

**Ecological zones rather than molecular forms predict genetic differentiation
in the malaria vector *Anopheles gambiae* s.s. in Ghana.**

Alexander E. Yawson *†‡ David Weetman* Michael D. Wilson† and Martin J. Donnelly*

*Vector Group

Liverpool School of Tropical Medicine

Liverpool, UK

†Noguchi Memorial Institute for Medical Research

University of Ghana

Legon, Accra, Ghana

‡Biotechnology and Nuclear Agricultural Research Institute,

Ghana Atomic Energy Commission

Kwabanya, Accra, Ghana.

Running head: Ecological barriers to gene flow in *Anopheles gambiae*

Keywords: Tropical Medicine, Diptera, Introgression

Corresponding author

Martin J. Donnelly, Vector Group, Liverpool School of Tropical Medicine,

Pembroke Place, Liverpool, L3 5QA, UK

Tel: +44 (0) 151 705 3296

Fax: +44 (0) 151 705 3369

Email: m.j.donnelly@liv.ac.uk

ABSTRACT

The malaria mosquito *Anopheles gambiae* s.s. is rapidly becoming a model for studies on the evolution of reproductive isolation. Debate has centered on the taxonomic status of two forms (denoted M and S) within the nominal taxon identified by point mutations in the X-linked rDNA region. Evidence is accumulating that there are significant barriers to gene flow between these forms, but that the barriers are not complete throughout the entire range of their distribution. We sampled populations from across Ghana and southern Burkina Faso, West Africa, from areas where the molecular forms occurred in both sympatry and allopatry. Neither Bayesian clustering methods nor F_{ST} -based analysis of microsatellite data found differentiation between the M and S molecular forms, but revealed strong differentiation among different ecological zones, irrespective of M/S status and with no detectable effect of geographical distance. Although no M/S hybrids were found in the samples, admixture analysis detected evidence of contemporary interform gene flow, arguably most pronounced in Southern Ghana where forms occur sympatrically. Thus, in the sampled area of West Africa lack of differentiation between M and S forms likely reflects substantial introgression, and ecological barriers appear to be of greater importance in restricting gene flow.

INTRODUCTION

When reproductive isolation between species is incomplete (i.e., when F1 hybrids are not completely sterile), genes may pass between species (MACHADO *et al.* 2002). Whether such interspecific hybridisation is a significant route for the transfer of genetic adaptation remains a major question in evolutionary biology (BARTON 2001). The African malaria vector *Anopheles gambiae* is a unique model for studies of the evolution of reproductive isolation and the importance of hybridisation for the transfer of adaptively advantageous genes in incipient species. Within *A. gambiae*, sequence variation in X-linked rDNA of populations from West Africa led to the description of two molecular forms (termed M and S), which differ in both the transcribed and the non-transcribed spacers in the rDNA repeat unit (DELLA TORRE *et al.* 2001; GENTILE *et al.* 2002). Reproductive isolation between the M and S forms has been supported by a lack of hybrids detected by PCR of the rDNA intergenic, and also partially by studies of the distribution of the *kdr* gene, a mutation involved in resistance to pyrethroid insecticides (CHANDRE *et al.* 1999). In general, in West African countries apart from Benin, the *kdr* resistance locus is at or near fixation in the S form populations but occurs in the M form at very low levels, even in sympatry (DELLA TORRE *et al.* 2001; YAWSON *et al.* 2004). Importantly, it has been proposed that the *kdr* mutation may have introgressed from the S form into the M form (WEILL *et al.* 2000) raising the possibility that there could be numerous adaptively and epidemiologically important genetic exchanges between forms.

Evidence from *Drosophila* has indicated that isolation does not occur simultaneously for the whole genome (MACHADO *et al.* 2002), and that the least likely parts to exchange are associated with hybrid sterility or mate choice (NOOR *et al.* 2001). This mosaic genome structure was first postulated for *Anopheles gambiae* by DELLATORRE *et al.* (1997) and more

recent evidence was provided by TURNER *et al.* (2005) who used a microarray approach to identified regions of the genome which they term 'islands of speciation.' One of the regions TURNER *et al.* identified does indeed appear to be intimately involved with the reproductive isolation of these forms (STUMP *et al.* 2005).

Outside the few areas of marked differentiation only slight differences have been observed in allele or haplotype arrays between the M and S forms using a range of genetic markers (DONNELLY *et al.* 2004; GENTILE *et al.* 2002; LEHMANN *et al.* 2003), although microsatellites from regions of the second chromosome associated with putatively selectively advantageous chromosomal inversions have been shown to differ between forms (LANZARO *et al.* 1998). The lack of marked differentiation throughout most of the genome has been attributed to the recent isolation of the forms whereby insufficient time has elapsed for even non-coding DNA to accrue differences (GENTILE *et al.* 2001). Premating isolation mechanisms apparently result in positive assortative mating when the forms are in sympatry, with 98.8% within form mating (TRIPET *et al.* 2001). Indeed, evidence from a study in Cameroon suggested that reinforcement between forms might occur in sympatry, with F_{ST} estimates of 0.035-0.052 between allopatric populations compared to 0.060 between sympatric populations (WONDI *et al.* 2002).

Given that there is strong positive assortative mating and molecular genetic suggestions of reinforcement in sympatry the question arises where is interform hybridisation occurring? Rates of hybridisation are unlikely to be equal across the sympatric distributions of the forms and therefore identifying the geographic regions where genetic exchange does occur is central to an understanding of the evolution of these incipient species. Previous evidence has suggested that the area between Cote d'Ivoire and Benin may be an area of transition between

populations that show strong positive assortative mating and populations that are more permissive of interform gene flow (BLACK IV and LANZARO 2001; DELLA TORRE *et al.* 2001). We used microsatellite loci to determine the intra and interform variation in M and S forms in Ghana and southern Burkina Faso. Our data suggest that interform hybridization occurs at significant levels, and that ecological barriers represent far more important barriers to gene flow in this area of West Africa.

MATERIALS AND METHODS

Sample sites and screening: sample sites are shown in Fig. 1, with a detailed description in YAWSON *et al.* (2004). In brief, M and S forms were sympatric in the mangrove swamps along the coast; in northern Ghana and Burkina Faso samples were overwhelmingly M form (80-86 % in northern Ghana and 80% in Burkina Faso) and the sample from the middle rainforest belt were all S forms (Fig.1). Specimens were identified to form using a combination of the approaches of SCOTT *et al.* (1993) and FANELLO *et al.* (2002). Seven microsatellite loci, two on chromosome X (AgXH7, AgXH99); one on chromosome two (Ag2H197); and four on chromosome three (Ag3H119, Ag3H812, Ag3H577, 33CI) were chosen based on physical and linkage maps of *A. gambiae* (ZHENG *et al.* 1996) and with the exception of Ag2H197 (Inversion 2Ra) were not contained within polymorphic inversions. Previous studies had been these markers to be highly variable and to amplify reliably (DONNELLY *et al.* 2001; PINTO *et al.* 2002; WONDJI *et al.* 2002). None of these markers are within the ‘genomic islands of speciation’ identified by other authors (TURNER *et al.* 2005) but some of the markers show marked differences in allele frequencies between molecular forms (PINTO *et al.* 2002; WONDJI *et al.* 2002). Thirty-one individuals were identified per site unless the forms occurred in sympatry, when 31 individuals of each form were identified.

Microsatellite genotyping methods followed (DONNELLY *et al.* 1999). In total genotypes were determined for 462 *A. gambiae* specimens from 11 sites in Ghana and one in Burkina Faso.

Statistical analysis: Tests of Hardy-Weinberg Equilibrium and linkage disequilibrium between all pairs of loci within populations were done in GENEPOP 3.4 (Raymond and Rousset 1995). Diversity per locus and per population was assessed using allelic richness (R_S), observed (H_O) and expected (H_E) heterozygosities, inbreeding coefficients (F_{IS}) and their significance were calculated using FSTAT 2.9.3.2 (GOUDET 2001). To explore the impact of putative null alleles on our analyses, we used the program MICRO-CHECKER (VAN OOSTERHOUT *et al.* 2004) to estimate null allele frequencies (using the 'Brookfield 2' estimator (BROOKFIELD 1996) for potentially affected loci that showed a consistent significant excess of homozygotes across sample sites. Using these estimated frequencies, an adjusted dataset, with nulls recoded as separate alleles was produced for comparison with the original dataset.

Genetic relationships within and between forms were assessed using three complementary approaches, two Bayesian clustering methods and an F_{ST} -based summary statistic method. A Bayesian clustering algorithm implemented in the program STRUCTURE (PRITCHARD *et al.* 2000) was applied to identify subgroups that have distinctive allele frequencies, without using prior knowledge of sampling sites. We used the admixture model with correlated allele frequencies between populations with a burn-in length of 100,000, and 100,000 Markov Chain Monte Carlo replications for each setting of K from 2 to 6 (20 replicate runs of each). To determine the most appropriate number of clusters we used the approach of (EVANNO *et al.* 2005) which is based upon an *ad hoc* quantity (ΔK) that evaluates the second order rate of change of the likelihood function with respect to K . For the STRUCTURE analysis we made

three predictions i) if there was extensive genetic exchange between the forms or if they had only recently become separated then we would expect M and S form samples to form a single cluster. ii) if sufficient time had elapsed for the forms to diverge then we would observe clusters corresponding with each of the forms. iii) if there were zones of hybridisation between the two forms then we would observe two clusters – one corresponding with each of the forms - and one or more additional clusters containing those samples where introgression was occurring.

A second Bayesian clustering analysis, implemented in the software BAPS 4.13 was also applied (CORANDER *et al.* 2003). By contrast to the individual-based algorithm applied in STRUCTURE, we used the group-level option in BAPS such that clusters are formed by merging whole samples. Since group-level analysis does not rely on the integrity of individual multilocus genotypes, we were able to use the method to compare solutions from the original and null allele-corrected datasets. The resulting clusters (from the original dataset) were saved and admixture proportions, estimated for each individual, again using BAPS 4.13 (CORANDER *et al.* 2004), with the posterior probability for the null hypothesis of pure ancestry computed by permutation. We used clusters, rather than individual sample sites as the units for the analysis because admixture results can be unreliable when the power to discriminate populations is low (J. Corander, personal communication). Thus whilst total contemporary admixture of an individual sample site could be determined, the source of the non-resident admixture proportion(s) could only be assigned to a cluster. Since our aim was to study patterns of admixture, rather than to identify specific potentially-migrant individuals, critical α for admixture was set at a posterior probability of 0.05, although we also determined individuals that showed admixture probabilities that remained significant following Bonferroni correction.

F_{ST} estimates and pairwise (permutation-based) tests of genic differentiation, which assumes unlinked markers but not Hardy-Weinberg equilibrium, were calculated using Genepop 3.4 (RAYMOND and ROUSSET 1995). Pairwise F_{ST} estimates were linearised following (SLATKIN 1995) and the matrix entered into the MEGA 3.1 program for neighbor-joining tree construction (KUMAR *et al.* 2004). We hypothesize that if introgression is ongoing in sympatry, F_{ST} values between sympatric populations will be lower than between interform allopatric comparisons, whereas the converse is expected if reinforcement occurs in sympatry. As an *ad hoc* test of these hypotheses, groups were resampled using 1000 bootstrap replicates to estimate 95% confidence intervals around the group mean F_{ST} using SAS (SAS-INSTITUTE-INC 1990). Isolation by distance was examined by testing the association of linearized F_{ST} with ln-transformed geographical distance using a Mantel test with 5000 permutations implemented by the POPTOOLS add-in for Microsoft[®] EXCEL[®] (written by Greg Hood, CSIRO; available from <http://www.cse.csiro.au/poptools>), Unless stated otherwise, where multiple tests were used we adjusted the significance level using a sequential Bonferroni procedure (HOLM 1979).

RESULTS

Hardy-Weinberg equilibrium and genotypic disequilibrium: An excess of homozygotes was very common in the dataset (Appendix 1), with 48 out of a total of 105 tests significant at $P < 0.05$, and 26 remaining so following sequential Bonferroni correction. Hardy-Weinberg deviations were not restricted to particular loci nor geographic samples (Appendix 1). Sixteen out of 315 tests for linkage disequilibrium were significant at $P < 0.05$, which is exactly the number expected by chance as Type I errors at this critical α level, and following sequential

Bonferroni correction, no tests remained significant in any sample or at any locus. Since within-population structure would be expected to cause both excess homozygosity and linkage disequilibrium, the widespread prevalence of the former but absence of the latter in our data suggested the presence of null alleles. Estimates of null allele frequencies per population and per locus were generated, assuming random mating, using MICRO-CHECKER (VAN OOSTERHOUT *et al.* 2004) wherever a significant excess of homozygotes was detected. The mean estimated null allele frequency of 8.7% was similar to the mean frequency of null alleles (5-8%) estimated for *A. gambiae* based on X-linked loci in males (BARNES *et al.* 2005; STUMP *et al.* 2005).

Intra- and interform genetic differentiation: Individual-based Bayesian clustering using STRUCTURE determined the most appropriate number of clusters as 2, and successive increases in K did not split the two major groupings into additional clusters (Supplementary materials). The ΔK -value provided support for the division into two groups with a marked peak at $K=2$ and a rapid decline at higher values. The clusters did not partition the molecular forms, but rather, grouped individuals by ecological zone, though not geographic proximity (see Fig. 1). One cluster comprised M and S form individuals from the coastal mangrove strand zone, with the other cluster containing all individuals from the other sites, sampled from Northern Sahel Savanna, Coastal Savanna and Deciduous Forest (Fig. 2a).

Population differentiation was widespread with 101 out of 105 pairwise tests significant following sequential Bonferroni correction (Appendix 2). The major division in an F_{ST} -based neighbor-joining tree was in accord with neither geographic proximity of the sample locations nor the molecular status of the samples, but represented a primary division between the populations in the mangrove strand zone of southern Ghana (in which M and S are sympatric)

and all remaining M and S populations from the north and south (Fig. 2b). Results from group-level Bayesian clustering analysis performed by BAPS also revealed a major split between the mangrove strand zone and other populations, but in addition partitioned the Kumasi (deciduous forest zone) separately (Fig. 2b). This optimal clustering solution was well supported since the next most likely solution - of two clusters, in which Kumasi grouped with all other non-mangrove zone samples - could be confidently rejected ($P < 0.0001$). Nevertheless, the Kumasi site was clearly far more distinct from the mangrove zone samples (mean pairwise $F_{ST} = 0.10$), than from the Northern and coastal Savanna sites ($F_{ST} = 0.04$).

There was limited evidence to suggest a contribution of molecular form to differentiation. For comparisons made either within or between ecological zones, interform mean F_{ST} values were only marginally higher than intraform values and none of these differences were significant (Fig. 3). However, since the difference between inter- and intraform comparisons was, if anything, greater in the non-mangrove cluster of allopatric M and S populations (categories B and E in Fig. 3) than in the mangrove cluster of sympatric M and S occurrence (categories C and F in Fig. 3) our *a priori* hypothesis of reinforcement in sympatry (see Materials and Methods) seems unlikely. Evidence for interform hybridisation in sympatry was more equivocal since both inter- and intraform differentiation within the mangrove strand cluster were lower than elsewhere. Such generally greater genetic homogeneity in the mangrove zone might be attributable to lower average pairwise separation (mangrove mean=61km; non-mangrove mean=336 km). Yet isolation by distance (analyzed via linearized F_{ST} vs. In geographical distance) was not significant, either over all samples (Mantel's test $P=0.14$) or within samples from outside of the mangrove zone ($P=0.092$); too few sites were available to permit separate analysis within the mangrove zone. Moreover, one of the sampling sites from the mangrove strand zone was actually closer geographically to those within the coastal

Savanna than to the other two sites within its own ecological zone (see Fig. 1), further suggesting that distance is not an important determinant of differentiation.

The concordant sample groupings observed in both the STRUCTURE analysis (Fig. 2a) and the F_{ST} neighbor-joining tree (Fig. 2b) were not unduly influenced by a single locus. The number of loci with F_{ST} values significantly different from zero was markedly different in comparisons from within the major clusters (mean=3.1 loci) compared to comparisons between clusters (mean=6.2 loci), (Appendix 2). Similarly null alleles exerted little influence on our results. A dataset in which null alleles were recoded according to their estimated frequencies showed a lower global F_{ST} (0.065, compared to 0.076 for the original dataset), but the neighbor-joining tree and group-level clustering analysis results were almost indistinguishable between datasets (Supplementary materials).

Genetic diversity and admixture: The seven microsatellites analysed were highly polymorphic in all populations (Appendix 3). Variation in genetic diversity among populations was quite moderate, whether measured as allelic richness (R_S , range 7.69-10.01) or expected heterozygosity (H_E , range 0.71-0.82). Both R_S and H_E differed significantly between the two major clusters identified above, with each higher in the mangrove strand zone than the cluster of other populations (FSTAT 2.9.3.2, 1000 permutations, both $P < 0.05$).

Since we have no evidence that the populations in this study are anywhere close to migration-drift equilibrium we applied a non-equilibrium admixture analysis. Admixture analysis allows us to determine whether individuals have recent ancestors originating from more than one sample population. We used this analysis to study patterns of very recent migration, using the three clusters identified by the BAPS group-level method (namely-Kumasi, Mangrove

strand zone and Savanna clusters) as potential source populations (see Materials and Methods). Of a total of 411 individuals, 33 were identified as having multilocus genotypes exhibiting significant evidence of population mixing (Appendix 4), representing a total multilocus admixture level of 4.3%. Individual admixture proportions rarely approached 1.0, that is no individual could be unequivocally assigned to a population other than the one in which it was sampled (Appendix 4), suggesting that our samples contained few first generation (i.e. F_0) immigrants. Four of the eight populations comprising the Savanna cluster contained no individuals showing significant evidence of mixed cluster ancestry, whereas all from the Mangrove cluster showed some degree of admixture (Table 2). Indeed, mean admixture proportions were significantly higher in the Mangrove than Savanna cluster (Mann-Whitney U -test, $U_{6,8}=3.5$, $P<0.01$). Interestingly, Kumasi, a single-population cluster from deciduous forest, contained no admixed individuals but was identified as the source for more admixture than either of the other clusters (Table 2). Moreover, since the Kumasi sample was composed entirely of the S molecular form, apparent introgression into M-form populations could be identified. Admixture proportions from Kumasi (S) did not differ between M or S recipient populations ($U_{6,8}=30.0$, $P=0.49$).

DISCUSSION

In this study we used microsatellites to screen *Anopheles gambiae* from Ghana and southern Burkina Faso, with samples from two Savanna zones and a deciduous forest zone, in which the M and S molecular forms are largely allopatric, and a coastal mangrove strand zone, in which M and S are sympatric. Significant genetic differentiation between sample sites was near ubiquitous, but the major genetic division, supported by both individual-level Bayesian cluster analysis and F_{ST} -based analysis was found between samples from the mangrove strand

zone vs. the other ecological zones. A group-level Bayesian analysis supported an additional partition between the single sample from the deciduous forest zone and the Savanna zones, though this was far less distinct from the Savanna cluster than the mangrove zone. Differentiation attributable to molecular form was limited and not significant. These patterns were found consistently across loci and were not the result of the major effect of a single locus, nor could they be attributed to null alleles, although these appeared to be common in the dataset.

Determinants of genetic structure. Such high differentiation ($F_{ST} > 0.1$ between the mangrove zone and other areas) is unusual in *Anopheles gambiae*. To date estimates of genetic differentiation have been low even between populations separated by many thousands of kilometres (DONNELLY *et al.* 2004; LEHMANN *et al.* 2003), unless there are major hydrographic or geographic barriers to migration (e.g. (LEHMANN *et al.* 2003; PINTO *et al.* 2002). Yet in the south of Ghana (all samples from south of Kumasi) there are no obvious extant or former geographic barriers to gene flow, with all samples from this region taken from a continuous coastal plain that does not rise more than 100m above sea level. Moreover, the effect of geographical distance on the patterns of differentiation in our study appears to be negligible.

High levels of differentiation are sometimes detected among the chromosomal forms of *A. gambiae*. For example using AFLP markers, (SLOTMAN *ET AL.* 2006) obtained estimates of Φ_{ST} (an F_{ST} analogue) between approximately 0.05 and 0.15 among BAMAKO, SAVANNA and MOPTI forms. Although estimates of differentiation from microsatellites and AFLPs may not be directly comparable because of the downward bias to the former caused by size homoplasy, SLOTMAN *et al.*s estimates seem compatible with those found among ecological zones in our study. Moreover, different chromosomal forms are often associated with

different habitats and are thought to play an important role in environmental adaptation (DELLA TORRE *et al.* 2001). However, studies of the presence and distribution of chromosomal forms in the area we sampled suggest that it is unlikely that high divergence results from the presence of different chromosomal forms in different ecological zones. APPAWU *et al.* 1994) reported that though the FOREST chromosomal form is dominant in the deciduous forest zone, the SAVANNA and MOPTI chromosomal forms were sympatric in both of the Savanna zones. Furthermore, MBOLE *et al.* (2004) found all three of these chromosomal forms within the mangrove strand zone. Since we do not have karyotypic information for the samples screened in the present study, and the relationship between molecular and chromosomal forms is often ambiguous (DELLA TORRE *et al.* 2001), we cannot rule out the possibility that the different habitat types are dominated by different and genetically divergent, chromosomal forms. Nevertheless, unless there are major chance, spatial, or temporal discontinuities between the chromosomal-forms represented in our samples and those obtained by APPAWU *et al.* (1994) and MBOLE *et al.* (2004), atypical chromosomal forms within a cluster should have been readily detectable via our clustering/ assignment analyses and would resemble first generation immigrants. By contrast, very few individuals in sample sites within each of the two clusters (STRUCTURE) or three clusters (BAPS) identified could be interpreted as immigrants from a different cluster (see Fig 2a).

Therefore, the genetic structure we have identified appears to be unrelated to physical barriers to gene flow, geographic distance, molecular form, and is unlikely to be attributable to habitat-specificity of chromosomal forms. A further possibility is that the very high differentiation is temporally unstable and results from a relatively recent colonisation (or re-colonisation) of, and spread within, the mangrove strand zone from a highly divergent source. Whilst considerable restriction of gene flow from the coastal Savanna zone would presumably be required for this scenario to persist for more than a few generations, the significantly lower

inter-population differentiation in the mangrove zone is consistent with this colonisation-expansion hypothesis. Yet, the significantly higher genetic diversity and admixture within the mangrove zone are the opposite to what would be expected under a colonisation-expansion scenario unless multiple differentiated source populations colonised the area, and this is at odds with the relatively low genetic structure therein. Consequently we think that recent colonisation can also be ruled out. Rather, the most likely explanation would seem that habitat type-specific selection against interzone immigrants represents a barrier to gene flow.

Gene flow between the molecular forms. Studies of gene flow between the molecular forms of *A. gambiae* have found differentiation between the two forms at four genomic regions (but see WONDJI *et al.* 2002; TURNER *et al.* 2005). One of these is on chromosome 2L, associated with insecticide resistance (GENTILE *et al.* 2004); and references therein). The other loci (including the rDNA) are at the proximal (centromeric) end of the X chromosome (LEHMANN *et al.* 2003; WANG *et al.* 2001). However, little differentiation has been found at loci elsewhere in the genome (GENTILE *et al.* 2001; MUKABAYIRE *et al.* 2001; TURNER *et al.* 2005; WANG *et al.* 2001). Debate is ongoing as to whether such generally limited differentiation primarily reflects insufficient time for differentiation to have occurred or ongoing introgression among forms. Neither the F_{ST} -based analysis nor the Bayesian clustering methods provided conclusive evidence to support or reject our main hypothesis, that interform hybridisation/ introgression would be more common when forms are sympatric than (primarily) allopatric. The alternative hypothesis, that reinforcement might occur between the forms in sympatry seems very unlikely given that the difference between the mean level of intra- and interform form comparisons was lower for sympatric than allopatric populations.

Nevertheless results from our admixture analysis suggest that genetic exchange between forms might be more frequent than previous molecular genetic and assortative mating studies

would imply (TRIPET *et al.* 2001; WONDJI *et al.* 2002). The Kumasi population, which is entirely composed of the S form (YAWSON *et al.* 2004) was sufficiently differentiated by the group-level cluster analysis to be identified as a source of admixture. Recipient M and S populations received low but very similar proportionate admixture from Kumasi, suggesting that intra- and interform gene flow is low but equally common from Kumasi. In high differentiation systems pairwise admixture proportions in the BAPS analysis that we applied appear to be insensitive to unsampled ‘ghost’ populations (see HANFLING and WEETMAN 2006). By contrast, since the M and S forms are virtually indistinguishable it is possible that admixture proportions allocated to M recipient populations from our Kumasi S sample might have originated from an unsampled ghost M form population closely resembling Kumasi. Whilst it is important to recognise this kind of limitation when applying any admixture or assignment analysis, there is no evidence for any M forms occurring within the deciduous forest zone.

A further line of evidence that higher interform exchange occurs in sympatry comes from typing of the pyrethroid insecticide knockdown resistance (*kdr*) genotype. This segregated with form: very low frequencies in the M form and near fixation in the S form; but the highest frequencies of the *kdr* genotype were all found in M forms from the sympatric M/S sample sites (YAWSON *et al.* 2004). The absence of any M/S hybrids in the 935 specimens from which the data presented here are drawn (YAWSON *et al.* 2004) suggest that at least during our collection period direct introgression between M and S forms is rare (TAYLOR *et al.* 2001; TOURE *et al.* 1998; TRIPET *et al.* 2003; TRIPET *et al.* 2001). Of course our data represent a short sampling period and there could be a seasonal component to introgression that we have not detected. Collections were made in the rainy season when mosquito numbers were high and it is possible that patterns of hybridisation differ during the dry season when mosquito numbers are low and either one of the two forms may predominate.

Conclusion. Although there is evidence to suggest that the M and S forms of *A. gambiae* may be separate entities that have been evolving under divergent selection pressures (STUMP *et al.* 2005), data from our study area suggest that interform gene flow is occurring at appreciable levels, perhaps most frequently in the southern mangrove strand zone of M/S sympatry. Our most striking and unexpected finding was of extremely strong genetic structure that correlated with ecological zones, with little, if any, contribution of distance, molecular form, and probably also chromosomal form. Results such as these suggest that climatic modeling and GIS based approaches will be fruitful in the prediction of gene flow in *A. gambiae*, the understanding of which is critical for malaria control in Africa.

ACKNOWLEDGEMENTS

We thank Bill Black, Harold Townson, Joao Pinto, Charles Wondji and Kwang Shik Choi for helpful discussions. This study was funded by grants from the Ford Foundation International Fellowships Program and the World Health Organization Special Programme in Research and Training.

LITERATURE CITED

- APPAWU, M. A., A. BAFFOE-WILMOT, E. A. AFARI, F. K. NKUMAH and V. PETRARCA, 1994 Species composition and inversion polymorphism of the *Anopheles gambiae* complex in some sites of Ghana, West-Africa. *Acta Tropica* **56**: 15-23.
- BARNES, M. J., N. F. LOBO, M. B. COULIBALY, N. SAGNON, C. COSTANTINI *et al.*, 2005 SINE insertion polymorphism on the X chromosome differentiates *Anopheles gambiae* molecular forms. *Insect Molecular Biology* **14**: 353-363.
- BARTON, N. H., 2001 The role of hybridization in evolution. *Molecular Ecology* **10**: 551-568.
- BLACK IV, W. C., and G. C. LANZARO, 2001 Distribution of genetic variation among chromosomal forms of *Anopheles gambiae* s.s.: introgressive hybridization, adaptive inversions, or recent reproductive isolation? *Insect Molecular Biology* **10**: 3-7.
- BROOKFIELD, J. F. Y., 1996 A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5**: 453-455.
- CHANDRE, F., F. DARRIET, L. MANGA, M. AKOGBETO, O. FAYE *et al.*, 1999 Status of pyrethroid resistance in *Anopheles gambiae sensu lato*. *Bulletin of the World Health Organization* **77**: 230-234.
- CORANDER, J., P. WALDMANN, P. MARTTINEN and M. J. SILLANPÄÄ, 2004 BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* **20**: 2363-2369.
- CORANDER, J., P. WALDMANN and M. J. SILLANPÄÄ, 2003 Bayesian analysis of genetic differentiation between populations. *Genetics* **163**: 367-374.
- DELLA TORRE, A., C. FANELLO, M. AKOGBETO, J. DOSSOU-YOVO, G. FAVIA *et al.*, 2001 Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Mol Biol* **10**: 9-18.
- 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.
- DONNELLY, M. J., N. CUAMBA, J. D. CHARLWOOD, F. H. COLLINS and H. TOWNSON, 1999 Population structure in the malaria vector, *Anopheles arabiensis* Patton, in East Africa. *Heredity* **83**: 408-417.
- DONNELLY, M. J., M. C. LICHT and T. LEHMANN, 2001 Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Molecular Biology and Evolution* **18**: 1353-1364.
- DONNELLY, M. J., J. PINTO, R. GIROD, N. J. BESANSKY and T. LEHMANN, 2004 Revisiting the role of introgression vs shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the *Anopheles gambiae* complex. *Heredity* **92**: 61-68.
- EVANNO, G., S. REGNAUT and J. GOUDET, 2005 Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611-2620.
- FANELLO, C., F. SANTOLAMAZZA and A. DELLA TORRE, 2002 Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Medical And Veterinary Entomology* **16**: 461-464.
- GENTILE, G., A. DELLA TORRE, B. MAEGGA, J. R. POWELL and A. CACCONE, 2002 Genetic differentiation in the African malaria vector, *Anopheles gambiae* s.s., and the problem of taxonomic status. *Genetics* **161**: 1561-1578.
- GENTILE, G., F. SANTOLAMAZZA, C. FANELLO, V. PETRARCA, A. CACCONE *et al.*, 2004 Variation in an intron sequence of the voltage-gated sodium channel gene correlates

- with genetic differentiation between *Anopheles gambiae s.s.* molecular forms. *Insect Molecular Biology* **13**: 371-377.
- GENTILE, G., M. SLOTMAN, V. KETMAIER, J. R. POWELL and A. CACCONE, 2001 Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae s.s.* *Insect Molecular Biology* **10**: 25-32.
- GOUDET, J., 2001 FSTAT, a program to estimate and test gene diversities and fixation indices (Version 2.9.3), pp.
- HANFLING, B., and D. WEETMAN, 2006 Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the River Sculpin, *Cottus gobio*. *Genetics* **173**: 1487-1501.
- HOLM, S., 1979 A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**: 65-70.
- KUMAR, S., K. TAMURA and M. NEI, 2004 MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Briefings in Bioinformatics*: 150-163.
- LANZARO, G. C., Y. T. TOURE, J. CARNAHAN, L. ZHENG, G. DOLO *et al.*, 1998 Complexities in the genetic structure of *Anopheles gambiae* populations in West Africa as revealed by microsatellite DNA analysis. *Proceedings of the National Academy of Science* **95**: 14260-14265.
- LEHMANN, T., M. LICHT, N. ELISSA, B. T. A. MAEGA, J. M. CHIMUMBWA *et al.*, 2003 Population structure of *Anopheles gambiae* in Africa. *Journal of Heredity* **94**: 133-147.
- MACHADO, C. A., R. M. KLIMAN, J. A. MARKERT and J. HEY, 2002 Inferring the history of speciation from multilocus DNA sequence data: The case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution* **19**: 472-488.
- MBOLE, N. M., D. A. BOAKYE, M. A. APPAWU and M. D. WILSON, 2004 The distribution of chromosomal forms of *Anopheles gambiae s.s.* (Diptera:Culicidae) in urban, periurban and rural locations in greater Accra. *American Journal of Tropical Medicine and Hygiene* **71**: 242.
- MUKABAYIRE, O., J. CARIDI, X. WANG, Y. T. TOURE, M. COLUZZI *et al.*, 2001 Patterns of DNA sequence variation in chromosomally recognized taxa of *Anopheles gambiae*: evidence from rDNA and single-copy loci. *Insect Molecular Biology* **10**: 33-46.
- NOOR, M. A. F., K. L. GRAMS, L. A. BERTUCCI and J. REILAND, 2001 Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 12084-12088.
- PINTO, J., M. J. DONNELLY, C. A. SOUSA, V. GIL, C. FERREIRA *et al.*, 2002 Genetic structure of *Anopheles gambiae* (Diptera: Culicidae) in Sao Tome and Principe (West Africa): implications for malaria control. *Molecular Ecology* **11**: 2183-2187.
- PRITCHARD, J. K., M. STEPHENS and P. DONNELLY, 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.
- RAYMOND, M., and F. ROUSSET, 1995 Genepop (Version-1.2) - Population-Genetics Software For Exact Tests And Ecumenicism. *Journal Of Heredity* **86**: 248-249.
- SAS-INSTITUTE-INC, 1990 *SAS language: references, version 6*. SAS Institute., Cary, NC.
- SCOTT, J. A., W. G. BROGDON and F. H. COLLINS, 1993 Identification of single specimens of the *Anopheles gambiae* complex by the Polymerase Chain Reaction. *American Journal of Tropical Medicine and Hygiene* **49**: 520-529.
- SLATKIN, M., 1995 A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457-462.
- SLOTMAN, M. A., M. M. MENDEZ, A. DELLA TORRE, G. DOLO, Y. T. TOURE *et al.*, 2006 Genetic differentiation between the Bamako and Savanna chromosomal forms of

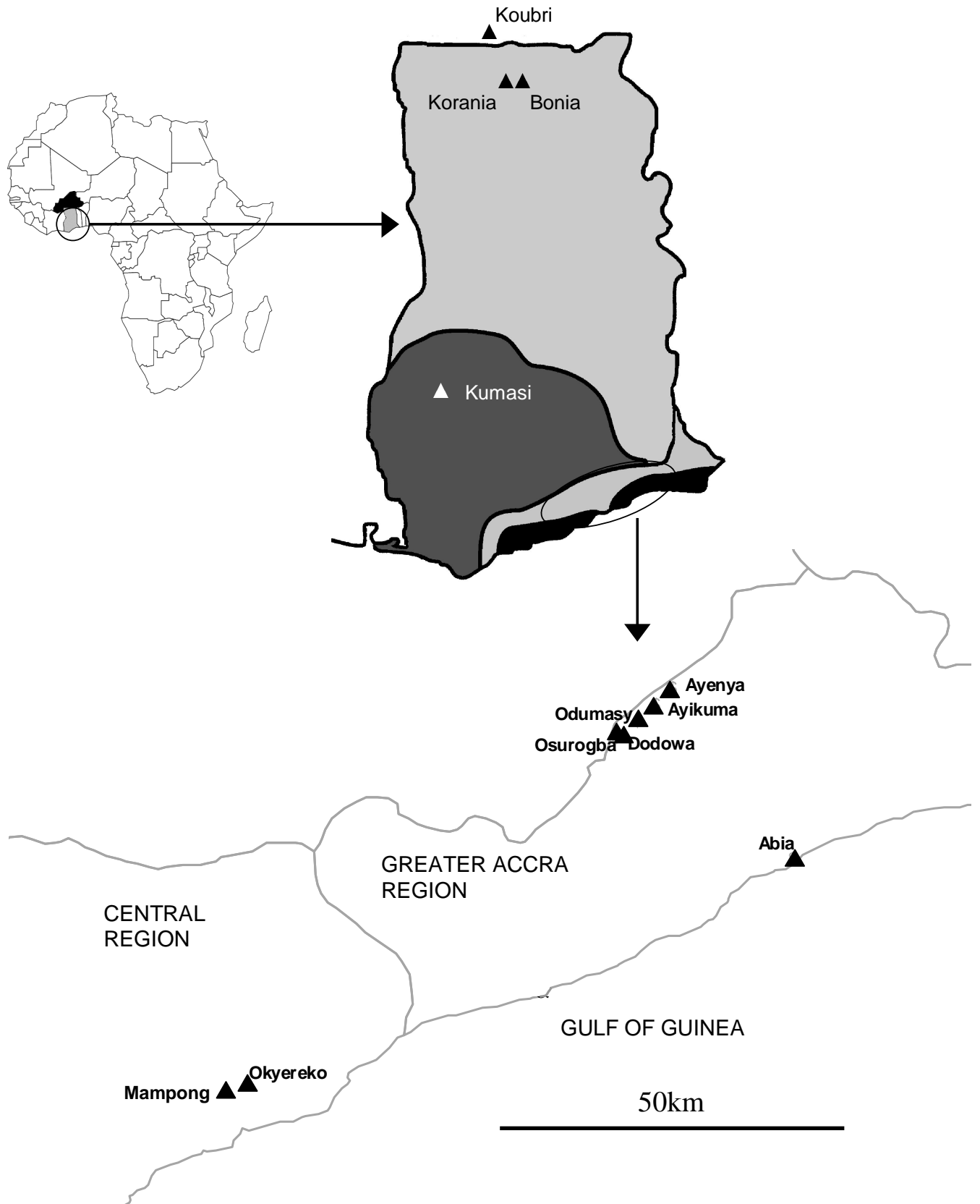
- Anopheles gambiae* as indicated by amplified fragment length polymorphism analysis. American Journal of Tropical Medicine and Hygiene **74**: 641-648.
- STUMP, A. D., J. A. SHOENER, C. COSTANTINI, N. F. SAGNON and N. J. BESANSKY, 2005 Sex-linked differentiation between incipient species of *Anopheles gambiae*. Genetics **169**: 1509-1519.
- TAYLOR, C. E., TOURE, Y.T., CARNAHAN, J., NORRIS, D.E., DOLO, G., TRAORE, S.F., EDILLO, F.E., LANZARO, G.C., 2001 Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, west Africa. Genetics **157**: 743-750.
- TOURÉ YT, P. V., TRAORÉ SF, COULIBALY A, MAIGA HM, SANKARÉ O, SOW M, DI DECO MA, COLUZZI M, 1998 Distribution of inversion polymorphism of chromosomally recognised taxa of the *Anopheles gambiae* complex in Mali, West Africa. Parasitologia **40**: 477-511.
- TRIPET, F., Y. T. TOURE, G. DOLO and G. C. LANZARO, 2003 Frequency of multiple inseminations in field-collected *Anopheles gambiae* females revealed by DNA analysis of transferred sperm. American Journal Of Tropical Medicine And Hygiene **68**: 1-5.
- TRIPET, F., Y. T. TOURE, C. E. TAYLOR, D. E. NORRIS, G. DOLO *et al.*, 2001 DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. Molecular Ecology **10**: 1725-1732.
- TURNER, T. L., M. W. HAHN and S. V. NUZHIDIN, 2005 Genomic islands of speciation in *Anopheles gambiae*. Plos Biology **3**: 1572-1578.
- VAN OOSTERHOUT, C., W. F. HUTCHINSON, D. P. M. WILLS and P. SHIPLEY, 2004 MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**: 535-538.
- WANG, R., L. ZHENG, Y. T. TOURE, T. DANDEKAR and F. C. KAFATOS, 2001 When genetic distance matters: Measuring genetic differentiation at microsatellite loci in whole-genome scans of recent and incipient mosquito species. Proceedings of the National Academy of Science **98**: 10769-10774.
- WEILL, M., F. CHANDRE, C. BRENGUES, S. MANGUIN, M. AKOGBETO *et al.*, 2000 The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae* s.s. through introgression. Insect Mol Biol **9**: 451-455.
- WONDJI, C., F. SIMARD and D. FONTENILLE, 2002 Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. Insect Molecular Biology **11**: 11-19.
- YAWSON, A. E., P. J. MCCALL, M. D. WILSON and M. J. DONNELLY, 2004 Species abundance and insecticide resistance of *Anopheles gambiae* in selected areas of Ghana and Burkina Faso. Medical And Veterinary Entomology **18**: 372-377.
- 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.

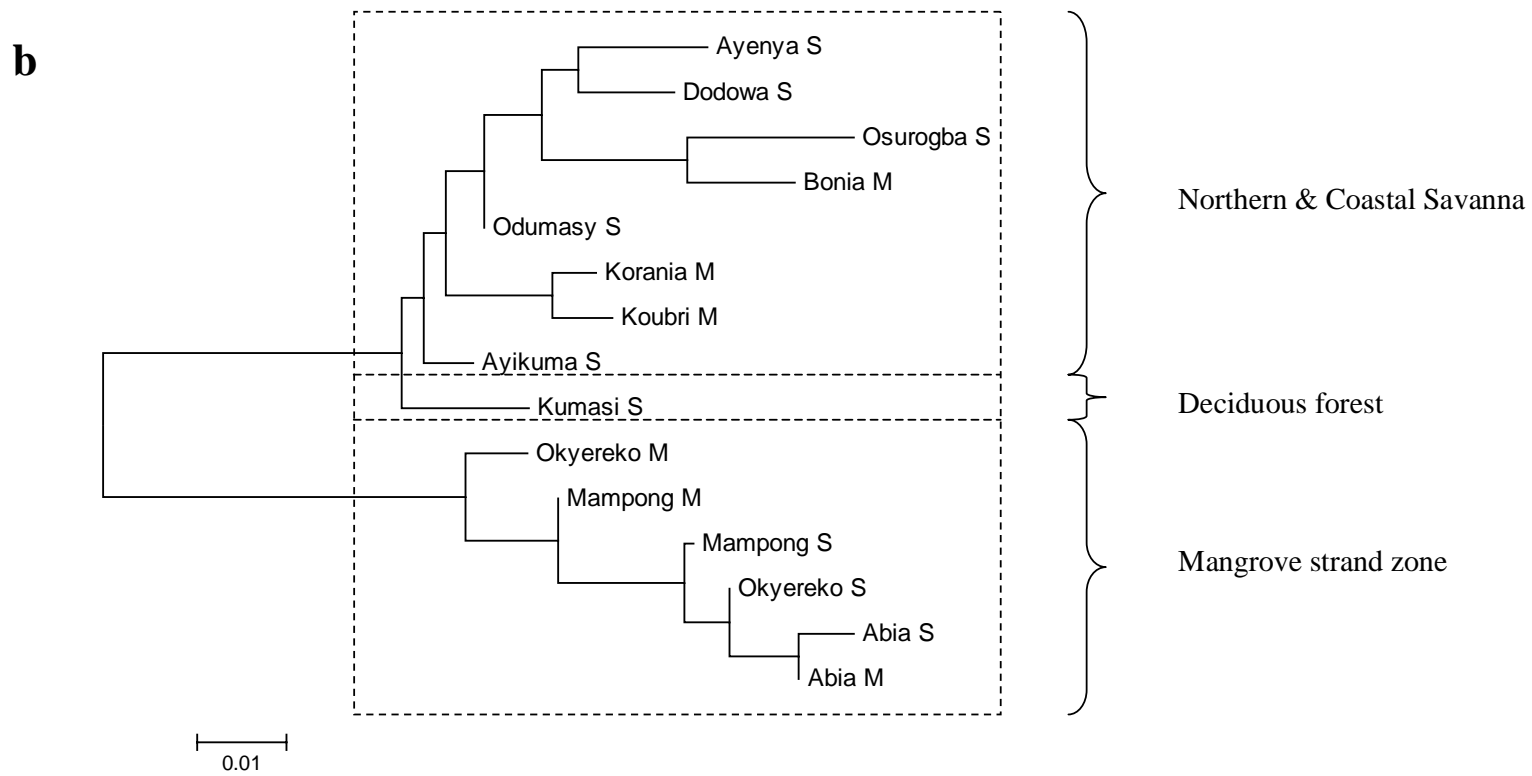
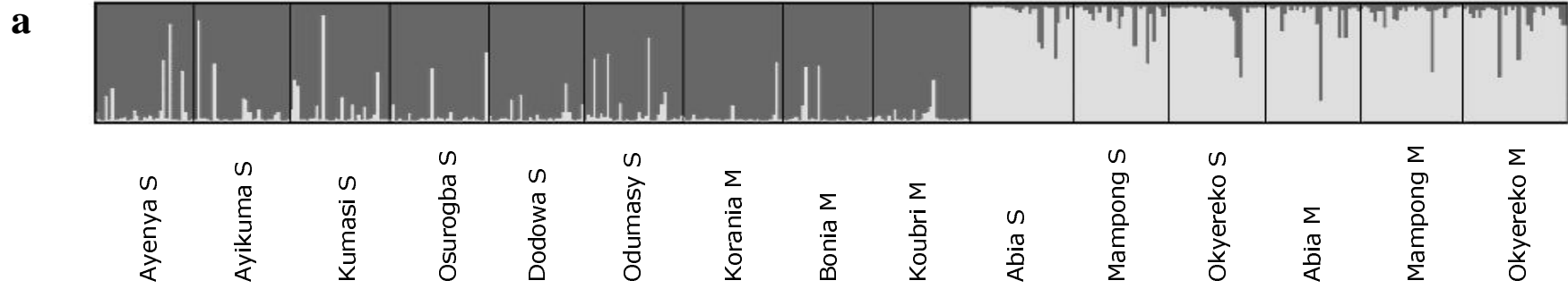
Figure legends

Figure 1. Sampling locations. Ghana and Burkina Faso are marked on the outline map of Africa in Grey and Black respectively. The more detailed map of Ghana shows the limits of the ecological zones within Ghana. Black=mangrove strand zone; Pale grey=coastal Savanna; Dark grey=deciduous forest, Mid grey (largest zone)=Sahel (Northern) Savanna. The sample from Koubri (Burkina Faso) is also within the Sahel Savanna. Sampling locations are marked by triangles. A large scale map of sampling locations within 50km of the coast is given.

Figure 2. a. Results from an individual-level Bayesian assignment/ cluster analysis in which the optimal number of clusters was 2. Vertical bars indicate the proportionate assignment of an individual to each cluster. b. Neighbor-joining tree based on linearized F_{ST} . Partitions among clusters of sample sites detected by a group-level Bayesian analysis (BAPS 4.13) are delineated by dashed boxes, with ecological zones from which samples originated indicated by parentheses. In both a & b the molecular form of the samples (M or S) is given after the location name.

Figure 3. Bootstrapped mean F_{ST} values 95% confidence intervals for comparisons made within and among molecular forms (M and S) and ecological zones. Comparison categories are: A=mangrove zone vs non-mangrove zones between form comparisons; B=non-mangrove zones all between form comparisons; C=mangrove zone all between form comparisons; D=mangrove zone vs non-mangrove zones within form comparisons; E=non-mangrove zones all within form comparisons; F=mangrove zone all within form comparisons.





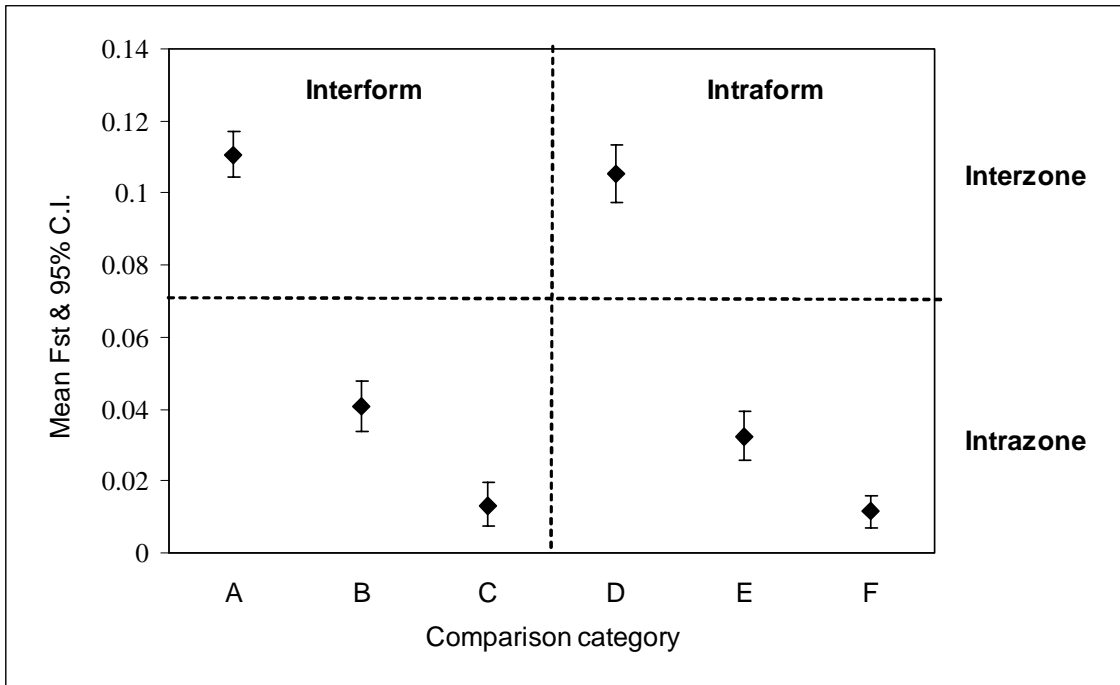


Table 1. Sampling details. Locations, habitat type and the percentage of each molecular form (M and S) are shown.

Locality	Ecological Zone	Latitude	Longitude	% S	% M
Dodowa	Coastal Savanna	5° 52.67' N	0° 06.36' W	94	6
Osurogba	Coastal Savanna	5° 52.75' N	0° 06.58' W	99	1
Odumasy	Coastal Savanna	5° 53.86' N	0° 04.72' W	81	19
Ayikuma	Coastal Savanna	5° 55.23' N	0° 03.20' W	97	3
Ayanya	Coastal Savanna	5° 56.59' N	0° 01.89' W	97	3
Okyereko	Mangrove	5° 24.87' N	0° 36.25' W	35	65
Mampong	Mangrove	5° 24.74' N	0° 36.95' W	69	31
Abia	Mangrove	5° 42.93' N	0° 07.69' W	61	39
Kumasi	Central forest	6° 41.00' N	01° 37.00' W	100	0
Korania	Northern Savanna	10° 53.16' N	01° 05.40' W	0	100
Bonia	Northern Savanna	10° 52.05' N	01° 07.25' W	0	100
Koubri	Northern Savanna	12° 11.56' N	01° 23.46' W	3	97

(Burkina Faso)

Table 2. Results from Bayesian admixture analysis assessing very recent migration among population clusters based on individuals assessed to show significant evidence of admixture.

Average proportions of resident (i.e. not admixed) multilocus DNA are shown in bold.

Underlined values show admixture proportions between molecular forms.

INTO (Site)	Form	FROM (Cluster)		
		Savanna M and S	Kumasi S	Mangrove M and S
<i>Savanna cluster</i>				
Ayanya S	S	0.97	0.051	0.045
Ayikuma S	S	0.94	0.057	0.018
Osuregba S	S	1.00	0	0
Dodowa S	S	1.00	0	0
Odumasy S	S	0.97	0.045	0.024
Korania M	M	0.98	<u>0.042</u>	0.001
Bonia M	M	1.00	0	0
Koubri M	M	1.00	0	0
<i>Kumasi (deciduous forest) cluster</i>				
Kumasi S	S	0	1.00	0
<i>Mangrove cluster</i>				
Abia S	S	0.034	0.018	0.93
Mampong S	S	0.029	0.014	0.94
Okyereko S	S	0.025	0.029	0.95
Abia M	M	0.021	<u>0.039</u>	0.93
Mampong M	M	0.017	<u>0.012</u>	0.97
Okyereko M	M	0.039	<u>0.019</u>	0.93

Appendix 1. Tests for deviation from Hardy-Weinberg equilibrium (HWE). F_{IS} values in bold are significantly different from zero ($P < 0.05$), and where underlined, remain significant following sequential Bonferroni correction.

Location	Molecular form	Abbreviation	Locus								Mean
			Ag3H119	Ag2H197	33CI	Ag3H812	Ag3H577	AgXH99	AgXH7		
Ayenia	S	Ae	<u>0.313</u>	0.053	0.289	0.167	<u>0.405</u>	0.144	0.193	<u>0.224</u>	
Ayikuma	S	Ai	0.096	0.069	<u>0.563</u>	0.348	0.312	<u>0.528</u>	0.271	<u>0.297</u>	
Kumasi	S	Km	0.104	0.126	<u>0.335</u>	0.235	<u>0.528</u>	<u>0.736</u>	<u>0.46</u>	<u>0.349</u>	
Osuogba	S	Os	0.035	0.112	<u>0.321</u>	0.095	0.116	-0.101	0.07	<u>0.116</u>	
Dodowa	S	Do	0.129	0.086	0.244	<u>0.538</u>	0.055	-0.067	<u>0.513</u>	<u>0.202</u>	
Odumasy	S	Od	<u>0.429</u>	0.178	0.212	0.189	<u>0.407</u>	0.236	0.176	<u>0.264</u>	
Korania	M	Kr	0.155	0.247	<u>0.339</u>	0.393	0.195	0.338	0.224	<u>0.267</u>	
Bonia	M	Bo	0.188	0.303	0.2	0.211	0.185	0.032	0.047	<u>0.182</u>	
Koubri	M	Ku	0.12	0.1	<u>0.355</u>	<u>0.344</u>	<u>0.328</u>	0.125	0.189	<u>0.218</u>	
Abia	S	Abx	0.065	0.195	0.1	0.146	<u>0.329</u>	0.109	0.335	<u>0.177</u>	
Mampong	S	Mnx	0.131	<u>0.412</u>	<u>0.403</u>	0.27	<u>0.223</u>	-0.012	0.259	<u>0.24</u>	
Okyereko	S	Okx	0.146	<u>0.152</u>	<u>0.224</u>	-0.078	0.302	0.306	0.256	<u>0.185</u>	
Abia	M	Abz	<u>0.414</u>	0.252	0.01	0.24	0.269	0.07	<u>0.435</u>	<u>0.244</u>	
Mampong	M	Mnz	0.247	<u>0.56</u>	0.182	0.24	0.101	0.166	<u>0.566</u>	<u>0.298</u>	
Okyereko	M	Okz	<u>0.277</u>	<u>0.295</u>	0.182	0.243	0.015	0.152	<u>0.683</u>	<u>0.264</u>	
N locations: $P < 0.05$			8	9	13	13	11	9	6		

Appendix 2. Population differentiation. Pairwise F_{ST} (below diagonal) and the number of loci $P < 0.05$ in genic tests of population differentiation.

F_{ST} values in italics correspond to genic tests combined across loci that are NOT significant following sequential Bonferroni correction.

	Ae	Ai	Km	Os	Do	Od	Kr	Bo	Ku	Abx	Mnx	Okx	Abz	Mnz	Okz
Ae		2	3	4	4	2	4	3	3	7	6	6	7	7	6
Ai	0.031		2	4	3	2	3	3	4	7	7	6	7	7	6
Km	0.047	0.023		3	3	2	3	4	6	7	5	6	7	7	6
Os	0.047	0.048	0.065		3	2	2	3	6	7	5	6	5	6	6
Do	0.025	0.041	0.041	0.042		2	3	4	6	7	6	7	7	7	7
Od	0.030	0.011	0.013	0.034	0.011		1	2	3	7	6	6	6	5	6
Kr	0.055	0.032	0.031	0.052	0.046	0.013		3	4	7	5	5	6	6	6
Bo	0.052	0.048	0.060	0.030	0.043	0.025	0.028		4	6	5	5	6	6	5
Ku	0.060	0.014	0.039	0.055	0.059	0.022	0.011	0.036		7	5	7	7	7	6
Abx	0.122	0.115	0.124	0.150	0.118	0.112	0.134	0.158	0.121		3	2	1	5	4
Mnx	0.106	0.097	0.102	0.138	0.106	0.095	0.110	0.135	0.100	0.007		3	1	4	3
Okx	0.105	0.095	0.103	0.149	0.109	0.094	0.114	0.144	0.103	0.008	0.008		2	4	4
Abz	0.101	0.107	0.108	0.146	0.103	0.097	0.123	0.148	0.119	<i>-0.001</i>	<i>0.012</i>	<i>0.003</i>		2	2
Mnz	0.084	0.082	0.090	0.118	0.088	0.076	0.101	0.125	0.092	0.021	0.016	0.012	<i>0.013</i>		3
Okz	0.085	0.080	0.085	0.123	0.091	0.079	0.103	0.124	0.091	0.030	0.034	0.019	0.013	0.021	

Appendix 3. Genetic diversity. Gene diversity (H_E) and Allelic richness (R_S) are shown for each population and locus combination.

	Ag2H11		Ag2H19		33C1		Ag3H81		Ag3H57		AgXH99		AgXH7		Mean	
	H_E	R_S	H_E	R_S	H_E	R_S	H_E	R_S	H_E	R_S	H_E	R_S	H_E	R_S	H_E	R_S
Ae	0.85	9.95	0.85	11.19	0.82	9.60	0.62	8.45	0.76	5.00	0.72	4.87	0.87	11.94	0.78	8.71
Ai	0.83	7.94	0.89	11.50	0.69	9.41	0.72	6.61	0.73	7.29	0.71	5.90	0.87	9.69	0.77	8.33
Km	0.85	10.16	0.84	9.74	0.73	7.97	0.72	5.72	0.82	6.93	0.63	4.53	0.84	9.41	0.77	7.78
Os	0.86	10.51	0.86	11.32	0.83	10.18	0.68	8.84	0.73	5.68	0.23	2.94	0.75	6.65	0.71	8.01
Do	0.81	9.51	0.91	13.94	0.79	9.41	0.51	7.38	0.78	5.77	0.66	4.72	0.82	7.89	0.75	8.37
Od	0.85	9.73	0.90	13.76	0.74	11.43	0.72	8.71	0.82	7.65	0.72	8.04	0.78	9.23	0.79	9.79
Kr	0.84	9.40	0.85	8.76	0.68	11.57	0.82	7.75	0.72	4.00	0.63	3.94	0.77	8.38	0.76	7.69
Bo	0.79	9.56	0.87	8.76	0.85	14.82	0.72	8.60	0.75	4.82	0.55	5.64	0.49	5.61	0.72	8.26
Ku	0.88	11.18	0.86	8.48	0.65	9.82	0.79	10.03	0.77	7.41	0.66	4.93	0.84	10.28	0.78	8.87
Abx	0.89	11.46	0.86	9.74	0.70	7.07	0.84	8.84	0.70	9.91	0.70	4.91	0.66	4.90	0.76	8.12
Mnx	0.90	10.00	0.85	12.62	0.78	9.14	0.82	9.81	0.69	10.21	0.79	7.30	0.65	5.75	0.78	9.26
Okx	0.89	11.33	0.87	11.40	0.73	9.50	0.77	9.78	0.76	11.16	0.77	8.42	0.81	8.48	0.80	10.01
Abz	0.91	13.09	0.88	12.04	0.77	8.82	0.75	8.81	0.68	10.13	0.75	6.48	0.71	6.69	0.78	9.44
Mnz	0.84	12.23	0.88	9.42	0.76	9.92	0.70	4.99	0.83	12.21	0.79	7.60	0.79	7.36	0.80	9.11
Okz	0.86	10.78	0.89	10.47	0.80	9.77	0.78	10.88	0.79	10.53	0.81	7.15	0.79	7.36	0.82	9.56
Mean	0.86	10.46	0.87	10.87	0.75	9.89	0.73	8.35	0.75	7.91	0.67	5.82	0.76	7.97		

Appendix 4. Bayesian admixture analysis (using BAPS 4.13) assessing very recent migration among population clusters. Each line shows an individual multilocus genotype that is significantly ($P < 0.05$) admixed, with the proportion assigned to each cluster: resident (i.e. not admixed) proportions are in bold. Underlined values show admixture proportions between molecular forms. To aid interpretation very low proportions (< 0.1) are shown in faint type.

INTO (Site)	Form	FROM (Cluster)		
		Savanna M and S	Kumasi S	Mangrove M and S
<i>Savanna cluster</i>				
Ayenia	S	0.36	0.62	0.02
Ayenia	S	0.46	0.01	0.53
Ayenia *	S	0.02	0.44	0.54
Ayenia *	S	0.18	0.51	0.31
Ayikuma	S	0.46	0.02	0.52
Ayikuma	S	0.47	0.52	0.01
Ayikuma	S	0.45	0.55	0
Ayikuma	S	0.39	0.61	0
Odumasy	S	0.39	0.17	0.44
Odumasy	S	0.35	0.35	0.30
Odumasy *	S	0.11	0.89	0
Korania	M	0.39	<u>0.59</u>	0.02
Korania *	M	0.30	<u>0.70</u>	0
<i>Mangrove cluster</i>				
Abia	S	0.05	0.36	0.59
Abia	S	0.43	0.01	0.56
Abia	S	0.43	0.01	0.56
Abia	S	0.18	0.18	0.64
Mampong	S	0	0.40	0.60
Mampong	S	0.35	0.03	0.62
Mampong	S	0.51	0	0.49
Okyereko	S	0.36	0.05	0.59
Okyereko *	S	0.38	0.23	0.39
Okyereko *	S	0	0.60	0.40
Abia	M	0.12	<u>0.21</u>	0.67
Abia *	M	0.28	<u>0.41</u>	0.31
Abia	M	0	<u>0.47</u>	0.53
Abia	M	0.24	<u>0.09</u>	0.67
Mampong	M	0	<u>0.34</u>	0.66
Mampong *	M	0.55	<u>0.05</u>	0.40
Okyereko *	M	0.46	<u>0.15</u>	0.39
Okyereko	M	0.42	<u>0.08</u>	0.50
Okyereko	M	0.02	<u>0.38</u>	0.60
Okyereko	M	0.34	0	0.66

*significant after Bonferroni correction