Journal of the American Mosquito Control Association, 14(3):266–272, 1998
Copyright © 1998 by the American Mosquito Control Association, Inc.

THE USE OF MICROSATELLITES TO STUDY GENE FLOW IN NATURAL POPULATIONS OF *ANOPHELES* MALARIA VECTORS IN AFRICA: POTENTIAL AND PITFALLS

C. WALTON,[1] N. J. THELWELL,[1] A. PRIESTMAN[2] AND R. K. BUTLIN[1]

ABSTRACT. The potential of microsatellites as population genetic markers in the malarial vectors *Anopheles gambiae* and *Anopheles arabiensis* was assessed using 4 loci. Substantial genetic divergence was found not only between these species but also between the Mopti and Forest chromosomal forms of *An. gambiae,* demonstrating that microsatellites do have the power to detect barriers to gene flow in these mosquitoes. However, application and interpretation of microsatellites was not necessarily straightforward. Despite the use of semiautomated fluorescent technology that enabled fragment sizes to be determined precisely, some difficulty was encountered in allele classification. Sequence analysis revealed insertions/deletions and base changes in the flanking regions of the microsatellite as the probable cause of this problem. The implications of this and other potential pitfalls in the use of microsatellites to study vector populations are discussed.

KEY WORDS *Anopheles gambiae, Anopheles arabiensis,* microsatellites, natural populations, null alleles, homoplasy, ecophenotypes

INTRODUCTION

Anopheles gambiae Giles, and to a lesser extent *Anopheles arabiensis* Patton, are major vectors of malaria in Africa that have largely eluded attempts at conventional vector control. Considerable effort is now being directed towards the possibility of introducing genes into the vectors that make them refractory to parasite transmission (e.g., Crampton et al. 1994). If this strategy is to be realized, such genes will need to be driven through natural populations. With this approach, as with other methods of vector control, an understanding of the genetic structure of natural populations of malaria vectors is essential.

Although the tolerance of *An. arabiensis* to aridity allows it to predominate over *An. gambiae* in drier habitats and during the dry season, both species can be found in sympatry over extensive areas of sub-Saharan Africa (Coluzzi et al. 1979). Introgression between *An. gambiae* and *An. arabiensis* is possible via female hybrids and such hybrids have been found at a rate of about 0.1% in natural populations (White 1970, Coluzzi et al. 1979). Evidence from sequence data suggests that nuclear as well as mitochondrial introgression has occurred in these species (Besansky et al. 1994, Caccone et al. 1996, García et al. 1996).

Both *An. gambiae* and *An. arabiensis* have a large number of paracentric inversion polymorphisms. In West Africa considerable heterogeneity exists within what is still currently recognized as the single species of *An. gambiae.* The association of different 2R karyotypes with particular environments led to the term ecophenotype for the different chromosomal forms of *An. gambiae* known as Savanna, Mopti, Bamako, Bissau, and Forest. No postmating barriers to reproduction exist between these forms (Di Deco et al. 1980). However, at some sympatric sites the observation that karyotype frequencies depart from Hardy–Weinberg expectations has been taken as evidence for incipient speciation (Coluzzi et al. 1985). This has recently been supported by molecular evidence (Favia et al. 1997).

The detection of gene flow within and between such closely related sibling species can be difficult. For this reason microsatellite loci, which are highly polymorphic within a species, are being used increasingly as genetic markers in population studies. Many microsatellites have already been developed in *An. gambiae* as a result of a genetic mapping program (Zheng et al. 1996). Microsatellites are simple sequence motifs that can vary in the number of repeats. They are thought to evolve primarily by an increase or decrease of one repeat unit so that most statistics used in their analysis are based on a stepwise mutation model (Goldstein et al. 1995, Slatkin 1995, Jarne and Lagoda 1996). Microsatellites offer many advantages over conventional markers: they are apparently neutral, distributed throughout the genome, and codominant. Their high levels of length polymorphism (generated by a high mutation rate) should give them greater power to resolve differentiation between populations.

Some of these microsatellites have already been shown to be more variable than allozymes within a natural population of *An. gambiae* (Lanzaro et al. 1995). However, when used in comparing populations of the Savanna karyotype of *An. gambiae,* surprisingly little variation was found between populations from eastern and western Africa (Lehmann et al. 1996b). In the latter case, the degree of divergence was comparable to allozymes that would be expected to be less powerful. Although this may be due to migration or recent population expansion,

[1] Ecology and Evolution Programme, School of Biology, University of Leeds, Miall Building, Clarendon Way, Leeds LS2 9JT, United Kingdom.

[2] Division of Biology, School of Science, Staffordshire University, College Road, Stoke-on-Trent, ST4 2DE, United Kingdom.

lack of resolving power due to size constraints on the microsatellites cannot be excluded.

If microsatellite markers are to be used successfully to address the issue of genetic differentiation of vector populations, it is important that we are fully aware of both their full capabilities and whatever pitfalls might be associated with their use. Previous studies have concentrated on looking at gene flow with respect to geographical distance using only the Savannah karyotype of *An. gambiae* (Lehmann et al. 1996b, 1997). The goal of this study was to further investigate the potential of microsatellites as population markers in both *An. gambiae* and *An. arabiensis,* and also to draw attention to some problems with the use of microsatellites. We have therefore used populations of *An. gambiae,* characterized according to paracentric inversions on chromosome 2R (Coluzzi et al. 1985), and populations of *An. arabiensis* from western Africa.

MATERIALS AND METHODS

Mosquito DNA samples: The mosquitoes were F_1 progeny of wild-caught females from western Africa that had been karyotyped using preparations of the ovarian nurse cells (Coluzzi et al. 1985, Milligan et al. 1993). The specimens came from 3 sites in Mali and 3 sites in Liberia (Table 1). The sites extend from Suakoko in Liberia along a line running approximately northeast to: at 100 km, Bonah and JDF (near Yackepa); 600 km further to Banambani and Moribabougou (near Bamako); and 600 km further to Diré (see Milligan et al. 1993 for further details). Despite being stored dry for more than 10 years before the study, it was possible to extract DNA from the mosquitoes using a silica-based DNA purification method (Höss and Pääbo 1993). To estimate the usefulness of the microsatellites as population markers only one individual was scored per family, resulting in a total of 41 unrelated karyotyped individuals being used in the analysis. Duplicates from the same family were performed in some cases and were found to be entirely consistent with Mendelian segregation.

Genotyping and DNA sequencing: Four dinucleotide repeat microsatellite loci, *AG2H143, AG2H147,* and *AG2H46* on chromosome 2 and *AGXH293* on the X chromosome were amplified (Zheng et al. 1996). A fluorescence-based semiautomated technology (Applied Biosystems, Warrington, United Kingdom) was used to determine the microsatellite genotypes. Polymerase chain reaction (PCR) amplification was performed in 25-μl volumes in 0.5-ml tubes in a Biometra (Maidstone, United Kingdom) thermal cycling machine. Each reaction contained 2 μl of genomic DNA (equivalent to $\frac{1}{75}$th of a DNA preparation from a whole mosquito), 40 μM dNTP, 1.5 mM $MgCl_2$, 0.25 units of Thermoprime plus polymerase (Advanced Biotechnologies Ltd., Epsom, United Kingdom), and 5 pmols of each primer. The forward primers were synthesized, labeled with a fluorescent dye (FAM or HEX) and purified by high-performance liquid chromatography (Oswel DNA, Southampton, United Kingdom). The cycling conditions for approximately 28 cycles (empirically determined) were denaturation at 95°C for 20 sec and annealing at 50°C for 30 sec. A defined extension step at 72°C is unnecessary to generate such short fragments and was therefore omitted from the cycling procedure. However, a final extension step at 72°C for 2 min was included before cooling to 10°C. A hot start procedure, where all the reaction components (except the enzyme) were heated to 96°C for 5 min before addition of the polymerase, was always used. The forward primers for *AG2H143* and *AG2H147* were differentially labeled, enabling these loci to be amplified and analyzed simultaneously.

The samples were combined with Genescan-350 TAMRA (Applied Biosystems) internal lane standards (50–350 base pairs [bp]). Electrophoretic separation was performed using 6% denaturing acrylamide gels and 12-cm Well-to-Read plates using a model 373A DNA Sequencer (Applied Biosystems) according to the manufacturer's instructions. Correct tracking of all the lanes was verified manually and peak sizes were estimated using the local Southern sizing option of Genescan 672 software. The amplified loci were directly sequenced using the microsatellite primers, *Taq* polymerase, and dye-terminator chemistry (Applied Biosystems).

RESULTS

Allele scoring and homoplasy: With the exception of *AGXH293,* the loci amplified well and gave distinct peaks of fluorescence that were easy to score. Locus *AGXH*293 was difficult to amplify, producing weak products with a high degree of background. Many of the alleles had highly stuttered peak profiles, presumably due to slippage of the *Taq* polymerase during amplification of the sometimes large number of repeats (inferred to be up to 43 repeats). This made them very difficult to score with confidence, particularly where heterozygotes were present for similar length alleles.

The inclusion of a differentially labeled set of size standards in each sample results in precise sizing of the fragments despite the inevitable lane-to-lane variability that occurs within a gel (Ziegle et al. 1992). Because fragment length is estimated by comparison to size standards that have a different base composition, which affects mobility, the estimated length is not necessarily an integer value. Fragment lengths and inferred genotypes for all individuals are given in Table 1. Several samples were run repeatedly on a number of gels and the largest within-sample standard deviation (SD) for fragment length was only 0.257 bp. Because the alleles are expected to fall into class sizes differing by 2 base pairs, the likelihood of misscoring an allele because an observation deviates from the

Table 1. The estimated allele sizes for the 4 loci showing the inferred numbers of repeats (excluding locus *AGXH293*) used in allele classification for the *G*-test analysis.[1]

Species, form, and site	*Ag2H46*		*AG2H147*		*AG2H143*		*AGXH293*
	Allele size (estimated bp)	Repeat number	Allele size (estimated bp)	Repeat number	Allele size (estimated bp)	Repeat number	Allele size (estimated bp)
Anopheles arabiensis							
Moribabougou, Mali	143.8	10, 10	173.9, 176.9	6, 7	157.1	3, 3	103.5
	135.3, 143.7	6, 10	176.7	7, 7	157.1	3, 3	us
	137.9, 140.0	7, 8	174.8, 176.7	6, 7	159.1, 167.0	4, 8	103.4
	144.1	10, 10	174.5, 178.6	6, 8	157.2, 167.1	3, 8	104.2
	146.3	11, 11	174.4, 179.4	6, 8	157.4, 167.8	3, 8	102.2
	139.9	8, 8	177.6, 181.7	7, 9	157.5, 167.6	3, 8	103.4
	146.6	11, 11	178.8, 181.7	8, 9	157.1, 167.2	3, 8	102.8
	146.9	11, 11	173.9, 178.9	6, 8	157.1, 167.5	3, 8	us
	149	12, 12	174.6, 178.7	6, 8	157.2	3, 3	103.6
Banambani, Mali	141.8, 144.0	9, 10	176.7	7, 7	157.2, 165.3	3, 7	103.2
	137.6, 142.1	7, 9	175.8, 179.9	?, 8	157.4, 159.1	3, 4	us
	136.8	7, 7	174.1	6, 6	159.3, 167.1	4, 8	103.5
	144.3	10, 10	174.4	6, 6	157.3, 159.1	3, 4	102.9
Anopheles gambiae							
Savanna form							
Banambani, Mali	135.8, 142.3	6, 9	176.7, 178.8	7, 8	159,1, 161.1	4, 5	99.5, 127.9
	146.8, 137.9	11, 7	176.6	7, 7	160.9, 162.9	5, 6	122.8, 148.4
Bamako form							
Banambani, Mali	137.8, 142.1	7, 9	172.3, 195.3	5, ?	159.4, 163.4	4, 6	115.3, 160.8
Moribabougou, Mali	146.3, 150.9	11, 13*	177.5, 183.3	7, 10	161.5, 163.4	5, 6	102.7, 117.0
Mopti form							
Banambani, Mali	137.3, 146.4	7, 11	174.4, 177.0	6, 7	165.2, 171.3	7, 10*	107.5, 124.5
	137.3, 139.3	7, 8	174.3, 177.3	6, 7	163.5, 165.3	6, 7	100.1
	137.2, 141.3	7, 9	174.2, 177.1	6, 7	163.3, 165.1	6, 7	us
Diré, Mali	146.6	11, 11	177.2	7, 7	161.2, 167.1	5, 8	us
	141.9	9, 9	171.9, 177.0	5, 7	165.1	7, 7	us
	144.3	10, 10	178.2	8, 8	161.3, 165.1	5, 7	102.0, 144.2
	135.1, 148.4	6, 12	177.2	7, 7	163.5, 165.2	6, 7	138.9, 108.7
Forest form							
Suakoko, Liberia	141.6, 135.3	9, 6	172.6, 177.7	5, 7	159.6	4, 4	134.8, 159.3
	146.3	11, 11	177.5, 183.5	7, 10*	161.5, 163.2	5, 6	96.1, 100.4
	137.4	7, 7	174.7, 176.8	6, 7	161.8	5, 5	111.8, 113.7
	144	10, 10	175.1	6, 6	160.5, 163.6	?, 6	104.1
	141.6	9, 9	174.7	6, 6	161.3	5, 5	98.2, 101.3
	146.2, 140.3	11, 8	171.4, 175.2	5, 6	160.4	?, ?	97.9, 105.7
	144.1	10, 10	174.9	6, 6	160.1	?, ?	101.5
	139	?	175	6, 6	161.2	5, 5	104.1, 146.3
	135.2, 141.6	6, 9	174.5	6, 6	161.2, 163.1	5, 6	106.3
JDF, Liberia	141.1, 148.0	9, 12	172.6, 184.8	5, 10*	157.1, 161.0	3, 5	111.5, 115.4
	137.5	7, 7	172.9, 174.6	5, 6	161.4	5, 5	122.6, 125.0
	139.9, 144.5	8, 10	176.9	7, 7	161.1, 165.1	5, 7	us
	139.8, 146.4	8, 11	173.3, 183.5	5, 10*	161.7	5, 5	122.6
	137.5	7, 7	173.5, 175.3	5, 6	161.6	5, 5	us
Bonah, Liberia	144.3, 148.8	10, 12	173.5, 175.4	5, 6	161.9	5, 5	115.2, 118.9
	142.0, 146.5	9, 11	172.2, 177.3	5, 7	161.4	5, 5	us
	136.0, 140.0	6, 8	174.9, 176.8	6, 7	161.1	5, 5	128.1, 139.6

[1] bp, base pairs; us, unable to score alleles; ?, unable to assign allele to size class; *, samples excluded from analysis to avoid low sample sizes.

mean size for an allele by more than 1 bp (4 SD) should therefore be extremely low. Given this, the variance within an allele size class was often unexpectedly large (e.g., SD value of 0.678 bp for the 5-repeat class, locus *AG2H147*; Fig. 1). This suggests that alleles are not homogeneous within an allele class and presumably differ from each other in the base composition of their flanking regions. This was confirmed by direct sequencing of 6 PCR products of *AG2H147* from homozygotes, which revealed a number of base substitutions in the flanking regions (Table 2).

The distribution of allele sizes for this locus was almost continuous (Fig. 1) and assignment to allele

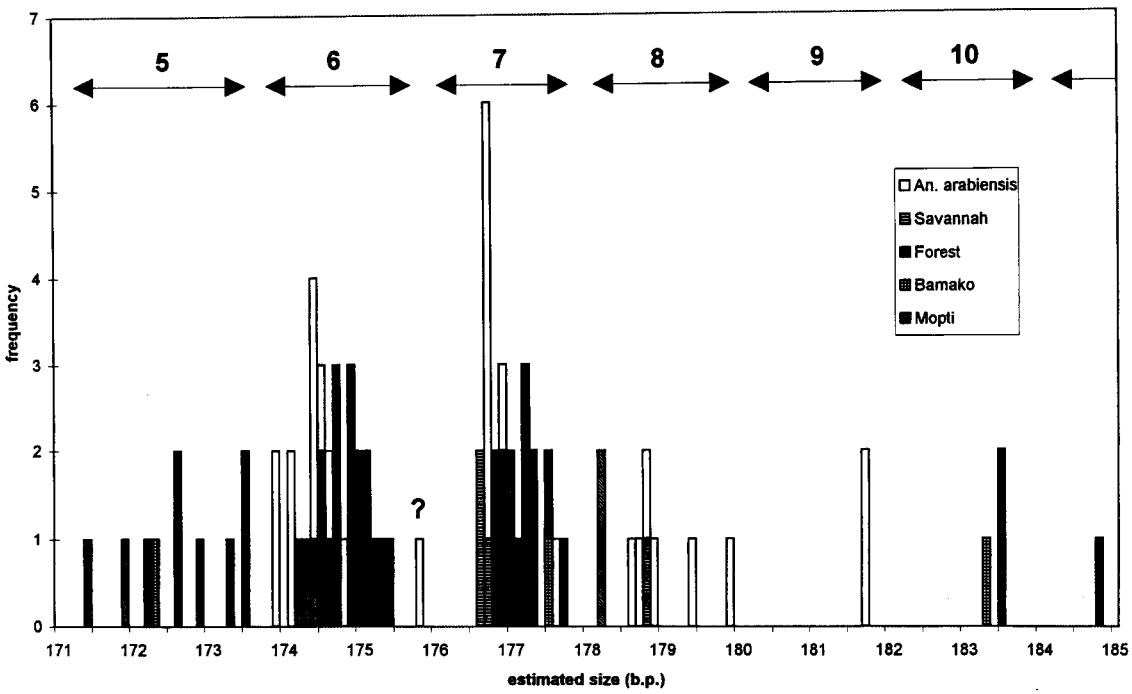

Fig. 1. Range of allele sizes at the *AG2H147* locus. An allele from an *Anopheles gambiae* Bamako individual that was estimated to be 195.3 base pairs is off this scale and not shown. The numbers above the arrows denote the inferred numbers of repeats in the allele classes.

classes had to be carried out subjectively. A suggestion of clustering of allele sizes according to ecophenotype or species also occurred, which would be consistent with differing base compositions of flanking sequences between these groups. In terms of inferring the correct number of repeats from allele size, an even greater problem was the presence of a 2-bp insertion in a run of Ts in 3 individuals of the Forest form from Suakoko, Liberia (Table 2). Such variation might be expected in interspecific comparisons but even within a chromosomal type observations occurred that differed by a noneven number of bases (Fig. 1).

Null alleles: Given the variation in flanking sequence described above it would not be surprising to find null alleles, that is differences at the priming site that prevent the amplification of one or both alleles. Because the samples were divided between many populations, testing for deviations from Hardy–Weinberg equilibria was not possible. With the *AGXH293* primers, there was no amplification from some individuals and an excess of apparent homozygotes. This might indicate the presence of null alleles, but it is also possible that it is due to the difficulty in amplifying and scoring this locus. However, null alleles have been reported in *AG2H46* (Lehmann et al. 1996a) and we had also rejected another locus, *AG2H161,* for which we

Table 2. Alignment of sequence excerpts of the polymerase chain reaction amplified fragments to illustrate compensatory length changes in the flanking regions.

Chromosomal form and site	Size (bp)	No. of GT repeats: Inferred	No. of GT repeats: Actual	Microsatellite and flanking sequence[1]
				* ** ** *
Forest; Suakoko, Liberia	174.93	6	5	AATTGTGTGTGTGT....AGCTT (N_{44}) TAGTTTTTTTTTTCTT
Forest; Suakoko, Liberia	174.71	6	5	AATTGTGTGTGTGT....AGCTT (N_{44}) TAGTTTTTTTTTTCTT
Forest; Suakoko, Liberia	175.08	6	5	AATTGTGTGTGTGT....AGCTT (N_{44}) TAGTTTTTTTTTTCTT
Forest; JDF, Liberia	176.86	7	7	AATTGTGTGTGTGTGTGTGGCTT (N_{44}) TAGTT..TTTTTCTT
Mopti; Diré, Mali	177.24	7	7	AATTGTGTGTGTGTGTGTGGCTT (N_{44}) TCGTT..TTCTTCTT
Savanna; Banambani, Mali	176.59	7	7	AATTGTGTGTGTGTGTGTGGCTT (N_{44}) GCGCG..TTTTTCTT

[1] The line indicates the core microsatellite, asterisks denote base changes, and N_{44} denotes the intervening number of bases.

Table 3. Heterozygosity and *G* values for loci *AG2H46, AG2H143,* and *AG2H147.*[1]

Locus	Chromosome and location in *Anonopheles gambiae*	Heterozygosity			*G* Values for pairwise comparisons of allele frequencies			
		Aa	M	F	df	Aa:F	Aa:M	M:F
AG2H46	II—tip of 2R	0.83	0.89	0.87	5	3.79	2.92	1.36
AG2H143	II—within *2La* inversion	0.64	0.76	0.56	5	58.32***	37.59***	23.86**
AG2H147	II—2L	0.72	0.65	0.69	3	19.64**	5.32	12.22*

[1] The heterozygosity was estimated according to Nei (1978). The data for the Savanna and Bamako forms and for low-frequency alleles were excluded from the *G*-test analysis to avoid low sample sizes. Probability values are *** $P \ll 0.001$; ** $P < 0.001$; * $P < 0.01$. df, degrees of freedom; Aa, *Anopheles arabiensis*; M, Mopti form of *An. gambiae*; F, Forest form of *An. gambiae.*

were not able to amplify products from some individuals. At locus *AG2H143* in *An. gambiae,* the deficiency of 5-repeat allele heterozygotes (only 4 when 8 might have been expected under Hardy–Weinberg equilibrium) would appear to indicate the presence of null alleles. However, when the frequency (*P*) of the 5-repeat allele at each of the 3 sites is taken into account ($P = 1.0$ at Bonah, $P = 0.8$ at JDF, and $P = 0.44$ at Suakoko), no significant departures from expected values occurred.

Population structure: The heterozygosities of the loci, not including *AGXH293* (Table 3), are consistent with the mean heterozygosity of 0.732 (±0.06) reported previously for other microsatellite markers in a population of *An. gambiae* from Banambani, Mali (Lanzaro et al. 1995). With the exception of *AG2H143,* the levels of heterozygosity were similar between the Mopti and Forest forms of *An. gambiae* and *An. arabiensis.* The low heterozygosity of the Forest sample at locus *AG2H143* was apparent from the preponderance of the 5-repeat allele (22 out of 34 alleles).

Although the difficulty in scoring all alleles at *AGXH293* meant that a meaningful analysis of this locus was precluded, the available data nevertheless indicate that the locus has interesting properties. In *An. gambiae*, 21 alleles were observed with alleles estimated to range from 12 to 43 repeats (median of 21). In striking contrast, in *An. arabiensis* the allele length was uniform (inferred repeat number of 14). Although this monomorphism could simply be due to changes in the structure of this locus in *An. arabiensis,* another possibility might be a hitchhiking effect on a linked locus that has recently experienced a selective sweep. The latter is another problem associated with microsatellites, particularly where there are inversions, which will extend the length of the genomic region influenced by selection on a locus within an inversion.

To assess the potential of the microsatellites to detect barriers to gene exchange, the allele frequency distributions for the loci (excluding *AGXH293*) have been examined using the *G* test for independence (Sokal and Rohlf 1995). A full-scale population analysis, which is not the aim of the present study, would obviously require larger sample sizes and further loci. The 3 pairwise comparisons were made between *An. arabiensis* and the Forest and Mopti forms of *An. gambiae* (Table 3). Because the samples were pooled according to species or chromosomal form, any additional component of variation due to site was lost in the analysis. Despite the difficulties at *AG2H147* discussed above, the majority of alleles were assigned to classes differing by 2 bp. At *AG2H143* 5 nonstandard alleles differing by 1 bp from the nearest class were found in the same population from Suakoko, Liberia. Because their sizes were not only the same but also were highly reproducible, it is likely that they are identical alleles found within this population. A total of 6 unclassifiable alleles were excluded from the analysis, although this was found to make no overall qualitative difference to the results.

Where differences in allele frequencies were found, they were all highly significant, demonstrating the potential of these microsatellites as population markers. Variation occurs not only in interspecific comparisons but also between the Mopti and Forest forms of *An. gambiae.* At *AG2H143,* marked differences occurred, with the most common allele differing between *An. arabiensis* and the Forest and Mopti forms of *An. gambiae.* These samples were collected from many sites with *An. arabiensis,* and to a lesser extent Mopti, being able to tolerate the drier sites in Mali, whereas all samples of the Forest ecophenotype, which has little tolerance of aridity, came from the forest regions of Liberia. Therefore, some of the divergence observed between *An. arabiensis* and the Mopti and Forest forms of *An. gambiae* possibly may be due to geographic isolation.

DISCUSSION

This study raises a number of concerns in the use of microsatellites to study gene flow within and between *An. gambiae* and *An. arabiensis.* As in all such studies of population structure, the correct application of data from microsatellites relies critically on a correct understanding of their rate and mode of evolution. This topic has received considerable attention in numerous theoretical papers (e.g., Nauta and Weissing 1996, Feldman et al. 1997; and see Jarne and Lagoda 1996 and references therein). As a general rule, statistics based on a stepwise mutation model (SMM) such as R_{ST} (Slatkin 1995) or

D_1 (Goldstein et al. 1995) seem more appropriate for microsatellite analysis than F statistics used for allozyme data (Wright 1978) that are based on an infinite alleles model. (The R_{ST} values reported for microsatellites in Lehmann et al. (1996b) were, as expected, generally higher than the corresponding F_{ST} values, although this made little difference to the interpretation in this case.) However, even with the SMM it is assumed that repeat number has no upper limit. If the allele size has limits, due to mutation bias or selection, they will lead to an underestimation of the differentiation between populations. Constraints on allele size at locus *AG2H46* have been suggested by Lehmann et al. (1996a) and the similarity in allele sizes despite variation in the flanking regions at locus *AG2H147* is also consistent with this.

The mutation rate of microsatellites is a major factor in determining the level of variability maintained in populations. This rate can be very high, for example about 1×10^{-3} per locus per gamete per generation in humans (Weber and Wong 1993). However, microsatellites in *An. gambiae* and *An. arabiensis* are more likely to have a relatively low mutation rate similar to that observed in *Drosophila.* The low average mutation rate of 6.3×10^{-6} per generation measured in *Drosophila* (Schug et al. 1997) is thought to be due to the small number of repeats present relative to human microsatellites, but that are comparable to the repeat numbers observed in *An. gambiae* (Zheng et al. 1996). Clearly the mutation/selection dynamics are not necessarily the same at all loci because at *AGXH293* the alleles have many repeats and cover a wide size range. Although this marker was difficult to score, similar loci, where little constraint appears to exist on allele size, have the potential to be highly informative.

Another concern was the allelic homoplasy, as indicated by the presence of base substitutions and insertion/deletions in the flanking regions of locus *AG2H147.* This has also been observed elsewhere (Estoup et al. 1995, Lehmann et al. 1996b, Grimaldi and Crouau-Roy 1997) and is probably a general problem that will become increasingly apparent as microsatellites are studied in more detail. Allelic homoplasy violates the assumption that differences between alleles are entirely due to differing numbers of repeat units. Assuming that the mutation rate for base substitutions in the flanking regions is substantially lower than that of the microsatellite itself, then some of these fragments of apparently the same size are clearly not identical by descent. This is not in itself a problem with distance measures such as R_{ST} (Slatkin 1995) and D_1 (Goldstein et al.1995) that are based on the SMM, but if alleles of the same size are not more likely to be closely related to each other than they are to alleles of other sizes then the extent of the differences between populations will effectively be concealed. Comparisons between species and ecophenotypes, or over large geographic ranges, are particularly likely to be prone to this problem.

Some evidence was found of null alleles in this study at loci *AGXH293* and *AG2H161* and one at locus *AG2H46* has been well documented previously (Lehmann et al. 1996a). The occurrence of null alleles is a common problem when sampling natural populations (Callen et al. 1993). If the sequence variation found in the flanking region of locus *AG2H147* is typical, null alleles may be fairly common. Encountering null alleles would be expected to be even more likely when studying species, or even ecophenotypes, that were not those from which the microsatellite loci were isolated.

Any gene flow between different karyotypes, whether within or between species of *An. gambiae* and *An. arabiensis,* would not necessarily be equal across the genome. In inversion heterokaryotypes, recombination will be inhibited around the break points or where multiple, close or overlapping inversions occur. Therefore, it is interesting to note that when the chromosomal position of the markers relative to inversions is considered, the differences in allele frequency distributions between *An. arabiensis* and the Mopti and Forest forms are not inconsistent with introgression events that are thought to have occurred in nature (Table 3).

For example, inversion *2La,* which is found in *An arabiensis* and appears to convey adaptation to aridity, is absent from the Forest form and is polymorphic in Mopti. Considerable differentiation occurs in allele frequencies at *AG2H143,* which maps to the *2La* inversion, and at the nearby locus, *AG2H147.* The notion that the *2La* inversion has introgressed into *An. gambiae* from *An. arabiensis* (Coluzzi et al. 1979) has recently been supported by laboratory crosses (della Torre et al. 1997). The most extreme difference observed between *An. arabiensis* and *An. gambiae* was at locus *AGXH293,* which is near the distal breakpoint of the *Xag* inversion. Sequence data have shown that the *Xag* inversion, which is fixed in *An. gambiae* and is absent from *An. arabiensis,* is monophyletic (García et al. 1996) and is unable to introgress in laboratory crosses (della Torre et al. 1997). By contrast, at *AG2H46,* distal to the complex inversions of 2R in all cases, no significant differences were found. These results would suggest that it should be possible to use multiple loci situated within and outside inversions to estimate levels of gene flow with respect to these inversions.

Despite the above concerns for the analysis and interpretation of microsatellite data, examination of the 4 loci studied here nevertheless demonstrates that microsatellites do have considerable potential in population studies of *An. gambiae* and *An. arabiensis* and could be used to examine gene flow relative to inversions. Highly significant variation not only occurred in the distribution of allele frequencies between *An. arabiensis* and *An. gambiae* but also between the Mopti and Forest forms of *An.*

gambiae. The problems encountered with null alleles, scoring stuttered loci and size homoplasy, and the differing characteristics between microsatellite loci demonstrate the importance of screening potential loci before embarking on a large-scale analysis. These points are also likely to apply to the study of other vector species using microsatellites.

ACKNOWLEDGMENTS

We thank M. Coluzzi for providing the karyotyped mosquito samples and L. Zheng for primer details. This work was funded by the World Health Organization.

REFERENCES CITED

Besansky, N. J., J. R. Powell, A. Caccone, D. M. Hamm, J. A. Scott and F. H. Collins. 1994. Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. Proc. Natl. Acad. Sci. USA 91:6885–6888.

Caccone, A., B. A. García and J. R. Powell. 1996. Evolution of the mitochondrial DNA control region in the *Anopheles gambiae* complex. Insect Mol. Biol. 5:51–59.

Callen, D. F., A. D. Thompson, Y. Shen, H. A. Phillips, R. I. Richards, J. C. Mulley and G. R. Sutherland. 1993. Incidence and origin of null alleles in the (AC)n microsatellite markers. Am. J. Hum. Genet. 52:922–927.

Coluzzi, M., V. Petrarca and M. Di Deco. 1985. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. Boll. Zool. Pubb. Unione Zool. Ital. 52:45–63.

Coluzzi, M., A. Sabatini, V. Petrarca and M. A. Di Deco. 1979. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. R. Soc. Trop. Med. Hyg. 73:483–498.

Crampton, J. M., A. Warren, G. J. Lycett, M. A. Hughes, I. L. Comley and P. Eggleston. 1994. Genetic manipulation of insect vectors as a strategy for the control of vector-borne disease. Ann. Trop. Med. Parasitol. 88:3–12.

della Torre, A., L. Merzagora, J. R. Powell and M. Coluzzi. 1997. Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. Genetics 146:239–244.

Di Deco, M. A., M. Petrarca, F. Villani and M. Coluzzi. 1980. Recombination and linkage disequilibria between chromosome-2 inversions in Anopheles gambiae s.s. Abstr. 87. Proc. 3rd Eur. Multicoll. Parasitol., September 7–13, 1980, Cambridge, United Kingdom.

Estoup, A., C. Tailliez, J.-M. Cornuet and M. Solignac. 1995. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol. Biol. Evol. 12(6):1074–1084.

Favia, G., A. della Torre, M. Bagayoko, A. Lanfrancotti, N. Sagnon, Y. T. Touré and M. Coluzzi. 1997. Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. Insect Mol. Biol. 6:377–383.

Feldman, M. W., A. Bergman, D. D. Pollock and D. B. Goldstein. 1997. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. Genetics 145:207–216.

García, B. A., A. Caccone, K. D. Mathiopoulos and J. R. Powell. 1996. Inversion monophyly in African anopheline malaria vectors. Genetics 143:1313–1320.

Goldstein, D. B., A. R. Linares, M. W. Feldman and L. L. Cavalli-Sforza. 1995. An evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463–471.

Grimaldi, M.-C. and B. Crouau-Roy. 1997. Microsatellite allelic homoplasy due to variable flanking sequences. J. Mol. Evol. 44:336–340.

Höss, M. and S. Pääbo. 1993. DNA extraction from Pleistocene bones by a silica-based purification method. Nucleic Acids Res. 21:3913–3914.

Jarne, P. and P. J. L. Lagoda. 1996. Microsatellites, from molecules to populations and back. Trends Ecol. Evol. 11:424–429.

Lanzaro, G. C., L. Zheng, Y. T. Toure, S. F. Traore, F. C. Kafatos and K. D. Vernick. 1995. Microsatellite DNA and isozyme variability in a West African population of *Anopheles gambiae*. Insect Mol. Biol. 4:105–112.

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

Lehmann, T., N. J. Besansky, W. A. Hawley, T. G. Fahey, L. Kamau, and F. H. Collins. 1997. Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. Mol. Ecol. 6:243–253.

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

Milligan, P. M. J., A. Phillips, G. Broomfield and D. H. Molyneux. 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.

Nauta, M. J. and F. J. Weissing. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics 143:1021–1032.

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

Schug, M. D., T. F. C. Mackay and C. F. Aquadro. 1997. Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nat. Genet. 15:99–102.

Slatkin, M. 1995. A measure of population sub-division based on microsatellite allele frequencies. Genetics 139: 457–462.

Sokal, R. R. and Rohlf, F. J. 1995. Biometry, 3rd ed. W. H. Freeman & Co., New York.

Weber, J. L. and C. Wong. 1993. Mutation of human short tandem repeats. Hum. Mol. Genet. 2:1123–1128.

White, G. B. 1970. Chromosomal evidence for natural interspecific hybridization by mosquitoes of the *Anopheles gambiae* complex. Nature 231:184–185.

Wright, S. 1978. Evolution and the genetics of populations, Volume 4, Variability within and among natural populations. Univ. Chicago Press, Chicago.

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

Ziegle, J. S., Y. Su, K. P. Corcoran, L. Nie, P. E. Mayrand, L. B. Hoff, L. J. McBride, M. N. Kronik and S. R. Diehl. 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics 14:1026–1031.