Heredity (2004) 92, 61–68

www.nature.com/hdy

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

MJ Donnelly[1,2], J Pinto[3], R Girod[4], NJ Besansky[5] and T Lehmann[2]

[1]*Vector Research Group, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK;* [2]*Centers for Disease Control and Prevention, MS F22, 4770, Buford Hwy, Chamblee, GA 30341, USA;* [3]*Centro de Malaria e outras Doencas Tropicais/Instituto de Higiene e Medicina Tropical, Rua da Junqueira 96, 1349-008 Lisbon, Portugal;* [4]*Mission de Prévention et de Lutte contre les Endémies Vectorielles, DRASS, BP 9, 97408 St Denis, La Réunion, France;* [5]*Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA*

The role of interspecific hybridisation in the evolution of pest species is poorly understood. In mosquito disease vectors this is of particular importance due to the evolution of insecticide resistance and the proposed release of transgenic strains that are refractory to the malaria parasite. In this study, we apply population genetic methods in a novel manner to determine whether mitochondrial DNA sequences have introgressed between the closely related African malaria vectors *Anopheles gambiae* and *A. arabiensis*. Our results suggest that speciation was geologically recent and ancestral haplotypes at the *ND5* locus are retained in both species. In addition, comparing haplotype frequencies in allopatric and sympatric populations, suggest locale specific unidirectional introgression of mitochondria from *A. arabiensis* into *A. gambiae*.
Heredity (2004) **92,** 61–68, advance online publication, 3 December 2003; doi:10.1038/sj.hdy.6800377

Keywords: hybridisation; speciation; disease vectors; malaria; filariasis

Introduction

A major question in evolutionary biology is whether interspecific hybridisation is a significant route for the transfer of genetic adaptations (Barton, 2001). Most studies of introgression have focused on plant systems (Martinsen *et al*, 2001). Where animal studies have been performed they have been in tractable systems such as hybrid zones (Evans *et al*, 2001), where species are morphologically distinguishable (Beaumont *et al*, 2001) or where exotic species have been introduced (Goodman *et al*, 1999). In these studies, the organisms have evolved separately for a long period and introgressed alleles can be readily identified due to mutational differences or differing allele distributions (Goodman *et al*, 1999; Roques *et al*, 2001).

The *Anopheles gambiae* complex of mosquito species forms the most important insect disease vector system. Of the seven recognised species within the complex, *A. gambiae s.s.* and *A. arabiensis* are the most abundant and widespread, occurring in sympatry over most of their distributions. In sub-Saharan Africa they are the primary malaria vectors, the main vectors of Bancroftian filariasis in rural areas and have localised importance in arbovirus transmission. In these and many other disease vector complexes, it is difficult to identify introgressed alleles, possibly because the vectors have only recently speciated and there has been insufficient time to accumulate mutations (Coluzzi, 1982; Powell *et al*, 1999; Walton *et al*, 2000). Hybrids occur naturally in the wild, in some regions at frequencies of up to 2 per 1000 (White *et al*, 1972; Tripet *et al*, 2001) but the evolutionary significance of these hybrids is unclear. It is not known if genes are actually introgressed into parent species by backcrossing of fertile F1 hybrid females (males are sterile). In laboratory experiments polymorphic chromosomal inversions can be transferred between species (della Torre *et al*, 1997). Some workers have postulated that certain chromosomal inversions, that are associated with aridity tolerance, may have introgressed from *A. arabiensis* into *A. gambiae* and have enabled this species to spread into novel habitats (Powell *et al*, 1999).

Phylogenetic approaches to resolving taxa relationships in these species have been hampered by high intraspecific and low interspecific variation (Besansky *et al*, 1994; Krzywinski *et al*, 2001). Previous studies of partial sequence of mitochondrial DNA have shown that large numbers of haplotypes are shared between *A. gambiae*, *A. arabiensis* and *A. bwambae* (Besansky *et al*, 1997; Thelwell *et al*, 2000; Donnelly *et al*, 2001), but whether this is a result of introgression or retention of ancestral polymorphisms is unresolved (Besansky *et al*, 1997; Thelwell *et al*, 2000). In this study, we examine patterns of intra and interspecific differentiation in ways that can allow us to distinguish between contemporary

Correspondence: MJ Donnelly, Vector Research Group, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK.
E-mail: mjames@liv.ac.uk
Received 21 January 2003; accepted 10 July 2003

introgression and retention of ancestral mitochondrial sequences. We propose that if introgression is ongoing, F_{ST} values between sympatric populations will be lower than between interspecific allopatric comparisons. Furthermore, this approach can reveal locale specific introgression and can determine if introgression occurs in a unidirectional or bidirectional manner.

The data in this study also serve as additional markers to compare with the large number of microsatellite-based studies that have investigated population structuring in these organisms on micro- and macrogeographic scales (Lehmann *et al*, 1997, 2003; Kamau *et al*, 1998; Lanzaro *et al*, 1998; Simard *et al*, 1999; Donnelly and Townson, 2000; Wondji *et al*, 2002).

Materials and methods

Mosquitoes were collected from 25 sites throughout sub-Saharan Africa (Figure 1 and Table 1). Most locations have been described in detail previously (Besansky *et al*, 1997; Lehmann *et al*, 1997, 1998, 2003; Donnelly and Townson, 2000; Pinto *et al*, 2002). Species identification, collection location and sample size are given in Table 1. Adult mosquitoes were obtained from houses within a village by resting and pyrethrum knockdown collections. A sample consisted of specimens collected in an area with a radius less than 1 km. Previous studies revealed no population subdivision within and among adjacent villages separated by 10–50 km and that mosquitoes within a house represent a random sample of the population (Petrarca and Beier, 1992; Besansky *et al*, 1997; Lehmann *et al*, 1997; Donnelly *et al*, 1999). To maximise the power of interspecific analyses specimens were pooled from the three regions where *A. gambiae* and *A. arabiensis* occur in sympatry, Kenya, Senegal and Malawi.

Recent work suggests that there are two forms of uncertain taxonomic status within *A. gambiae*, which may be characterised by their rDNA molecular type (Favia *et al*, 1997; Wondji *et al*, 2002). All the samples used in this study are the S rDNA type except samples from Senegal and Ghana, which are from populations that only exhibit the M form (Lehmann *et al*, 2003) and the sample from Sao Tome and Principe which is also M form (Pinto *et al*, 2003). There is no evidence for reproductively isolated forms within *A. arabiensis*. After DNA extraction (Lehmann *et al*, 1997; Donnelly *et al*, 1999) and PCR-based species identification (Scott *et al*, 1993) individuals were sequenced over a 650 bp stretch of the *ND5* region of the mitochondrial genome following the protocols of Besansky *et al* (1997). DNA sequencing was performed on ABI 377 or ABI 3100 machines (Perkin-Elmer) using fluorescent labelling technology and standard analytical protocols. Sequences were unambiguously aligned using the Clustal W option in Sequence Navigator (Version 1.0.1; ABI Systems, CA, USA).

F_{ST}-based methods

We estimated intra- and interspecific differentiation between samples using F_{ST} (Hudson *et al*, 1992). A weighted means approach was used to account for differences in sample size. Significance of F_{ST} values was estimated using a permutation program written in the SAS language with 2000 replications (SAS Institute, 1990). Pairwise estimates of F_{ST} were used as distance measures to generate a neighbour-joining (NJ) tree using Mega V2.1 (Kumar *et al*, 2001).

Phylogeography

A haplotype tree was constructed using the statistical parsimony algorithm of Templeton *et al* (1992). The TCS 1.13 program of Clement *et al* (2000) was used to estimate the haplotype tree.

Results

Complete and unambiguous sequence of a 650 bp section of the *ND5* region of the mitochondrial genome was obtained for a total of 331 specimens (Table 1). In addition to newly generated sequences of 52 *A. gambiae* and 61 *A. arabiensis* specimens, we used data from 68 *A. gambiae* and 50 *A. arabiensis* published by Besansky *et al* (1997) and 45 *A. gambiae* published by Lehmann *et al* (1997) and 55 *A. arabiensis* for which summary statistics were published previously (Donnelly *et al*, 2001). Reference sequences have been deposited in Genbank (accession numbers AY312090-AY312142). No nuclear/nonfunctional copies of the gene were present as evidenced by the very low pairwise divergence, the absence of stop codons, the predominance of synonymous substitutions and the unambiguous electropherograms.

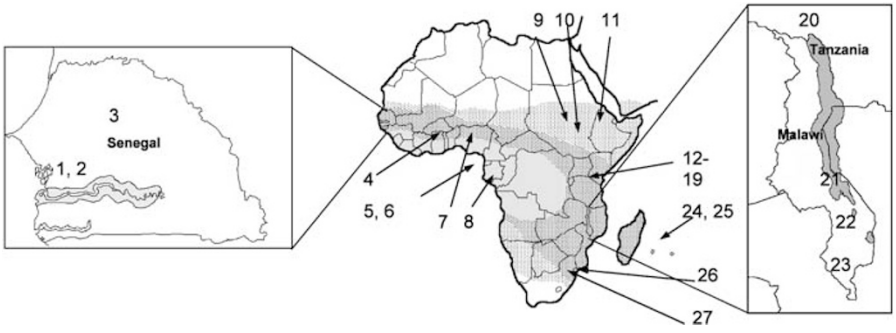

Figure 1 Location of sample sites. Species ranges are marked for *A. arabiensis* (stippled) and for *A. gambiae* (grey). More detailed information for sampling locations 12–19 can be found in Besansky *et al* (1997). The key to locations is indicated by the location superscripts in Table 1.

Table 1 Geographic distribution and details of *ND5* haplotype for each sample

Location	*Anopheles gambiae*		*Anopheles arabiensis*	
	Sample size	*Haplotypes*	*Sample size*	*Haplotypes*
Senegal				
Dielmo (M)[1]	10[a]	**2**(4), 3(2), **4**, 16, **35**, 39	9[a]	5, 17 18, 19, 20, 21, 22, 23, 26
Ndiop (M)[2]	9[a]	**2**, 3(2), 6, 7, 24, 25, 27, **34**	9[a]	**1**(2), 10, **11**, 12, 13, **14**, 15, **37**
Barkedji (M)[3]	9[a]+4	**2**, 3(5), 9, 28, 29, **33**(2), 38, 106	8[a]	**1**, **2**, 8, **33**(2), **34**, **35**, **36**
Ghana				
Navrongo (M)[4]	6	3(4), **33**, 52		
Sao Tome and Principe				
Praia Burras (M)[5]	2	79(2)		
Porto Allegre (M)[6]	1	3		
Nigeria				
Gwamlar (S)[7]	9	**14**, **33**(2), 59, 63, 64, 89, 91, 104	2	**1**, **2**
Gabon				
Dienga(S)[8]	2	**11**, **32**		
Sudan				
Wad Awad[9]			9	**1**, **14**, 20, **33**, **41**, 71, 85, 92, 96,
Um Rakuba[10]			13	**1**(5), 5, **11**(2), 60, 66, 70, 88, 105
Ethiopia				
Harosha[11]			28[b]	**1**(14), **11**(7), **33**(2), **41**, 61, 72, 84(2)
Kenya				
Asembo (S)[12]	45[b]	**1**(2), **2**(3), 3, **14**(2), **33**(8), **35**, **36**, **37**, 48(2), 49, 52, 56, 59(5), 62, 63, 65(2), 67, 68, 78, 80, 81, 87, 90, 94(2), 95, 102, 103,	15	**1**(5), **11**(5), **33**, **41**(3), 77,
Escarpment (S)[13, d]	6[a]	**33**, **42**, **50**, 51, 52, 53	2[a]	**1**, **41**
Ahero (S)[14]			4[a]	**1**, **11**, **42**, **43**
Kisian (S)[15]	4[a]	**2**, **11**, **32**, **54**	2[a]	**1**, 44
Wathrego (S)[16]	6[a] +11[c]	**2**, **11**, **33**(5), 48, 52, 53, 107, 108, 109, 110 111, 112, 113	5[a]	**1**(2), **11**, **43**, 45
Nyakoch (S)[17]	4[a]	**2**, **11**, 55, 56		
Muhroni (S)[18]	4[a]	**33**(2), 58, 59	4[a]	**1**, **2**, **41**, 46
Jego (S)[19]	5[a]	**32**(2), **41**(2), 57	2[a]	**1**(2)
Tanzania				
Kyela (S)[20]	9	**1**(2), **11**, **32**, **37**, **41**, **54**, 82, 83,		
Malawi				
Mkali (S)[21]	2	**32**, **37**	27[b]	**1**, **2**, **11**(2), 30, **32**(7), **43**, 44, 45(5), **54**(4), 73, 74, 99, **100**
Thyolo (S)[22]	16	**1**, **11**(2), **31**, **32**(5), **43**, 75, 83, 93, 98(2), **100**		
Seseo (S)[23]	1	76	13	30, **31**, **43**, 45(2), **50**, **54**(2), 69, 86(2), 97, 101
Reunion				
Riv. des Galets[24]			1	**11**
Riv. du Mat[25]			5	**11**(5)
Mozambique				
Maputo[26]			3	**11**(3)
South Africa				
Malahlapanga[27]			5[a]	**11**, 30(3), **32**

[a]These sequences are from Besansky *et al* (1997).
[b]Summary statistics for these sequences were previously published in Donnelly *et al* (2001).
[c]These sequences are from Lehmann *et al* (1997).
[d]Refers to whether the samples are from an M or S form population of *A. gambiae*.
Superscript numbers refer to the sampling locations detailed in Figure 1. Haplotype numbers are the same as stated in Besansky *et al* (1997) up to haplotype 59. Haplotypes marked in bold are shared between both species. Figures in brackets are frequencies for each haplotype.

Intraspecific estimates of differentiation

A. arabiensis: Pairwise estimates of differentiation as inferred from F_{ST} statistics were generally low (Figure 2a; mean $\pm$ SE $= 0.098 \pm 0.027$; Data matrix available from authors). Strong evidence for structuring was observed between the sample from the island of Reunion and continental populations. Only a single haplotype was present in the Reunion sample ($n = 6$) probably reflecting a lower effective population size (N_e) on the island. Therefore, as the lower N_e will complicate interpretation of differentiation this sample was excluded from the analysis. Remaining pairwise F_{ST} values were used to construct an NJ tree (Figure 2a). Three clusters were observed: Sudan and Senegal (cluster 1); Western Kenya and Ethiopia (cluster 2) and Malawi (cluster 3).

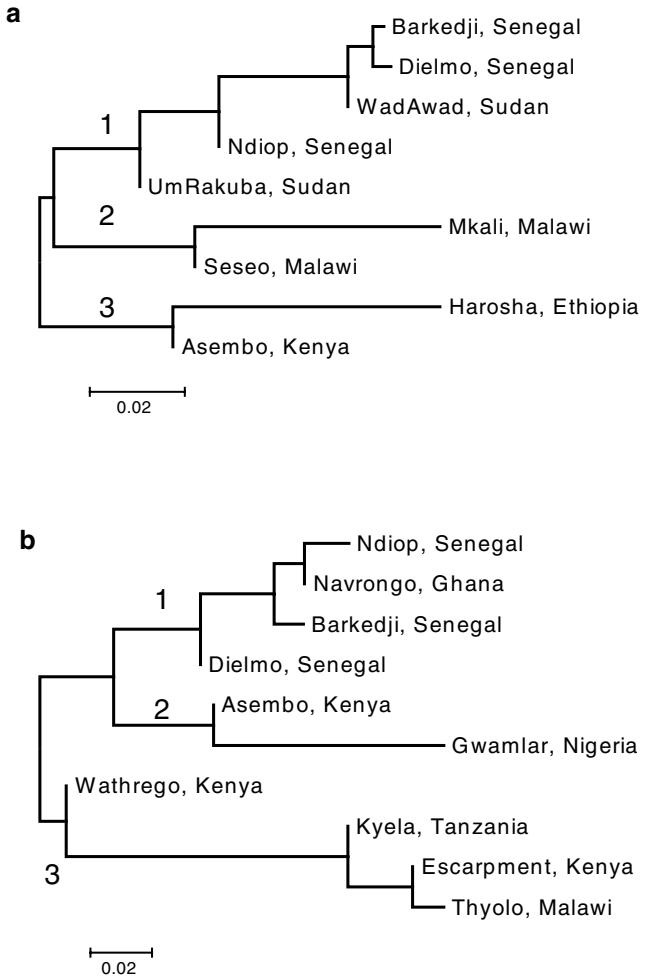

Figure 2 F_{ST} distance-based trees for *A. arabiensis* (**a**) and *A. gambiae* (**b**) recovered from partial 650 bp *ND5* mitochondrial DNA sequences. The trees were constructed using an NJ method based upon a pairwise interpopulation values of F_{ST}. Branch lengths (see scale bar) are proportional to the mean pairwise F_{ST} values. For the *A. arabiensis* data there were no significant pairwise comparisons of F_{ST} within each of the clusters ($P > 0.05$). When samples from different clusters were compared, F_{ST} values were significantly different ($P < 0.05$) for 30% (clusters 1 and 3), 40% (clusters 1 and 2) or 50% (clusters 2 and 3) of pairwise comparisons. For the *A. gambiae* data when samples from different clusters were compared, F_{ST} values were significantly different ($P < 0.05$) for 56% (clusters 1 and 3), 62% (clusters 1 and 2) or 50% (clusters 2 and 3) of pairwise comparisons.

A. gambiae: The NJ tree suggested a primary division between samples from West African/western Kenya (clusters 1 and 2) with samples from Tanzania and Malawi (cluster 3) (Figure 2b). The genetic division corresponds with the topographic division of populations by the Rift Valley complex. Ghanaian and Senegalese populations grouped closely (cluster 1), while Nigerian specimens grouped more closely with populations from Western Kenya (cluster 2). These groupings run contrary to the geographical distance. Notably, cluster 1 comprises M form specimens whereas S form specimens are found in clusters 2 and 3. There was only one significant pair wise comparison of F_{ST} within the clusters (cluster 3 Thyolo-Wathrego) but over half of the pairwise comparisons between clusters were significant (Figure 2). The Wathrego sample and the Escarpment sample are both from Western Kenya and the grouping with Tanzanian and Malawian samples from east of the Rift Valley Complex is likely to be a result of the lower sample size of these Kenyan samples (Lehmann *et al*, 2000). Pairwise values of F_{ST} were high for comparisons involving Wathrego and Escarpment but resampling tests were insignificant (data matrix available from authors).

Haplotype networks

For these data haplotypes separated by up to 11 mutational steps have a probability of $\geqslant 0.95$ of being connected in a parsimonious manner. The statistical parsimony algorithm produced networks for both species within the 11-step limit of parsimony but there were a number of ambiguities within the tree, apparently due to homoplasy within the mitochondrial sequences. These ambiguities were resolved following the suggestions of Templeton *et al* (1992) and Crandall and Templeton (1993) (Figures 3 and 4). In general this resulted in the haplotype being placed externally and linked to one of the high-frequency internal alleles. The high levels of homoplasy observed in our data are common in *Anopheles* (Walton *et al*, 2000). At present the evidence suggests that the homoplasy reflects mutational hot spots rather than recombination. There was no evidence for heteroplasmy in our electropherograms and of the 72 polymorphic sites within the data set there were 10 sites with three variants and one site with four variants suggesting large variation in mutation rates across the locus.

Introgression and ancestral retention

There were no fixed nucleotide differences between the species and 17 out of 113 haplotypes were shared between species (Table 1). Mean F_{ST} ($\pm$95% CI) for sympatric ($n = 3$) and allopatric ($n = 6$) comparisons were 0.070 ($\pm$0.104) and 0.121 ($\pm$0.068), respectively (Table 2) ($P > 0.05$). While estimates of F_{ST} involving *A. gambiae* from Malawi were nonsignificant ($P > 0.05$) the two allopatric comparisons involving *A. arabiensis* from the same sample locations were highly significant ($P < 0.001$) (Table 2).

Of the shared haplotypes only two (Nos. 2 and 33 Table 1) were common to *A. arabiensis* and *A. gambiae* M and S forms, and were internal to both species haplotype trees (Figures 3 and 4). The remaining shared haplotypes were found predominantly in *A. arabiensis* and S form

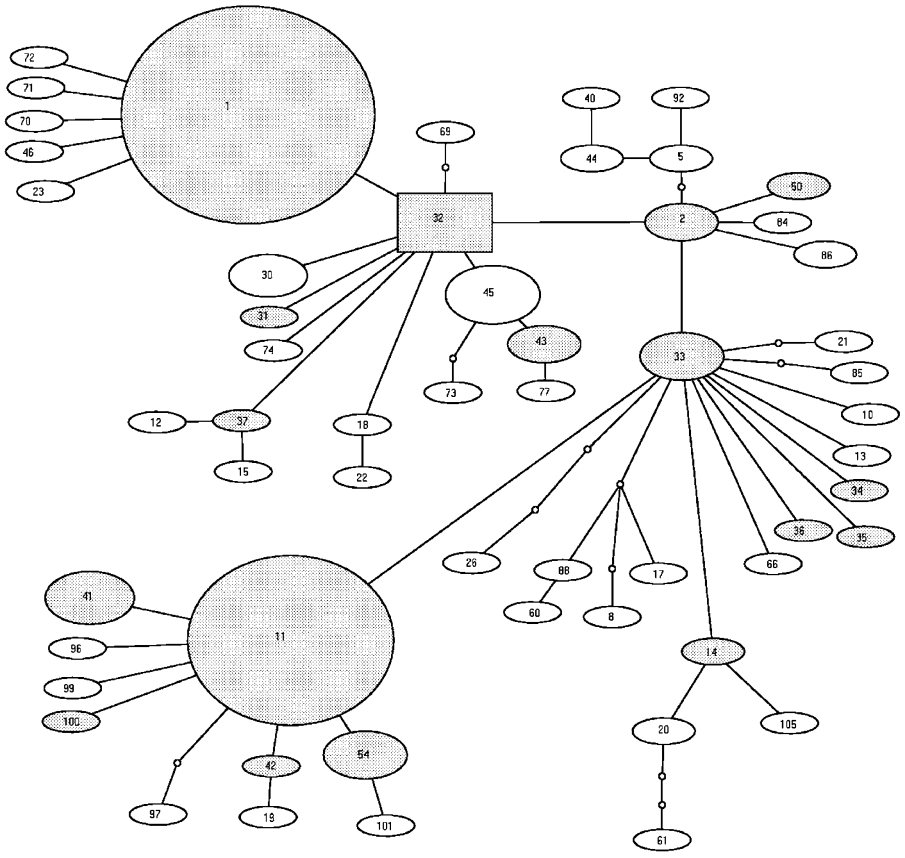

Figure 3 Statistical parsimony network for *A. arabiensis* data set. Haplotype node area is proportional to the number of specimens contained. Links between nodes are all single mutational steps regardless of length. Shaded nodes are haplotypes that are found in both species.

comparisons ($n = 11$) rather than between *A. arabiensis* and M form comparisons ($n = 2$). The shared haplotypes are found throughout the species distribution of *A. gambiae* although there is an apparent clustering of shared haplotypes in certain populations. Kyela, Tanzania; Mkali, Malawi; Kisian, Kenya and Dienga, Gabon. Five shared haplotypes were found in samples of *A. arabiensis* from Ethiopia, Sudan and Reunion (Table 1). *A. gambiae* is absent from the samples collected from these countries and therefore the presence of these haplotypes must reflect noncontemporary processes such as incomplete lineage sorting between the two species or historical introgression events. As would be predicted if ancestral polymorphisms were retained, the majority of shared haplotypes in these populations (four of five; Figure 3 and Table 1) were internal to the network. Similarly in the single species collection from Gabon, where *A. arabiensis* is absent, there were also shared haplotypes, again internal to the haplotype network (Figure 4 and Table 1) and reflecting a noncontemporary process. In *A. gambiae* except for two populations the proportion and frequency of haplotypes that were common to both species was between 22–50 and 17–50%, respectively. However, in samples from Tanzania and Malawi these figures rose to 75–78 and 67–68%, respectively.

Discussion

These data provide a corollary of the results of microsatellite-based macrogeographic studies of differentiation in both species (Lehmann *et al*, 1997, 2003; Kamau *et al*, 1998; Lanzaro *et al*, 1998; Simard *et al*, 1999; Donnelly and Townson, 2000). In general, as in microsatellite-based studies the levels of population differentiation were lower in *A. arabiensis* (Figure 2) (Donnelly and Townson, 2000; Lehmann *et al*, 2003) and the only large estimates of F_{ST} involved an island population which is likely to have experienced founding effects/genetic bottleneck (Simard *et al*, 1999; Donnelly *et al*, 2002). Surprisingly, samples of *A. arabiensis* from Senegal and Sudan were grouped in a cluster, which runs contrary to geographic distance. Similar patterns have been observed in *A. gambiae* and were thought to reflect similarities in ecological zones and the absence of topographic barriers to gene flow such as those that may isolate Sudanese samples from those to the south (Lehmann *et al*, 2003).

In *A. gambiae* there was a clear distinction between populations of M and S rDNA forms and within the S-form comparisons between samples from the West and East of the Rift Valley complex. These data are in accordance with the studies of Wondji *et al* (2002) and Lehmann *et al* (2003) that provided evidence for a degree of genetic isolation between M and S forms. The failure of other studies to detect differences between M and S forms reflects how recent the disruption of gene flow must have been (Gentile *et al*, 2001) or conversely that ongoing gene flow may be homogenising allele arrays in both forms, outside certain regions of the genome. This study utilised haplotype frequency-based approaches,

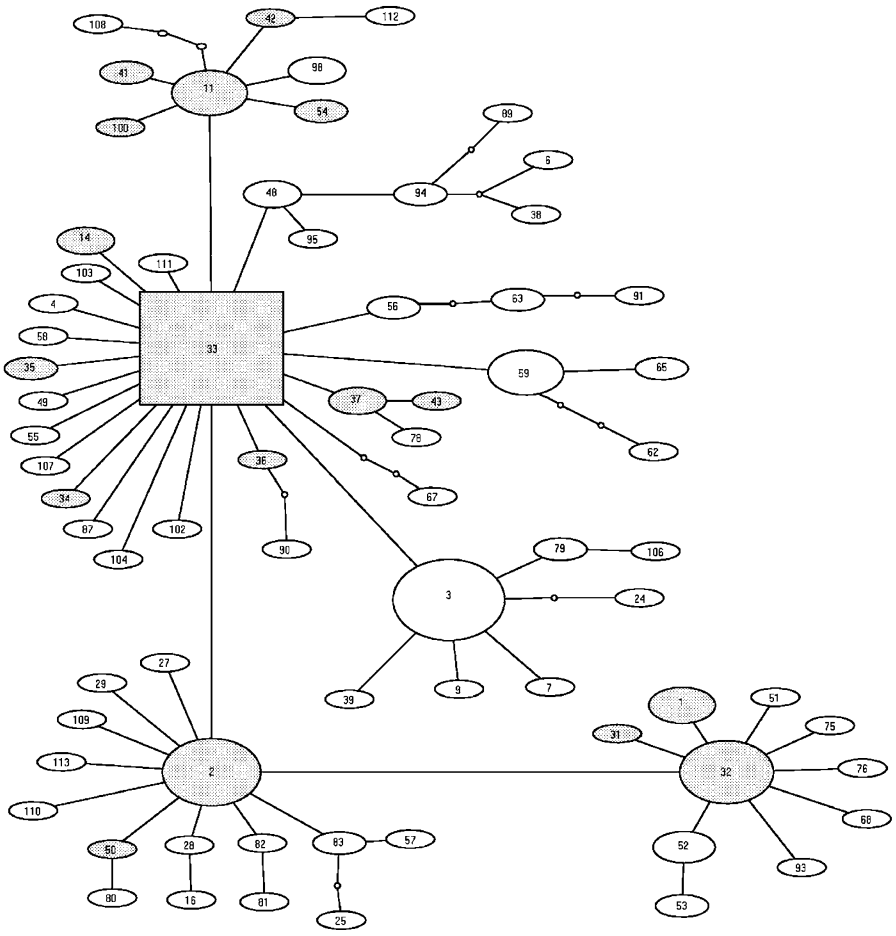

Figure 4 Statistical parsimony network for *A. gambiae* data set. Haplotype node area is proportional to the number of specimens contained. Links between nodes are all single mutational steps regardless of length. Shaded nodes are haplotypes that are found in both species.

which are more sensitive to more recent separation events since they are not reliant upon the accumulation of infrequent mutation events.

Introgression and ancestral retention

When contemporary introgression can be discounted, due to absence of one of the species from a locale, there are large numbers of haplotypes shared between species. Whether these shared sequences are true ancestral retentions or traces of recent introgression events cannot be determined and caution against definitive statements. These data suggest that for closely related taxa even very extensive mitochondrial DNA data sets may have insufficient power to conclusively resolve between the conflicting hypotheses of ancestral retention and contemporary introgression in certain populations.

Mitochondrial introgression is the most parsimonius explanation for the similarity in haplotype arrays between some sympatric populations of *A. gambiae* and *A. arabiensis* but introgression is apparently not an ubiquitous phenomenon. If introgression was occurring between all sympatric populations of *A. gambiae* and *A. arabiensis* then we would expect interspecific estimates of F_{ST} to be significantly lower in sympatric rather than allopatric comparisons. This was not observed in these data. However, interspecific F_{ST} analyses involving *A. gambiae* from Malawi were all nonsignificant whereas those interspecific comparisons involving Malawian *A. arabiensis* and *A. gambiae* from Kenya and Senegal were both highly significant (Table 2). This suggests that, since the haplotype distributions in Malawian *A. gambiae* are similar to all three *A. arabiensis* distributions, introgression in Malawi is likely to be a unidirectional process from *A. arabiensis* into *A. gambiae*. However, it should be noted that sample size was lowest in the sample of *A. gambiae* from Malawi and that mtDNA-based phylogenies using colonised specimens of *A. gambiae* and *A. arabiensis* suggested that introgression may have occurred in the opposite direction (Caccone *et al*, 1996). However, these two species have a far wider species distribution than other members of the complex. This is likely to result in higher effective population size and therefore since genetic drift will be lower there may well be a greater retention of ancestral haplotypes in these species than in other members of the complex, despite the possible closer phylogenetic proximity of different species pairs. There were insufficient data to apply similar tests to samples from Tanzania but a large number and proportion of haplotypes were shared between sympatric populations in Tanzania suggesting that introgression may be occurring in this region as well. *A. arabiensis* occurs at much higher frequencies than *A. gambiae* in the study sites in Tanzania and Malawi

Table 2 Interspecific pairwise estimates of F_{ST}

		A. arabiensis		
		Kenya (34)	Malawi (40)	Senegal (26)
A. gambiae	Kenya (85)	*0.1318*	*0.1398*	*0.0760*
	Malawi (19)	0.0318 NS	−0.0356 NS	0.0421 NS
	Senegal (32)	*0.2309*	*0.2061*	*0.1137*

F_{ST} calculated using the formulas of Hudson *et al* (1992) with weighting for population size. Significance of F_{ST} estimates was evaluated against the results from 2000 random permutations executed by a program written in the SAS language (SAS Institutes, 1990). Italic numerals indicate $P<0.001$. NS, not significant.

(Charlwood *et al*, 2000; Spiers *et al*, 2002) and differing species' density is thought to be one of the major determinants of introgression (Avise and Saunders, 1984). Whether introgression is frequent enough to play a role in the differentiation within *A. gambiae* S form that is observed either side of the Rift Valley remains to be investigated.

An analysis of the relative position of shared haplotypes on the species networks, based on the assumption that on average the haplotypes found interior to the haplotype network should be older than those found at the tips, was also suggestive of ongoing introgression. If there is contemporary introgression both old and more recently derived haplotypes are equally likely to cross the species barrier and therefore shared haplotypes will not preferentially occur at internal nodes. An exact test based upon the location (internal/external) of shared haplotypes in each species tree also showed no evidence for significant differences ($P>0.05$). This approach, although far from conclusive for our data given the number of ambiguities within the species trees, may be a powerful way of detecting introgression in those species with more robust and deeper networks.

Recent studies have demonstrated the importance of unidirectional introgression events in evolution associated with environmental changes (Grant and Grant, 2002). In the *A.gambiae* complex the interspecific transfer of DNA we observed at mitochondrial loci is likely occur at nuclear loci (Besansky *et al*, 2003) but it is unknown if this process is important for the acquisition of selectively advantageous genes. However, the evidence for introgression in natural populations and, in particular, from *A. arabiensis* into *A. gambiae*, lends credence to the hypothesis of Powell *et al* (1999) that some selectively advantageous genes may have moved from *A. arabiensis* into *A. gambiae*. Furthermore, while *A. gambiae* has been thought to be undergoing incipient speciation in certain locations the converse may actually be occurring and the M and S forms and *A. arabiensis* may be converging to form a hybrid swarm. This is of particular concern given widespread insecticide resistance and possibilities of transgenic release in these highly pernicious malaria vectors.

Acknowledgements

MJD was supported by a post-doctoral fellowship from the American Society of Microbiology. This work received financial support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases and from the NIH Grant (A140631-01). JP was funded by Praxis XXI/FCT (BD/15754/98).

References

Avise JC, Saunders NC (1984). Hybridization and introgression among species of sunfish (*Lepomis*): analysis by mitochondrial DNA and allozyme markers. *Genetics* **108**: 237–255.

Barton NH (2001). The role of hybridization in evolution. *Mol Ecol* **10**: 551–568.

Beaumont M, Barratt EM, Gottelli D, Kitchener AC, Daniels MJ, Pritchard JK *et al* (2001). Genetic diversity and introgression in the Scottish wildcat. *Mol Ecol* **10**: 319–336.

Besansky NJ, Krzywinski J, Lehmann T, Simard F, Kern M, Mukabayire O *et al* (2003). Semiperrmeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proc Natl Acc USA* **100**: 10818–10823.

Besansky NJ, Lehmann T, Fahey GT, Fontenille D, Braak LEO, Hawley WA *et al* (1997). Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *A. arabiensis*, suggest extensive gene flow. *Genetics* **147**: 1817–1828.

Besansky NJ, Powell JR, Caccone A, Hamm DH, Scott JA, Collins FH (1994). Molecular phylogeny of the *Anopheles gambiae* suggests genetic introgression between the principle malaria vectors. *Proc Natl Acad Sci USA* **91**: 6885–6888.

Caccone A, García BA, Powell JR (1996). Evolution of the mitochondrial DNA control region in the *Anopheles gambiae* complex. *Insect Mol Biol* **5**: 51–59.

Charlwood JD, Vij R, Billingsley PF (2000). Dry season refugia of malaria-transmitting mosquitoes in a dry savannah zone of east Africa. *Am J Trop Med Hyg* **62**: 726–732.

Clement M, Posada D, Crandall KA (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657–1659.

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

Crandall KA, Templeton AR (1993). Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134**: 959–969.

Della Torre A, Merzagora L, Powell JR, Coluzzi M (1997). Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. *Genetics* **146**: 239–244.

Donnelly MJ, Cuamba N, Charlwood JD, Collins FH, Townson H (1999). Population structure in the malaria vector, *Anopheles arabiensis* Patton, in East Africa. *Heredity* **83**: 408–417.

Donnelly MJ, Licht MC, Lehmann T (2001). Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Mol Biol Evol* **18**: 1353–1364.

Donnelly MJ, Simard F, Lehmann T (2002). Evolutionary studies of malaria vectors. *Trends Parasitol* **18**: 75–80.

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

Evans BJ, Supriatna J, Melnick DJ (2001). Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. *Evolution* **55**: 1686–1702.

Favia G, Della Torre A, Bagayoko M, Lanfrancotti A, Sagnon NF, Touré YT *et al* (1997). Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Mol Biol* **6**: 377–383.

Gentile G, Slotman M, Ketmaier V, Powell JR, Caccone A (2001). Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae* s.s. *Insect Mol Biol* **10**: 25–32.

Goodman SJ, Barton NH, Swanson G, Abernethy K, Pemberton JM (1999). Introgression through rare hybridization: a genetic study of a hybrid zone between red and sika deer (genus *Cervus*) in Argyll, Scotland. *Genetics* **152**: 355–371.

Grant PR, Grant BR (2002). Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**: 707–711.

Hudson RR, Boos DD, Kaplan NL (1992). Statistical test for detecting geographic subdivision. *Mol Biol Evol* **9**: 138–151.

Kamau L, Lehmann T, Hawley WA, Orago ASS, Collins FH (1998). Microgeographic genetic differentiation of *Anopheles gambiae* mosquitoes from Asembo Bay, Western Kenya: a comparison with Kilifi in coastal Kenya. *Am J Trop Med Hyg* **58**: 64–69.

Krzywinski J, Wilkerson RC, Besansky NJ (2001). Evolution of mitochondrial and ribosomal gene sequences in Anophelinae (Diptera:Culicidae): implications for phylogeny reconstruction. *Mol Phylo Evol* **18**: 479–487.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244–1245.

Lanzaro GC, Toure YT, Carnahan J, Zheng L, Dolo G, Traoré S *et al* (1998). Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proc Natl Acad Sci USA* **95**: 14260–14265.

Lehmann T, Besansky NJ, Hawley WA, Fahey TG, Kamau L, Collins FH (1997). Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. *Mol Ecol* **6**: 243–253.

Lehmann T, Blackston CR, Besansky NJ, Escalante AA, Collins FH, Hawley WA (2000). The effect of the Rift Valley on gene flow between *Anopheles gambiae* populations in Kenya: the mtDNA perspective. *J Hered* **91**: 165–168.

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

Lehmann T, Licht MC, Elissa N, Maega BTA, Chimumbwa JM, Watsenga FT *et al* (2003). Population structure of *Anopheles gambiae* in Africa. *J Hered* **94**: 133–147.

Martinsen GD, Whitham TG, Turek RJ, Keim P (2001). Hybrid populations selectively filter gene introgression between species. *Evolution* **55**: 1325–1335.

Petrarca V, Beier JC (1992). Intraspecific chromosomal 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.

Pinto J, Donnelly MJ, Malta-Vacas J, Sousa CA, Gil V, Ferreira C *et al* (2003). An island within an island: genetic differentiation of *Anopheles gambiae* in São Tomé, West Africa, and its relevance to malaria vector control. *Heredity* **91**: 407–414.

Pinto J, Donnelly MJ, Sousa CA, Gil V, Ferreira C, Elissa N *et al* (2002). Genetic structure of *Anopheles gambiae* (Diptera: Culicidae) in São Tomé and Príncipe (West Africa): implications for malaria control. *Mol Ecol* **11**: 2183–2187.

Powell JR, Petrarca V, Della Torre A, Merzagora L, Caccone A, Coluzzi M (1999). Population structure, speciation and introgression in the *Anopheles gambiae* complex. *Parassitologia* **41**: 101–114.

Roques S, Sevigny JM, Bernatchez L (2001). Evidence for broadscale introgressive hybridization between two redfish (genus *Sebastes*) in the North-west Atlantic: a rare marine example. *Mol Ecol* **10**: 149–165.

SAS Institute Inc (1990). *SAS Language: References*, Version 6, 1st edn. Cary, NC.

Scott JA, Brogdon WG, Collins FH (1993). Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am J Trop Med Hyg* **49**: 520–529.

Simard F, Fontenille D, Lehmann T, Girod R, Brutus L, Gopaul R *et al* (1999). High amounts of genetic differentiation between populations of the malaria vector *Anopheles arabiensis* from west Africa and eastern outer islands. *Am J Trop Med Hyg* **60**: 1000–1009.

Spiers AA, Mzilahowa T, Atkinson D, McCall PJ (2002). The malaria vectors of the Lower Shire Valley, Malawi. *Malawi Med J* **14**: 4–7.

Templeton AR, Crandall KA, Sing CF (1992). A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA-sequence data III Cladogram estimation. *Genetics* **132**: 619–633.

Thelwell NJ, Huisman RA, Harbach RE, Butlin RK (2000). Evidence for mitochondrial introgression between *Anopheles bwambae* and *Anopheles gambiae*. *Insect Mol Biol* **9**: 203–210.

Tripet F, Toure YT, Taylor CE, Norris DE, Dolo G, Lanzaro GC (2001). DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Mol Ecol* **10**: 1725–1732.

Walton C, Handley JM, Tun-Lin W, Collins FH, Harbach RE, Baimai V *et al* (2000). Population structure and population history of *Anopheles dirus* mosquitoes in southeast Asia. *Mol Biol Evol* **17**: 962–974.

White GB, Magayuka SA, Boreham PFL (1972). Comparative studies on sibling species of the *Anopheles gambiae* Giles complex (Dipt., Culicidae): bionomics and vectorial activity of species A and species B at Segera, Tanzania. *Bull Ent Res* **62**: 295–317.

Wondji C, Simard F, Fontenille D (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 Mol Biol* **11**: 11–19.