Published in final edited form as:

Mol Ecol. 2011 December ; 20(23): 4983–4994. doi:10.1111/j.1365-294X.2011.05339.x.

Asymmetric introgression between the M and S forms of the malaria vector, *Anopheles gambiae*, maintains divergence despite extensive hybridisation

Clare D. Marsden¹, Yoosook Lee¹, Catelyn C. Nieman¹, Michelle R. Sanford¹, Joao Dinis², Cesario Martins², Amabelia Rodrigues², Anthony J. Cornel³, and Gregory C. Lanzaro^{1,*}

¹Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California – Davis, California, USA, 95616

²National Institute of Public Health (INASA), CP 1013, Bissau, Guinea-Bissau

³Mosquito Control Research Laboratory, Department of Entomology, University of California - Davis, California, USA, 95616

Abstract

The suggestion that genetic divergence can arise and/or be maintained in the face of gene flow, has been contentious since first proposed. This controversy and a rarity of good examples has limited our understanding of this process. Partially reproductively isolated taxa have been highlighted as offering unique opportunities for identifying the mechanisms underlying divergence with gene flow. The African malaria vector, *Anopheles gambiae* s.s., is widely regarded as consisting of two sympatric forms, thought by many to represent incipient species, the M and S molecular forms. However, there has been much debate about the extent of reproductive isolation between M and S, with one view positing that divergence may have arisen and is being maintained in the presence of gene flow, and the other proposing a more advanced speciation process with little realised gene flow due to low hybrid fitness. These hypotheses have been difficult to address because hybrids are typically rare (<1%). Here, we assess samples from an area of high hybridisation and demonstrate that hybrids are fit and responsible for extensive introgression. Nonetheless, we show that strong divergent selection at a subset of loci combined with highly asymmetric introgression has enabled M and S to remain genetically differentiated despite extensive gene flow. We propose the extent of reproductive isolation between M and S varies across West Africa resulting in a “geographic mosaic of reproductive isolation”; a finding which adds further complexity to our understanding of divergence in this taxon and which has considerable implications for transgenic control strategies.

Keywords

Asymmetric introgression; sympatric speciation; malaria; ecological speciation; reproductive isolation

INTRODUCTION

The occurrence of genetic divergence between sympatric populations may involve processes whereby populations diverge, or divergence is maintained, despite on-going gene flow. This can range from the most extreme scenario of sympatric speciation where gene flow occurs

*Corresponding author: Gregory C. Lanzaro, gclanzaro@ucdavis.edu, Fax: +1 - 530-754-0299.

throughout the divergence process, to secondary contact where gene flow proceeds a period of divergence in allopatry, as well as intermediate scenarios with periodic gene flow (Nosil 2008; Pinho & Hey 2010). This process, hereafter divergence with gene flow, was once a highly debated and contentious topic but is becoming more accepted due to accumulating evidence from both theoretical and empirical studies (Pinho & Hey 2010). However, a general paucity of demonstrative examples in natural populations has resulted in few opportunities to study species that have undergone this process. An alternative avenue of research that has been proposed to improve our understanding of how population divergence may arise or be maintained in sympatry is to assess taxa that are partially reproductively isolated (Via 2009). A particular advantage of this approach is that it enables the detection of mechanisms contributing to restricted gene flow during the early stages of this process; factors that may become confounded by changes post-speciation (Via 2009).

Models of the elements inherent to divergence with gene flow imply that several discrete factors are necessary to drive this process, including strong divergent selection and assortative mating (Gavrilets 2006; Maynard Smith 1966; Pinho & Hey 2010). During the beginning stages of this process, when reproductive isolation is incomplete, it is predicted that the genomes of diverging taxa will be homogenized except in those regions involved in ecological differentiation and reproductive isolation, where gene flow is predicted to be restricted by selection (Wu 2001; Pinho & Hey 2010). However, as divergence becomes more advanced and reproductive isolation between taxa becomes stronger, divergence should become widespread across the genome.

The African malaria vector, *Anopheles gambiae* s.s., has become a widely studied and influential subject for research on the genetics of speciation (Lawniczak *et al.* 2010; Turner & Hahn 2010). Two morphologically indistinguishable molecular forms (M and S) have been recognised based on fixed differences at an X linked ribosomal DNA marker (Fanello *et al.* 2003; Favia *et al.* 1997). Extensive surveys have detected very few M-S hybrids in most areas of West Africa where M and S occur in sympatry (<1%, della Torre *et al.* 2005; Simard *et al.* 2009; Tripet *et al.* 2001). To date, no post-zygotic reproductive barriers have been found in artificially produced hybrids (Diabate *et al.* 2007), but spatial segregation of mating swarms (Diabate *et al.* 2009) and assortative mating (Tripet *et al.* 2001) via harmonic convergence of wing beat frequency (Pennetier *et al.* 2010) appear to reduce hybridisation in wild populations. Furthermore, and consistent with most models of divergence with gene flow (Bolnick & Fitzpatrick 2007; Pinho & Hey 2010; Schluter 2001; Wu 2001), evidence of niche differentiation suggests that ecologically based divergent selection may be driving the evolutionary trajectories of M and S (Costantini *et al.* 2009; Gimonneau *et al.* 2011; Simard *et al.* 2009), which could cause ecological pre-zygotic isolation and/or environmentally-dependent post-zygotic isolation. Based on these findings, M and S are regarded as incipient species diverging in the face of considerable gene flow.

Contradictory findings from genomic studies conducted on samples from different parts of the species distribution where the forms are sympatric, has led to considerable debate about the extent of reproductive isolation between M and S (e.g. Lawniczak *et al.* 2010; Turner & Hahn 2010; White *et al.* 2010). Two competing hypotheses concerning divergence between these two taxa have been generated to account for the differences observed in patterns of genetic differentiation. The first of these is the speciation-with-gene-flow model as described by Turner *et al.* (2005), which proposes that gene flow from hybridisation results in low differentiation across the genome except at gene regions involved in driving divergence where selection prevents introgression resulting in a genome consisting of a mosaic of high and low diverged regions. By contrast, the speciation-without-gene-flow model described by White *et al.* (2010) and Lawniczak *et al.* (2010) proposes a more advanced speciation process where low fitness of F₁ hybrids results in little realised gene

flow from hybridisation, and widespread divergence across the genome. The general rarity of hybrids (<1%, della Torre *et al.* 2005; Simard *et al.* 2009; Tripet *et al.* 2001) makes it difficult to assess these conflicting views. However, high hybridisation frequencies reported in some regions of West Africa (Caputo *et al.* 2008; Oliveira *et al.* 2008) may offer unique opportunities to distinguish between these hypotheses and improve our understanding of how between-form divergence is maintained. To this end, we assessed introgression and genetic divergence between M and S in an area of high hybridisation rates in Guinea-Bissau.

MATERIALS AND METHODS

Sample collection, DNA extraction

Specimens of *A. gambiae* were collected from five sites in the Republic of Guinea-Bissau between October and November 2009 (Fig. 1). Wherever possible, specimens were collected as blood fed indoor resting adults. At sites where adults were rare or absent, we collected larvae from road side pools. All samples were stored in individual tubes containing 70% ethanol. DNA was extracted from specimens using DNeasy blood and tissue kit (Qiagen, Valencia, CA) using the Qiagen Biosprint 96 DNA extraction system. The DNA samples were subsequently amplified with whole genome amplification Repli-G kits (Qiagen, Valencia, CA) to raise DNA concentrations to 50µg/µl for downstream analyses.

Taxonomic status and *Wolbachia* testing

The Scott *et al.* (1993) PCR assay was used to identify *A. gambiae s.s.* from other members of the morphologically indistinguishable *A. gambiae s.l.* species complex. A PCR method using male-specific primers was used to determine the sex of larvae (Ng'habi *et al.* 2007); male samples were not used in further analyses, except for *Wolbachia* testing (described below). The molecular form of 418 female samples was determined using standard diagnostic PCR assays (Fanello *et al.* 2003; Favia *et al.* 1997) which target a pair of adjacent SNPs within the inter-genic spacer region on the X chromosome. However, sequencing of this diagnostic region in a subset of samples (n=69, ABI 3730 sequencer) revealed inaccuracies in these PCR-based diagnostics; a problem that has been recorded elsewhere (Caputo *et al.* 2011; Oliveira *et al.* 2008). Therefore, we further verified the molecular form of all samples by SNP genotyping the pair of SNPs targeted by the Fanello *et al.* (2003) and Favia *et al.* (1997) diagnostics, using a Sequenom iPLEX[®] Gold assay (Sequenom, Inc., San Diego, CA, USA, van den Boom & Ehrich 2007). It has previously been shown that the first of the two diagnostic SNPs is not 'form specific' in all populations (Gentile *et al.* 2002), including Guinea-Bissau (Oliveira *et al.* 2008), which is what we found. As such, in this study, determination of molecular form of an individual was based on the second diagnostic SNP site (position 581bp according to Scott *et al.* 1993), hereafter ScottX: S form - C homozygous; M form - T homozygous; hybrid - C/T heterozygous. As such, any reference to M, S or hybrids refers to designations based on ScottX, unless otherwise stated. We tested for *Wolbachia* infection, an intracellular bacterium that can alter patterns of species hybridisation in other mosquito species (Dean & Dobson 2004), in a subset of 56 male and 56 female samples according to the protocol outlined in Sakamoto *et al.* (2006).

SNP genotyping

Ninety-six SNPs were selected from published sequences of which multiple isolates of *A. gambiae s.s.* have been sequenced (Cohuet *et al.* 2008; Harris *et al.* 2010; Lehmann *et al.* 2009; Mendes *et al.* 2008; Morlais *et al.* 2004; Obbard *et al.* 2009; Parmakelis *et al.* 2008; Santolamazza *et al.* 2008; Slotman *et al.* 2007; Turner *et al.* 2005; Turner & Hahn 2007; White *et al.* 2009; White *et al.* 2010; White *et al.* 2007). Based on these sequences, primers were designed for use in a customised Illumina[®] Golden Gate assay. SNP genotyping was conducted on all female samples available from Eiticoga, Bruce and Abu, and a subset of

female samples from Antula (63/124) and Prabris (64/98). Genotyping was conducted on the Illumina Bead Station 500G Golden Gate genotyping platform (Illumina, San Diego, USA) at the University of California, Davis DNA Technologies Core Facility using protocols provided by the manufacturer. Data was analysed using BeadStudio v3.2 software (Illumina, San Diego, USA) which normalises the raw hybridisation intensity data before generating genotype calls. Genotype clusters were subsequently adjusted manually.

Genetic diversity

Based on the SNP data, genetic diversity of M, S and hybrid forms was measured as the number of private alleles (PA), and observed (H_o) and expected heterozygosity (H_e) as calculated in GenALEX6.3 (Peakall & Smouse 2006), and fixation index (F_{IS}) and allelic richness standardised for sample size (R_S) as calculated in FSTAT 2.9.3.2 (Goudet 1995).

Associations between the Islands of Speciation

We tested for evidence of non-random associations (gametic disequilibrium) between the islands of speciation on the 2nd, 3rd and X chromosomes by examining a single SNP from each of these regions. Specifically we used the ScottX diagnostic on the X chromosome, and for the 2nd and 3rd chromosome we selected loci Ag2L-2422079 and Ag3L-413944 respectively, from the data set of 52 loci, as these loci have been shown to exhibit fixed differences between M and S samples from Mali and Cameroon (Turner *et al.* 2005; White *et al.* 2010). It is noteworthy that locus Ag2L-2422079 on chromosome 2 is strictly speaking located adjacent to, rather than within, the speciation island. We tested for non-random associations between these three SNPs, hereafter island SNPs, in Arlequin v.3.5.1.2 using a likelihood ratio test, with 50,000 permutations and 5 initial EM conditions using unphased genotypic data.

Assessing hybrid class

We used posterior probability assignment tests, implemented through the USEPOPINFO option in STRUCTURE v 2.3.3 (Pritchard *et al.* 2000), to determine the class of hybrids (i.e. F_1 , $M \times F_1$ backcross, etc) present amongst those individuals in our samples with M/S (hybrid) genotypes. We first simulated a set of hybrids in HYBRIDLAB v1.0 (Nielsen *et al.* 2006) using all M and S samples with non-admixed genotypes (Fig. 2b) as the parental populations: M and S samples were defined as non-admixed, if they showed <5% admixture in STRUCTURE analyses of the complete set of 52 SNP loci (described below, Fig. 2b). We then simulated 50 hybrids for each of the following categories; F_1 ($M \times S$), F_2 ($F_1 \times F_1$), backcross F_{2M} ($F_1 \times M$), backcross F_{2S} ($F_1 \times S$), backcross F_{3M} ($F_2 \times M$), backcross F_{3S} ($F_2 \times S$), backcross F_{4M} ($F_3 \times M$), backcross F_{4S} ($F_3 \times S$). We then ran STRUCTURE with the simulated hybrids and parental M and S samples to determine the expected range of ancestry proportions to the M and S clusters for the different hybrid classes. These values were then compared with those obtained for our field collected hybrids in a subsequent STRUCTURE analysis. In all cases, STRUCTURE was run with the non-admixed M and S samples used as training samples, assuming correlated allele frequencies, admixture, a migration rate of 0.01, 200,000 burn-in cycles and 1,000,000 Markov Chain Monte Carlo runs (MCMC) runs.

Genetic differentiation and *Wolbachia* infection

To estimate genetic differentiation based on SNP loci, we calculated population and locus specific pairwise F_{ST} for M, S and hybrids using Arlequin 3.5.1.2 (Excoffier & Lischer 2010). Significance was assessed after Bonferroni correction using $\alpha=0.05$. We then assessed genetic structure using Principle Coordinates Analysis (PCoA) calculated in GenALEX (Peakall & Smouse 2006) and Bayesian clustering analysis implemented through STRUCTURE v2.3.3. STRUCTURE was run five times at $K = 1-7$ assuming no prior

population information, with correlated allele frequencies and admixture, 200,000 burn-in cycles and 1,000,000 MCMC. The value of K that best fit our data was selected using the ΔK statistic (Evanno *et al.* 2005). STRUCTURE was first run on the complete SNP dataset, and then using only SNP loci on chromosome 3 to test for evidence of a cryptic exophilic subpopulation of *A. gambiae* recently described in larval populations in Burkina Faso by Riehle *et al.* (2011).

RESULTS

Quantifying hybridisation rates

We collected 418 adult and larval *A. gambiae* from five sites in Guinea-Bissau and verified their molecular form (Fig. 1). Similar to other studies in this region (Caputo *et al.* 2011; Oliveira *et al.* 2008), sequencing validation of a subset of samples highlighted that traditional diagnostic methods were associated with high numbers of misclassifications (9% Fanello n=67, 14% Favia n=56). These errors appear to be associated with the multicopy nature of the IGS rDNA region which results in unequal proportions of M and S type rDNA in backcrossed hybrids that causes amplification biases to the extent that the less frequent rDNA type is undetectable on an agarose gel (Caputo *et al.* 2011; Gentile *et al.* 2002). As such, we verified the molecular form of samples through sequencing and/or SNP genotyping of the M-S diagnostic rDNA locus (see methods), and found this methodology to have a high sensitivity for detecting hybrids not visible on gels. Based on this approach, we found high hybridisation rates in excess of 20% across all five sites and in both adult (n = 137) and larval (n = 281) samples (23.5–35.4%; Fig. 1, Table 1). These high hybrid frequencies cannot be explained by sampling of larvae, as collections from three of the sites (ANT, PRA, ABU) consisted only of adults (Table S1, Supplementary Information).

Genetic diversity

We typed 323 of the 418 samples at 96 SNP loci using a customized Illumina® Golden Gate assay. After excluding loci that were non-informative, i.e. those that exhibited high failure rates, indistinguishable clusters, no polymorphism or deviations from Hardy Weinberg equilibrium within the M or S form, 52 loci remained for analyses. These 52 loci were distributed across all three chromosomes with 8–14 SNPs per chromosomal arm, and included 14 loci within the divergence islands, also known as the “islands of speciation” (Table S2, Supplementary Information).

Associations between the Islands of Speciation

Previous studies have demonstrated complete gametic disequilibrium between loci on the islands of speciation on the X, 2L and 3L chromosomes using samples from Mali, Cameroon and Burkina Faso (White *et al.* 2010). By contrast, in Guinea-Bissau, we found all but one of the possible pairwise combinations of SNPs (Table S3, Supplementary Information), which is consistent with a similar assessment conducted by Caputo *et al.* 2011 (Antula, Guinea-Bissau). However, some pairwise combinations were very rare (n<10), and overall, pair-wise associations between the three SNPs located within the islands of speciation on the X, 2L and 3L showed evidence of non-random associations (i.e. gametic disequilibrium): 2L-3L, $\chi^2 = 22.59$, d.f. = 4, $p < 0.001$; 2L-X, $\chi^2 = 21.93$, d.f. = 2, $p < 0.0001$; 3L-X, $\chi^2 = 13.28$, d.f. = 2, $p < 0.01$. Closer examination of the data revealed stronger associations in the M form than the S form: 40/45 M form samples (89%) were homozygous for M form SNPs at both the 2L and 3L islands, whereas only 8/170 (<5%) S form samples were homozygous for S form SNPs at both the 2L and 3L islands. Moreover, 30/170 (18%) S form samples were homozygous for M form SNPs at the 2L and 3L islands. Together these data indicate that despite high levels of hybridisation, an association between the X, 2L and 3L islands of speciation has been maintained in Guinea-Bissau, although it is clearly

weaker than the association described by White *et al.* 2010. These results do, however, show that M and S are not panmictic in Guinea-Bissau.

Assessing hybrid class

We conducted posterior probability testing in STRUCTURE to identify the class of hybrids that we had sampled (i.e. F_1 , $F_1 \times M$ backcross etc). We first determined the expected ancestry values (i.e. proportion of their genome derived from M or S) for a set of simulated hybrids generated using the M and S samples as parental types. These values were then compared with the values STRUCTURE generated for the wild hybrids. Due to overlap in assignment probability ranges of the different classes of simulated hybrids it was not possible to assign the sampled hybrids to specific hybrid classes (Table S4, Supplementary Information). For example, the proportion of S ancestry derived from a set of simulated $F_1 \times S$ and $F_2 \times S$ hybrids were very similar (0.51–0.95 and 0.55–0.95 respectively). Nonetheless, it is noteworthy that 70% (76/108) of the sampled hybrids had ancestry values outside the range of F_1 hybrids and 61% (66/108) were outside the range expected for F_1 or $F_2(F_1 \times F_1)$ or M hybrid backcrosses ($S > 0.63$, $M < 0.37$), thus indicating them to be hybrid $\times$ S backcrosses (Fig. 3). Only three hybrids had ancestry values within the range expected for M backcrosses (M assignment values 0.57–0.68), but these three individuals were also consistent with classification as F_1 and $F_2(F_1 \times F_1)$ hybrids (Fig. 3).

Evidence of repeated backcrossing with the S form raises the question of when to classify a backcrossed hybrid as S. Given that the diagnostic assay is based on a multi-copy gene, it is possible that we detected hybrids derived from multiple generations of backcrossing with the S form, that some might classify as S. Likewise it is possible that the diagnostic assay was unable to detect backcrosses with very few S rDNA copies, and thus resulted in underestimates of hybrid frequencies. To address this issue, it would be useful to develop an assay that can distinguish between early and late generation hybrids.

Genetic diversity and differentiation

Based on 52 SNP markers we found genetic diversity metrics (R_S , H_e) to be consistently higher, and inbreeding metrics (H_o , F_{IS}) to be consistently lower, in the S form and hybrids than the M form (Table 2). Furthermore, only the S form had private alleles (Table 2).

We found genetic differentiation to be low between S and hybrids (F_{ST} 0.031, $p < 0.0001$), but high between M and S (F_{ST} 0.348, $p < 0.0001$), and M and hybrids (F_{ST} 0.221, $p < 0.0001$). The PCoA identified three clusters (Fig. 2a): 1) all M samples; 2) approximately half of the hybrids; 3) all S samples and the remaining hybrids. Similarly, STRUCTURE analyses found M and S to form largely independent clusters with hybrids exhibiting either an admixed M-S or largely S genotype (Fig. 2b, Fig. S1, Supplementary Information). To verify that these results were not biased by the presence of eight pairs of linked loci within our dataset, we removed one of each of these pairs of eight loci and repeated F_{ST} , PCoA and STRUCTURE analyses on the 44 remaining loci. We found no evidence of a bias, as the results from the 44 loci data set were virtually identical (Fig. S2, Supplementary Information). Together these results suggest that differentiation between the M and S forms has been maintained despite extensive gene flow. However, it is also evident from these data that hybrids are more genetically similar to the S form, than the M form. This indicates that backcrossing has been highly asymmetric towards the S form.

In light of this unexpected result, we evaluated two additional hypotheses that arose following analysis. Mosquito populations containing individuals infected with bacteria in the genus *Wolbachia* may exhibit asymmetric introgression due to cytoplasmic incompatibility caused by these organisms. We evaluated a subset of individuals (56♀ : 56♂) from our

study populations using a PCR diagnostic and found no evidence of *Wolbachia* infection. In addition, we tested for evidence a cryptic exophilic subpopulation of *A. gambiae* characterised by high hybridisation rates that was recently described in larval specimens from Burkina Faso (Riehle *et al.* 2011). However, we found no evidence for this in our data set (Fig. S3, Supplementary Information).

To assess whether the level of divergence between M and S was consistent across the genome, we calculated F_{ST} values at loci within (14 loci) and outside (38 loci) previously identified regions of divergence (islands of speciation, as defined in Lawniczak *et al.* 2010; Turner *et al.* 2005), at loci on autosomes (42 loci) and the X chromosome (10 loci), and lastly at each locus separately. Differentiation between M and S was found to be greater at loci on the X chromosome (F_{ST} 0.824, $p < 0.0001$), than at autosomal loci (F_{ST} 0.070, $p < 0.0001$; Table S5, Supplementary Information). Furthermore, locus specific F_{ST} values show that 14/52 loci were significantly differentiated between M and S, and these were largely those loci located within or near the islands of speciation and/or centromeres. Overall, F_{ST} values were an order of magnitude greater within the islands of speciation (F_{ST} 0.616), than outside (F_{ST} 0.056), despite being significant between M and S both within and outside the islands of speciation (Table S5, Supplementary Information). As such, we add the caveat that our overall estimates of differentiation likely provide an upwardly biased estimate of genome wide differentiation because our data set was enriched with loci from the islands of speciation.

DISCUSSION

A. gambiae s.s. has become an unexpected but ideal model for research into the genetics of species divergence. Recently however, a high profile debate has arisen about the degree of reproductive isolation between the two forms, with one view positing that divergence is occurring in the presence of gene flow (Turner *et al.* 2005), and the other proposing a more advanced speciation process with little gene flow due to low fitness of hybrids (Lawniczak *et al.* 2010; White *et al.* 2010). Resolving this dichotomy is not only important for understanding divergence in this system, but also for genetic vector control strategies for *A. gambiae* which will rely on detailed knowledge of patterns of gene flow between these incipient taxa. Here, we present data suggesting that the degree to which reproductive isolation is complete in *A. gambiae*, and the mechanisms underlying divergence, may vary over the sympatric range.

Hybridisation and genetic differentiation

In contrast to most other regions of West Africa where hybridisation rates between the M and S molecular forms have been typically estimated at $<1\%$ (e.g. della Torre *et al.* 2005; Simard *et al.* 2009; Tripet *et al.* 2001), we found atypically high hybridisation across all five sites in western Guinea-Bissau (Fig. 1; Table 1). These data in combination with evidence of similar hybrid frequencies at one collection site in previous years (Antula, 1995, 1996, 2007, 19.4–24.1% Caputo *et al.* 2011; Oliveira *et al.* 2008), is indicative of either a less advanced speciation process or significantly altered pattern of assortative mating (possibly due to secondary contact) in this region of West Africa.

Central to the debate concerning reproductive isolation between M and S, is whether hybrids have sufficient fitness to result in introgression. Here, we found that 60% of field-collected hybrids were backcrossed hybrids, thus providing strong evidence that hybridisation yields viable and fecund F_1 hybrids and considerable gene flow (Fig. 3). As such, our data disputes the notion that M-S hybrids are less viable in nature (White *et al.* 2010). Nonetheless, contrary to what we would expect based on the extent of backcrossing observed, we found high genetic differentiation between M and S (F_{ST} 0.348, $p < 0.0001$) and both the PCoA and

STRUCTURE analyses showed M and S to form two clear and largely independent clusters (Fig. 2). Furthermore, despite high levels of gene flow, we found evidence of non-random associations between the islands of speciation on the X, 2L and 3L chromosomes. These data indicate that M and S are not panmictic in Guinea-Bissau, and that genetic divergence has been maintained between M and S despite extensive gene flow which is consistent the speciation-with-gene flow model as proposed by Turner *et al.* (2005).

How is divergence maintained?

In addition to the observed population structure, F_{ST} and clustering analyses of our data also showed greater genetic similarity between hybrids and the S form vs hybrids and the M form (Fig. 2, Table S5, Supplementary Information). These data suggest that there has been strong directionality in hybrid backcrossing towards the S form; a finding that is consistent with assignment tests which found that of the 76 backcrossed hybrids, a maximum of 3, and minimum of 0, were potential M backcrosses. To further verify the occurrence of asymmetric introgression, we examined data for the 2L and 3L island SNPs. As predicted, the majority of M form (>89%, $n = 45$) displayed the M form SNPs, whereas <5% of S form samples ($n=170$) were homozygous for S form SNPs. We hypothesise that asymmetric introgression is contributing to the maintenance of differentiation between M and S. Unidirectional movement of nuclear genes from the M into the S form would prevent homogenisation of the M and S gene pools due to the conservation of unique polymorphism within the S form and a lack of admixture in M. Consistent with this hypothesis, private alleles were only present in the S form, and genetic diversity was considerably lower in the M than the S form (Table 2); a pattern that has been reported elsewhere in West Africa (specific genes, Turner *et al.* 2005; 2R, Turner & Hahn 2007; 3L, White *et al.* 2010).

Where divergence occurs in the face of ongoing gene flow, the genomes of the diverging taxa are expected to consist of a mosaic of regions of low and high divergence, reflecting locations where gene flow occurs freely or is restricted by selection (Wu 2001; Pinho & Hey 2010). In *A. gambiae s.s.*, a small number of regions of the genome exhibiting differentiation (speciation islands) have been identified and are implicated in driving divergence between M and S despite gene flow (Turner *et al.* 2005). Evidence of non-random associations between SNPs located in these speciation islands on the X, 2L and 3L indicates the role of strong selection and/or mate choice in maintaining divergence between M and S. Furthermore, based on locus specific F_{ST} values, we found a mosaic pattern of introgression across the genome, indicating that strong divergent selection has restricted gene flow at a subset of loci, thus enabling the M and S forms to remain genetically differentiated despite extensive introgression. Specifically, we found that the majority (75%) of SNPs do not show significant differentiation between M and S, whereas other loci, particularly those close to the centromeres and in/near the putative islands of speciation, were often highly differentiated (Table S5, Supplementary Information). For example, of the 14 loci on chromosome 3L, only three loci exhibited significant differentiation (F_{ST}) and these were all located within the speciation island. Overall, the highest divergence was found at SNP loci on the X chromosome, which is similar to other studies (e.g. Neafsey *et al.* 2010; White *et al.* 2010), and may suggest that genes on the X chromosome are particularly important for promoting divergence in this taxon (large X effect, Coyne & Orr 2004). Assuming recent divergence of the M and S forms, it is also possible that the low divergence between M and S observed at 75% of loci represents shared ancestral polymorphism and not contemporary gene flow. Distinguishing between signals of ancestral polymorphism and gene flow is notoriously difficult (Muir & Schlötterer 2005). However, we propose that the recovery of so many hybrids with high and variable admixture proportions (Fig. 2b, Fig. 3) provides very strong evidence of recent introgression (Lexer *et al.* 2006). Furthermore the large number of individuals with multi-locus genotypes consistent with F_1 hybrid status ($n=32$,

Fig. 3) would be very difficult to explain if these shared alleles represent ancestral polymorphisms.

In summary, we postulate that a combination of asymmetric gene flow and strong divergent selection has maintained genetic divergence between M and S despite extensive hybridisation; although it remains to be shown whether high hybridisation rates are the result of a less advanced speciation process or secondary contact. Below we discuss the factors that may have contributed to high hybridisation rates and asymmetric gene flow.

Geographic mosaic of reproductive isolation

There has been considerable debate about the strength of reproductive isolation between M and S. However, reproductive isolation results from the evolution of traits associated with pre and/or postzygotic isolation (Borge *et al.* 2005); traits which can evolve locally (Martin & Willis 2010), and because they are not fixed in populations with incomplete reproductive isolation, may vary regionally (Borge *et al.* 2005) or temporally (Grant & Grant 2002) in response to different selection pressures (Seehausen *et al.* 2008). This results in differences in the strength and/or type of reproductive barriers amongst geographic areas (Seehausen *et al.* 2008). A geographic mosaic of reproductive isolation such as this could account for the geographic differences in hybridisation rates observed in *A. gambiae* which range from classically observed low rates of <1%, to areas where hybridisation is absent (e.g. Cameroon, Simard *et al.* 2009), or high (e.g. Guinea-Bissau and The Gambia, Caputo *et al.* 2008). Consistent with this hypothesis, geographic differences in the strength of pre-zygotic barriers have been identified (Diabate *et al.* 2009) and some of the genomic regions thought to underlie reproductive isolation have been found to be unique to specific geographic areas; sections of chromosomes 2R in Cameroon (Turner & Hahn 2007) and 3L in Mali (White *et al.* 2010). Further research is required to investigate niche differentiation and mating barriers as well as to examine broader regions of genomic divergence in Guinea-Bissau so to identify specific differences from locations with low hybridisation.

Asymmetric introgression

Asymmetric introgression, whereby one of two hybridising species acts as the genetic donor and the other as the genetic recipient, has been documented in a variety of taxa (e.g. Borge *et al.* 2005; Gomes *et al.* 2009), and has been shown to be important for adaptive evolutionary change (Grant & Grant 2002; Pfennig 2007). Whilst this is the first study to demonstrate asymmetric introgression within *A. gambiae*, this phenomenon has been proposed previously in this taxa (Oliveira *et al.* 2008) and is consistent with findings of other genetic studies in Guinea-Bissau (Caputo *et al.* 2011; Oliveira *et al.* 2008). Furthermore, it has been recorded between *Anopheles* species in Asia (Morgan *et al.* 2010), as well as within the *A. gambiae* s.l. species complex (Donnelly *et al.* 2003). The cause of asymmetric introgression in *A. gambiae* is currently unclear, but may be a consequence of the differences in relative abundance of the two taxa (S was more common than M in all sites we assessed, as much as 43 S : 2 M in Bruce, Table 1), which has been shown to result in hybrids backcrossing more frequently with the more common species (e.g. Borge *et al.* 2005). Similarly, where range expansions occur, the direction of introgression is predicted to occur from the local species into the invading species (Curat *et al.* 2008; Excoffier *et al.* 2009). This would be consistent with the hypothesis proposed by Caputo *et al.* (2011) that Guinea-Bissau represents a secondary contact zone, rather than a region where the speciation process is less advanced. However, insufficient data is available to conclusively distinguish between this pair of hypotheses. Alternatively, the pattern may reflect asymmetric behavioural isolation (Egger *et al.* 2010), differential fitness of reciprocal crosses in the local environment (Morgan *et al.* 2010), or selection for introgression of advantageous M genes into the S form, and/or selection against introgression in the opposite

direction (Morgan *et al.* 2010). However, we did rule out cytoplasmic incompatibility via *Wolbachia* sp. infection, as we found no evidence of infection in a subset of our samples. Given that asymmetric introgression could impede the movement of transgenes within this taxon, knowledge about this mechanism is highly significant for the design and implementation of malaria control programmes using genetically modified mosquitoes. As such, it would be valuable to assess whether introgression is asymmetric elsewhere in West Africa, particularly in other areas with elevated hybridisation rates such as areas of The Gambia (e.g. Sare Illo Buya, 17%, Caputo *et al.* 2008) and Burkina Faso (e.g. Koukoulou, 57%, Costantini *et al.* 2009).

CONCLUSION

In Guinea-Bissau we found atypically high rates of hybridisation and evidence of backcrossing indicating that assortative mating in this region differs from those locations where the incipient nature of these species was initially described. Nonetheless, we found that asymmetric introgression and strong divergent selection at a subset of loci have enabled the M and S forms to remain genetically differentiated despite extensive gene flow, which is consistent with Turner *et al.* (2005). Overall, the data presented here suggest that the mechanisms of reproductive isolation, and the degree to which reproductive isolation is complete, may vary over the sympatric range of the molecular forms of *A. gambiae*. We have called this “the geographic mosaic of reproductive isolation”, which may be a characteristic associated with divergence in other taxa also. As such these data highlight the need for widespread sampling, as well as the importance of ecological complexity and associated local mechanisms, in studies of divergence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Charles E. Taylor (Dept. Ecology and Evolutionary Biology, University of California - Los Angeles) for help with field sample collections, Charles Nicolet and Vanessa Rashbrook (DNA technology Core, University of California - Davis) for their support and assistance with Illumina SNP genotyping, and three anonymous reviewers for comments on an earlier version of this manuscript. This research was supported by NIH grants R21AI062929 and T32AI074550. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIAID or NIH.

REFERENCES

- Bolnick DI, Fitzpatrick BM. Sympatric speciation: Models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*. 2007; 38:459–487.
- Borge T, Lindroos K, Nádvorník P, Syvänen AC, Sætre GP. Amount of introgression in flycatcher hybrid zones reflects regional differences in pre and post-zygotic barriers to gene exchange. *Journal of Evolutionary Biology*. 2005; 18:1416–1424. [PubMed: 16313454]
- Caputo B, Nwakanma D, Jawara M, et al. *Anopheles gambiae* complex along The Gambia river, with particular reference to the molecular forms of *An. gambiae* s.s. *Malaria Journal*. 2008; 7:182. [PubMed: 18803885]
- Caputo B, Santolamazza F, Vicente JL, et al. The “far-west” of *Anopheles gambiae* molecular forms. *PLoS ONE*. 2011; 6:e16415. [PubMed: 21347223]
- Cohuet A, Krishnakumar S, Simard F, et al. SNP discovery and molecular evolution in *Anopheles gambiae*, with special emphasis on innate immune system. *BMC Genomics*. 2008; 9:227. [PubMed: 18489733]

- Costantini C, Ayala D, Guelbeogo W, et al. Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecology*. 2009; 9:16. [PubMed: 19460144]
- Coyne, JA.; Orr, HA. Speciation. Sunderland, Massachusetts, USA: Sinauer Associates; 2004.
- Currat M, Ruedi M, Petit RJ, Excoffier L. The hidden side of invasions: Massive introgression by local genes. *Evolution*. 2008; 62:1908–1920. [PubMed: 18452573]
- Dean JL, Dobson SL. Characterization of *Wolbachia* infections and interspecific crosses of *Aedes (Stegomyia) polynesiensis* and *Ae. (Stegomyia) riversi* (Diptera: Culicidae). *Journal of Medical Entomology*. 2004; 41:894–900. [PubMed: 15535618]
- della Torre A, Tu Z, Petrarca V. On the distribution and genetic differentiation of *Anopheles gambiae* s.s. molecular forms. *Insect Biochemistry and Molecular Biology*. 2005; 35:755–769. [PubMed: 15894192]
- Diabate A, Dabire R, Millogo N, Lehmann T. Evaluating the effect of postmating isolation between molecular forms of *Anopheles gambiae* (Diptera: Culicidae). *Journal of Medical Entomology*. 2007; 44:60–64. [PubMed: 17294921]
- Diabate A, Dao A, Yaro AS, et al. Spatial swarm segregation and reproductive isolation between the molecular forms of *Anopheles gambiae*. *Proceedings of the Royal Society B: Biological Sciences*. 2009; 276:4215–4222.
- Donnelly MJ, Pinto J, Girod R, Besansky NJ, Lehmann T. 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*. 2003; 92:61–68. [PubMed: 14666125]
- Egger B, Mattersdorfer K, Sefc KM. Variable discrimination and asymmetric preferences in laboratory tests of reproductive isolation between cichlid colour morphs. *Journal of Evolutionary Biology*. 2010; 23:433–439. [PubMed: 20002244]
- Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. 2005; 14:2611–2620. [PubMed: 15969739]
- Excoffier L, Foll M, Petit RJ. Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics*. 2009; 40:481–501.
- Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 2010; 10:564–567. [PubMed: 21565059]
- Fanello C, Petrarca V, della Torre A, et al. The pyrethroid knock-down resistance gene in the *Anopheles gambiae* complex in Mali and further indication of incipient speciation within *An. gambiae* s.s. *Insect Molecular Biology*. 2003; 12:241–245. [PubMed: 12752657]
- Favia G, Della Torre A, Bagayoko M, et al. Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Molecular Biology*. 1997; 6:377–383. [PubMed: 9359579]
- Gavrilets S. The Maynard Smith model of sympatric speciation. *Journal of Theoretical Biology*. 2006; 239:172–182. [PubMed: 16242727]
- Gentile G, della T, Maegga B, Powell J, Caccone A. Genetic differentiation in the African malaria vector, *Anopheles gambiae* s.s., and the problem of taxonomic status. *Genetics*. 2002; 161:1561–1578. [PubMed: 12196401]
- Gimonneau G, Pombi M, Choisy M, et al. Larval habitat segregation between the molecular forms of the mosquito *Anopheles gambiae* in a rice field area of Burkina Faso, West Africa. *Medical and Veterinary Entomology*. 2011 In press.
- Gomes B, Sousa C, Novo M, et al. Asymmetric introgression between sympatric molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in the Comporta region, Portugal. *BMC Evolutionary Biology*. 2009; 9:262. [PubMed: 19895687]
- Goudet J. FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*. 1995; 86:485–486.
- Grant PR, Grant BR. Unpredictable Evolution in a 30-Year Study of Darwin's Finches. *Science*. 2002; 296:707–711. [PubMed: 11976447]

- Harris C, Rousset F, Morlais I, Fontenille D, Cohuet A. Low linkage disequilibrium in wild *Anopheles gambiae* s.l. populations. *BMC Genetics*. 2010; 11:81. [PubMed: 20843306]
- Lawniczak MKN, Emrich SJ, Holloway AK, et al. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*. 2010; 330:512–514. [PubMed: 20966253]
- Lehmann T, Hume JCC, Licht M, et al. Molecular evolution of immune genes in the malaria mosquito *Anopheles gambiae*. *PLoS ONE*. 2009; 4:e4549. [PubMed: 19234606]
- Lexer C, Kremer A, Petit RJ. COMMENT: Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology*. 2006; 15:2007–2012. [PubMed: 16689915]
- Martin NH, Willis JH. Geographical variation in postzygotic isolation and its genetic basis within and between two *Mimulus* species. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2010; 365:2469–2478.
- Maynard Smith J. Sympatric speciation. *American Naturalist*. 1966; 100:637–650.
- Mayr E. Ecological factors in speciation. *Evolution*. 1947; 1:263–288.
- Mendes AM, Schlegelmilch T, Cohuet A, et al. Conserved mosquito/parasite interactions affect development of *Plasmodium falciparum* in Africa. *Plos Pathogens*. 2008; 4:e1000069. [PubMed: 18483558]
- Morgan K, Linton Y-M, Somboon P, et al. Inter-specific gene flow dynamics during the Pleistocene-dated speciation of forest-dependent mosquitoes in Southeast Asia. *Molecular Ecology*. 2010; 19:2269–2285. [PubMed: 20444081]
- Morlais I, Poncon N, Simard F, Cohuet A, Fontenille D. Intraspecific nucleotide variation in *Anopheles gambiae*: New insights into the biology of malaria vectors. *American Journal of Tropical Medicine and Hygiene*. 2004; 71:795–802. [PubMed: 15642974]
- Muir G, Schlötterer C. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology*. 2005; 14:549–561. [PubMed: 15660945]
- Neafsey DE, Lawniczak MKN, Park DJ, et al. SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science*. 2010; 330:514–517. [PubMed: 20966254]
- Ng'habi KR, Horton A, Knols BGJ, Lanzaro GC. A new robust diagnostic polymerase chain reaction for determining the mating status of female *Anopheles gambiae* mosquitoes. *American Journal of Tropical Medicine and Hygiene*. 2007; 77:485–487. [PubMed: 17827364]
- Nielsen EE, Bach LA, Kotlicki P. HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Resources*. 2006; 6:971–973.
- Nosil P. Speciation with gene flow could be common. *Molecular Ecology*. 2008; 17:2103–2106. [PubMed: 18410295]
- Obbard DJ, Welch JJ, Little TJ. Inferring selection in the *Anopheles gambiae* species complex: an example from immune-related serine protease inhibitors. *Malaria Journal*. 2009; 8
- Oliveira E, Salgueiro P, Palsson K, et al. High levels of hybridization between molecular forms of *Anopheles gambiae* from Guinea Bissau. *Journal of Medical Entomology*. 2008; 45:1057–1063. [PubMed: 19058629]
- Parmakelis A, Slotman MA, Marshall JC, et al. The molecular evolution of four anti-malarial immune genes in the *Anopheles gambiae* species complex. *BMC Evolutionary Biology*. 2008; 8:79. [PubMed: 18325105]
- Peakall ROD, Smouse PE. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Resources*. 2006; 6:288–295.
- Pennetier C, Warren B, Dabiré KR, Russell IJ, Gibson G. "Singing on the wing" as a mechanism for species recognition in the malarial mosquito *Anopheles gambiae*. *Current Biology*. 2010; 20:131–136. [PubMed: 20045329]
- Pfennig KS. Facultative mate choice drives adaptive hybridization. *Science*. 2007; 318:965–967. [PubMed: 17991861]
- Pinho C, Hey J. Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics*. 2010; 41

- Pritchard J, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945–959. [PubMed: 10835412]
- Riehle MM, Guelbeogo WM, Gneme A, et al. A cryptic subgroup of *Anopheles gambiae* is highly susceptible to human malaria parasites. *Science*. 2011; 331:596–598. [PubMed: 21292978]
- Sakamoto JM, Feinstein J, Rasgon JL. *Wolbachia* infections in the Cimicidae: Museum specimens as an untapped resource for endosymbiont surveys. *Applied and Environmental Microbiology*. 2006; 72:3161–3167. [PubMed: 16672453]
- Santolamazza F, Mancini E, Simard F, et al. Insertion polymorphisms of SINE200 retrotransposons within speciation islands of *Anopheles gambiae* molecular forms. *Malaria Journal*. 2008; 7:163. [PubMed: 18724871]
- Schluter D. Ecology and the origin of species. *Trends in Ecology & Evolution*. 2001; 16:372–380. [PubMed: 11403870]
- Scott J, Brogdon W, Collins F. Identification of single specimens of the *Anopheles gambiae* complex by PCR. *American Journal of Tropical Medicine and Hygiene*. 1993; 49:520–529. [PubMed: 8214283]
- Seehausen OLE, Takimoto G, Roy D, Jokela J. Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*. 2008; 17:30–44. [PubMed: 18034800]
- Simard F, Ayala D, Kamdem G, et al. Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. *BMC Ecology*. 2009; 9:17. [PubMed: 19460146]
- Slotman MA, Tripet F, Cornel AJ, et al. Evidence for subdivision within the M molecular form of *Anopheles gambiae*. *Molecular Ecology*. 2007; 16:639–649. [PubMed: 17257119]
- Tripet F, Toure Y, Taylor C, et al. DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Molecular Ecology*. 2001; 10:1725–1732. [PubMed: 11472539]
- Turner T, Hahn M, Nuzhdin S. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*. 2005; 3:e285. [PubMed: 16076241]
- Turner TL, Hahn MW. Locus- and population-specific selection and differentiation between incipient species of *Anopheles gambiae*. *Molecular Biology and Evolution*. 2007; 24:2132–2138. [PubMed: 17636041]
- Turner TL, Hahn MW. Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology*. 2010; 19:848–850. [PubMed: 20456221]
- van den Boom D, Ehrich M. Discovery and identification of sequence polymorphisms and mutations with MALDI-TOF MS. *Methods in Molecular Biology*. 2007; 2007:287–306. [PubMed: 17568131]
- Via S. Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*. 2009; 106:9939–9946.
- White BJ, Cheng C, Sangare D, et al. The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. *Genetics*. 2009; 183:275–288. [PubMed: 19581444]
- White BJ, Cheng C, Simard F, Costantini C, Besansky NJ. Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology*. 2010; 19:925–939. [PubMed: 20149091]
- White BJ, Hahn MW, Pombi M, et al. Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. *Plos Genetics*. 2007; 3:e217. [PubMed: 18069896]
- Wu C-I. The genic view of the process of speciation. *Journal of Evolutionary Biology*. 2001; 14:851–865.

Fig. 1.
Distribution of molecular forms in Guinea-Bissau. Land cover is coloured according to the University of Maryland Land Cover Scheme of 2008 remote sensing data

Fig. 2.

Clustering analyses of SNP genotype data for M (N=45), S (N=170) and hybrid form samples (H, N=108) as defined by Scott diagnostic SNP⁵⁸¹. a) PCoA analysis identified three clusters: 1 - all M samples; 2 – approximately half of the hybrids; 3 - all S samples and the remaining hybrids. b) Clustering using STRUCTURE. Individuals are represented by columns, with colours showing the proportion of their genome assigned to the different clusters. Shown is the most likely number of clusters, K=2, (Fig. S1a, Supplementary Information), which demonstrates that M and S form largely independent clusters with hybrids exhibiting either an admixed M-S or largely S genotype. Similar results were found at higher K values (Fig. S1b, Supplementary Information)

Fig. 3.

The proportion of S ancestry in hybrids as identified by STRUCTURE. Plotted are the S ancestry proportions of field collected hybrids (x), the values of which have been ranked, and then plotted against those ranks. The shaded area depicts the range of S ancestry proportions expected for F₁ hybrids based on assessments of simulated hybrids, and dotted lines the range for Sx F_{1+n} and Mx F_{1+n} backcrosses (Table S4, Supplementary Information).

Table 1

Observed frequencies of M, S, and hybrid (H) forms at five sites in Guinea-Bissau

Village	Lat	Long	N	M (%)	H (%)	S (%)
Abu (ABU)	11.46144	-15.91411	48	9 (18.75)	17 (35.4)	22 (45.8)
Antula (ANT)	11.91005	-15.58374	124 [#]	17 (13.7)	41 (33.1)	66 (53.2)
Bruce (BRU)	11.30474	-15.83737	60	2 (3.3)	15 (25.0)	43 (71.7)
Eticoga (ETI)	11.15525	-16.14029	88	0 (0)	21 (23.9)	67 (76.1)
Prabis (PRA)	11.80066	-15.74332	98 [#]	29 (29.6)	23 (23.5)	46 (46.9)

[#] SNP typing was conducted on all samples from Abu, Bruce and Eticoga, and 64/98 samples from Prabis, and 63/124 samples from Antula

Table 2

Diversity statistics for the M and S molecular forms, and hybrids (H)

Diversity statistics	M (n=45)	H (n=108)	S (n=170)
R_S	1.722	1.871	1.849
PA (SD)	0	0	0.038 (0.027)
H_o	0.149	0.248	0.199
H_e	0.141	0.226	0.188
F_{IS}	-0.043	-0.093	-0.057

R_S - allelic richness standardized for sample size; PA – private alleles; H_o – Observed heterozygosity; H_e – Expected heterozygosity; F_{IS} – Fixation index