

Breakdown in the Process of Incipient Speciation in *Anopheles gambiae*

Davis C. Nwakanma,* Daniel E. Neafsey,[†] Musa Jawara,* Majidah Adiamoh,* Emily Lund,*
Amabelia Rodrigues,[§] Kovana M. Loua,** Lassana Konate,** Ngayo Sy,** Ibrahima Dia,^{§§}
T. Samson Awolola,*** Marc A. T. Muskavitch,^{*,†,†††} and David J. Conway^{*,†††,1}

*Medical Research Council Laboratories, Banjul, The Gambia, [†]Broad Institute, Cambridge, Massachusetts 02467, [‡]Harvard School of Public Health, Boston, Massachusetts 02115, [§]Bandim Health Project, Bissau, Guinea-Bissau, **National Institute of Public Health, Conakry, Republic of Guinea, ^{††}Universite Cheikh Anta Diop, Dakar, Senegal, ^{†††}Service de Lutte Antiparasitaire, Thies, Senegal, ^{§§}Institut Pasteur, Dakar, Senegal, ***Nigerian Institute of Medical Research, Yaba, Lagos, Nigeria, ^{†††}Boston College, Chestnut Hill, Massachusetts 02467, and ^{†††}London School of Hygiene and Tropical Medicine, London, WC1E 7HT United Kingdom

ABSTRACT Understanding genetic causes and effects of speciation in sympatric populations of sexually reproducing eukaryotes is challenging, controversial, and of practical importance for controlling rapidly evolving pests and pathogens. The major African malaria vector mosquito *Anopheles gambiae sensu stricto* (s.s.) is considered to contain two incipient species with strong reproductive isolation, hybrids between the M and S molecular forms being very rare. Following recent observations of higher proportions of hybrid forms at a few sites in West Africa, we conducted new surveys of 12 sites in four contiguous countries (The Gambia, Senegal, Guinea-Bissau, and Republic of Guinea). Identification and genotyping of 3499 *A. gambiae* s.s. revealed high frequencies of M/S hybrid forms at each site, ranging from 5 to 42%, and a large spectrum of inbreeding coefficient values from 0.11 to 0.76, spanning most of the range expected between the alternative extremes of panmixia and assortative mating. Year-round sampling over 2 years at one of the sites in The Gambia showed that M/S hybrid forms had similar relative frequencies throughout periods of marked seasonal variation in mosquito breeding and abundance. Genome-wide scans with an Affymetrix high-density single-nucleotide polymorphism (SNP) microarray enabled replicate comparisons of pools of different molecular forms, in three separate populations. These showed strong differentiation between M and S forms only in the pericentromeric region of the X chromosome that contains the molecular form-specific marker locus, with only a few other loci showing minor differences. In the X chromosome, the M/S hybrid forms were more differentiated from M than from S forms, supporting a hypothesis of asymmetric introgression and backcrossing.

THE major malaria vector mosquito *Anopheles gambiae sensu stricto* (s.s.) exists throughout most of Sub-Saharan Africa, but there are many polymorphisms including chromosomal inversions that appear to be involved in the adaptation of subpopulations to different environments (Coluzzi *et al.* 2002), and molecular forms (M and S) have been identified that appear to be reproductively isolated (della Torre *et al.* 2001). The S form is distributed widely throughout the *A. gambiae* species range, whereas the M form is common but

restricted to western parts of Africa, and hybridization between them is rare in most areas of sympatry (Della Torre *et al.* 2005). Scanning of large numbers of polymorphic markers initially indicated the molecular forms were very highly differentiated at only a small number of discrete regions within the two autosomes and the pericentromeric region of the X chromosome, which contains the form-specific markers in the ribosomal (r)RNA genes (Turner *et al.* 2005). However, a finding that transposable element insertion sites showed marked differentiation between the forms indicated that reproductive isolation might affect a larger portion of the genome (Esnault *et al.* 2008). A high-density array to type 400,000 single-nucleotide polymorphisms (SNPs) has recently enabled more thorough genome-wide analyses, confirming that the molecular forms are differentiated at many loci throughout the genome in addition to those previously shown (Lawniczak *et al.* 2010; Neafsey *et al.* 2010; Reidenbach *et al.* 2012).

Copyright © 2013 by the Genetics Society of America
doi: 10.1534/genetics.112.148718

Manuscript received September 19, 2012; accepted for publication January 13, 2013
Available freely online through the author-supported open access option.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.148718/-/DC1>.

¹Corresponding author: Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, Keppel St., London WC1E 7HT, United Kingdom.
E-mail: david.conway@lshtm.ac.uk

It is vital to understand such differentiation and genetic subdivision and its importance in vector evolution, as this complexity can affect malaria control, including resistance to insecticides (Dabire *et al.* 2009; Djogbenou *et al.* 2010; Lynd *et al.* 2010) and susceptibility to malaria parasites and other infections (Rottschaefer *et al.* 2011; White *et al.* 2011). Differences in the adult mosquito transcriptome between the M and S molecular forms appear to be minimal (Cassone *et al.* 2008; Aguilar *et al.* 2010), but evidence is accumulating that the larval stages are differentially adapted to particular features of breeding sites, with the M forms being generally more common in large areas of irrigation for crop cultivation (Diabate *et al.* 2008; Gimonneau *et al.* 2011). Studies of mosquito behavior suggest assortative mating may be the means of preventing natural hybridization between the forms, as most mating swarms consisted of single forms in an area of sympatry in Mali (Diabate *et al.* 2009), although some mixed swarms were seen in Burkina Faso (Diabate *et al.* 2006), and very close-range barriers to interbreeding may also operate (Pennetier *et al.* 2010; Sanford *et al.* 2011).

Surveys of populations in the extreme west of Africa show substantial local variation in relative frequencies of *A. gambiae* s.s. M and S molecular forms (della Torre *et al.* 2005). For example, the M form predominates in most of The Gambia, but <100 km farther inland in eastern Senegal almost all members of the species are S form (Caputo *et al.* 2008). Of particular interest are areas in which both forms exist, as the occurrence of natural hybrids may be possible. Elsewhere in Africa, areas of sympatry consistently exhibit very low frequencies of hybrid forms, but a few years ago exceptionally high proportions of M/S hybrids were reported at a few sites near the west coast, with 3% at Dielmo in Senegal (Ndiath *et al.* 2008), 7% at Njabakunda in The Gambia (Caputo *et al.* 2008), and 24% at Antula in Guinea-Bissau (Oliveira *et al.* 2008). Analysis of an insertion polymorphism linked to the form-specific marker on the X chromosome, as well as an unlinked SNP on chromosome 3, in a subset of the samples from The Gambia and Guinea-Bissau suggested that there may be introgression between the forms (Caputo *et al.* 2011), and separate studies using broader scans of genomic polymorphisms in samples from Guinea-Bissau have supported this inference (Marsden *et al.* 2011; Weetman *et al.* 2012).

To thoroughly discover and map the area with high frequencies of hybrids and assess its impact on genomic differentiation, we conducted new surveys of diverse sites in the coastal areas of four contiguous countries (The Gambia, Senegal, Guinea-Bissau, and Republic of Guinea). To investigate the stability of molecular form and hybrid frequencies over time throughout different seasons, we conducted a longitudinal survey over a 2-year period in the Njabakunda area of The Gambia. We performed high-density genome-wide scans for differentiation between sympatric M and S forms, as well as M/S hybrids, at the Njabakunda site and at single sites from each of Senegal and Guinea-Bissau. We

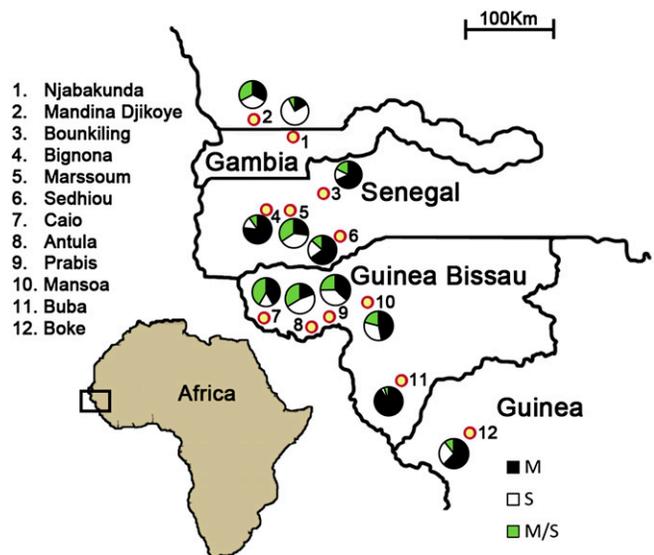


Figure 1 Map of sampling locations for surveys of *Anopheles gambiae* s.s. molecular forms between 2007 and 2009 in The Gambia, Senegal, Guinea-Bissau, and Republic of Guinea. Each location is indicated by a yellow circle, and the frequencies of the M form, the S form and M/S hybrid forms sampled at each are indicated by an adjacent pie chart (sample sizes and exact frequencies are given in Table 1).

employed a pooled hybridization approach for expediency, allowing use of a genotyping array with much higher marker density than that of previous surveys for introgression in West Africa (Marsden *et al.* 2011; Weetman *et al.* 2012). Results show extensive hybridization throughout the area and that mosquitoes with M and S form-specific markers are largely homogenized throughout the genome, with evidence supporting a hypothesis of asymmetric hybridization and frequent backcrossing particularly between the M/S hybrid forms and the S forms.

Materials and Methods

Sampling sites for *A. gambiae* mosquitoes

Twelve sites were surveyed in four countries (Figure 1). These were chosen to encompass the few areas where unusually high frequencies of M/S hybrids had previously been noted and to extend the sampling to the north and south of these, within 100 km of the west coast of Africa. The northern part of the sampling range has Sudan Savannah-type vegetation, with most annual rainfall of 600–1000 mm normally occurring within 4 months (June to September). The south of the sampling range has Guinea Savannah-type vegetation including more woodland, with annual rainfall of up to 2000 mm extending over a longer season (May to October). In all areas, the majority of *A. gambiae* s.s. breeding occurs during and at the end of the annual rains, and highest densities of adult mosquitoes are found between August and October. The 12 sites are numbered and described below (sites 1 and 2 represent Sudan Savannah, sites 3–6 represent northern Guinea Savannah, and sites 7–12 represent southern Guinea Savannah).

Table 1 Proportions of M/S hybrid heterozygote forms and M and S homozygote forms of *Anopheles gambiae* s.s. in each of 12 sites surveyed from four West African countries

Country	Site	No. mosquitoes	M	S	M/S	H_{exp}	F_{IS}
Gambia	Njabakunda	1474	0.164	0.760	0.076	0.322	0.76
Senegal	Madina Djikoye	77	0.325	0.350	0.325	0.500	0.35
Senegal	Boukiling	145	0.683	0.138	0.179	0.354	0.49
Senegal	Bignona	30	0.767	0.133	0.100	0.295	0.66
Senegal	Marsassoum	434	0.270	0.387	0.343	0.493	0.30
Senegal	Sedhiou	389	0.653	0.211	0.136	0.402	0.66
Guinea-Bissau	Caio	12	0.417	0.167	0.417	0.469	0.11
Guinea-Bissau	Antula	464	0.194	0.470	0.336	0.461	0.27
Guinea-Bissau	Prabis	91	0.363	0.385	0.253	0.500	0.49
Guinea-Bissau	Buba	194	0.907	0.041	0.052	0.130	0.60
Guinea-Bissau	Mansoa	152	0.467	0.322	0.210	0.490	0.57
Guinea	Boke	37	0.622	0.270	0.108	0.435	0.75

H_{exp} , expected proportion of M/S heterozygotes under Hardy–Weinberg equilibrium, assuming random mating. There was a significant heterozygote deficiency ($P < 0.005$) in each of the populations except for Caio. F_{IS} , inbreeding coefficient (proportion of heterozygotes missing compared to expectations under random mating).

The Gambia: Site 1, the Njabakunda village area (13°33'N, 15°54'W) in the North Bank region of the country (30 km west of Farafenni town), was sampled, as previous surveys (in 2005 and 2006) had shown this area to contain a higher frequency of M/S form hybrids than elsewhere in The Gambia (Caputo *et al.* 2008). The site is ~4 km away from the Gambia River on free-draining laterite soil, covered with open woodland savannah and farmland mainly for cultivation of subsistence and cash crops. Four hamlets (Maria Samba Nyado, Sare Illo Buya, Kerr Birom Kardo, and Kerr Sama Kuma) around Njabakunda and within 1 km of each other were selected for longitudinal sampling for a 2-year period between April 2007 and March 2009. Fortnightly (mid-month and end of the month) collections were conducted from June to December covering the rainy season period, and monthly (end of the month) collections were done from January to May covering the dry season.

Senegal: Five sites were sampled in Senegal. One of these (site 2), Madina Djikoye (clustered with nearby villages Kerr Samba Gueye and Kerr Ousainou Dieng) (13°37'N, 16°18'W) is in the Kaolack region immediately to the north of the Gambian border. The other four sites are south of The Gambia in the lower Cassamance region: (site 3) Sedhiou (12°43'N, 15°4'W) adjacent to the Cassamance River, (site 4) Boukiling (13°2'N, 15°42'W), (site 5) Bignona (12°47'N, 16°14'W), and (site 6) Marsassoum (12°50'N, 15°58'W). The sites have free-draining laterite sand or alluvial soil and are covered with open woodland savannah or farmland for groundnut and cereal cultivation. These sites were sampled in August 2009, except for Sedhiou, which was sampled in October 2008.

Guinea-Bissau: Five sites were sampled in Guinea-Bissau: (site 7) Caio village (11°55'N, 16°12'W) is in a coastal area of rice and cashew plantations mixed with woodland with traditional ritual significance, (site 8) Antula (11°53'N, 15°35'W) is a village east of the capital Bissau, (site 9) Prabis

(11°48'N, 15°44'W) is a village in a peri-urban area of Bissau with marsh vegetation and rice cultivation, (site 10) Mansoa (12°4'N, 15°19'W) is a small town in the north of the country in a rice cultivation area, and (site 11) Buba (11°35'N, 15°0'W) is in the south of the country on the River Rio Grande de Buba near the National Park Contanhez containing a canopy of tropical rain forest, within which the village of Banta Furu was sampled. These sites were sampled in August 2009, and Prabis was also sampled in October 2008.

Republic of Guinea: One area in Guinea was sampled in August 2009, to the south of the border with Guinea-Bissau: (site 12) near Boke town (10°56'N, 14°18'W) four villages (Korera, Dabaya, Balangdoungou, and Bintoumodia) in terrain of flat plains interspersed with hilly slopes divided by fast-flowing streams, fed by a higher rainfall than at any of the other sites studied here.

Collection and identification of mosquitoes

Mosquitoes were sampled from houses or inhabited huts in each site, using indoor pyrethrum spray collections (PSC) in most cases for all the sites, with some sampling by indoor overnight light trap collections (LTC) in addition to PSC in Guinea-Bissau and Republic of Guinea sites. All mosquitoes that were collected were identified morphologically in the field, and *A. gambiae sensu lato* specimens were stored individually in 1.5-ml polyethylene tubes with desiccant or in Carnoy's solution if they were semigravid, for subsequent DNA typing of species and molecular forms. DNA was extracted from individual mosquito specimens, using the Corbett Robotics X-Tractor Gene automated DNA extraction platform (QIAGEN, Crawley, UK) according to the manufacturer's protocol for tissue extraction. Whole mosquito material was extracted, except for a few legs of each specimen retained for confirmatory PCR testing in case the initial result indicated a hybrid molecular form. Each of the three species of the *A. gambiae* complex that exist in West Africa (*A. gambiae* s.s., *A. arabiensis*, and *A. melas*) were discriminated in the laboratory

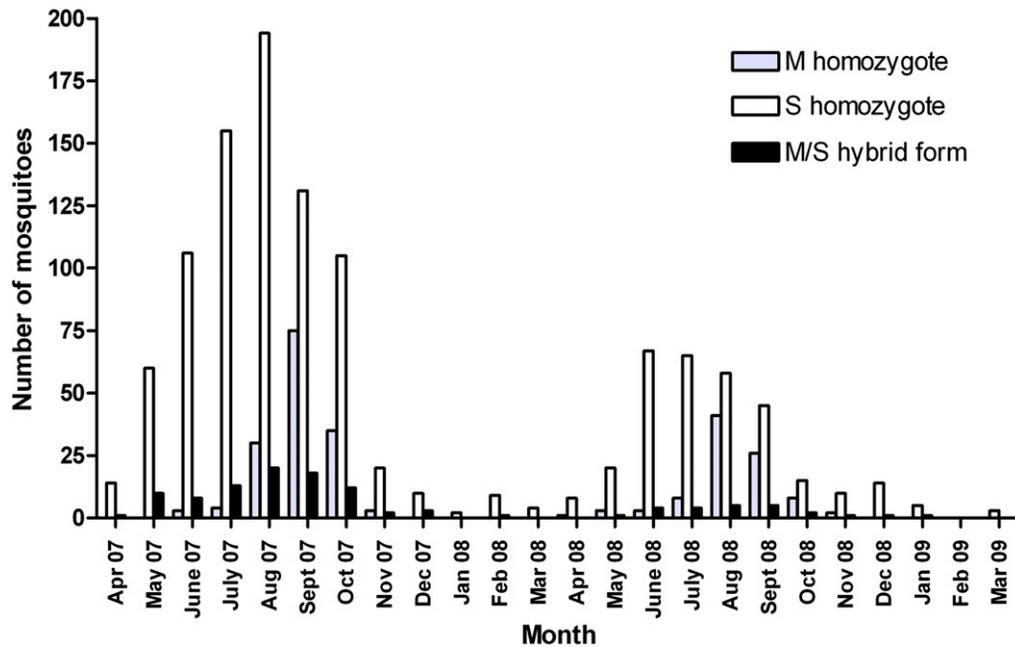


Figure 2 Numbers of molecular forms M and S and M/S hybrid forms sampled in each month from a total of 1474 *Anopheles gambiae* s.s. individuals collected in the Njabakunda area of The Gambia over a 2-year period from April 2007 until March 2009.

by species-specific PCR of sequences within the rRNA gene locus on the X chromosome (Scott *et al.* 1993), and *A. gambiae* s.s. mosquitoes were differentiated into molecular forms M and S (and M/S hybrids) by form-specific restriction enzyme digestion of the amplified fragment (Fanello *et al.* 2002).

Genome-wide scans for differentiation between pools of *A. gambiae* s.s. mosquitoes

A custom Affymetrix oligonucleotide array for genome-wide typing of 400,071 *A. gambiae* s.s. SNPs (Neafsey *et al.* 2010), termed the 400K array, was used for population genotyping. DNA hybridization was performed with separate pooled samples of 20 homozygous M-form and 20 homozygous S-form *A. gambiae* s.s. mosquitoes from each of three sites sampled (Njabakunda in The Gambia, Sedhiou in Senegal, and Antula in Guinea-Bissau). The Njabakunda site was the primary one selected for this genome-wide SNP analysis, as we studied this site intensively. Therefore, it had the largest sample size and most accurate estimation of frequencies. We added the other two sites to include one from Senegal and one from Guinea-Bissau, each of which had large sample sizes with higher proportions of hybrids. In addition, a pool of DNA from 20 M/S hybrid mosquitoes was tested and compared with the homozygous samples, for each of these three sites. Each of the 180 individual mosquitoes that contributed to the pooled arrays (20 M form, 20 S form, and 20 M/S forms, selected from each of the three sites noted) was also genotyped for the SINE200x6.1 transposon insertion site polymorphism that is linked to the molecular form marker on the X chromosome (Santolamazza *et al.* 2011). Affymetrix oligonucleotide array hybridization was performed following methods already described (Neafsey *et al.* 2010), as previous analysis revealed a very high correlation between aggregate

individual hybridizations and pooled hybridizations for pools of 20 individuals (Pearson's $r^2 = 0.96$, Supporting Information, Figure S1), leading to the repetition of that assay design here. Following Robust Multi-array Average (RMA) background correction and quantile normalization using Affymetrix Power Tools, allelic balance at assayed loci within the pools was inferred using the Contrast statistic generated by BRLMM-P, using a K value of 1 (Rabbee and Speed 2007). Such plots result in a modal background mean Contrast difference of ~ 0.25 rather than 0, which is the result of noise in the signal rather than a genome-wide divergence signal. Plots of local divergence along the chromosomes were generated using a stepping window approach, by calculating the mean of the absolute value of the arithmetic difference in Contrast for 50 adjacent assays. The mean physical distance spanned by 50 adjacent assays was 28.2 kb (range = 9.1–280 kb, standard deviation = 16.6 kb). Thresholds for statistical significance ($\alpha = 0.05$ after Bonferroni correction) for these windowed analyses were calculated using a bootstrapping approach, in which mean Contrast difference was calculated for each pairwise comparison for 50 SNPs that were sampled randomly from the genome a total of 10 million times.

Results

Quantifying proportions of molecular forms and hybrids at different sites

Molecular form typing was performed on a total of 3499 *A. gambiae* s.s. mosquitoes sampled from the 12 sites surveyed in the four countries (Senegal, The Gambia, Guinea-Bissau, and Republic of Guinea) (Figure 1). In each site, M and S molecular forms coexisted sympatrically and substantial proportions of M/S hybrid forms were found (from 5 to 42%) (Table 1). The

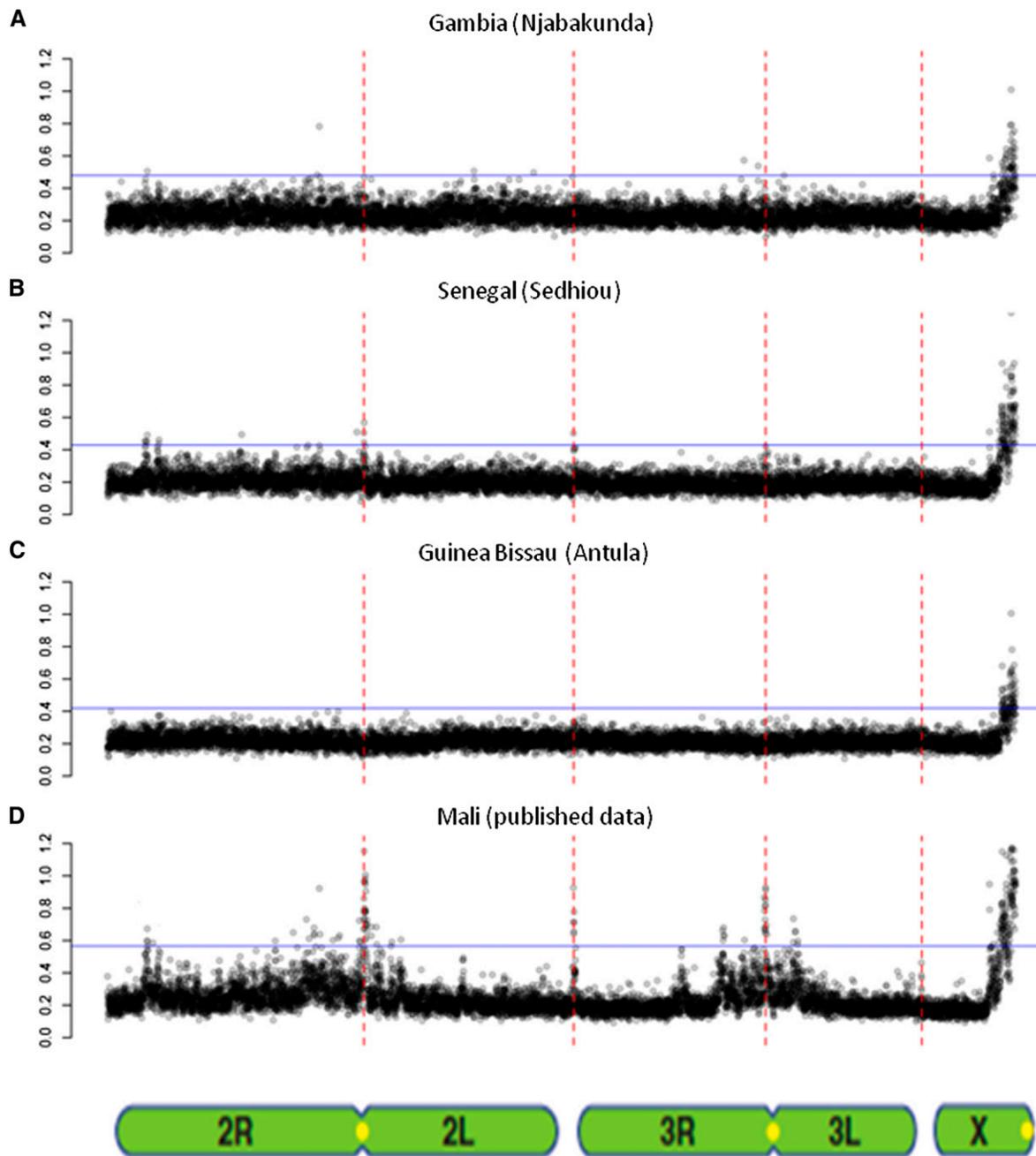


Figure 3 Test for genomic differentiation between pooled DNA from samples of 20 homozygous M-form and 20 homozygous S-form *Anopheles gambiae* s.s. from each of three sites in this study: (A) Njabakunda in The Gambia, (B) Sedhiou in Senegal, and (C) Antula in Guinea-Bissau. The vertical axis represents divergence measured as the absolute value of mean Contrast difference for stepping windows of 50 assays (this has a modal background of ~ 0.25 that is due to noise rather than a genome-wide divergence signal), using data from all SNPs on the 400K genome-wide array. The horizontal blue lines indicate the Bonferroni-corrected threshold for statistical significance obtained via bootstrapping. Only the X-chromosomal pericentromeric region that contains the M and S molecular form markers exhibits differentiation. For comparison, D shows differentiation between similarly pooled DNA from M and S forms in Mali (Neafsey *et al.* 2010).

estimated inbreeding coefficients for this rRNA gene locus, which has previously been considered a putative speciation marker, showed a wide spectrum of values (from 0.11 to 0.76) (Table 1). Seven of the sites each had samples of >100 mosquitoes analyzed, allowing highly accurate estimations of frequencies, and we note that these also show a wide range of M/S hybrid proportions (from 5 to 34%) and estimated inbreeding coefficients (from 0.27 to 0.76). Thus, the pro-

portions of M/S hybrids were less than expected under Hardy-Weinberg equilibrium with complete panmixia, but in some cases not much less.

Analysis of proportions of molecular forms and hybrids at one site throughout a 2-year period

To investigate the stability of frequencies over time, we sampled the Njabakunda site in The Gambia throughout a 2-year period

Table 2 Number of 50-SNP windows exhibiting significant ($P < 0.05$) mean contrast differences according to pairwise pool comparison and chromosomal location

Total windows	2R:	2L:	3R:	3L:	X:	Chi-square P -value: X vs. autosomes (compared with Mali)
	$n = 2256$	$n = 1846$	$n = 1692$	$n = 1374$	$n = 829$	
Gambia M vs. S	3	2	2	1	42	1.70×10^{-04}
Senegal M vs. S	9	3	1	0	98	7.17×10^{-11}
Guinea-Bissau M vs. S	0	0	0	0	58	7.20×10^{-11}
Mali M vs. S	20	30	21	11	112	Comparator
Gambia M vs. M/S	5	3	1	0	37	1.83×10^{-03}
Gambia S vs. M/S	0	1	0	0	0	Not applicable

(annual periods extending from April to the following March to capture full seasonal dynamics due to rainfall), yielding 1474 typed mosquitoes from this site (1048 in the first annual period and 426 in the second). The relative frequencies of M/S hybrid forms were stable in both years, accounting for 8.3% in the first and 5.6% in the second year (chi-square test, $P = 0.07$) (Figure 2). During each year, S-form mosquito numbers peaked early in the rainy season, and the peak of the less abundant M form occurred later in the season, but M/S hybrid forms were seen throughout. In 2007, M/S forms had seasonally stable frequencies of 8.6% (32/374) in April–July and 8.1% (50/620) in August–October (chi-square test, $P = 0.78$), while proportions of M forms increased from 1.9% (7/374) to 22.6% (140/620) during those respective periods (chi-square test, $P < 10^{-7}$); in 2008, M/S forms had frequencies of 4.9% (9/184) in April–July and 5.9% (12/205) in August–October (chi-square test, $P = 0.67$), while proportions of M forms increased from 8.1% (15/184) to 36.6% (75/205) in those periods (chi-square test, $P = 10^{-7}$).

Genome-wide scans for differentiation between molecular forms within sympatric samples

To identify whether marked genomic differentiation between M and S forms could be seen, we randomly selected 20 homozygous M-form and 20 homozygous S-form mosquitoes from the Njabakunda population for a pooled-population genome-wide scan of single-nucleotide polymorphism differentiation, using the custom Affymetrix 400K array. Remarkably, with the clear exception of the X chromosome pericentromeric region that contains the specific polymorphism used to stratify the molecular forms, there were no loci that exhibited strong differentiation between the M and S molecular forms (Figure 3A). To replicate the test, genome-wide comparison between M and S forms was conducted for pools of 20 M-form and 20 S-form mosquitoes from each of two other sites that also exhibit high frequencies of M/S hybrids, at Sedhiou in Senegal (Figure 3B) and Antula in Guinea-Bissau (Figure 3C). Both of these sites similarly showed a lack of substantial differentiation except in the X-chromosomal pericentromeric region. This contrasts strongly with the observation of differentiation at pericentromeric autosomal loci and elsewhere throughout the genome of M and S forms sampled from

other African populations (Figure 3D shows published data from Mali for comparison). A summary of the number of nonoverlapping windows of 50 adjacent SNPs from each chromosomal arm exhibiting significant differentiation in each pairwise comparison is presented in Table 2. M vs. S comparisons for each of the three separate sites exhibit a significantly higher ratio of divergent windows on the X chromosome compared to the autosomes, relative to the more canonical interform divergence profile exhibited by the Mali comparison (chi-square tests on the ratio at each of the sites compared with Mali: $P = 1.7 \times 10^{-4}$ for the Gambian site and $P < 10^{-10}$ for the sites in Senegal and Guinea-Bissau) (Table 2). (Table S1 includes the Contrast Difference datafile).

To test whether minor signals of differentiation may have been being obscured by our windowed analysis approach, we examined the region between 28 and 42 Mb on chromosome 2R in closer detail, at the level of individual SNP assays, for the comparison between M- and S-form pools in Guinea-Bissau (Figure S2). This region was focused on as it contains loci that showed differentiation between the M and S forms in a recent study, using a panel of several hundred SNPs with a sample from Guinea-Bissau (Weetman *et al.* 2012). Although some individual SNP assays in the region showed high Contrast differences between the pools, the Contrast distributions were similar to those observed in a control region of chromosome 3R (19–32 Mb) and much lower than those observed in the highly differentiated X centromeric region (Figure S2). To test whether signals of differentiation might have been obscured by noise due to poorly performing SNP genotyping on the array, we repeated the genome-wide analysis with a subset of 65,974 SNPs that had previously been validated and shown to be in Hardy–Weinberg equilibrium for individual molecular forms in Mali (Neafsey *et al.* 2010). This confirmed that, outside of the pericentromeric X-chromosomal region, there was no major differentiation between the forms in the populations studied here, in contrast to the interform differentiation previously seen in Mali (Figure S3). We cannot fully reconcile this observation with reports of very narrow regions of differentiation on 2R in another population sample from Guinea-Bissau (Weetman *et al.* 2012), as differences in findings between the studies may result from fine differences in spatial or temporal sampling or in the sensitivity of the respective assays.

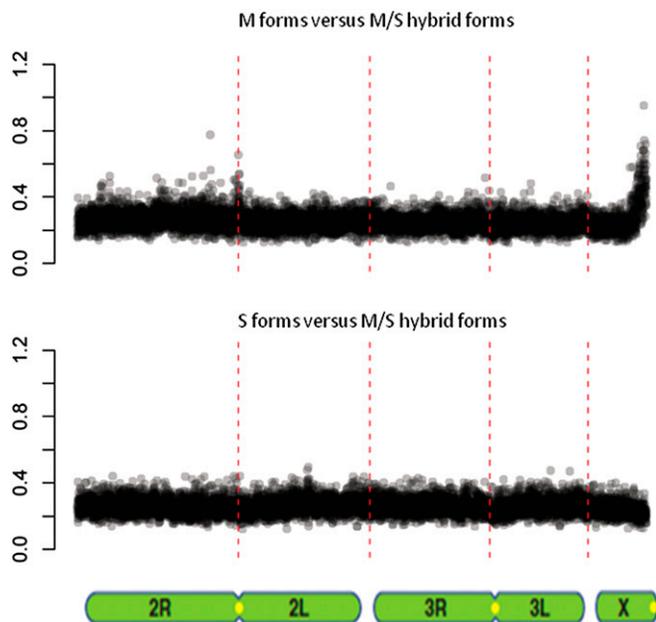


Figure 4 Genome-wide comparison of an M/S pool of 20 hybrid form mosquitoes with M and S homozygous pools from Njabakunda in The Gambia, using data from all SNPs on the 400K genome-wide array: (top) M homozygotes vs. M/S hybrid forms; (bottom) S homozygotes vs. M/S hybrid forms. The more complete lack of differentiation between S and M/S suggests asymmetric gene flow between the forms or selection to maintain the residual differentiation in the M form. (Replicate comparisons between the pooled hybrid forms and each of the major forms are shown for two separate populations in Figure S4.)

Asymmetric minimal differentiation between hybrid forms and homozygous molecular forms

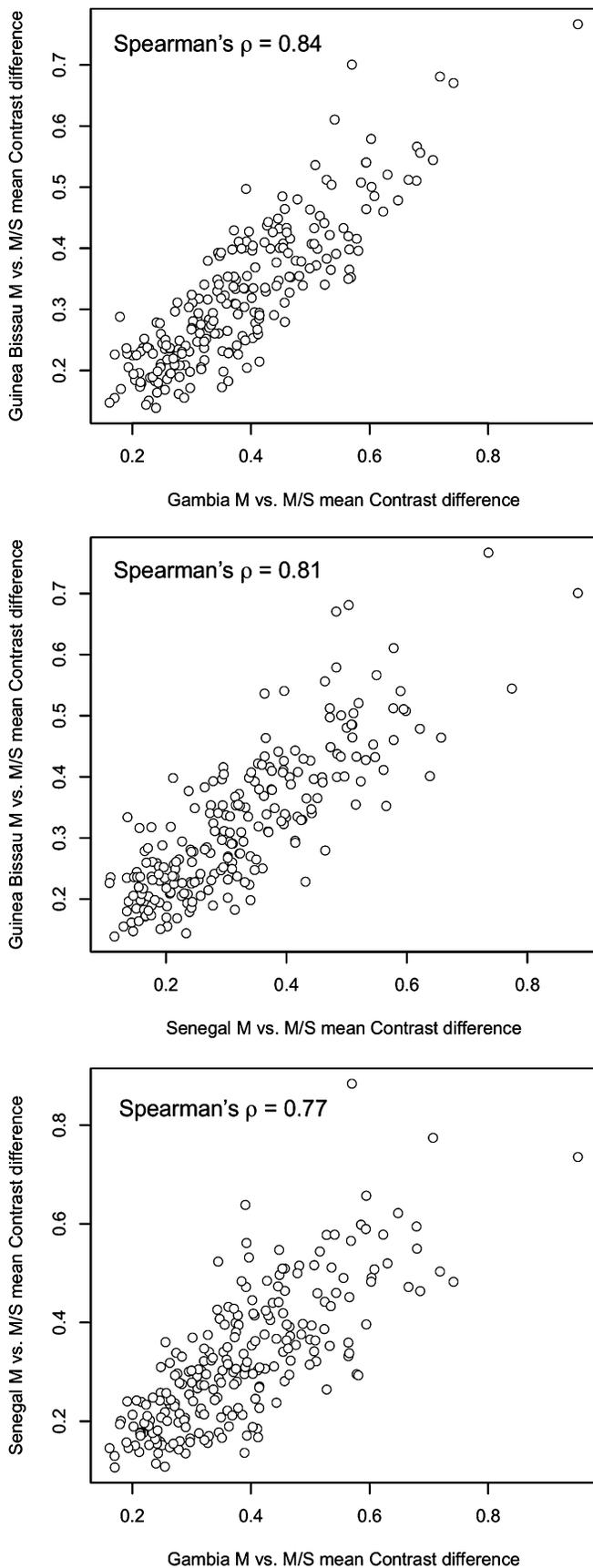
We also performed pooled array hybridizations, using 20 mosquitoes genotyped as M/S heterozygote “hybrid forms” from each site and compared these with the M and S homozygote pools. Results for the Gambian population sample (Figure 4, Table 2) show no divergence in the X pericentromeric region between S and M/S, whereas some divergence remains between M and M/S. The degree of differentiation observed in the X pericentromeric region (megabases 15–24.2) is significantly higher in the M vs. M/S comparison relative to the S vs. M/S comparison (Wilcoxon’s sign-rank test, $V = 25,996$, $P = 2.2 \times 10^{-16}$). Replication of these comparisons for samples from Sedhiou in Senegal and Antula in Guinea-Bissau showed similar results (Figure S4), except that the S and M/S pools showed some detectable X pericentromeric differentiation in Senegal (less than observed for the comparison of M and M/S). This suggests that although the forms are now highly homogenized, the speed of the process has been asymmetric, with partial isolation or a lower recombination rate allowing maintenance of greater X chromosome pericentromeric divergence in the M form. Genotyping of the SINE200 transposon insertion site polymorphism in the 180 mosquitoes contributing to the pooled arrays (20 M form, 20 S form, and 20 M/S forms, in each of the three selected sites) gave

results that were concordant with the molecular form typing in 159 (88%) of the individuals. While there was 98% (59/60) concordance for the M form and 100% (60/60) for the S form, confirming appropriate sample selection and robust comparison of the forms here, the lower concordance of 67% (40/60) observed for the M/S forms confirms that many of these are not F₁ hybrids, but that backcrossing must have occurred over multiple generations. The profile of windows with significant differentiation in M vs. M/S comparisons is highly similar among the three countries sampled (Figure 5). Across the pairs of countries, M vs. M/S comparisons exhibit Spearman’s rank correlations ranging between 0.77 and 0.84, whereas S vs. M/S comparisons have Spearman’s rank correlations of only 0.15–0.42 between the same pairs (Table S1).

Discussion

It is widely considered that the major malaria vector *A. gambiae* s.s. is undergoing a process of speciation into at least two reproductive units defined as the M and S molecular forms (della Torre *et al.* 2001). Although the divergence of these forms is likely proceeding in many parts of west and central Africa (Reidenbach *et al.* 2012), our results and those of other recent surveys of populations in the extreme west of Africa (Caputo *et al.* 2011; Marsden *et al.* 2011; Oliveira *et al.* 2008; Weetman *et al.* 2012) indicate that hybridization is common. A contiguous region is now outlined, covering ecologically diverse sites sampled in a contiguous area within 100 km of the West African coast within four countries in a zone of <400 km from north to south, where frequencies of M/S hybrid forms are consistently much higher than elsewhere, and there is minimal interform genomic differentiation outside of the pericentromeric region of the X chromosome. This contrasts significantly with the results obtained from populations farther east, using the same methods (Reidenbach *et al.* 2012), and is important for understanding evolution and implementing control of this malaria vector and for investigation of the processes of speciation.

A wide range of proportions of M/S hybrid forms was detected within both Guinea-Bissau and Senegal, with the Gambian site in the north and the Guinean site in the south both showing lower proportions than many of the intervening sites. It appears likely that the area we have surveyed, extending across these four countries, probably covers most of the current geographical range of the *A. gambiae* s.s. populations that exhibit extensive hybridization. Surveys immediately to the south in Guinea have shown mostly S-form mosquitoes near the coast, and both M and S forms, with a very low frequency of hybrids, further inland (della Torre *et al.* 2005; Vezzenegho *et al.* 2009; Carnevale *et al.* 2010). In northern Senegal, *A. gambiae* s.s. is less common than the related vector species *A. arabiensis* and exists mostly as the M form (della Torre *et al.* 2005; Dia *et al.* 2008), while surveys in the eastern part of Senegal have shown the S



form to be more abundant than the M form, with hybrid forms very rarely detected (della Torre *et al.* 2005; Caputo *et al.* 2008), as has been observed in Mali and further eastward (della Torre *et al.* 2005; Diabate *et al.* 2009; Aboagye-Antwi *et al.* 2010).

It is possible that our results reflect consequences of the reversal of a previous process that promotes speciation, with the genomic integrity of one or both of the molecular forms now being substantially compromised by introgression within these western populations. It was previously suggested that areas might exist within which there is extensive gene flow between the molecular forms of *A. gambiae s.s.*, while reproductive barriers are maintained at other sites (Black and Lanzaro 2001; Lehmann *et al.* 2003). A more recent view, also supported by pooled array hybridizations, is that sporadic bouts of hybridization mediated by changing ecological conditions or a very low background rate of gene flow between M and S forms could be responsible for the relative homogeneity of M and S genomes outside of the “speciation islands” (Reidenbach *et al.* 2012). Such processes would not be expected to disrupt the divergence at well-established pericentromeric islands of speciation, however. Our results enable sites to be specifically selected for population genomic studies to investigate mechanisms of reproductive isolation. Studies of male mating swarm composition and behavior will be important to conduct in these areas, to identify whether there are proximal determinants of hybridization (Diabate *et al.* 2006, 2009). Further sampling and measurement of multiple ecological parameters at different times of the year are also needed to characterize these and additional sites more fully, which could enable exploration of correlations between environmental variables, population genetic structure, and reproductive isolation.

If the M and S forms were undergoing genome-wide reciprocal introgression, it is expected that the pericentromeric divergence on the X chromosome should be gradually eliminated in both forms by recombination. The observation that the X pericentromeric region remains relatively distinct in M-form populations in this area, as also reported recently by others sampling in Guinea-Bissau (Marsden *et al.* 2011), supports a hypothesis that partial reproductive isolation or selective pressure preserves a narrow genomic island linked to the M-form-specific marker (Caputo *et al.* 2011; Marsden *et al.* 2011; Weetman *et al.* 2012). A high degree of similarity in the specific pericentromeric divergence profile among sites (Figure 5, Table S2) further indicates the likely role of

Figure 5 Comparison of M vs. M/S divergence among geographic sites in the X pericentromeric region (megabases 15–24.2). Each point represents a 50-SNP window. The profile of differentiation is maintained in a highly similar manner in all pairwise comparisons of geographic sites: (top) Gambia vs. Guinea-Bissau, (middle) Senegal vs. Guinea-Bissau, and (bottom) Gambia vs. Senegal. The locations sampled in each country for these comparisons were Njabakunda (Gambia), Sedhiou (Senegal), and Antula (Guinea-Bissau).

selection in limiting gene flow in this pericentromeric region. This also illustrates that mosquitoes typed as M/S heterozygotes in our study generally do not represent F₁ hybrids between the molecular forms, but are members of a thoroughly backcrossed and homogenized population, together with S/S homozygotes. Elsewhere in the range of *A. gambiae* s.s., major introgression between M- and S-form mosquitoes appears to have been very rare, although it is strongly suggested by the distribution of insecticide resistance genotypes. For example, the *kdr* mutation conferring resistance to pyrethroids has apparently introgressed from the S form into the M form (Weill *et al.* 2000; Diabate *et al.* 2003), and the *ace-1(R)* mutation has been observed in both M- and S-form populations in several West African countries, although it is unclear in which population this latter mutation originated (Djogbenou *et al.* 2008).

Reinforcement might occur during the speciation process by strengthening premating isolation mechanisms, in a situation where postmating isolation is already operating due to reduced fitness of hybrids (Wallace 1889). In areas where M and S forms have long been in contact, it is possible that they have evolved premating isolation mechanisms to reduce the frequency of hybridization, such as selective swarm formation (Diabate *et al.* 2009) or male–female flight-tone matching (Pennetier *et al.* 2010). As the S form is more widespread, it is possible that S-form populations that have had no contact with M populations would have had no opportunity to evolve such reinforcement. If M-form populations are gradually expanding or shifting their range over time due to ecological changes and carrying reinforcement mechanisms developed through preceding contact with members of S-form populations, then it may be expected that the M form would be generally less susceptible to introgression. The demographic and geographic range history of the forms is currently unclear, however. Behavioral mechanisms, such as the relative importance of male or female mate choice in determining prezygotic isolation, could also play an important role. Further investigation of the ecology and range-wide demographic history of these forms may be required to understand the nature of the observed introgression.

Reproductive isolation in this most important African malaria vector species is complex and might not orient primarily around the commonly typed M and S forms within *A. gambiae* s.s. There is evidence of some genetic substructure within each of the forms (Slotman *et al.* 2007; Weetman *et al.* 2010), of ecologically important variation and genomic divergence due to inversion polymorphisms that are not routinely typed (Coluzzi *et al.* 2002; White *et al.* 2007, 2009), and of some discrepancies among different methods used for molecular typing of the forms (Santolamazza *et al.* 2011). Recent findings even suggest that there may be other subpopulations of *A. gambiae* s.s. that have not yet been adequately sampled (Riehle *et al.* 2011). Implications of our findings for the evolution and spread of genetic traits relevant to malaria control will warrant further investigation, within the known geographical region of hybridization and through-

out the species range. The application of whole-genome resequencing to such studies promises to further increase the power to map loci under selection and track ongoing changes in population genetic structure (Cheng *et al.* 2012).

Acknowledgments

We thank the staff of our institutes who were involved in entomological field collections, particularly Sainey Kanteh, Lamin Camara, Alhadji Kaba Sylla, Tolno Gnouma Benoit, and Aboubacarr Conteh, and those who assisted in the laboratory, transport, or administration. We appreciate the encouragement and strong support from institute and departmental heads including Tumani Corrah and Moussa Dieng Sarr. We acknowledge the expert efforts of members of the Biological Samples Platform and Genetic Analysis Platform of the Broad Institute. The genome-wide SNP array was supported by joint funding from several sources, including Burroughs Wellcome Fund Request 1008238, the Broad Institute Director's Fund, Harvard School of Public Health Department of Immunology and Infectious Diseases under direction of D. F. Wirth, Wellcome Trust Programme grant 077229/Z/05/Z to F. C. Kafatos and G. K. Christophides, and the DeLuca Professorship from Boston College to M.A.T.M. This work was enabled by funding from the United Kingdom Medical Research Council (MRC) and Foundation for the National Institutes of Health (FNIH)/Gates Grand Challenges in Global Health to the MRC Unit in The Gambia.

Literature Cited

- Aboagye-Antwi, F., A. Guindo, A. S. Traore, H. Hurd, M. Coulibaly *et al.*, 2010 Hydric stress-dependent effects of *Plasmodium falciparum* infection on the survival of wild-caught *Anopheles gambiae* female mosquitoes. *Malar. J.* 9: 243.
- Aguilar, R., F. Simard, C. Kamdem, T. Shields, G. E. Glass *et al.*, 2010 Genome-wide analysis of transcriptomic divergence between laboratory colony and field *Anopheles gambiae* mosquitoes of the M and S molecular forms. *Insect Mol. Biol.* 19: 695–705.
- Black, 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 Mol. Biol.* 10: 3–7.
- Caputo, B., D. Nwakanma, M. Jawara, M. Adiamoh, I. Dia *et al.*, 2008 *Anopheles gambiae* complex along The Gambia river, with particular reference to the molecular forms of *An. gambiae* s.s. *Malar. J.* 7: 182.
- Caputo, B., F. Santolamazza, J. L. Vicente, D. C. Nwakanma, M. Jawara *et al.*, 2011 The “far-west” of *Anopheles gambiae* molecular forms. *PLoS ONE* 6: e16415.
- Carnevale, P., J. C. Toto, P. Guibert, M. Keita, and S. Manguin, 2010 Entomological survey and report of a knockdown resistance mutation in the malaria vector *Anopheles gambiae* from the Republic of Guinea. *Trans. R. Soc. Trop. Med. Hyg.* 104: 484–489.
- Cassone, B. J., K. Mouline, M. W. Hahn, B. J. White, M. Pompi *et al.*, 2008 Differential gene expression in incipient species of *Anopheles gambiae*. *Mol. Ecol.* 17: 2491–2504.
- Cheng, C., B. J. White, C. Kamdem, K. Mockaitis, C. Costantini *et al.*, 2012 Ecological genomics of *Anopheles gambiae* along

- a latitudinal cline: a population-resequencing approach. *Genetics* 190: 1417–1432.
- Coluzzi, M., A. Sabatini, A. della Torre, M. A. Di Deco, and V. Petrarca, 2002 A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* 298: 1415–1418.
- Dabire, K. R., A. Diabate, M. Namontougou, L. Djogbenou, P. Kengne *et al.*, 2009 Distribution of insensitive acetylcholinesterase (ace-1R) in *Anopheles gambiae* s.l. populations from Burkina Faso (West Africa). *Trop. Med. Int. Health* 14: 396–403.
- 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., Z. Tu, and V. Petrarca, 2005 On the distribution and genetic differentiation of *Anopheles gambiae* s.s. molecular forms. *Insect Biochem. Mol. Biol.* 35: 755–769.
- Dia, I., L. Konate, B. Samb, J. B. Sarr, A. Diop *et al.*, 2008 Bionomics of malaria vectors and relationship with malaria transmission and epidemiology in three physiographic zones in the Senegal River Basin. *Acta Trop.* 105: 145–153.
- Diabate, A., T. Baldet, C. Chandre, K. R. Dabire, P. Kengne *et al.*, 2003 KDR mutation, a genetic marker to assess events of introgression between the molecular M and S forms of *Anopheles gambiae* (Diptera: Culicidae) in the tropical savannah area of West Africa. *J. Med. Entomol.* 40: 195–198.
- Diabate, A., R. K. Dabire, P. Kengne, C. Brengues, T. Baldet *et al.*, 2006 Mixed swarms of the molecular M and S forms of *Anopheles gambiae* (Diptera: Culicidae) in sympatric area from Burkina Faso. *J. Med. Entomol.* 43: 480–483.
- Diabate, A., R. K. Dabire, K. Heidenberger, J. Crawford, W. O. Lamp *et al.*, 2008 Evidence for divergent selection between the molecular forms of *Anopheles gambiae*: role of predation. *BMC Evol. Biol.* 8: 5.
- Diabate, A., A. Dao, A. S. Yaro, A. Adamou, R. Gonzalez *et al.*, 2009 Spatial swarm segregation and reproductive isolation between the molecular forms of *Anopheles gambiae*. *Proc. Biol. Sci.* 276: 4215–4222.
- Djogbenou, L., F. Chandre, A. Berthomieu, R. Dabire, A. Koffi *et al.*, 2008 Evidence of introgression of the ace-1(R) mutation and of the ace-1 duplication in West African *Anopheles gambiae* s.s. *PLoS ONE* 3: e2172.
- Djogbenou, L., N. Pasteur, M. Akogbeto, M. Weill, and F. Chandre, 2010 Insecticide resistance in the *Anopheles gambiae* complex in Benin: a nationwide survey. *Med. Vet. Entomol.* 25: 256–267.
- Esnault, C., M. Boulesteix, J. B. Duchemin, A. A. Koffi, F. Chandre *et al.*, 2008 High genetic differentiation between the M and S molecular forms of *Anopheles gambiae* in Africa. *PLoS One* 3: e1968.
- Fanello, C., F. Santolamazza, and A. Della Torre, 2002 Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med. Vet. Entomol.* 16: 461–464.
- Gimonneau, G., M. Pombi, M. Choisy, S. Morand, R. K. Dabire *et al.*, 2011 Larval habitat segregation between the molecular forms of the mosquito *Anopheles gambiae* in a rice field area of Burkina Faso, West Africa. *Med. Vet. Entomol.* 26: 9–17.
- Lawniczak, M. K., S. J. Emrich, A. K. Holloway, A. P. Regier, M. Olson *et al.*, 2010 Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* 330: 512–514.
- Lehmann, T., M. Licht, N. Elissa, B. T. Maega, J. M. Chimumbwa *et al.*, 2003 Population Structure of *Anopheles gambiae* in Africa. *J. Hered.* 94: 133–147.
- Lynd, A., D. Weetman, S. Barbosa, A. Egyir Yawson, S. Mitchell *et al.*, 2010 Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in *Anopheles gambiae* s.s. *Mol. Biol. Evol.* 27: 1117–1125.
- Marsden, C. D., Y. Lee, C. C. Nieman, M. R. Sanford, J. Dinis *et al.*, 2011 Asymmetric introgression between the M and S forms of the malaria vector, *Anopheles gambiae*, maintains divergence despite extensive hybridization. *Mol. Ecol.* 20: 4983–4994.
- Ndiath, M. O., C. Brengues, L. Konate, C. Sokhna, C. Boudin *et al.*, 2008 Dynamics of transmission of *Plasmodium falciparum* by *Anopheles arabiensis* and the molecular forms M and S of *Anopheles gambiae* in Dielmo, Senegal. *Malar. J.* 7: 136.
- Neafsey, D. E., M. K. Lawniczak, D. J. Park, S. N. Redmond, M. B. Coulibaly *et al.*, 2010 SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science* 330: 514–517.
- Oliveira, E., P. Salgueiro, K. Palsson, J. L. Vicente, A. P. Arez *et al.*, 2008 High levels of hybridization between molecular forms of *Anopheles gambiae* from Guinea Bissau. *J. Med. Entomol.* 45: 1057–1063.
- Pennetier, C., B. Warren, K. R. Dabire, I. J. Russell, and G. Gibson, 2010 “Singing on the wing” as a mechanism for species recognition in the malarial mosquito *Anopheles gambiae*. *Curr. Biol.* 20: 131–136.
- Rabbee, N., and T. P. Speed, 2007 BRLMM-P: a genotype calling method for the SNP 5.0 array. Available at http://media.affymetrix.com/support/technical/whitepapers/brlmmpp_whitepaper.pdf.
- Reidenbach, K. R., D. E. Neafsey, C. Costantini, N. Sagnon, F. Simard *et al.*, 2012 Patterns of genomic differentiation between ecologically differentiated M and S forms of *Anopheles gambiae* in West and Central Africa. *Genome Biol. Evol.* 4: 1202–1212.
- Riehle, M. M., W. M. Guelbeogo, A. Gneme, K. Eiglmeier, I. Holm *et al.*, 2011 A cryptic subgroup of *Anopheles gambiae* is highly susceptible to human malaria parasites. *Science* 331: 596–598.
- Rottschaefer, S. M., M. M. Riehle, B. Coulibaly, M. Sacko, O. Niare *et al.*, 2011 Exceptional diversity, maintenance of polymorphism, and recent directional selection on the APL1 malaria resistance genes of *Anopheles gambiae*. *PLoS Biol.* 9: e1000600.
- Sanford, M. R., B. Demirci, C. D. Marsden, Y. Lee, A. J. Cornel *et al.*, 2011 Morphological differentiation may mediate mate-choice between incipient species of *Anopheles gambiae* s.s. *PLoS ONE* 6: e27920.
- Santolamazza, F., B. Caputo, M. Calzetta, J. L. Vicente, E. Mancini *et al.*, 2011 Comparative analyses reveal discrepancies among results of commonly used methods for *Anopheles gambiae* molecular form identification. *Malar. J.* 10: 215.
- 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. *Am. J. Trop. Med. Hyg.* 49: 520–529.
- Slotman, M. A., F. Tripet, A. J. Cornel, C. R. Meneses, Y. Lee *et al.*, 2007 Evidence for subdivision within the M molecular form of *Anopheles gambiae*. *Mol. Ecol.* 16: 639–649.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin, 2005 Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3: e285.
- Vezenegho, S. B., B. D. Brooke, R. H. Hunt, M. Coetzee, and L. L. Koekemoer, 2009 Malaria vector composition and insecticide susceptibility status in Guinea Conakry, West Africa. *Med. Vet. Entomol.* 23: 326–334.
- Wallace, A. R., 1889 *Darwinism: An Exposition of the Theory of Natural Selection with Some of its Applications*. Macmillan, London/New York.
- Weetman, D., C. S. Wilding, K. Steen, J. C. Morgan, F. Simard *et al.*, 2010 Association mapping of insecticide resistance in wild *Anopheles gambiae* populations: major variants identified in a low-linkage disequilibrium genome. *PLoS ONE* 5: e13140.

- Weetman, D., C. S. Wilding, K. Steen, J. Pinto, and M. J. Donnelly, 2012 Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S forms. *Mol. Biol. Evol.* 29: 279–291.
- 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.
- White, B. J., M. W. Hahn, M. Pombi, B. J. Cassone, N. F. Lobo *et al.*, 2007 Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. *PLoS Genet.* 3: e217.
- White, B. J., C. Cheng, D. Sangare, N. F. Lobo, F. H. Collins *et al.*, 2009 The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. *Genetics* 183: 275–288.
- White, B. J., M. K. Lawniczak, C. Cheng, M. B. Coulibaly, M. D. Wilson *et al.*, 2011 Adaptive divergence between incipient species of *Anopheles gambiae* increases resistance to *Plasmodium*. *Proc. Natl. Acad. Sci. USA* 108: 244–249.

Communicating editor: B. P. Lazzaro

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.148718/-/DC1>

Breakdown in the Process of Incipient Speciation in *Anopheles gambiae*

Davis C. Nwakanma, Daniel E. Neafsey, Musa Jawara, Majidah Adiamoh, Emily Lund,
Amabelia Rodrigues, Kovana M. Loua, Lassana Konate, Ngayo Sy, Ibrahima Dia,
T. Samson Awolola, Marc A. T. Muskavitch, and David J. Conway

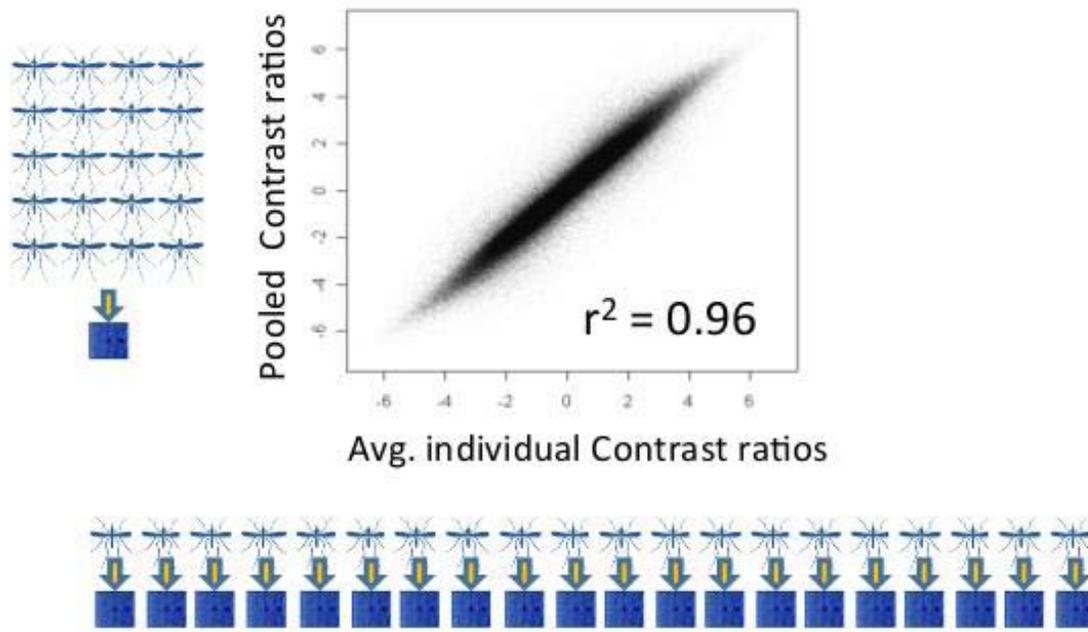


Figure S1 Scatterplot showing correlation of the distribution of values of pooled Contrast ratios for a pool of DNA from 20 mosquitoes, compared with the average values across the 20 DNA samples from the individual mosquitoes, with each point corresponding to one feature of the 400K SNP array, for a pilot assay with 20 *An gambiae s.s.* mosquitos from Mali (Neafsey *et al.* 2010).

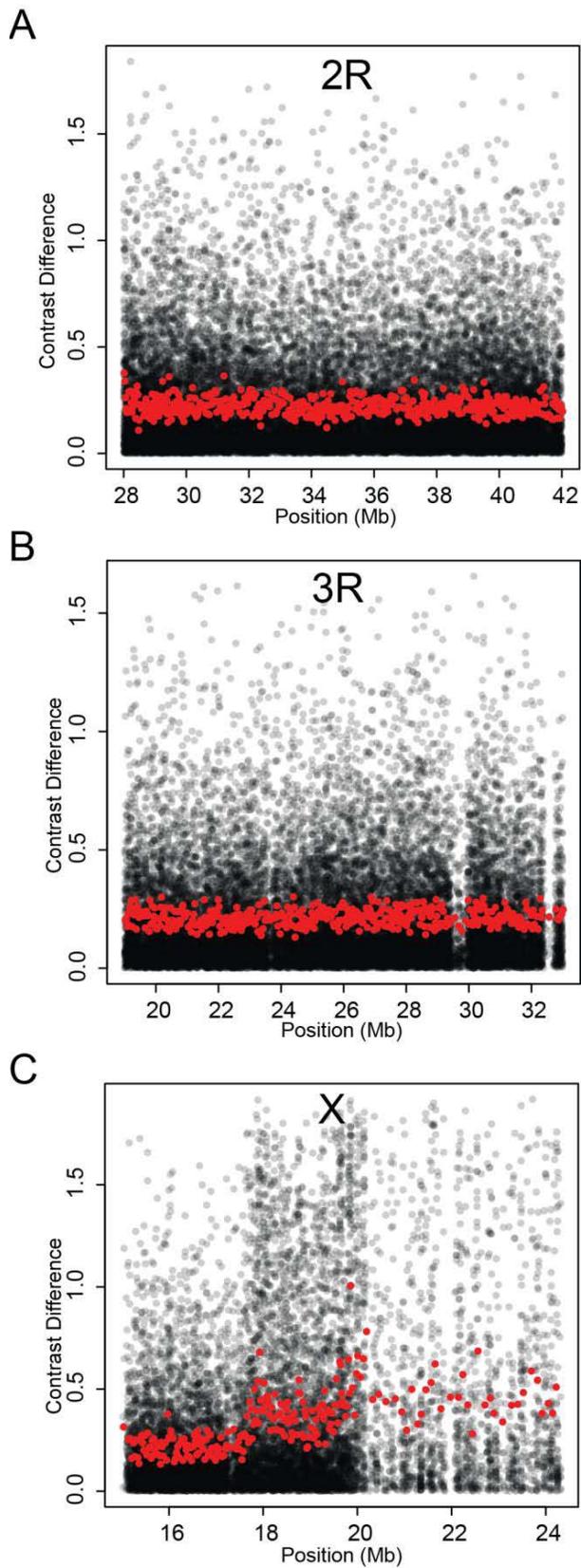


Figure S2 Focused profile of differentiation between M and S pools from Guinea Bissau on chromosome 2R (A), 3R (B), and X (C). Black dots indicate Contrast differences between the pools for individual SNP assays. Red dots indicate mean Contrast differences for windows of 50 adjacent SNPs.

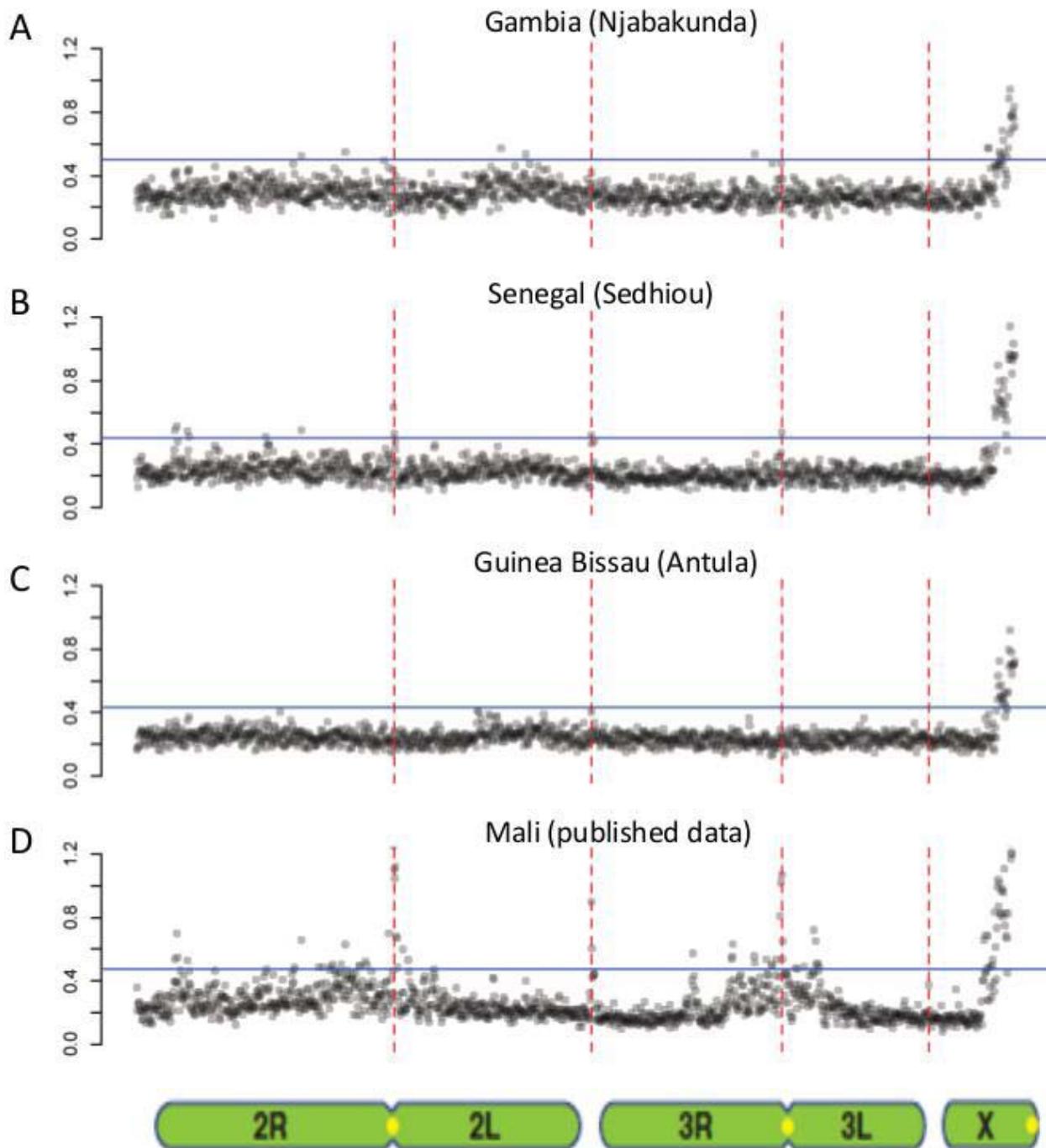
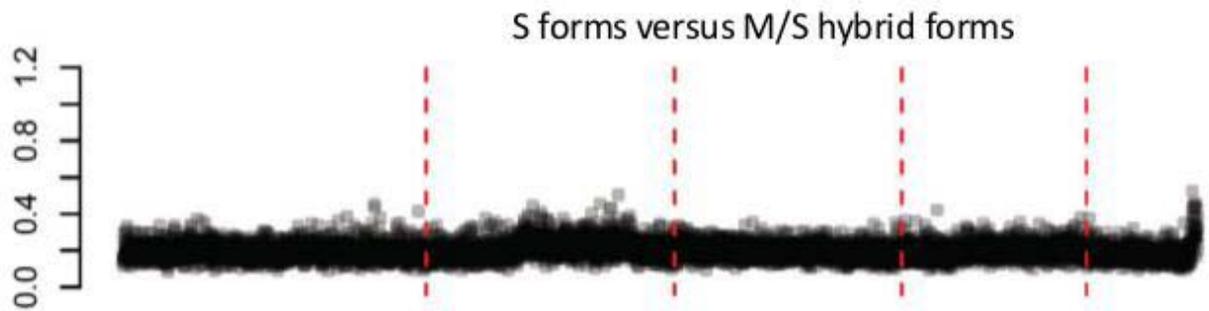
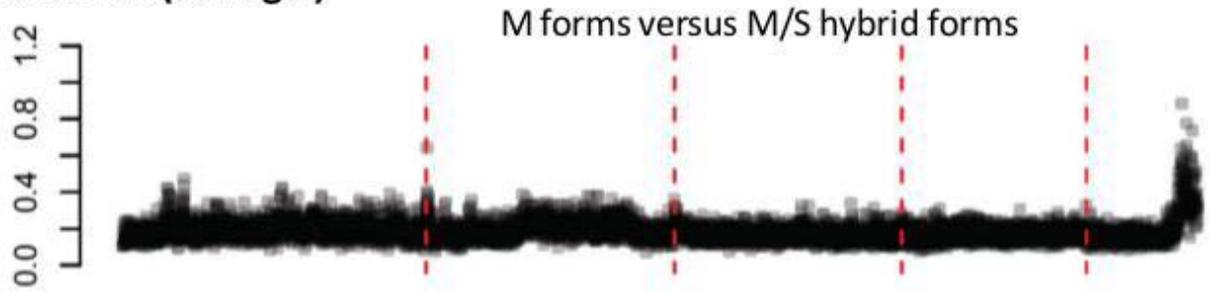


Figure S3 Test for genomic differentiation between pooled DNA from samples of 20 homozygous M form and 20 homozygous S form *Anopheles gambiae s.s.* from each of three sites in this study: (A) Njabakunda in The Gambia, (B) Sedhiou in Senegal, (C) Antula in Guinea Bissau, using data from the 66K sub-set of SNPs that had previously shown Hardy-Weinberg equilibrium within molecular forms in Mali (Neafsey *et al.* 2010). The vertical axis represents divergence measured as the absolute value of mean Contrast difference for stepping windows of 50 assays (this has a modal background of approximately 0.25 which is due to noise rather than a genome-wide divergence signal). The horizontal blue lines indicate the Bonferroni-corrected threshold for statistical significance obtained via bootstrapping. For comparison, panel (D) shows differentiation between similarly pooled DNA from M and S forms in Mali (Neafsey *et al.* 2010).

Sedhiou (Senegal)



Antula (Guinea Bissau)

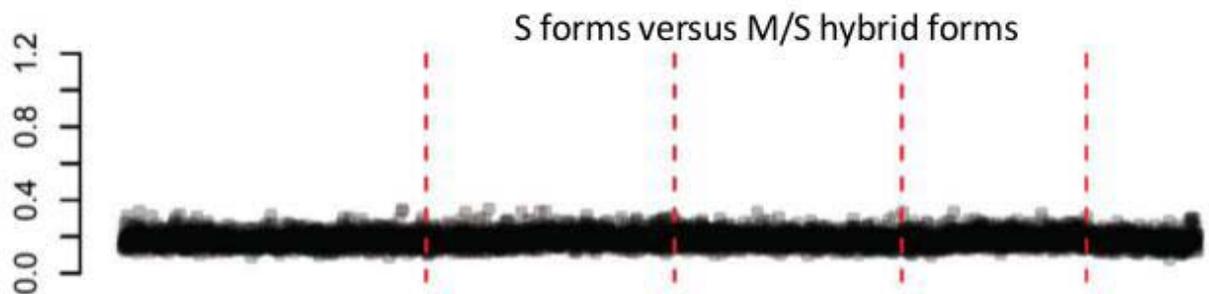
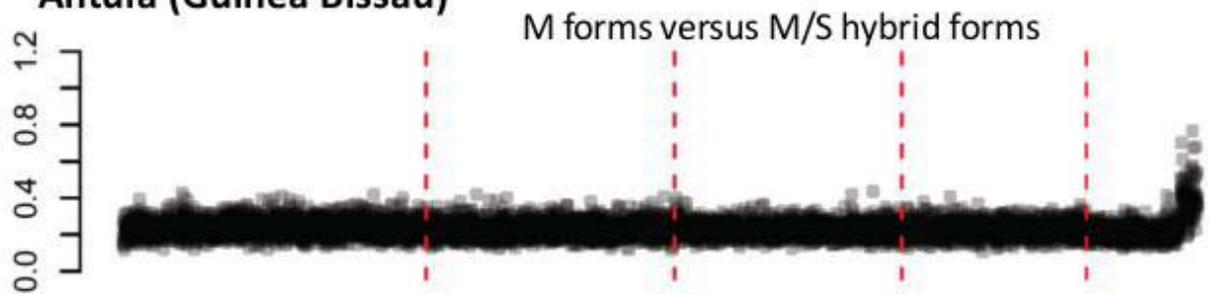


Figure S4 Comparison of M and S homozygous genotypes with heterozygous genotypes from two study sites: (A) Sedhiou M homozygotes vs. M/S heterozygotes, (B) Sedhiou S homozygotes vs. M/S heterozygotes, (C) Antula M homozygotes vs. M/S heterozygotes, (D) Antula S homozygotes vs. M/S heterozygotes.

Table S1 Spearman's rank correlation coefficients between pairwise pool comparisons of mean Contrast difference in the X pericentromeric region (Mb 15 – 24.2).

Table S1 is available for download at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.148718/-/DC1>.