Insect Molecular Biology (2002) **11**(1), 11–19

Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry

C. Wondji[1], F. Simard[1,2] and D. Fontenille[1,2]

[1]*Laboratoire de l'Institut de Recherche pour le Développement (IRD), Organisation de Coordination pour la Lutte Contre les Endémies en Afrique Centrale (OCEAC), Yaoundé, Cameroun; and* [2]*Laboratoire de Lutte contre les Insectes Nuisibles (LIN), Institut de Recherche pour le Développement (IRD), Montpellier, France*

Abstract

We studied genetic variation at ten microsatellite DNA loci in *Anopheles gambiae* populations from the Forest chromosomal form collected in four villages in Cameroon (Central Africa). Both recently described M and S molecular forms occur in sympatry in this area. Geographic differentiation within form was low (*Fst* < 0.017) despite geographical distance between collection sites ranging from 35 to 350 km. However, higher (*Fst* > 0.035) and statistically significant levels of genetic differentiation were observed between forms, being the highest between sympatric M and S populations collected within the same village. Results were consistent across all loci spread throughout the genome, therefore reflecting a genome-wide pattern. Considering previous findings of strong assortative mating within forms and general lack of hybrids in areas of sympatry, we propose that there is now sufficient direct and indirect evidence to consider both M and S molecular forms of *An. gambiae* as distinct species that have probably speciated recently.

Keywords: *Anopheles gambiae*, molecular forms, microsatellites, malaria, Africa.

Introduction

Reproductive isolation and incipient speciation within the species *Anopheles gambiae*, the most efficient malaria vector in Sub-Saharan Africa, has been the focus of much research for more than two decades. The first evidence for genetic heterogeneity within this mosquito species came from cytological observations of the banding pattern of polytene chromosomes that revealed a complex system of polymorphic paracentric inversions leading to different chromosomal arrangements (Coluzzi & Sabatini, 1967; Coluzzi *et al*., 1979). Frequencies of alternative arrangements, especially involving inversions on chromosome 2, were shown to correlate with ecological/climatic factors such as the degree of aridity of the environment, suggesting an adaptive potential for inversions, different combinations favouring survival under a variety of environmental conditions (Coluzzi *et al*., 1979; Coluzzi, 1982; Touré *et al*., 1994, 1998). Furthermore, extensive studies of karyotype distributions in natural *An. gambiae* populations often revealed strong and persistent deviations from Hardy–Weinberg equilibrium due to a deficit or even complete absence of certain heterokaryotypes, providing ample evidence against panmixia. These results led to the designation, in West Africa, of five 'chromosomal forms' named under a non-Linnean nomenclature as Forest, Savanna, Mopti, Bamako and Bissau (Coluzzi *et al*., 1985; Touré *et al*., 1998). These forms have been characterized based on the presence or absence of certain diagnostic inversions combinations (e.g. inversion 2Rb in Savanna, 2Rj in Bamako, 2Rbc/u in Mopti, 2Rd in Bissau), different relative frequencies of polymorphic arrangements (e.g. the standard arrangement 2R+/2L+ is almost fixed in Forest, but rare in the other forms), geographical distribution and ecological data (e.g. Mopti is adapted to dryer environments where it breeds all year long in irrigated fields, while Forest is exclusively found in more humid forested areas). They appear more or less genetically isolated in the field, presumably through prezygotic barriers because viable and fertile hybrids have been obtained in the laboratory (Di Deco *et al*., 1980). Cytogenetic analysis alone, however, does not allow precise evaluation of the degree of hybridization between forms because of the presence of 'cryptic' heterokaryotypes impossible to identify with confidence, shared arrangements between forms (especially the standard uninverted karyotype 2R+) and the adaptive nature of inversions strongly

Received 11 July 2001; accepted after revision 4 September 2001. Correspondence: Frédéric Simard, Laboratoire de Lutte contre les Insectes Nuisibles (LIN), Institut de Recherche pour le Développement (IRD), 911 Avenue Agropolis, BP 5045, 34 032 Montpellier Cedex 1, France. Tel.: +33 (0)4 67 04 32 24; fax: +33 (0)4 67 54 20 44; e-mail: oceac@camnet.cm

exposed to selection, making them poor markers for studies of reproductive isolation (Touré *et al.*, 1998; della Torre *et al.*, 2001).

Based on the rationale that reproductive isolation would lead to genome-wide heterogeneity between chromosomal forms, recent efforts have focused on the pattern of variation observed with molecular markers that are not linked to inversions, in order to help resolving the taxonomic status of these forms (Favia *et al.*, 1994, 1997, 2001; Lanzaro *et al.*, 1998; Gentile *et al.*, 2001; Mukabayire *et al.*, 2001; della Torre *et al.*, 2001). All these studies reported negligible amounts of genetic differentiation between forms with the exception of loci in the ribosomal DNA (rDNA) cluster, on chromosome X. rDNA loci, however, did not support the subdivision of *An. gambiae* natural populations into five different incipient taxa, but suggested the existence of only two different entities, now referred to as molecular forms M and S (della Torre *et al.*, 2001). In Mali and Burkina Faso, exact correspondence between molecular and chromosomal forms has been found, all Mopti individuals belonging to the molecular form M while sympatric Savanna and Bamako specimens showed the S pattern (Favia *et al.*, 1997). This association breaks down in other areas of Africa, but reproductive isolation of the M and S forms is maintained (Chandre *et al.*, 1999; della Torre *et al.*, 2001). Thus, mate recognition seems to be independent of chromosomal inversions involved in ecotypic adaptation.

Accordingly, in forested areas of Central Africa, both M and S share the standard chromosome 2 arrangement, which is typical of the Forest chromosomal form (della Torre *et al.*, 2001; C. Wondji *et al.*, unpublished). Thus, the Forest form of *An. gambiae*, previously described as a single panmictic unit based on chromosomal arrangements frequencies, appears to be a variant common to both molecular forms. If M and S are effectively reproductively isolated from one another, genetic discontinuity within the Forest form should be observed in areas where both forms occur in sympatry. On the other hand, if environmental selection on inversions is the main force driving genetic variability within and between *An. gambiae* populations, genetic homogeneity is expected within the Forest chromosomal form.

To test this hypothesis, we studied genetic variation at ten microsatellite loci in five *An. gambiae* populations from the Forest chromosomal form, collected in four villages in South Cameroon. Sampling within the Forest form allowed us to assess the level of genetic differentiation between molecular forms while overcoming the effect of both geographical distance between populations (both forms exist in sympatry in this area), and differential selective pressure on inversions (both forms are homosequential in this environment, presenting only the standard, uninverted, chromosome 2 arrangement).

Results and discussion

Genotypes at ten microsatellite loci were determined for 242 *An. gambiae* specimens collected in four villages in Cameroon. In this area, the Forest chromosomal form is the only one encountered (or nearly so, Robert *et al.*, 1993; della Torre *et al.*, 2001; C. Wondji *et al.*, unpublished). All loci were highly polymorphic showing between 8 and 16 alleles per locus. Similar levels of variability were observed in all populations with a mean number of alleles per locus ranging from 8.3 to 10.0 and average unbiased estimates of heterozygosity across all loci approximately 0.80 (Table 1).

Heterogeneity of the gene pool

We applied a hierarchical design to detect population subdivision by testing for deviations from Hardy–Weinberg equilibrium (HWE) based upon a null hypothesis of heterozygote deficit.

First, we considered that all specimens belonged to the same, panmictic population. In this case, eight out of ten loci showed positive *Fis* values, heterozygote deficit being statistically significant in six cases (Table 2, 'pooled populations'). This finding, suggesting heterogeneity of the gene pool, was strengthened by analysis of linkage disequilibrium between loci because twelve out of forty-five tests were significant ($P < 0.001$, binomial test with 0.05 'success' rate) and involved loci that were not necessarily on the same chromosome. Altogether, these results are indicative of a Wahlund effect (mixing of different gene pools).

Geographical distance between populations may lead to genetic differentiation (isolation by distance), therefore we considered each village as a basic reproductive unit. In Nkoteng, Dschang and Mbalmayo, where all individuals analysed belonged to either one of the molecular forms, most of the loci conformed to Hardy–Weinberg expectations (Table 2). Significant deficit in heterozygotes were observed at less than three loci out of ten tested in any of these populations, and linkage disequilibrium analysis within each population resulted in one to four significant tests out of forty-five ($P > 0.18$). However, in Simbock, where both molecular forms S and M occur in sympatry, strong and highly significant deficit in heterozygotes was observed at six out of ten loci suggesting further separation of the gene pool in this location. Thus, we defined two subpopulations within Simbock, according to the molecular form (M and S) and tested for goodness of fit to Hardy–Weinberg expectations. In the M and S forms, respectively, two and three loci out of ten showed significant excess of homozygotes (loci Ag3H93 and Ag3H750 in the M form, and loci Ag3H93, Ag3H170 and Ag3H750 in the S form), which falls within the range observed in the other samples. No significant linkage disequilibrium was observed in the M population ($P > 0.39$) while six out of forty-five tests were statistically significant in the S population ($P = 0.024$).

Table 1. Genetic variability at ten microsatellite loci in *An. gambiae* from Cameroon

Locus		Populations					
		Nkoteng (2*n* = 106)	Dschang (2*n* = 108)	Simbock S (2*n* = 66)	Simbock M (2*n* = 98)	Mbalmayo (2*n* = 106)	All (2*n* = 484)
AgXH80	Nall	12	16	9	9	9	16
(X: 2B)*	Hobs	0.87	0.85	0.71	0.96	0.88	0.85
	Hexp	0.83	0.89	0.82	0.82	0.82	0.84
AgXH99	Nall	6	7	5	6	7	8
(X: 2C)	Hobs	0.74	0.72	0.64	0.59	0.79	0.70
	Hexp	0.80	0.74	0.67	0.70	0.76	0.74
AgXH49	Nall	10	9	7	8	4	11
(X: 1D)	Hobs	0.89	0.80	0.74	0.57	0.69	0.74
	Hexp	0.80	0.81	0.77	0.64	0.58	0.72
Ag2H417	Nall	6	8	5	7	7	10
(II: 7)	Hobs	0.65	0.65	0.68	0.76	0.86	0.72
	Hexp	0.56	0.71	0.62	0.69	0.70	0.66
Ag2H769	Nall	12	12	12	14	11	16
(II: 9B)	Hobs	0.77	0.88	0.80	0.76	0.87	0.82
	Hexp	0.90	0.88	0.88	0.86	0.88	0.88
Ag2H325	Nall	8	8	7	7	8	9
(II: 23C-D)	Hobs	0.66	0.75	0.78	0.77	0.64	0.72
	Hexp	0.80	0.76	0.77	0.71	0.65	0.74
Ag3H93	Nall	10	11	7	7	7	11
(III: 29A)	Hobs	0.79	0.85	0.43	0.55	0.75	0.67
	Hexp	0.85	0.82	0.76	0.78	0.73	0.79
Ag3H555	Nall	9	9	10	8	8	11
(III: 35B-D)	Hobs	0.83	0.81	0.73	0.73	0.87	0.80
	Hexp	0.81	0.84	0.81	0.72	0.77	0.79
Ag3H170	Nall	11	10	10	15	14	16
(III: 35)	Hobs	0.76	0.64	0.58	0.90	0.75	0.73
	Hexp	0.84	0.84	0.88	0.92	0.88	0.87
Ag3H750	Nall	13	10	11	10	11	13
(III: 39A)	Hobs	0.59	0.69	0.42	0.70	0.69	0.62
	Hexp	0.88	0.86	0.85	0.85	0.88	0.86
Mean across	Nall	9.7	10	8.3	9.1	8.6	12.1
all loci	Hobs	0.76	0.76	0.65	0.73	0.78	0.74
	Hexp	0.81	0.81	0.78	0.77	0.77	0.79

*Cytological location of the locus; 'All' refers to populations pooled. 2*n*, number of chromosomes scored; Nall, number of alleles per locus; Hobs, observed heterozygosity (direct count); Hexp, expected heterozygosity under Hardy–Weinberg equilibrium (Nei, 1978).

Although each population exhibited heterozygote deficit at some loci, it is more likely the result of locus-specific constraints such as null alleles (Callen *et al*., 1993), limited allelic range (Epplen *et al*., 1993) or preferential amplification of one allele in heterozygotes (Wattier *et al*., 1998), rather than population substructure or inbreeding. Deviations from HWE are a common finding in population genetics studies that have been conducted so far using microsatellite loci in *An. gambiae* and/or its sibling species, *An. arabiensis* (Lehmann *et al*., 1996a; Lanzaro *et al*., 1998; Simard *et al*., 2000; Donnelly & Townson, 2000) and direct evidence for null alleles have been reported (Lehmann *et al*., 1996b; Walton *et al*., 1998; Lanzaro *et al*., 1998). Several evidences indeed suggest locus-specific effects in our dataset. First, loci Ag3H750 and Ag3H170 showed positive *Fis* values in all five populations tested and heterozygote deficits reached statistical significance in four and two cases, respectively. High and significant positive *Fis* values were also observed at locus Ag3H93 in both populations from Simbock. However, when all populations are pooled or when the M and S forms are not set apart in Simbock, these deviations are strengthened by heterozygote deficits at other loci and the lack of heterozygotes appears as a genome-wide pattern, involving most of the loci (six out of ten in the pooled population, as well as in Simbock). Second, these loci appear in a state of linkage equilibrium at every level of the analysis, suggesting that the different alleles are distributed evenly between populations and the deviation observed does not result from genetic substructure. On the basis of considerably higher levels of heterogeneity observed when populations are pooled, we regarded each of the five populations we defined as panmictic units. However, because the effect of deviations from HWE on *Fst* estimators is unclear, we apply them with caution and

Table 2. Test for deviation form Hardy–Weinberg equilibrium: *Fis* and significance level for goodness of fit tests to Hardy–Weinberg equilibrium, based upon a null hypothesis of heterozygote deficit, for all *An. gambiae* populations

Population			Locus									
Molecular form	Location		AgXH80	AgXH99	AgXH49	Ag2H417	Ag2H769	Ag2H325	Ag3H93	Ag3H555	Ag3H170	Ag3H750	Mean across loci*
Pooled populations		(2*n* = 484)	−0.010	**+0.091**	+0.027	−0.078	**+0.085**	+0.044	**+0.126**	**+0.015**	**+0.164**	**+0.279**	**+0.074**
	Mbalmayo	(2*n* = 106)	−0.061	−0.045	−0.201	−0.237	+0.010	+0.012	−0.029	−0.122	+0.145	+0.212	−0.031
M	Simbock	(2*n* = 98)	−0.163	+0.159	+0.107	−0.091	+0.115	−0.081	**+0.289**	−0.024	+0.024	**+0.184**	+0.051
	All	(2*n* = 204)	−0.111	+0.064	−0.042	−0.164	**+0.066**	−0.030	+0.148	−0.058	+0.088	**+0.204**	+0.016
Area of sympatry M + S	Simbock	(2*n* = 164)	−0.018	**+0.163**	**+0.147**	−0.094	+0.126	−0.058	**+0.356**	**+0.080**	**+0.155**	**+0.340**	**+0.119**
	Simbock	(2*n* = 66)	+0.137	+0.052	+0.031	−0.093	+0.091	−0.022	**+0.438**	+0.097	**+0.337**	**+0.505**	**+0.157**
S	Nkoteng	(2*n* = 106)	−0.053	+0.082	−0.113	−0.154	+0.143	+0.179	+0.063	−0.023	+0.096	**+0.319**	+0.054
	Dschang	(2*n* = 108)	+0.047	**+0.033**	+0.017	+0.083	−0.002	+0.017	−0.029	+0.032	**+0.232**	**+0.195**	**+0.062**
	All	(2*n* = 280)	+0.039	+0.072	−0.022	−0.018	**+0.081**	+0.068	+0.101	**+0.028**	**+0.211**	**+0.313**	**+0.088**

Fis was calculated according to Weir & Cockerham (1984). Note that when populations are pooled, *Fis* = *Fit*. Significance levels were corrected according to the Bonferroni procedure to take into account multiple tests (ten tests per population). Bolded values $P < 0.05$; bolded underlined values $P < 0.01$. 2*n*, number of chromosome scored. *Mean *Fis* across loci and its statistical significance was estimated by bootstrapping over loci (5000 replicates).

consistently pay attention to the homogeneity of the results across loci.

It is worth noting that, in our study, loci Ag2H769 and Ag2H325 are located within chromosomal inversions that are polymorphic in *An. gambiae* (Ag2H769 is within inversion 2Rj and Ag2H325 is within 2La). Although selection on inversions may influence the genotype at linked microsatellites loci by genetic hitchhiking (Lanzaro *et al*., 1998), it is unlikely in our case because an overwhelming majority of specimens from this area belong to the Forest chromosomal form (see Experimental procedures), which is characterized by standard (uninverted) homozygous karyotypes. However, no karyotypic information was available for the samples we have tested and putative influence of the cytological location of the locus (within or outside inversions) will be further considered.

Geographic differentiation within molecular forms

Levels of genetic differentiation between populations were assessed by estimating *Fst* (Weir & Cockerham, 1984) and tested using the exact test of genotypic differentiation (Goudet *et al*., 1996).

Genotype distribution was similar at all loci in the three villages within the S form and between both M populations: only two out of thirty pair-wise tests showed significant heterogeneity within the S form (between Nkoteng and Simbock at loci AgXH99 and Ag3H93), and three out of ten tests were significant between M populations from Simbock and Mbalmayo (Table 3A). Single-locus *Fst* estimates were very low and did not significantly differ from 0, except in the above-mentioned cases (Table 3A). However, when information from all loci is combined, mean *Fst* estimates reached statistical significance except between populations of the S form collected in Dschang and Simbock. Despite very low values ($0.007 < Fst < 0.017$), significant *Fst* estimates suggest that populations are indeed genetically differentiated, probably as a result of isolation by distance. This finding, however, needs further consideration and additional samples are actually being processed to ascertain this hypothesis.

Genetic differentiation between molecular forms

Strikingly higher levels of differentiation were observed when comparing populations belonging to alternative molecular forms than when geographical differentiation was investigated within each molecular form (Table 3). Except for locus Ag2H417 and, to a lower extent Ag2H325 that showed limited amounts of genetic differentiation between forms, all pair-wise *Fst* estimates were highly homogenous across loci and were statistically significant ($P < 0.01$ in most cases, Table 3B). Removing from the analysis either loci that deviated from HWE or loci mapped within polymorphic inversions did not change the results obtained (Table 3). Thus, the pattern observed reflects a genome-wide characteristic

Table 3. Measure of genetic differentiation (*Fst*) between *An. gambiae* populations of the Forest chromosomal form
(**A**) Within molecular forms

	Within the molecular form S			Within the molecular form M
Locus	Nkoteng vs. Dschang	Nkoteng vs. Simbock S	Dschang vs. Simbock S	Simbock M vs. Mbalmayo
AgXH80	0.008	0.003	0.006	0.001
AgXH99	0.012	**0.047**	0.014	**0.031**
AgXH49	0.002	0.019	0.024	0.003
Chromosome X	**0.007**	**0.023**	0.014	**0.012**
Ag2H417	0.041	0.000	0.006	0.008
Ag2H769	0.012	0.021	0.000	0.009
Ag2H325	0.000	0.000	0.004	0.012
Chromosome 2	0.014	0.008	0.003	**0.010**
Ag3H93	0.000	**0.008**	0.000	**0.047**
Ag3H555	0.000	0.006	0.004	0.034
Ag3H170	0.018	0.020	0.000	0.004
Ag3h750	0.000	0.003	0.000	**0.015**
Chromosome 3	0.002	**0.009**	0.000	**0.024**
All loci	**0.007**	**0.013**	0.005	**0.017**
[95% C.I.]*	[0–0.016]	[0.004–0.026]	[0–0.010]	[0.008–0.026]
Mean across seven loci†	***0.009***	***0.014***	*0.008*	***0.014***
Loci outside inversions‡	***0.008***	***0.014***	*0.006*	***0.018***

(**B**) Between molecular forms

Locus	(S Form) (M Form)	Nkoteng vs. Simbock M	Nkoteng vs. Mbalmayo	Dschang vs. Simbock M	Dschang vs. Mbalmayo	Simbock S vs. Simbock M	Simbock S vs. Mbalmayo	All S vs. All M
AgXH80		**0.030**	**0.016**	**0.037**	**0.023**	**0.041**	0.014	**0.025**
AgXH99		**0.044**	**0.029**	0.029	**0.063**	**0.102**	**0.140**	**0.047**
AgXH49		**0.068**	**0.094**	**0.095**	**0.131**	**0.151**	**0.167**	**0.104**
Chromosome X		**0.047**	**0.046**	**0.054**	**0.072**	**0.096**	**0.106**	**0.058**
Ag2H417		0.012	**0.022**	0.001	**0.035**	0.000	**0.021**	**0.008**
Ag2H769		**0.046**	**0.035**	**0.018**	**0.013**	**0.046**	**0.019**	**0.023**
Ag2H325		**0.023**	**0.059**	0.021	**0.061**	0.000	0.032	**0.032**
Chromosome 2		**0.029**	**0.039**	**0.014**	**0.035**	**0.016**	**0.024**	**0.022**
Ag3H93		**0.037**	**0.021**	**0.040**	**0.008**	**0.039**	0.000	**0.013**
Ag3H555		**0.049**	**0.044**	**0.049**	**0.025**	**0.105**	**0.074**	**0.044**
Ag3H170		**0.033**	**0.013**	**0.035**	**0.011**	**0.027**	**0.017**	**0.017**
Ag3h750		**0.034**	**0.015**	**0.043**	**0.011**	**0.078**	0.026	**0.030**
Chromosome 3		**0.038**	**0.023**	**0.042**	**0.014**	**0.063**	**0.030**	**0.026**
All loci		**0.038**	**0.035**	**0.038**	**0.037**	**0.060**	**0.052**	**0.035**
[95% C.I.]*		[0.030–0.047]	[0.022–0.050]	[0.025–0.053]	[0.018–0.062]	[0.032–0.088]	[0.020–0.087]	[0.021–0.052]
Mean across seven loci†		***0.040***	***0.043***	***0.037***	***0.050***	***0.065***	***0.068***	***0.041***
Loci outside inversions‡		***0.040***	***0.032***	***0.042***	***0.038***	***0.069***	***0.058***	***0.037***

Bolded values: $P < 0.05$; Bolded underlined values: $P < 0.01$; *95% confidence intervals (C.I.) around *Fst* estimates based on all loci information were built by bootstrapping over loci (1000 permutations); †excluding loci showing significant heterozygote deficits at the population level (i.e. loci Ag3H93, Ag3H170 and Ag3H750); ‡excluding loci within polymorphic inversions (i.e. loci Ag2H769 and Ag2H325).

and led to overall *Fst* estimates of 0.035–0.052 between allopatric populations and *Fst* = 0.060 between sympatric populations in Simbock ($P < 0.01$), providing evidence for limited gene exchange between the M and S molecular forms of *An. gambiae*.

It is interesting to note that sympatric populations showed the highest level of genetic differentiation. Although differences are not statistically significant (95% confidence intervals of *Fst* estimates do overlap to a large extent, Table 3B), this pattern is very similar to what has been observed with cuticular hydrocarbons profiles. Milligan *et al.* (1993) indeed reported diagnostic differences in cuticular hydrocarbons composition between *An. gambiae* chromosomal forms when found in sympatry, yet correct assignment of individuals to their respective form (and even species within the *An. gambiae* complex) was impossible between allopatric populations. Cuticular hydrocarbons are well known to play a role in species (mate or kin) recognition in different arthropods (see Flechter & Michener, 1987) and the pattern observed by Milligan *et al.* (1993) may thus reflect reproductive character displacement as a consequence of natural selection of a premating mechanism for the reinforcement of mate recognition in areas of sympatry. Reproductive character displacement towards differential cuticular hydrocarbons profiles has indeed been demonstrated in *Drosophila serrata* when sympatric with its closely

related species *D. birchii* (Higgie *et al*., 2000), but it is unclear how selection acting specifically on cuticular hydrocarbons genes (or their putative regulation factors) could translate into a genome-wide pattern such as observed in the present study.

Taxonomic status of the molecular forms of An. gambiae

Our results clearly demonstrated the existence of a barrier to gene flow between homokaryotypic (Forest) populations of the S and M forms of *An. gambiae* in an area of sympatry. Although values of *Fst* estimated from allelic and genotypic frequencies distributions at ten microsatellite loci do not provide clear-cut evidence for complete reproductive isolation between molecular forms, the results presented above coupled with previous evidence for complete lack of M/S hybrids in Cameroon (out of a total of 950 specimens typed throughout the country; C. Wondji *et al*., unpublished) represents a very strong argument for the species status of the M and S forms.

Our results are consistent with those initially obtained by Favia *et al*. (1997) using PCR-RFLP in 203 field specimens collected in Mali, in an area of sympatry between Mopti, Savanna and Bamako, who reported complete absence of M/S hybrid patterns despite biased sampling towards inclusion of putative Mopti/Savanna hybrid karyotypes. Female hybrids produced in the laboratory, however, showed clear M/S patterns by PCR-RFLP as well as by the recently developed PCR diagnostic (Favia *et al*., 1997, 2001). Further support for reproductive isolation of the M and S forms was provided by Chandre *et al*. (1999) through an extensive study of the distribution of the pyrethroid resistance gene *kdr*. In an area of sympatry in Ivory Coast, the authors found the *kdr* mutation (sometimes reaching frequency as high as 0.96) in the S form only; it was not detected in sympatric and synchronous M populations.

The occurrence of M/S hybrids in the field, however, has been reported several times. Della Torre *et al*. (2001) found three hybrid patterns out of 1161 *An. gambiae* adult mosquitoes tested from throughout Africa (hybrid frequency = 3/1161 = 0.26%), but two of them were observed in localities where the S form only had been recorded. The authors therefore mentioned contamination as a possible cause of the hybrid patterns they reported, but recognized that cross-mating between *An. gambiae* and its sibling species *An. arabiensis* also occurs in areas of sympatry, although at a lower frequency of approximately 0.1–0.2‰ (Touré *et al*., 1998). Further evidence for the viability of M/S hybrids in the wild was provided by Taylor *et al*. (2001), who reported occurrence of M/S hybrid larvae at a frequency of 0–1.29% (average over 3 years = 0.71%) in Banambani village, Mali. In a nearby village from the same area, Tripet *et al*. (2001) identified an inseminated female showing the M/S pattern, which demonstrated that M/S hybrids can be produced in the field, survive up to the adult stage and are reproductively active. Determination of the genotype of this female at four microsatellite loci ruled out the possibility of contamination and led the authors to estimate adult hybrid frequency at 1/329 = 0.3%. Recent findings of the highly adaptive *kdr* mutation in M mosquitoes from Benin and subsequent molecular analysis of the DNA sequence of a large upstream intron provided direct evidence for introgression between molecular forms (Weill *et al*., 2000). Such an event, however, must be rare and recent because no M mosquitoes displaying the *kdr* mutation were found outside Benin despite thorough sampling throughout West and Central Africa (Chandre *et al*., 1999; Weill *et al*., 2000; della Torre *et al*., 2001).

Thus, moderate estimates of *Fst* as observed in the present study may reflect shared ancestral polymorphism rather than ongoing gene flow. There is indeed strong evidence for recent dramatic changes in the demography and history of these taxa, mostly due to anthropic environmental changes (Coluzzi, 1999; Powell *et al*., 1999), which suggest that the pattern of differentiation observed may relate to a transient state of incipient speciation, with *Fst* values that have not yet reached their equilibrium. Moreover, at a global (species) scale, both M and S forms may not have spent the $4Ne$ generations (since initial splitting) necessary for distinct phylogenetic lineages to establish (Takahata, 1992). Thus, attempts to differentiate populations through classical population genetics and/or phylogenetics would undoubtedly reveal conflicting patterns, even if gene flow between molecular forms is currently extremely restricted and exclusively limited to occasional events of introgression driven by selection on adaptive genes.

Conclusion

The primary objective of our study was to estimate levels of genetic differentiation between the M and S molecular forms of *An. gambiae* while overcoming confounding effects of both geographical distance between populations and environmental selection on chromosomal inversions. This was achieved by sampling at a local scale (both forms were found in sympatry in the same village) within the Forest chromosomal form of *An. gambiae* (standard chromosome 2 karyotype). We found highly significant genetic differentiation between molecular forms, whereas geographical isolates within each form were much more homogenous. These results are consistent with previous findings of strong assortative mating within both M and S forms and general lack of hybrids in areas of sympatry. Taking into account possible limitations of equilibrium models of genetic differentiation (i.e. Rousset, 1996), we would emphasize that there is now sufficient support to consider that the molecular forms of *An. gambiae* may represent distinct species that have speciated in a recent past.

Experimental procedures

Study sites and sampling

Mosquitoes were collected either after landing on human volunteers or by indoor pyrethrum spraying, in four localities of Cameroon: Simbock, Mbalmayo, Nkoteng and Dschang.

Simbock (3°51′N, 11°30′E) is located approximately 5 km from the capital city Yaoundé, in a forested area within the guineo-congolese bio-climatic domain. Mbalmayo (3°30′N, 11°30′E) is located 35 km south of Simbock in similar ecological settings. Nkoteng (4°30′N, 12°03′E) is about 140 km north-east of Simbock, in a forest-savanna transition area and Dschang (5°25′N, 10°10′E), approximately 350 km north-west from Yaoundé, is located in forest-savanna highlands.

A detailed analysis of the geographical distribution and relative prevalence of *An. gambiae* molecular and chromosomal forms in this area will be given elsewhere (C. Wondji *et al.*, unpublished). Briefly, in Nkoteng and Dschang, prevalence of the S form was more than 98%. Mosquitoes belonging to the M molecular form were predominant in the other two localities considered in this paper, representing up to 97% of the specimens collected in Mbalmayo. However, in Simbock, the sympatric S population accounted for 9.5% of the total number of *An. gambiae* mosquitoes tested. Chromosomal identification was reported for twenty females from Simbock ($n = 8$), Nkoteng ($n = 8$) and Dschang ($n = 4$) and additional cytogenetics data can be found in Robert *et al.* (1993) and della Torre *et al.* (2001). All females examined showed typical Forest arrangements, except one specimen from Nkoteng, chromosomally identified as Savanna. The homozygous standard karyotype (2R+/2L+) was by far the most frequent, although two 2Rb/+ heterokaryotypes were found in Dschang, and one female 2Rb/+ and two 2La/+ carriers were found in Nkoteng. In Simbock, all the specimens presented the standard uninverted homokaryotype 2R+/2L+.

Collections were conducted between December 1998 and January 2000 in Simbock, between December 1998 and April 2000 in Nkoteng, in June 2000 in Mbalmayo and in August 2000 in Dschang. Mosquitoes from the *An. gambiae* complex were visually sorted from other anophelines according to morphological identification keys (Gillies & Coetzee, 1987). All specimens were stored individually and kept at –20 °C until further analysis.

DNA extraction and genotype scoring

Genomic DNA was extracted from wings or legs of each individual mosquito as described by Cornel & Collins (1996) and resuspended in 100 µL of TE buffer. Only female *An. gambiae* specimens were included in the analysis after species identification was carried out by the diagnostic PCR described by Scott *et al.* (1993). Determination of the molecular form (M/S) of each specimen was done by the diagnostic PCR-based assay recently described by Favia *et al.* (2001), after validation of the technique for the study of field specimens collected in Cameroon (C. Wondji *et al.*, unpublished). Each specimen was unequivocally assigned to one or the other of the M or S forms and no hybrid pattern was observed.

Ten microsatellite loci were selected from published *An. gambiae* sequence data (Zheng *et al.*, 1996; Wang *et al.*, 1999) based on high polymorphism, mean allele sizes ranging from 80 to 150 bp, and cytological location spread over the three chromosome pairs. We used loci AgXH80, AgXH99 and AgXH49 on chromosome X, loci Ag2H417, Ag2H769 and Ag2H325 on chromosome 2 and loci Ag3H93, Ag3H555, Ag3H170 and Ag3H750 on chromosome 3. Microsatellite alleles were PCR amplified as previously described (Simard *et al.*, 1999, 2000). PCR products were loaded on 15% non-denaturing polyacrylamide gels and the allelic bands were visualized after rapid silver staining (Sanguinetti *et al.*, 1994).

A set of approximately fifty specimens per population was investigated but, due to the strong predominance of one molecular form to the other in each of the locations sampled, only specimens from the S form were further considered from the villages of Dschang and Nkoteng, while only the M population from Mbalmayo could be investigated. Intensive sampling in Simbock, however, allowed us to consider one sample of each molecular form, although only thirty-three specimens of the S form were available.

Data analysis

Microsatellite-based genotype frequencies were tested against Hardy–Weinberg expectations for each locus in the pooled population, in each location, and within M and S populations from Simbock. Statistical significance was assessed by the exact probability test available in GENEPOP 3.2 (Raymond & Rousset, 1995). Exact tests of linkage disequilibrium between pairs of loci were computed in the pooled population and within each population, using GENEPOP 3.2. The sequential Bonferroni procedure (Holm, 1979) was applied to evaluate significance when multiple tests were performed.

Differentiation between populations was assessed by *F*-statistics (Wright, 1978), calculated according to Weir & Cockerham (1984). Significance of *Fis* and *Fst* was assessed using the exact probability test (Rousset & Raymond, 1995) and the G-based exact test of genotypic differentiation (Goudet *et al.*, 1996), respectively (tests available in GENEPOP 3.2). Where mentioned, 95% confidence intervals around *Fst* estimates were built by bootstrapping over loci (1000 randomizations).

Acknowledgements

This work was supported in part by grant N°980101 from MIM/TDR/WHO and by the French Institut de Recherche pour le Développement (IRD) – Research unit R016. We are indebted to T. Tchuinkam, C. A. Nkondjio, P. Awono-Ambene, J.-C. Toto, J.-Y. Meunier, S. Zebaze-Kemleu and R. Nyambam and for valuable help in the field and useful technical support in the lab. We thank F. Rousset, M. Raymond and three anonymous reviewers for critical comments that greatly improved earlier versions of the manuscript.

References

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.

Chandre, F., Manguin, S., Brengues, C., Dossou-Yovo, J., Darriet, F., Diabaté, A., Carnevale, P. and Guillet, P. (1999) Current distribution of pyrethroid resistance gene (*kdr*) in *Anopheles gambiae* complex from West Africa and further evidence for reproductive isolation of the Mopti form. *Parassitologia* **41**: 319–322.

Coluzzi, M. (1982) Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes. In: *Mechanisms of Speciation* (ed. C. Barigozzi), pp. 143–153. A.R. Liss, Inc., New York.

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

Coluzzi, M., Petrarca, V. and Di Deco, M.A. (1985) Chromosomal inversion intergradation an 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 environments in the *Anopheles gambiae* complex. *Trans R Soc Trop Med Hyg* **73**: 483–497.

Cornel, A. and Collins, F.H. (1996) PCR of the Ribosomal DNA Intergenic Spacer Regions as a method for identifying mosquitoes in the *Anopheles gambiae* complex. *Meth Mol Biol* **50**: 321–332.

Di Deco, M.A., Petrarca, V., Villani, F. and Coluzzi, M. (1980) Polimorfismo cromosomico da inversioni paracentriche ed eccesso degli eterocariotipi in ceppi di *Anopheles* allevati in laboratorio. *Parassitologia* **22**: 304–306.

Donnelly, M.J. and Townson, H. (2000) Evidence for extensive genetic differentiation among populations of the malaria vector *Anopheles arabiensis* in Eastern Africa. *Insect Mol Biol* **9**: 357–367.

Epplen, C., Melmer, G., Siedlaczck, I., Schwaiger, F.W., Mäueler, W. and Epplen, J.T. (1993) On the essence of 'meaningless' simple repetitive DNA in Eucaryotes genomes. In: *DNA Fingerprints: State of the Science* (eds S. D. J. Pena, R. Chakraborty, J. T. Epplen and A. J. Jeffreys), pp. 29–45. Birkhauser Verlag, Basel, Switzerland.

Favia, G., della Torre, A., Bagayoko, M., Lanfrancotti, A., Sagnon, N'F., Touré, Y.T. and Coluzzi, M. (1997) Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Mol Biol* **6**: 377–383.

Favia, G., Dimopoulos, G., della Torre, A., Touré, Y.T., Coluzzi, M. and Louis, C. (1994) Polymorphism detected by random PCR distinguish between different chromosomal forms of *Anopheles gambiae*. *Proc Natl Acad Sci USA* **91**: 10315–10319.

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.

Flechter, D.J.C. and Michener, C.D. (1987) *Kin Recognition in Animals*. John Wiley & Sons, New York.

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*, 2nd edn. Pub South Afr Inst Med Res, Johannesburg, South Africa. 55, p. 143.

Goudet, J., Raymond, M., De Meeüs, T. and Rousset, F. (1996) Testing differentiation in diploid populations. *Genetics* **144**: 1933–1940.

Higgie, M., Chenoweth, S. and Blows, M.W. (2000) Natural selection and the reinforcement of mate recognition. *Science* **290**: 519–521.

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

Lanzaro, G.C., Touré, Y.T., Carnahan, J., Zheng, L., Dolo, G., Traoré, S., 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**: 14260–14265.

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

Lehmann, T., Hawley, W.A., Kamau, L., Fontenille, D., Simard, F. and Collins, F.H. (1996a) Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. *Heredity* **77**: 192–200.

Milligan, P.J.M., Phillips, A., Broomfield, G., Molyneux, D.H., Touré, Y.T. and Colluzzi, M. (1993) A Study of the use of gas chromatography of cuticular hydrocarbons for identifying members of the Anopheles gambiae (Diptera: Culicidae) complex. *Bull Entomol Res* **83**: 613–624.

Mukabayire, O., Caridi, J., Wang, X., Touré, 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.

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

Powell, J.R., Petrarca, V., della Torre, A., Caccone, A. and Coluzzi, M. (1999) Population structure, speciation and introgression in the *Anopheles gambiae* complex. *Parassitologia* **41**: 101–113.

Raymond, M. and Rousset, F. (1995) Genepop (version 1.2), a population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.

Robert, V., Petrarca, V., Le Goff, G. and Manga, L. (1993) Quelques données cytogénétiques sur le complexe *Anopheles gambiae* au Sud-Cameroun. *Bulletin de liaison et de documentation de l'OCEAC* **26**: 99–101.

Rousset, F. (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* **142**: 1357–1362.

Rousset, F. and Raymond, M. (1995) Testing heterozygote excess and deficiency. *Genetics* **140**: 1413–1419.

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**: 915–919.

Scott, J.A., Brogdon, W.G. and Collins, F.H. (1993) Identification of single specimens of the *Anophles 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**: 1000–1009.

Simard, F., Lehmann, T., Lemasson, J.J., Diatta, M. and Fontenille, D. (2000) Persistence of *Anopheles arabiensis* during the severe dry season conditions in Senegal: an indirect approach using microsatellite loci. *Insect Mol Biol* **9**: 467–479.

Takahata, N. (1992) Evolutionary genetics of human paleopopulations. In: *Mechanisms of Molecular Evolution: Introduction to Molecular Biology* (eds N. Takahata and A. G. Clark), pp. 1–21. Sinauer Associates Inc, Sunderland, MA.

Taylor, C., Touré, Y.T., Carnahan, J., Norris, D.E., Dolo, G., Traoré, S., Edillo, F.E. and Lanzaro, G.C. (2001) Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics* **157**: 743–759.

della Torre, A., Fanello, C., Akogbeto, M., Dossou-Yovo, J., Favia, G., Petrarca, V. and Coluzzi, M. (2001) Molecular evidence of

incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Mol Biol* **10**: 9–18.

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. 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. (1998) Distribution and inversion polymorphism of chromosomally recognised taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia* **40**: 477–511.

Tripet, F., Touré, Y.T., Taylor, C.E., Norris, D.E., Dolo, G. and Lanzaro, G. (2001) DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Mol Ecol* **10**: 1725–1732.

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**: 266–272.

Wang, R., Kafatos, F.C. and Zheng, L. (1999) Microsatellite markers and genotyping procedures for *Anopheles gambiae*. *Parasitol Today* **15**: 33–37.

Wattier, R., Engel, C.R., Saumitou-Laprade, P. and Valero, M. (1998) Short allele dominance as a source of heterozygote deficiency at microsattellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Mol Ecol* **7**: 1569–1573.

Weill, M., Chandre, F., Brengues, C., Manguin, S., Akogbeto, M., Pasteur, N., Guillet, P. and Raymond, M. (2000) The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae* s.s. through introgression. *Insect Mol Biol* **9**: 451–455.

Weir, B.S. and Cockerham, C.C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

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 human malaria vector mosquito, *Anopheles gambiae*. *Genetics* **143**: 941–952.