miller & Kapuscinski, 1997; Tessier & Bernatchez, 1999). Numeric estimates of the effective population size appear higher across the dry season (December–August) than across the rainy season (August–December, Table 4). This apparently contradictory result may reflect noise (i.e. small sample size effect) or a biological reason. Because Ne across a series of generations approximates the smallest generation's Ne and because each sample was part of the estimation of Ne in both rainy and dry season intervals we expected that these values would be similar to each other (or that the Ne across the rainy season would be *slightly* larger). A higher Ne for dry season, however, may reflect that the number of generations between December and August was actually lower than eight; in other words that mosquitoes estivated during the dry season. Indeed, assuming that there were only two generations (instead of eight), the estimates become similar. Undoubtedly, finding adults during the midst of the dry season would provide direct evidence for the presence of *An. arabiensis* in Barkedji during this period. Indeed, after intensive searching and trapping, thirteen females were collected in human-bated traps in March 1996 and eleven were collected by the same method in March 1997 (F. Simard and D. Fontenille, unpublished data). At this time no breeding sites were found in the area. Unfortunately, the very low numbers precluded genetic analysis and no information was obtained on their physiological state with respect to estivation (i.e. gonotrophic dissociation). Were these individuals immigrants from adjacent locations where permanent breeding was possible or estivating females? This question remains to be answered, but it is noteworthy that in February and March, the densities in 'permanent sites' such as Dielmo is very low (Fig. 1b) and thus migration during this period is unlikely.

Permanent population with Ne in the thousands was reported for two *An. gambiae* (savanna cytotype) populations in Kenya (Lehmann *et al.*, 1998), but the dry season in those locations is milder and shorter. Using chromosomal inversions as markers in *An. arabiensis* from Mali, Nigeria and Burkina Faso, Taylor *et al.* (1993) concluded that these populations were maintained continuously. Furthermore, their estimates were also in the hundreds or few thousands. As mentioned earlier, certain inversions are clearly under selection pressure. Selection on inversions may influence variation at microsatellite loci if they happen to be genetically linked to inversions. We found (i) no significant linkage disequilibrium between loci on chromosome II, (ii) no cyclical change in allele frequency between August and December and similar Ne estimates

whether samples were collected under different (August and December) or identical environmental conditions and (iii) slightly higher estimates of Ne for loci on chromosome II (or within inversions) than for loci on the other chromosomes (or outside inversions, see Table 4). Altogether, these findings suggest that selection on inversions is unlikely to confound our results. However, information based on additional loci from chromosome X and chromosome III would allow a more precise assessment of the influence of inversions on our Ne estimates. Nevertheless, excluding all chromosome II loci would not change our conclusions.

The temporal method to estimate Ne is probably the most robust (indirect) method available today, but it is also subject to assumptions and simplifications. Recognizing these limitations, our inferences were not focused on the exact Ne value but on its order of magnitude. Below we list the assumptions we made and consider the effects of violating these assumptions on Ne estimates. The temporal method assumes random sampling of a homogenous gene pool, selective neutrality of the genetic markers, negligible mutation and negligible migration (Waples, 1989b, 1991). Microsatellite loci that are found in non-coding regions, devoid of known biological function are generally considered as neutral. Nevertheless, evidence suggests the existence of constraints on allele size or biased mutation rates on these loci (e.g. Bowcock *et al*., 1994; Garza *et al*., 1995; Lehmann *et al*., 1996a). Such constraints could have biased our estimates. Deviations from Hardy–Weinberg equilibrium were evident in five of our loci. As pointed out earlier, these deviations probably reflected null alleles rather than evidence for subdivision or selection mediated by inversions. Importantly, single-locus Ne estimates show no relationship to whether the locus deviated from Hardy–Weinberg or not (Table 4). This together with no evidence for linkage disequilibrium between loci within Barkedji, suggests homogeneity of the gene pool and random sampling (see Results for more details). However, our samples were collected by indoor pyrethrum catches and could therefore represent only the endophilic fraction of the population. Previous population genetic studies have always revealed panmictic conditions for *An. arabiensis* with respect to behavioural heterogeneity (Coluzzi *et al*., 1979; Smits *et al*., 1996) suggesting such a subdivision is unlikely. Nevertheless, our estimates hold at least for this fraction of the population and Ne for the total *An. arabiensis* population may be even larger. Mutations are unlikely to have a significant effect on allele frequencies during 4 years (twenty to forty generations). Further support for the negligible effect of mutation was provided by the low and insignificant R_{ST} values between temporal samples. The assumption of negligible migration probably does not hold, as suggested by the high migration index revealed between Barkedji and Dielmo. Nevertheless, the effect of migration may be limited because allele frequencies and distributions appeared very similar at all the loci we studied between these two sampling sites. The assumption of discrete generations, which is unlikely to hold true, has a minimal effect, given that the time period under study encompasses many generations (Waples, 1989b; Jorde & Ryman, 1995).

Anopheles arabiensis is the most widespread species among the seven members of *the An. gambiae* complex and its behavioural plasticity with respect to feeding and resting choices (Gillies & De Meillon, 1968; Chauvet & Rajaonarivelo, 1973; Gillies & Coetzee, 1987) reflects its high adaptive potential. These traits retarded the efficiency of malaria control efforts by indoor spraying of insecticides. Extensive gene exchange as measured between Barkedji and Dielmo requires special attention to prevent emergence of insecticide resistance, which under selection could spread rapidly. On the other hand, high rate of gene exchange across large distances could be exploited to spread genes with favourable effects on disease transmission, such as proposed by the genetic approach for malaria control (e.g. Collins & Besansky, 1994; Crampton *et al*., 1994).

Experimental procedures

Study sites and sampling

The village of Barkedji (15°17′N, 14°53′W), with 700 inhabitants, is situated in the fossil valley river bed of the Ferlo, in the Sahelian region of Senegal (Fig. 1a). The rainy season is short, extending from July to October (Fig. 1b). Rainfall varies from one year to another ranging from 200 to 400 mm per year. Barkedji is surrounded by clay hollows which collect water as soon as the rains start and dry up around January. These temporary ponds are the only anopheline breeding sites encountered in the area. *Anopheles gambiae* and *An. arabiensis* are very abundant in the wet season. In the dry season, numbers of mosquitoes drop to virtually zero. They reappear just after the onset of rains and vector density gradually increases up to its maximum in October–November (Fig. 1b). The population dynamics of these vectors has been previously described (Lemasson *et al*., 1997). *Anopheles gambiae s.l.* specimens were collected after indoor pyrethrum spraying early in the morning and stored individually in numbered tubes with desiccant for laboratory processing in Dakar. The same houses were repeatedly sampled from 1994 to 1997. Each year, collections were conducted twice: in August, when temporary breeding sites get flooded and in December, just before they dry out (see Fig. 1b).

Using the same collection technique, an additional sample was taken from the village of Dielmo (13°45′N, 16°25′W), 250 km south-west of Barkedji, in August 1996 (Figure 1). While human density *per* se is fairly low, the area between Dielmo and Barkedji is more or less continuously inhabited. This part of Central to Northern Senegal is totally devoid of geographical disruption because no mountains or even hills are encountered along the distance between both locations. Vegetation mostly consists of

sparsely dispersed shrubby trees and baobabs that do not cluster to form dense forests. A detailed entomological description of the site of Dielmo was published recently (Fontenille *et al.*, 1997). Malaria transmission is due to *An. arabiensis*, *An. gambiae* and *An. funestus*. Unlike Barkedji, Dielmo is located on the marshy bank of a small permanent stream that permits the persistence of anophelines larval development sites and year round malaria transmission.

Microsatellite genotype scoring

DNA was extracted from the legs or wings of each individual mosquito using the technique described by Collins *et al.* (1987). Only female *An. arabiensis* specimens were included in the analysis, after species identification was carried out by the diagnostic PCR described by Scott *et al.* (1993).

Nine microsatellite loci (24D, 147, 141, 26 and 803 on chromosome II; 45C and 29C on chromosome III; and 49 and 7 on chromosome X) were analysed as previously described (Simard *et al.*, 1999), the only change being that PCR reactions were performed in 12.5 µl instead of 25 µl. PCR products were loaded on 10% non-denaturing polyacrylamide gels and the allelic bands were visualized after rapid silver staining (Sanguinetti *et al.*, 1994). Multilocus genotypes of approximately fifty specimens per sampling date and location were scored.

Data analysis

For each population and sampling date, we calculated the allelic frequencies, the mean number of alleles per locus, the observed and expected heterozygosity under Hardy–Weinberg's equilibrium using BIOSYS-1 software (Swofford & Selander, 1989). Tests for deviations from Hardy–Weinberg expectations and for linkage disequilibrium between loci were computed using exact tests available in GENEPOP 3.1 (Raymond & Rousset, 1995). The sequential Bonferroni procedure (Holm, 1979) was applied to evaluate significance when multiple tests were performed. Exact tests were also used to perform pair-wise comparisons of allelic frequencies distribution at each locus between two consecutive sampling dates in Barkedji.

Genetic differentiation among the Barkedji population at each sampling date and between geographical isolates was examined by *F* statistics (Wright, 1978), calculated according to Weir & Cockerham (1984), and their microsatellite's equivalent *R* statistics (Slatkin, 1995b). Estimates of F_{ST} and R_{ST} were computed using the ARLEQUIN 1.1 software package (available at <http://anthropologie.unige.ch/arlequin/>), developed by Schneider *et al.* (1997). Both F_{ST} and R_{ST} were tested for statistical significance by permuting individual genotypes among populations. Estimates of the gene exchange index, *Nm*, were derived from F_{ST} and R_{ST} according to equations 15b and 15a in Slatkin (1995b), respectively. *Nm* values derived from loci on the X-chromosome, were adjusted for their lower effective population size (assuming three-quarters that of the autosomes).

Estimation of the effective population size

We estimated Ne of the *An. arabiensis* population from Barkedji based on temporal variations in allelic frequencies. Because specimens were sampled from an *a priori* very large population (based on direct estimates), and not replaced, we followed the methods drawn for sampling prior to reproduction (Nei & Tajima, 1981; Waples, 1989b). Thus, estimates of Ne were obtained using equation 11 in Waples (1989b):

$$N_e = \frac{t}{2\left(F - \frac{1}{2S_0} - \frac{1}{2S_t}\right)}$$

where S_0 and S_t represent sample sizes (number of individuals) at generation 0 and t, respectively; and *F* estimates the standardized variance of allele frequency change. Several methods of computing *F* have been proposed (reviewed in Waples, 1989b). These methods generally lead to very similar results (Waples, 1989b; Taylor *et al.*, 1993; Miller & Kapuscinski, 1997; Lehmann *et al.*, 1998). We also verified this accordance, therefore only results based on F_k (Pollak, 1983) will be presented. Pollak's estimator for the change in allele frequency from one time period to another at one locus is:

$$F_k = \frac{1}{K-1}\sum_{i=1}^{K}\frac{(x_i - y_i)^2}{(x_i + y_i)/2}$$

where *K* is the number of alleles, and x_i and y_i represent the frequency of allele *i* at generation 0 and t, respectively. Waples (1989b) advocated that extreme allele frequencies could introduce a bias in the estimation of F_k leading to an overestimation of Ne. Therefore, alleles with frequencies lower than 2% at both time points of the considered interval were pooled into one class (Lehmann *et al.*, 1998). For estimation of F_k over all loci, we computed weighted means of single-locus values as:

$$F_{kall} = \Sigma(K_j - 1)F_{k_j}/\Sigma(K_j - 1)$$

where *j* stands for the different loci. This weighted mean was then used to estimate Ne based on all loci information combination. Calculation of the 95% confidence intervals (CI) followed equation 16 of Waples (1989b).

Like previous studies (Taylor *et al.*, 1993; Lehmann *et al.*, 1998), we conservatively assumed that twelve (discrete) generations occurred yearly for *An. arabiensis* in Barkedji. Estimates of Ne are also given for the total time period (between August 1994 and December 1997) assuming four generations between August and December (rainy season) and only two generations during the dry season (between December and August), based on the hypothesis that populations are effectively maintained by estivating females.

Acknowledgements

F. Simard gratefully acknowledges financial support from François Lacoste. We are grateful to Sadjo SOW and all the inhabitants of Barkedji and Dielmo for their help in collecting mosquitoes. We thank Habibatou Diallo-Agne and Teddy Diop for technical assistance in the laboratory. We also acknowledge M. J. Donnelly and two anonymous reviewers whose comments on an earlier draft appreciably improved this paper. This work was supported by the French Institut de Recherche pour le Développement (IRD), as part of grant no. CNRS 96/C/08, and MIM project ID 980101/WHO TDR 053S.2 'Population Structure of *Anopheles gambiae* and *An. funestus* in Kenya and West Africa'.

References

Avise, J.C. (1994) *Molecular Markers, Natural History and Evolution*. Chapman and Hall, London.

Beier, J.C., Copeland, R., Oyaro, C., Masinya, A., Odago, W.O., Oduor, S., Koech, D.K. and Roberts, C.R. (1990) *Anopheles gambiae* complex egg-stage survival in dry soil from larval development sites in western Kenya. *J Am Mosq Control Assoc* **6** (1): 105–109.

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.

Brookfield, J.F.Y. (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol Ecol* **5**: 453–455.

Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H.A., Richards, R.I., Mulley, J.C. and Sutherland, G.R. (1993) Incidence and origin of 'null' alleles in the (AC) n microsatellite markers. *Am J Hum Gen* **22**: 1–10.

Charlwood, J.D., Kihonda, J., Sama, S., Billingsley, P.F., Heiz, B. and Takken, W. (1995) The rise and fall of *Anopheles arabiensis* (Diptera: Culicidae) in a Tanzanian village. *Bull Ent Res* **85**: 37–44.

Chauvet, G. and Rajaonarivelo, E. (1973) 'Modification de comportement d'une espèce et variation de la composition d'une population appartenant au complexe *Anopheles gambiae* pendant et après les pulvérisations d'insecticides domiciliaires dans les environs de Tananarive (Madagascar). *Cah ORSTOM , Sér Ent Méd et Parasitol* **11** (3): 155–167.

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 *Anopheles gambiae* complex. *Am J Trop Med Hyg* **37**: 37–41.

Coluzzi, M., Sabatini, A., Petrarca, V. and Di Deco, M.A. (1979) Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Trans R Soc Trop Med Hyg* **73** (5): 483–497.

Costantini, C., Li, S.G., Della Torre, A., Sagnon N'F., Coluzzi, M. and Taylor, C.E. (1996) Density, survival and dispersal of *Anopheles gambiae* complex mosquitoes in a West African Sudan savanna village. *Med Vet Entomol* **10**: 203–219.

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.

Donnelly, M.J., Cuamba, N., Charlwood, J.D., Collins, F.H. and Townson, H. (1999) Populations structure in the malaria vector, *Anopheles arabiensis* Patton, in East Africa. *Heredity* **83**: 408–417.

Fontenille, D., Lochouarn, L., Diagne, N., Sokhna, C., Lemasson, J.J., Diatta, M., Konate, L., Faye, F., Rogier, C. and Trape, J.F. (1997) High annual and seasonal variations in malaria transmission by anophelines and vector species composition in Dielmo, a holoendemic area in Senegal. *Am J Trop Med Hyg* **56** (3): 247–253.

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** (4): 594–603.

Gillies, M.T. and Coetzee, M. (1987) A supplement to the Anophelinae of Africa south of the Sahara, 2nd edn. *Pub. South Afr. Inst. Med. Res.* **55**: 143 pp.

Gillies, M.T. and De Meillon, B. (1968) The Anophelinae of Africa south of the Sahara. *Pub. South Afr. Inst. Med. Res.* **54**: 343 pp.

Gupta, S. and Day, K.P. (1994) A strain theory of malaria transmission. *Parasitol Today* **10**: 476–481.

Hartl, D.L. and Clark, A.G. (1997) *Principles of Population Genetics*, 3rd edn. Sinauer Associates, Inc. Sunderland, Mass.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scand J Statistics* **6**: 65–70.

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

Jorde, P.E. and Ryman, N. (1995) Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change in allele frequencies. *Genetics* **143**: 1369–1381.

Kamau, L., Lehmann, T., Hawley, W.A., Orago, A.S.S. and Collins, F.H. (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** (1): 64–69.

Lanzaro, G.C., Touré, Y.T., Carnahan, J., Zheng, L., Dolo, G., Traoré, S.F., Petrarca, V., Vernick, K.D. and Taylor, C.E. (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**: 14 260–14 265.

Lanzaro, G.C., Zheng, L., Touré, Y.T., Traoré, S.F., 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** (2): 105–112.

Lehmann, T., Hawley, W.A. and Collins, F.H. (1996a) An evaluation of evolutionary constraints on microsatellite loci using null alleles. *Genetics* **144**: 1155–1163.

Lehmann, T., Hawley, W.A., Grebert, H. and Collins, F.H. (1998) The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. *Mol Biol Evol* **15** (3): 264–276.

Lehmann, T., Hawley, W.A., Kamau, L., Fontenille, D., Simard, F. and Collins, F.H. (1996b) Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: a comparison of microsatellite and allozyme loci. *Heredity* **77**: 192–208.

Lemasson, J.J., Fontenille, D., Lochouarn, L., Dia, I., Simard, F., Ba, K., Diop, A., Diatta, M. and Molez, J.F. (1997) Comparison of behaviour and vector efficiency of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a sahelian area of Senegal. *J Med Entomol* **34**(4): 396–403.

Miller, L.M. and Kapuscinski, A.R. (1997) Historical analysis of genetic variation reveals low effective population size in a Northern Pike (*Esox lucius*) population. *Genetics* **147**: 1249–1258.

Nei, M. (1978) Estimation of the average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Nei, M. and Tajima, F. (1981) Genetic drift and estimation of effective population size. *Genetics* **98**: 625–640.

Omer, S.M. and Cloudsley-Thompson, J.L. (1968) Dry season biology of *Anopheles gambiae* Giles in the Sudan. *Nature* **217**: 879–880.

Omer, S.M. and Cloudsley-Thompson, J.L. (1970) Survival of female *Anopheles gambiae* Giles through a 9-month dry season in Sudan. *Bull Wld Hlth Org* **42**: 319–330.

Petrarca, V., Vercruysse, J. and Coluzzi, M. (1987) Observations on the *Anopheles gambiae* complex in the Senegal river basin, West Africa. *Med Vet Entomol* **1**: 303–312.

Pollak, E. (1983) A new method for estimating effective population size from allele frequency changes. *Genetics* **104**: 531–548.

Raymond, M. and Rousset, F. (1995) GENEPOP, Version 1.2. A population genetics software for exact tests and ecumenicism. *J Heredity* **86**: 248–249.

Richards, C. and Leberg, P.L. (1996) Temporal changes in allele frequencies and a population's history of severe bottlenecks. *Conservation Biol* **10**: 832–839.

Sanguinetti, C.J., Neto, E.D. and Simpson, A.J.G. (1994) Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques* **17** (5): 915–919.

Schneider, S., Kueffer, J.M., Roessli, D. and Excoffier, L. (1997) *ARLEQUIN, a Software Package for Population Genetics*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

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

Simard, F., Fontenille, D., Lehmann, T., Girod, R., Brutus, L., Gopaul, R., Dournon, C. and Collins, F.H. (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** (1): 1000–1009.

Slatkin, M. (1995a) Hitchhiking and associative overdominance at a microsatellite locus. *Mol Biol Evol* **12** (3): 473–480.

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

Smits, A., Roelants, P., Van Bortel, W. and Coosemans, M. (1996) Enzyme polymorphisms in the *Anopheles gambiae* (Diptera: Culicidae) complex related to feeding and resting behavior in the Imbo valley, Burundi. *J Med Entomol* **33** (4): 545–553.

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.

Taylor, C.E., Toure, Y.T., Coluzzi, M. and Petrarca, V. (1993) Effective population size and persistence of *Anopheles arabiensis* during the dry season in West Africa. *Med Vet Entomol* **7**: 351–357.

Tessier, N. and Bernatchez, L. (1999) Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L). *Mol Ecol* **8**: 169–179.

Thomson, M.C., Connor, S.J., Quinones, M.L., Jawara, M., Todd, J. and Greenwood, B.M. (1995) Movement of *Anopheles gambiae s.l.* malaria vectors between village in the Gambia. *Med Vet Entomol* **9**: 413–419.

Touré, Y.T., Dolo, G., Petrarca, V., Traoré, S.F., Bouaré, M., Dao, A., Carnahan, J. and Taylor, C.E. (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.

Touré, Y.T., Petrarca, V., Traoré, S.F., Coulibaly, A., Maiga, H.M., Sankaré, O., Sow, M., Di Deco, M.A. and Coluzzi, M. (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., Petrarca, V., Traoré, S.F., Coulibaly, A., Maiga, H.M., Sankaré, O., Sow, M., Di Deco, M.A. and Coluzzi, M. (1998a) The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia* **40** (4): 477–511.

Walton, C., Thelwell, N.J., Priestman, A. and Butlin, R.K. (1998) The use of microsatellites to study gene flow in natural populations of Anopheles malaria vectors in Africa: potential and pitfalls. *J Am Mosq Control Assoc* **14**(3): 266–272.

Waples, R.S. (1989a) Temporal variation in allelic frequencies: testing the right hypothesis. *Evolution* **43**(6): 1236–1251.

Waples, R.S. (1989b) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**: 379–391.

Waples, R.S. (1991) Genetic methods for estimating the effective population size of cetacean populations. In: *Genetic Ecology of Whales and Dolphins* (Hoelzel, A.R., ed.), pp. 279–300. International Whaling Commission (Special Issue no. 13), Cambridge, UK.

Waples, R.S. and Teel, D.J. (1990) Conservation biology of Pacific salmon. I. Temporal changes in allele frequency. *Conservation Biol* **4**: 144–156.

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.

Wright, S. (1938) Size of populations and breeding structure in relation to evolution. *Science* **87**: 430–431.

Wright, S. (1978) *Evolution and Genetics of Populations, Vol. 4. Variability Within and Among Natural Populations*. University of Chicago Press. Chicago, IL.

Zheng, L., Benedict, M.Q., Cornel, A.J., Collins, F.H. and Kafatos, F.C. (1996) An integrated genetic map of the African malaria vector mosquito, *Anopheles gambiae*. *Genetics* **143**: 941–952.