Use of short tandem repeats for the analysis of genetic variability in sympatric populations of *Anopheles gambiae* and *Anopheles arabiensis*

LUNA KAMAU†‡*, WILLIAM A. HAWLEY†§, TOVI LEHMANN§, ALLOYS S. S. ORAGO‡, ANTHONY CORNEL§, ZHAOXI KE§ & FRANK H. COLLINS§

†*Kenya Medical Research Institute, PO Box 54840, Nairobi, Kenya, ‡Department of Zoology, Kenyatta University, PO Box 43844, Nairobi, Kenya, and §Division of Parasitic Diseases, Centres for Disease Control and Prevention, Mailstop F22, 4770 Buford Highway, Chamblee, GA, U.S.A.*

Anopheles gambiae and *An. arabiensis* were analysed at 30 short tandem repeat (STR) loci originally developed for use in *An. gambiae*. All specimens were collected from the same village in Kilifi district, coastal Kenya. All 30 loci were amplified in the *An. gambiae* specimens, whereas 25 out of 30 loci (83.3%) were successfully amplified in the *An. arabiensis* specimens. Both species had similar levels of polymorphism for the loci that were amplified (93.3% for *An. gambiae* and 92% for *An. arabiensis*). Median F_{ST} and R_{ST} values between the two species were 0.249 and 0.197, respectively, corresponding to *Nm* values of 0.75 and 0.51, respectively, and suggesting limited interchange of genes between these species. These, together with the relatively high Nei unbiased genetic distance (0.202) between the two sibling species, are consistent with the occurrence of sympatric species with limited gene flow. F_{ST}/R_{ST} values for individual loci varied greatly (F_{ST} range 0.00–0.87; R_{ST} range 0.00–0.73), indicating that the loci differ in their ability to measure levels of differentiation between these two species. Location of loci within paracentric inversions seems to be an important factor affecting levels of differentiation measured by the different loci.

Keywords: *Anopheles gambiae*, genetic variability, microsatellites, paracentric inversions, sympatric species.

Introduction

The genetic relationship between the two most important malaria vectors in the *Anopheles gambiae* complex, *An. gambiae* and *An. arabiensis*, has been examined on the basis of inversion karyotypes (Coluzzi *et al*., 1985), allozymes (Miles, 1978; Collins *et al*., 1988) and ribosomal DNA (Collins *et al*., 1987). Although fixed inversions may be used to differentiate these species reliably, other methods have yielded results that are consistent with a relatively recent divergence of the two lineages. In the present study, we conducted a population genetic analysis using short tandem repeats (STRs) to compare sympatric *An. gambiae* and *An. arabiensis* populations from coastal Kenya.

STRs or microsatellites are relatively short (<100 bp) tracts of tandemly repeated DNA with repeat lengths of 6 bp or less (Beckman & Weber, 1992). Their exceptionally high mutation rates make them informative for working out relationships among closely related species as well as among subpopulations of a single species (Bowcock *et al*., 1994). In addition, they are co-dominant and relatively easy to score. A survey of STR loci in *An. gambiae* (Lanzaro *et al*., 1995) demonstrated that such loci are superior to allozymes for population genetic studies, because they revealed greater variability. Similar conclusions were drawn by Estoup *et al*. (1995), who observed a high level of divergence

between populations within lineages of the honey bee *Apis mellifera* using microsatellites but virtually none using allozymes. In contrast, differentiation between *An. gambiae* populations from Kenya and Senegal measured by STRs was not higher than that measured by allozymes (Lehmann *et al.*, 1996).

In the present study, genetic variability is compared in *An. gambiae* and *An. arabiensis*, with two specific goals: first, to determine the overall level of genetic divergence between these species, as measured using this novel category of genetic marker; and, secondly, to determine whether the location of loci in the mosquito genome, i.e. chromosomal location or location within paracentric inversions, affects the level of differentiation measured. Pairs of chromosomes may carry different chromosomal inversion arrangements, referred to as standard or inverted forms. Paracentric inversions are said to be fixed in a species if only one of the arrangements is found in that species and polymorphic if both arrangements may be found. We suspect that markers with greater power to distinguish these closely related species may be more useful for future studies of intraspecific differentiation. Such markers are less likely to have constraints on allele size (Deka *et al.*, 1994).

Materials and methods

Specimen collection and microsatellite analysis

Mosquitoes were collected using the pyrethrum spray collection method on 16–17 February 1996 from Paziani village, 20 km south of Malindi town (3°10′S; 40°7′E) in Kilifi District, coastal Kenya. Sixteen *An. gambiae* and 16 *An. arabiensis* specimens were collected over a period of 3 days. Genomic DNA was extracted from mosquito abdomens by the alcohol precipitation method, described by Collins *et al.* (1987), and identified as either *An. gambiae* or *An. arabiensis* by the method of Scott *et al.* (1993). Microsatellite amplification was performed using standard polymerase chain reaction (PCR) in 8-μL reaction volumes run in a Perkin-Elmer 9600 Cetus Gene Amp thermocycler with primers obtained from Zheng *et al.* (1996). The thermal cycling conditions were an initial hold at 94°C for 5 min, followed by 30 cycles of 94°C for 15 s, 55°C for 20 s and 72°C for 30 s and a final extension at 72°C for 5 min. Analysis of the PCR product was performed using the high-resolution horizontal polyacrylamide gel electrophoresis technique described by Budowle & Allen (1991). Bands were visualized after silver staining

(Cairns & Murray, 1994). A total of 30 microsatellite loci were analysed (Table 1).

Data analysis

Conformance to Hardy–Weinberg expectations was tested by exact goodness-of-fit tests using GENEPOP 1.2 (Raymond & Rousset, 1995). The sequential

Table 1 Location of loci and F_{ST}/R_{ST} values for comparisons between *Anopheles gambiae* and *An. arabiensis*

Locus	Cytological location	Inversion region	F_{ST}	R_{ST}
AGXH37	X (1C)	FI	*	*
1D1†	X (1D)	FI	0.87	0.68
AGXH80	X (2B)	FI	0.06	0.00
AGXH99	X (2C)	FI	0.74	0.16
AGXH19	X (3)	FI	*	*
AGXH180	X (3A)	FI	0.26	0.51
AGXH1002	X (3C)	FI	*	*
AGXH503	X (4B)	FI	*	*
AGXH77	X	FI	0.14	0.07
AGXH293	X	FI	0.42	0.33
AGXH7	X	FI	0.34	0.44
AG2H46	II (7A)	OI	0.06	0.02
9C†	II (9C)	OI	**	**
AG2H141	II (14B)	2Rd	0.07	0.20
AG2H91	II (15C)	2Rd	0.72	0.19
22C†	II (22C)	OI	0.05	0.05
AG2H147	II	OI	0.00	0.06
AG2H143	II	2La	0.32	0.18
AG2H161	II	2La	0.51	0.73
AG2H815	II	OI	*	*
29C†	III (29C)	OI	0.00	0.00
33C†	III (33C)	3Ra	0.07	0.17
AG3H750	III (39)	OI	0.08	0.04
AG3H577	III (42A)	OI	0.13	0.16
45C†	III (45C)	OI	−0.01	0.00
AG3H119	III (35B)	OI	0.33	0.42
AG3H88	III	OI	0.32	0.11
AG3H83	III	OI	0.29	0.14
AG3H746	III	OI	0.14	0.07

FI, fixed inversion; OI, outside inversion.
*Did not amplify for *An. arabiensis*.
**Monomorphic in both species.
†These microsatellites have been mapped within certain genes: 1D1 is within the actin 1D gene (GeneBank accession number U02930), 9C within adult cDNA-69, 22C within an unidentified esterase gene, 29C within the xanthine dehydrogenase gene and 45C within the ATPase subunit 2-like cDNA (F. H. Collins, unpubl. data); 33C is within the dopa decarboxylase gene (P. Romans, unpubl. data).

Bonferroni procedure (Holms, 1979) and the binomial probability test were used to evaluate the overall significance of the values obtained. F-statistics were calculated using the method of Weir & Cockerham (1984), whereas R_{ST} was calculated according to Slatkin (1995). Differences between F_{ST}/R_{ST} values for different loci combinations were tested using the Wilcoxon rank sum test and the median test. Migration indices (Nm) were calculated from F_{ST} values by rearranging the formula of Wright (1969), $F_{ST} = 1/(4Nm + 1)$, and from R_{ST} by substituting R_{ST} for F_{ST} in the same formula. A measure of genetic distance between the two populations was obtained using Nei's unbiased genetic distance (Nei, 1978).

Results

For *An. gambiae*, all 30 loci were amplified for most of the 16 specimens. Fourteen loci were amplified for all 16 specimens, eight loci were amplified in 15 out of 16 specimens and the remaining eight loci were amplified in at least 10 specimens. In contrast, 25 out of 30 loci were amplified for *An. arabiensis* in most specimens, whereas five loci (37, 1002, 503, 19 and 815) could not be amplified satisfactorily even after two further attempts.

Allele frequency distribution

Loci were considered polymorphic if they had more than one allele present. Conversely, a monomorphic locus was one that had only one allele. *Anopheles arabiensis* was polymorphic at 92% of the loci that were amplified and *An. gambiae* at 93.33% of all loci. Locus 9c was monomorphic in both species, whereas loci 127 and 91 were monomorphic in *An. arabiensis* and *An. gambiae*, respectively.

Allele frequencies were calculated from genotypes scored for each of the 32 mosquitoes studied. Allele frequency distribution between the two sibling species varied from locus to locus, with some loci having similar distributions and others very varied patterns (Fig. 1).

Conformance to Hardy–Weinberg equilibrium

Excluding five loci that could not be amplified and two that were monomorphic in *An. arabiensis*, 23 out of 30 loci were analysed for conformance to Hardy–Weinberg equilibrium. For *An. gambiae*, two loci were monomorphic, so 28 loci were analysed for the Hardy–Weinberg test as described above. Both the

Table 2 Genetic variability at the 25 loci studied in *Anopheles arabiensis* and *An. gambiae* (standard errors in parentheses)

	An. arabiensis	*An. gambiae*
Mean sample size	14.6 (0.3)	14.9 (0.3)
Mean no. of alleles per locus	3.9 (0.4)	3.4 (0.3)
Per cent loci polymorphic	92.0	93.3
Mean heterozygosity		
Direct count	0.382 (0.050)	0.378 (0.051)
Hardy–Weinberg expected	0.478 (0.056)	0.512 (0.045)

sequential Bonferroni test and the binomial test showed significant deficiencies in heterozygotes in both species. Such deviations indicate violation of one or more assumptions of Hardy–Weinberg.

Deficiency of heterozygotes may be caused by the presence of null alleles. Whether the effect of null alleles skewed the distribution of alleles in these two populations to any significant extent was tested by correlating the F_{IS}, which is a measure of the reduction in heterozygosity, with the number of specimens that could not be amplified. Lack of PCR product was assumed to be caused by the presence of either one null allele in the homozygous form or two different null alleles at a given locus. No significant correlation was found ($r = 0.29$; d.f. = 26; $P = 0.1361$), suggesting lack of sufficient evidence to attribute the observed heterozygote deficiencies to the presence of null alleles.

Genetic variability and differentiation

Genetic variability and differentiation measures were calculated based on the 24 loci that were amplified in both species. Locus 9C was monomorphic in both species and was therefore excluded from analyses of differentiation. Genetic variability measures for these loci in each species are presented in Table 2.

Locus-specific F_{ST}/R_{ST} values varied greatly, but there was general agreement between the two indices (Table 1). The mean F_{ST} and R_{ST} values were higher than their corresponding median values (F_{ST} mean = 0.249, median = 0.136; R_{ST} mean = 0.197, median = 0.153), indicating that, although most values were low, the high values were considerably higher, with a significant influence on the mean. To avoid this bias, median values rather than mean values were analysed. The median F_{ST} and R_{ST} values were partitioned in two ways: first by the location of loci on the three chromosomes and then by the loca-

Fig. 1 Allele frequency distribution in the *Anopheles gambiae* and *An. arabiensis* populations studied. Alleles increase in size from allele 'A' to allele 'I'. The actual size of the alleles was not determined. DNA from the specimens can be made available to interested individuals.
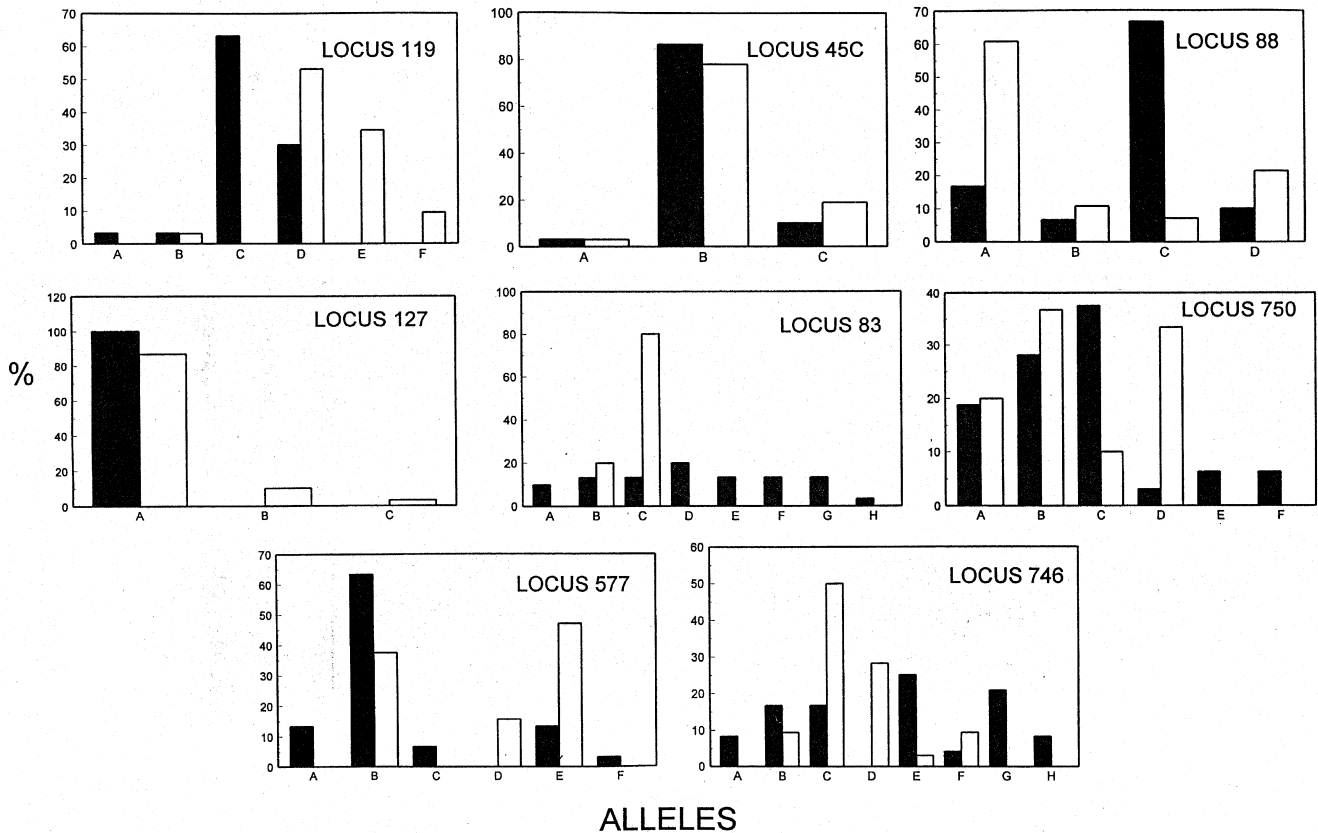
Fig. 1 *Continued*

tion inside or outside chromosomal inversions. Among loci used were those located on the X chromosome, which has different fixed inversions in the two species (Coluzzi *et al.*, 1979), and those within inversions 2La, 2Rd and 3Ra. Inversion 2La has been found to be polymorphic in *An. gambiae* and fixed in *An. arabiensis* from coastal and western Kenya and from West Africa, whereas inversion 3Ra was polymorphic in *An. arabiensis* and fixed in *An. gambiae* from western Kenya and West Africa (Coluzzi *et al.*, 1979; Mosha & Petrarca, 1983; Petrarca & Beier, 1992). Although the 2Rd inversion has not been documented in coastal or western Kenya, it has been shown to be polymorphic in both species in West Africa (Coluzzi *et al.*, 1979). Loci located within chromosomal inversions were first analysed without distinguishing between the fixed and polymorphic inversions and then after this distinction (Table 3). An inversion was considered fixed only if both species were fixed for the inversion and polymorphic if at least one of the species was polymorphic. We tested whether differences between loci located on different chromosomes were

significant by using the Wilcoxon rank sum test and the median test. Differences between chromosomes were insignificant (based on F_{ST}: Wilcoxon's test $P = 0.1497$, median test $P = 0.0784$; based on R_{ST}: Wilcoxon's test $P = 0.1405$, median test $P = 0.2343$). However, significant results were obtained for both tests when inversions were considered ($P < 0.05$).

Table 3 Median F_{ST}/R_{ST} values between the two species of *Anopheles* for the different loci

	F_{ST}	R_{ST}
Analysis by chromosome		
Chromosome X	0.34	0.33
Chromosome II	0.07	0.18
Chromosome III	0.10	0.09
Analysis by inversion		
Fixed inversions	0.34	0.33
Polymorphic inversions	0.32	0.19
All inversions	0.33	0.20
Not in inversions	0.08	0.05

Significant differences were observed when loci were compared in three groups — those within fixed inversions, those within polymorphic inversions and those outside inversions — and also when loci were compared in two groups, i.e. those within inversions vs. those outside inversions. Results remained significant irrespective of whether R_{ST} or F_{ST} values were used. Whether loci on the X chromosome (which contains the fixed inversions) were significantly different from those on chromosomes II and III (which contain some polymorphic inversions) was then examined. Differences were significant with the median test ($P = 0.0279$; Wilcoxon's test $P = 0.0527$). These results suggest that, although the chromosomal location of loci in itself does not affect the level of differentiation measured between these two species, chromosomal inversions are important.

When median Nm values were compared, only the Nm value for loci outside the inversion region was greater than 1 (median Nm for loci outside inversion regions based on $F_{ST} = 1.5$ and based on $R_{ST} = 2.3$). This suggests a significant level of exchange of genes, because genetic differentiation results from genetic drift only if $Nm < 1$ but not if $Nm > 1$ (Slatkin, 1987). Genetic drift would be expected to have a stronger effect on loci located on the X chromosome because of the smaller N_e of these loci (3/4) associated with a single copy of this chromosome in males. We tested whether the differences in levels of differentiation are attributable to inversions and not to the smaller N_e of X-linked loci by adjusting Nm values for loci on the X chromosome upwards by a factor of 1/4. When loci within inversions were compared with those outside inversions after this adjustment, differences were significant for R_{ST} with both the Wilcoxon and median tests and for F_{ST} with the Wilcoxon test. It is therefore concluded that differences in gene flow are not attributable to differences in N_e.

Discussion

Anopheles gambiae and *An. arabiensis* are two of the morphologically similar six members of the *An. gambiae* complex. The use of microsatellite loci across species or sibling species depends on the conservation of priming sites within flanking sequences to enable amplification and the maintenance of repeated arrays long enough to promote polymorphism (Weber *et al.*, 1990). STRs have previously been shown to be conserved in closely related mammalian species. Bowcock *et al.* (1994), for example, found that many human microsatellites are also present in the great apes. Even in more divergent families, such as primates and rodents, persistence of STR arrays has been reported (Stallings *et al.*, 1991). But the proportion of loci developed for a certain species that can amplify in another decreases rapidly with increasing evolutionary divergence (Irwin *et al.*, 1991). In the present study, 83.8% of the loci that were amplified in *An. gambiae* were also amplified in *An. arabiensis*, providing a strong indication that the two species are closely related. The very similar percentage of polymorphic loci (93.3% in *An. gambiae* and 92% in *An. arabiensis*) further supports this notion.

Allele frequencies at several loci did not conform to Hardy–Weinberg expectations. The number of loci that did not conform was greater than that expected by chance alone, indicating violation of some Hardy–Weinberg equilibrium assumptions. Lack of conformance was associated with heterozygote deficiencies, indicating excess homozygosity. Such excess homozygosity could be caused by the Wahlund effect, i.e. sampling individuals from different demes. As mosquitoes were collected from the same village during the same time period, this seems unlikely. Lehmann *et al.* (1996) have suggested that the minimum distance associated with a deme is more than 50 km in diameter. Another possible cause of excess homozygosity is the presence of null alleles (Chakraborty *et al.*, 1992). Although F_{IS} values were not significantly correlated with the number of specimens that did not amplify, we believe that null alleles may have contributed to heterozygote deficiency, and our test did not detect this because of low statistical power associated with small sample sizes.

Genetic distance between any two species or sibling species is a measure of the genetic differences that have accumulated in the two groups since their divergence. Nei's unbiased genetic distance between *An. gambiae* and *An. arabiensis* in this study was 0.202. This value is similar to that published for the two sibling species using allozymes (Nei's $D = 0.15$; Cianchi *et al.*, 1983) and is consistent with published values for sibling species in other insect groups, which often have Nei genetic distances between 0.142 and 0.437 (Brussard *et al.*, 1985).

The low estimate of Nm between the two species based on all loci studied signifies extremely low or no interchange of genes. Indeed, in wild sympatric populations of *An. gambiae* complex mosquitoes, interbreeding rarely occurs, with between 0.1% and 0.2% hybrids being found in nature (White, 1971). This is thought to be caused by the existence of some form of physiological or behavioural barrier preventing cross-mating. Laboratory crossings

between the two species produce sterile hybrid males (Davidson *et al.*, 1967), suggesting partial reproductive isolation.

In enzyme electrophoresis studies, Gillespie (1991) found that certain groups of loci are highly polymorphic, whereas others are almost always monomorphic. It therefore follows that using loci from one group may reflect the genome poorly and may give a less accurate level of differentiation. Our data concur with this notion because F_{ST}/R_{ST} values obtained in the present study span a very wide range (F_{ST} 0.00–0.87; R_{ST} 0.00–0.73), indicating that some microsatellite loci are better at differentiating between these two species than others. Our results suggest that loci within chromosomal inversions have greater power to distinguish between these species than loci outside the inversion regions. Chromosomal polymorphism caused by paracentric inversions is common in *An. gambiae* and *An. arabiensis* (Coluzzi *et al.*, 1979). Because chromosomal inversions consist of gene sequences protected from recombination, they may preserve certain gene associations. It is thought that such genes may be subject to different evolutionary processes, leading to the development of isolation between populations carrying different arrangements. The present study supports this line of thought, because higher levels of differentiation were observed for loci within inversions compared with loci outside the inversions. We found no evidence of gene flow between the two species based on loci within inversions, suggesting that the two species may not be very closely related. Such an inference was made by Coluzzi *et al.* (1979) based on chromosomal inversion studies. However, recent results based on DNA sequence analysis indicate that these two species are much closer and suggest that gene flow does occur between them (Besansky *et al.*, 1994).

Loci outside inversion regions suggested a significant level of gene flow between the two species. The observed gene flow may be that which occurred between the two species in recent evolutionary times because of common descent, or continuing gene flow/introgression may play an important part in the evolution of these species. Such possibilities may be used to explain the similar and even identical paracentric inversions between these species (Coluzzi *et al.*, 1979).

Our data suggest that the location of loci within paracentric inversions influences the level of differentiation they can measure, at least for interspecific studies of *An. gambiae* and *An. arabiensis*. As no karyotype information on the specimens studied was available, we were unable to tie our conclusion to particular inversions. Nevertheless, these results have serious implications on the selection of loci for genetic differentiation studies of *An. gambiae* populations. Loci within inversion regions may give different, even conflicting, results from loci outside inversion regions.

Acknowledgements

References

BECKMAN, J. S. AND WEBER, J. L. 1992. Survey of human and rat microsatellites. *Genomics*, **12**, 627–631.

BESANSKY, N. J., POWELL, J. R., CACCONE, A., HAMM, D. M., SCOTT, J. A. AND COLLINS, F. H. 1994. Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between the principal malaria vectors. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 6885–6888.

BOWCOCK, A. M., ROES-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.

BRUSSARD, P. F., EHRLICH, P. R., MURPHY, D. D., WILCOX, B. A. AND WRIGHT, J. 1985. Genetic distances and the taxonomy of checkerspot butterflies (Nymphalidae: Nymphalinae). *J. Kans. Entomol. Soc.*, **58**, 403–412.

BUDOWLE, B. AND ALLEN, R. C. 1991. Discontinuous polyacrylamide gel electrophoresis of DNA fragments. In: Mathew, C. (ed.) *Methods in Molecular Biology: Protocols in Human Molecular Genetics*, Vol. 9, pp. 123–132. The Humana Press, Clifton, NJ.

CAIRNS, M. J. AND MURRAY, V. 1994. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques*, **17**, 915–919.

CHAKRABORTY, R., DE ANDRADE, M., DIAGER, S. P. AND BUDOWLE, B. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implication in forensic applications. *Ann. Hum. Genet.*, **56**, 45–57.

CIANCHI, R., URBANELLI, S., VILLANI, F., SABATINI, A. AND BULLINI, L. 1983. Electrophoretic studies in mosquitoes: recent advances. *Parasitologia*, **27**, 157–167.

COLLINS, F. H., MENDEZ, M. A., RASMUSSEN, M. O., MEHAFFEY, P. C., BESANSKY, N. J. AND FINNERTY, V. 1987. A ribosomal RNA gene probe differentiates member

species of the *Anopheles gambiae* complex. *Am. J. Trop. Med. Hyg.*, **37**, 37–41.

COLLINS, F. H., MEHAFFEY, P. C., RASMUSSEN, M. O., BRANDLING-BENNET, A. D., ODERA, J. S. AND FINNERTY, V. 1988. Comparison of DNA-probe and isozyme methods for differentiating *Anopheles gambiae* and *Anopheles arabiensis* (Diptera: Culicidae). *J. Med. Entomol.*, **25**, 116–120.

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.

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

DAVIDSON, G., PATERSON, H. E., COLUZZI, M., MASON, G. F. AND MICKS, D. W. 1967. The *Anopheles gambiae* complex. In: Wright, J. W. and Pal, R. (eds) *Genetics of Insect Vectors of Disease*, pp. 211–301. Elsevier, Amsterdam.

DEKA, R., SHIVER, M. D., YU, L. M., JIN, L., ASTON, C. E., CHAKRABORTY, R. ET AL.. 1994. Conservation of human chromosome 13 polymorphic microsatellite $(CA)_n$ repeats in chimpanzees. *Genomics*, **22**, 226–230.

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

GILLESPIE, J. H. 1991. *The Causes of Molecular Evolution*. Oxford University Press, New York.

HOLMS, S. 1979. A simple sequential rejection multiple test procedure. *Scand. J. Stat.*, **6**, 65–70.

IRWIN, D. M., KOCHER, T. D. AND WILSON, A. C. 1991. Evolution of the cytochrome b gene of mammals. *J. Mol. Evol.*, **32**, 128–144.

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

LEHMANN, T., HAWLEY, W. A., KAMAU, L., FONTENILLE, D., SIMARD, F. AND COLLINS, F. H. 1996. Genetic differentiation of *Anopheles gambiae* populations from east and west Africa: comparison of microsatellite and allozyme loci. *Heredity*, **77**, 192–200.

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

MOSHA, F. W. AND PETRARCA, V. 1983. Ecological studies on *Anopheles gambiae* complex sibling species on the Kenya coast. *Trans. R. Soc. Trop. Med. Hyg.*, **77**, 344–345.

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

PETRARCA, V. AND BEIER, J. C. 1992. Intraspecific polymorphism in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. *Am. J. Trop. Med. Hyg.*, **46**, 229–237.

RAYMOND, M. AND ROUSSET, F. 1995. An exact test for population differentiation. *Evolution*, **49**, 1280–1283.

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

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

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

STALLINGS, R. L., FORD, A. F., NELSON, D., TORNEY, C. D., HILDEBRAND, C. E. AND MOYZIS, R. K. 1991. Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics*, **10**, 807–815.

WEBER, J. L., KWITEK, A. E., MAY, P. E. AND KILLARY, A. M. 1990. Dinucleotide repeat polymorphism at the, D4s, 174-locus. *Nucleic Acids Res.*, **18**, 4636–4638.

WHITE, G. B. 1971. Chromosomal evidence for natural interspecific hybridization by mosquitoes of the *Anopheles gambiae* complex. *Nature*, **231**, 184–185.

WEIR, B. S. AND COCKERHAM, C. C. 1984. Estimating *F*-statistics for analysis of population structure. *Evolution*, **38**, 1358–1370.

WRIGHT, S. 1969. *Evolution and the Genetics of Populations:* Vol. 2, *The Theory of Gene Frequencies*. University of Chicago Press, Chicago.

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.