

# Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, *Anopheles gambiae*

Yoosook Lee<sup>a,b</sup>, Clare D. Marsden<sup>a,b</sup>, Laura C. Norris<sup>a,c</sup>, Travis C. Collier<sup>a,b</sup>, Bradley J. Main<sup>a,b</sup>, Abdrahamane Fofana<sup>d</sup>, Anthony J. Cornel<sup>a,c</sup>, and Gregory C. Lanzaro<sup>a,b,1</sup>

<sup>a</sup>Vector Genetics Laboratory, <sup>b</sup>Department of Pathology, Microbiology and Immunology, and <sup>c</sup>Department of Entomology and Nematology, University of California, Davis, CA 95616; and <sup>d</sup>Malaria Research and Training Center, University of Bamako (BP 1805), Bamako, Mali

Edited\* by Tilahun D. Yilma, University of California, Davis, CA, and approved October 28, 2013 (received for review September 6, 2013)

**The M and S forms of *Anopheles gambiae* have been the focus of intense study by malaria researchers and evolutionary biologists interested in ecological speciation. Divergence occurs at three discrete islands in genomes that are otherwise nearly identical. An “islands of speciation” model proposes that diverged regions contain genes that are maintained by selection in the face of gene flow. An alternative “incidental island” model maintains that gene flow between M and S is effectively zero and that divergence islands are unrelated to speciation. A “divergence island SNP” assay was used to explore the spatial and temporal distributions of hybrid genotypes. Results revealed that hybrid individuals occur at frequencies ranging between 5% and 97% in every population examined. A temporal analysis revealed that assortative mating is unstable and periodically breaks down, resulting in extensive hybridization. Results suggest that hybrids suffer a fitness disadvantage, but at least some hybrid genotypes are viable. Stable introgression of the 2L speciation island occurred at one site following a hybridization event.**

reproductive isolation | population structure | *Anopheles coluzzii*

The M and S forms of the African malaria vector *Anopheles gambiae* have been the subject of intense study over the past decade. The focus has centered on models of the evolution and maintenance of genetic divergence between the two forms in relation to speciation (reviewed in ref. 1). *A. gambiae* has become a model, described in a number of recent reviews on speciation (2–5).

The two forms occur in sympatry throughout West and Central Africa (6). They were initially described on the basis of several single nucleotide polymorphisms (SNPs) in the X-linked ribosomal DNA locus (7, 8). Heterozygotes were rarely found in nature and studies of reproductive isolation (RI) confirmed strong assortative mating with interform matings occurring at a frequency of ~1% (9). Progeny of laboratory crosses and backcrosses show no signs of reduced fitness (10). However, it is widely held that, in nature, some degree of ecologically dependent postzygotic isolation, in addition to assortative mating, contributes to divergence between the two forms (11, 12).

Studies of the genetic structure of M and S based on microsatellite markers revealed little between-form differentiation outside the centromeric region of the X chromosome and a few regions associated with inversions (13–15). This overall lack of divergence was attributed to the homogenizing effect of gene flow between the forms (16).

In 2008, several reports of much higher frequencies of M/S hybrids, as high as 24%, appeared (17–19). These were all observed in populations in coastal West Africa, an area now thought to represent a zone of secondary contact (20, 21). These reports resulted in the emergence of this species as the focus of research aimed at exploring the evolution and maintenance of genetic divergence with gene flow (11, 12, 22, 23).

The first genome-wide comparison of the M and S forms by Turner et al. (24) was consistent with earlier observations that divergence overall is low, but there are small, discrete regions of divergence representing about 3% of the genome. They identified three diverged regions: one near the centromere on the X chromosome, one on the left arm of chromosome 2 (2L), and one on the right arm of chromosome 2 (2R). A number of genome-wide scans comparing M and S have been conducted since. These have applied several methods, including the same microarray used by Turner et al. (25, 26), high-density SNP arrays (12, 21, 23, 27), and whole-genome sequencing (28). These studies likewise revealed little divergence except in a few discrete regions of the genome characterized by high levels of differentiation (islands of divergence).

This body of work has culminated in two opposing models aimed at describing the evolution of M and S (1, 11). The “islands of speciation” model supposes that (i) small regions of divergence contain genes responsible for RI because they are directly associated with assortative mating and/or are under strong ecologically dependent divergent selection, and (ii) the rest of the genome is either neutral with respect to differentiation or close enough to neutral so that contemporary gene flow overwhelms selection. The alternative “incidental island” model recognizes the presence of islands of divergence, but suggests that (i) these are not related to RI and the remainder of the genome is less differentiated due to segregating ancestral polymorphism, not gene

## Significance

**Populations of the African malaria vector, *Anopheles gambiae*, are structured into M and S forms. All current work assumes the two rarely hybridize. Here we show this assumption is false. We demonstrate (i) significant exchange of genes between the two forms, even though (ii) hybrids have reduced fitness and (iii) the gene exchange process is spatially and temporally dynamic. For malaria, it is important to determine if genes for traits like insecticide resistance are shared between forms. For evolutionary biologists, this work confirms that this mosquito is a good model for studying if and how species may evolve in cases where there is ongoing gene flow.**

Author contributions: Y.L., A.J.C., and G.C.L. designed research; Y.L., C.D.M., L.C.N., A.F., and A.J.C. performed research; Y.L. contributed new reagents/analytic tools; Y.L., C.D.M., T.C.C., and B.J.M. analyzed data; and Y.L., C.D.M., L.C.N., B.J.M., and G.C.L. wrote the paper.

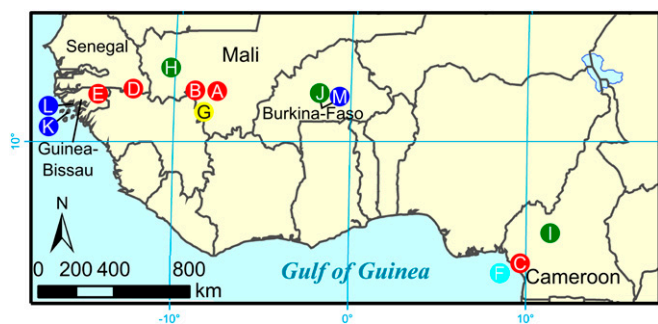
The authors declare no conflict of interest.

\*This Direct Submission article had a prearranged editor.

Data deposition: The raw genotype data reported in this paper have been deposited in the online individual level population genomic database Popl, with OpenProject id AgDIS, <http://popi.ucdavis.edu>.

<sup>1</sup>To whom correspondence should be addressed. E-mail: [gclanzaro@ucdavis.edu](mailto:gclanzaro@ucdavis.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1316851110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1316851110/-DCSupplemental).



**Fig. 1.** Collection site map. Site index follows Table 1. Sites are marked with colors that group populations by genetic structure as follows: red, past introgression evidence for ongoing gene flow; cyan, past introgression and loss of divergence on 2L; green, ongoing gene flow and introgression; blue, complete or nearly complete introgression; and yellow, Selinkenyi where a longitudinal study was conducted.

flow, and (ii)  $F_1$  hybrids are effectively sterile, therefore the amount of “realized” gene flow between M and S is near zero and that M and S are in fact “good species.” Indeed, the M form has recently been elevated to species status and provided the formal species name *Anopheles coluzzii* (29). We continue to refer to M and S forms to facilitate discussion with reference to the recent literature.

Limitations in the genome-wide scans described above may have contributed to disparate views of the evolution of divergence between *A. gambiae* subgroups. All comparisons of natural populations to date used single-locus, X-linked genotypes (7, 30–32) to identify M and S form specimens used in downstream analyses. These single-locus assays misidentify a significant proportion of backcross individuals. In addition, assessment of hybridization frequencies and comprehensive tests for introgression are precluded from these studies because DNA pools were used (21, 23, 27), sample sizes were too small (24, 26), or such assessments may be irrelevant because M and S laboratory colony mosquitoes were used (28). Finally, there are strong limitations in relying on genome scans alone for detecting selection, gene flow, and recombination that occurs during speciation (33, 34).

In this study, we used a simple multilocus SNP genotype approach to distinguish M, S,  $F_1$  hybrids, and backcross individuals (35), which we call the “divergence island SNP” (DIS) assay (*SI Appendix*). We describe the spatial and temporal distribution of hybridization frequencies, the extent to which hybridization results in introgression, and the fitness of hybrid genotypes in nature. Assumptions concerning these between-form interactions have played a central role in the interpretation of comparative genomics data as applied to understanding speciation in this system. The results we report challenge a number of these assumptions and they provide unique information about temporal dynamics in the way in which M and S forms interact that suggests new avenues for future research.

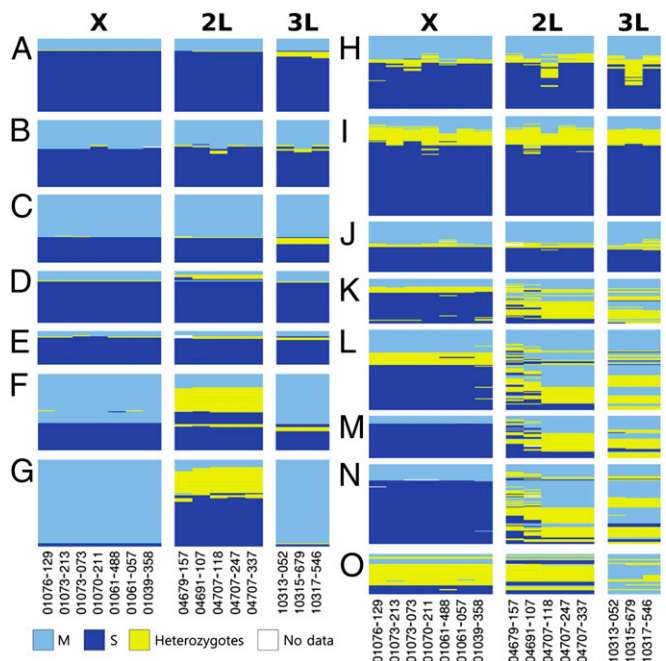
## Results

**Multilocus, Single-Copy DIS Assay.** A few studies have used multiple restriction fragment length polymorphism assays and/or mismatch primer-based SNP assays to genotype one SNP in each of the three (X, 2L, and 3L) islands of divergence (26, 36). We extended this approach with the development of the DIS assay (35) to describe *A. gambiae* populations throughout West Africa (Fig. 1 and *SI Appendix, Table S3*). The DIS assay includes 7 SNPs spanning 4.4 Mbp on the X chromosome, 5 spanning 2.2 Mbp on chromosome 2L, and 3 spanning 117 kb on chromosome 3L (35). Genotypes at these 15 SNPs were used to designate individuals as M form, S form (homozygous for form-specific alleles at a minimum of 13 of the 15 SNPs),  $F_1$  hybrids (hetero-

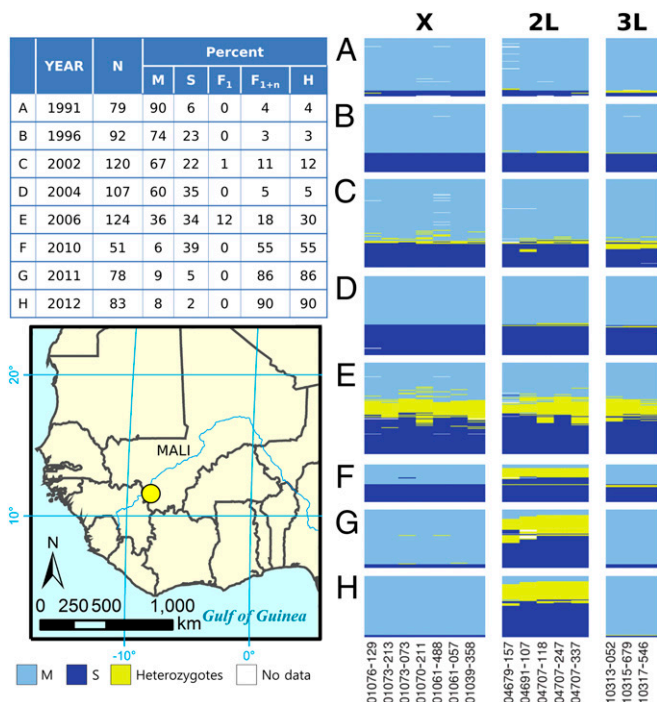
zygous at all 15 SNPs), and backcross (individuals with mixed M, S, and heterozygous genotypes at 3 or more SNPs). The GOUNDRY and ENDO populations were defined on the basis of genotypes at eight microsatellite loci as described previously (37). We used the method described in ref. 35 to illustrate the distribution of DIS genotypes among each study population, using what we refer to as “DIS maps” (Figs. 2 and 3 and *SI Appendix, Figs. S3–S7*).

**Analysis of Allopatric M and S Populations.** Five allopatric populations, two “pure” M forms and three pure S forms (pure = predominant form occurs at frequency >0.96), were analyzed to establish that molecular-form-specific SNPs are not polymorphic within forms. These sites covered a broad geographic area, from central Mali to southern Cameroon (*SI Appendix, Fig. S3*). Only 12 of 466 carried mixed genotypes. Heterozygotes occurred at SNPs distributed over all three chromosomes. Ten of the 14 individuals were heterozygous at multiple, physically linked, SNPs. Four were heterozygous at a single SNP, but each of these was only observed in a single population, suggesting these were due to rare past migration events and do not represent shared ancestral polymorphisms.

**Analysis of Sympatric M and S Populations.** Thirteen populations in which the M and S forms occur in sympatry were analyzed (Fig. 1). These included the ENDO and GOUNDRY populations from Burkina Faso (37), four populations from Mali, three from Guinea-Bissau, two from Cameroon, one population from Equatorial Guinea, and one from Senegal. A total of 825 individuals were sampled (Table 1). In every population studied, individuals with hybrid genotypes occurred at a frequency in excess of 5%, ranging from 5.2% to 96.9%. The genetic structure of these populations formed a continuum, from populations with evidence of past introgression, but clear boundaries between M and S, to those in which the boundaries between M and S are completely absent.



**Fig. 2.** DIS maps for sympatric populations. Site index follows Table 1. A, Banambani; B, Founia; C, Tiko; D, Nathia; E, Canjufa; F, Bioko; G, Selinkenyi; H, Bantinggougou; I, Njigalap; J, ENDO; K, Abu females; L, Prabis females; M, Abu males; N, Prabis males; O, GOUNDRY.



**Fig. 3.** DIS maps for sympatric populations from the town of Selinkenyi, Mali collected at eight time points over a 21-y period, 1991–2012 (A–H).

**Populations with past introgression and no ongoing gene flow.** Populations at Banambani and Founia, Tiko, Nathia, and Canjufa included backcross individuals at frequencies ranging from 5% to 15%, but of the 312 individuals genotyped, no F<sub>1</sub> hybrids were detected (Fig. 2 A–E and *SI Appendix*, Fig. S4). Genotype frequencies at all SNPs departed significantly from Hardy–Weinberg equilibrium (HWE) with a deficiency of heterozygotes in every case. Association among all three unlinked islands of divergence was strong, but nowhere near complete (mean  $r^2 = 0.83$ ; *SI Appendix*, Table S4). Recombination was more common in autosomal islands than for X chromosome islands (*SI Appendix*, Table S4). Introgressed SNPs in both the homozygous and heterozygous states were present, suggesting backcrossing for multiple generations. Excluding the possibility of ancestral polymorphism (as justified above), the distribution of genotype frequencies can be explained by hybridization, either ongoing at low frequencies (too low for the detection of F<sub>1</sub> genotypes) or in the recent past, followed by the reestablishment of strong premating barriers. In either case, introgression has occurred with backcross genotypes maintained at nonequilibrium frequencies, likely due to incomplete assortative mating and/or selection too weak to eliminate them (at least between the time of their formation by hybridization and the time they were sampled).

**Populations with past introgression and loss of divergence on 2L.** Populations at Bioko and Selinkenyi collected in 2012 were composed of a large number of backcross individuals, 54.9–89.2% (Table 1, Fig. 2 F and G, and *SI Appendix*, Fig. S5). As in populations described above, the absence of F<sub>1</sub> hybrids suggests a lack of ongoing gene flow in these populations. Presence of S form chromosome 2 SNPs in otherwise typical M form genomes accounted for nearly all introgression (110 of 116 recombinant genotypes). Genotype frequencies at all SNPs on both 3L and X departed significantly from HWE, with a deficiency of heterozygotes in all cases and strong association of form-specific SNPs across these two islands of divergence. SNPs on 3L and X are associated (mean  $r^2 > 0.81$ ; *SI Appendix*, Table S4), whereas SNPs on 2L were not associated with X or 3L ( $r^2 < 0.23$ ).

Genotype frequencies for all five SNPs on 2L, in both populations, were in HWE, suggesting stable introgression of the 2L island. Introgression of X-linked SNPs was absent (0/116).

**Populations with ongoing gene flow and introgression.** Populations at Bantinnougou, Njigalap, and the ENDO population included F<sub>1</sub> hybrids at frequencies ranging from 3.3% to 11.8% and backcross genotypes from 10.0% to 25.3% (Table 1, Fig. 2 H–J, and *SI Appendix*, Fig. S6). Of 45 tests for compliance of genotype frequencies to HWE, all but 2 departed from expectations, with a deficiency of heterozygotes in every case. The proportion of backcross genotypes relative to F<sub>1</sub>s is quite low (in Njigalap the frequency of F<sub>1</sub> genotypes exceeds backcrosses by 50%) and in nearly every case the introgressed SNP is heterozygous (56/62), indicating these are the product of first generation backcrosses (F<sub>1</sub> × parental). This suggests that hybridization is consistently high in this population with strong selection against later generation backcross genotypes or that at the time these populations were sampled, they were experiencing a very recent breakdown in assortative mating (isolate breaking) (38). Overall form-specific islands remained associated (mean  $r^2 = 0.82$ ).

**Populations with complete or nearly complete introgression.** Populations at Abu and Prabris consist of backcross individuals at frequencies of 82.0% and 73.9%, respectively (Fig. 2 K–N and *SI Appendix*, Fig. S7). Introgression is strongly asymmetric, with the M form introgressed into the S (22). Our analysis supports this observation and illustrates that introgression into the S form has proceeded to the extent that pure S form genotypes are completely absent, whereas pure M form individuals persist. Genotype frequencies for all eight SNPs on chromosomes 2 and 3 are in HWE; however, strong linkage disequilibrium (LD) is maintained for all SNPs on the X chromosome ( $r^2 > 0.93$ ), with a significant deficiency of heterozygotes at all seven SNPs. A separate analysis of males reveals the strength of the asymmetric introgression (Fig. 2 M and N and *SI Appendix*, Fig. S7 C and D). Males that were heterozygous for all eight autosomal SNPs (males are hemizygous for X-linked SNPs) were presumed to be F<sub>1</sub> hybrids. Of 19 F<sub>1</sub> hybrid males, 18 were hemizygous for S-specific SNPs on the X chromosome, suggesting S form females exhibit weak assortative mating.

The GOUNDRY population exhibits the most extreme level of introgression, with a frequency of backcross genotypes of 96.9% (Fig. 2 O and *SI Appendix*, Fig. S7 E). Only a single individual of 32 sampled was homozygous at all SNPs. Genotype frequencies at all

**Table 1.** M, S, F<sub>1</sub>, and backcross (F<sub>1+n</sub>) frequencies

Index	Site	Year	N	%M	%S	%F <sub>1</sub>	%F <sub>1+n</sub>
Populations with past introgression and no ongoing gene flow							
A	Banambani	2005	87	15	73	0	12
B	Founia	2006	77	39	56	0	5
C	Tiko	2003	81	16	2	0	82
D	Nathia	2009	41	7	78	0	15
E	Canjufa	2009	26	15	73	0	12
Populations with past introgression and loss of divergence on 2L							
F	Bioko	2002	95	0	96	0	4
G	Selinkenyi	2012	83	8	2	0	90
Populations with ongoing gene flow and introgression							
H	Bantinnougou	2006	87	24	46	5	25
I	Njigalap	2006	119	10	70	12	8
J	ENDO	2006	60	42	45	3	10
Populations with complete or nearly complete introgression							
K	Abu (♀♀)	2009	50	98	0	0	2
L	Prabris (♀♀)	2009	92	25	0	1	74
M	Abu (♂♂)	2009	48	19	0	19	62
N	Prabris (♂♂)	2009	90	13	0	15	72
O	Goundry	2006	32	3	0	0	97



15 SNPs are in HWE. Within each island, all pairs of SNPs were in significant LD ( $r^2 > 0.60$ ,  $P = 0.00$ ). No association among form-specific islands was evident ( $r^2 < 0.17$ ,  $P > 0.05$ ).

**Temporal Variation in Hybridization.** Patterns of hybridization and introgression vary spatially, probably due to differences among local environments (39–41). These differences may include environmental cues used in mate selection and/or factors that influence hybrid fitness. If true, it is likely that hybridization and introgression also have a temporal dynamic in places that experience environmental fluctuations over time. To examine this possibility we conducted a longitudinal study of sympatric M and S populations spanning a period of 21 y at Selinkenyi. Rainfall in this area is controlled by movement of the intertropical convergence zone (ITCZ), which oscillates between the northern and southern tropics annually. Variations in the latitudinal movement of the ITCZ from year to year cause large interannual variability in wet-season (June–October) rainfall (42).

The distribution of DISs in populations at Selinkenyi between 1991 and 2012 are illustrated in Fig. 3. All collections were made during the rainy season (July–October). In 1991, M was the predominant form, representing 89.9% of individuals sampled. A total of 3.8% of individuals had mixed, backcross genotypes, but no  $F_1$  hybrids were observed (Fig. 3A). The relative proportion of forms changed in 1996 with S form increasing to 22.8%, but the proportion of backcross hybrids remained roughly as it was in 1991 (3.3%, Fig. 3B). As in 1991, no  $F_1$  hybrids were collected. Temporal changes in the relative abundance of the two forms have been reported in the past (43, 44). The presence of backcross individuals indicates that some degree of hybridization had occurred in the past, but a lack of  $F_1$  hybrids suggests that there was no ongoing hybridization at the time these samples were collected.

A change occurred in 2002 when the relative abundance of the two forms remained similar to 1996 (66.7% M and 21.7% S); however, the frequency of backcross hybrids increased to 10.8% and  $F_1$  hybrids were observed at a frequency of 0.8% (Fig. 3C). The occurrence of  $F_1$  hybrids and accompanying increase in backcross frequencies suggests this sample was taken at a time when hybridization between the two forms was occurring and that hybrid individuals were actively mating, resulting in introgression of the two forms. Two years later (2004) near complete assortative mating appears to have been reestablished, evident from an absence of  $F_1$  hybrids and a decrease in the frequency of backcross hybrids (from 11.6% in 2002 to 5.6% in 2004, Fig. 3D).

A dramatic change occurred in 2006 when the frequency of  $F_1$  hybrids jumped to 12.1% and backcross hybrids increased to 18.5% (Fig. 3E). An apparent breakdown of assortative mating in 2006 precipitated a profound change in the genetic structure of this population such that the frequency of hybrids in samples collected in 2010, 2011, and 2012 increased to 54.9%, 85.9%, and 89.2%, respectively (Fig. 3F–H). Nearly all backcross individuals in the 2010–2012 populations had mixed genotypes due to introgression of SNPs within the 2L island, resulting in the complete loss of 2L association with the X and 3L islands. The X and 3L SNPs remained in tight LD ( $r^2 > 0.83$ , *SI Appendix, Table S4*). Genotype frequencies for all five SNPs on 2L remained at similar frequencies subsequently (2010–2012), so it appears that divergence in this region of the genome has been permanently lost.

The nearly complete disappearance of heterozygotes at all SNPs following the episode of hybridization that occurred in 2002 (Fig. 3C), and for heterozygotes on 3L and X following the hybridization event in 2006 (Fig. 3E), provides strong evidence for selection against hybrids in nature.

**Patterns of Introgression Among Islands of Divergence.** White et al. (26) reported nearly complete LD between the three islands of divergence in natural populations. Hahn et al. (36) used laboratory crosses to demonstrate a lack of intrinsic incompatibilities

that might explain this association. Based on these observations, it was proposed that the association among the three islands is maintained by near-complete RI between M and S forms in nature. RI is presumed to be the consequence of strong pre-mating isolation and that those few hybrids that may be formed rarely survive to reproduce (26). This idea is central to the incidental islands model described above. The notion that association is maintained by strong selection against mixed genotypes is weakened by reports of linkage equilibrium in coastal populations where hybridization rates are high (20). We included two coastal populations in this study, the villages of Abu and Prabis in Guinea-Bissau, and we likewise found low LD among the three divergence islands ( $r^2 < 0.17$ ). The pattern we observed is complicated by the fact that the extent of introgression among the three islands is not equivalent. Genotypes for all eight SNPs on chromosomes 2 and 3 are at equilibrium frequencies so that introgression is complete. This suggests that selection against mixed genotypes at these SNPs is very low.

The pattern is quite different for the X chromosome where LD among the seven SNPs is quite strong ( $r^2 > 0.93$ ) and genotype frequencies depart significantly from HWE with a deficiency of heterozygotes at all seven SNPs. A pattern of lower introgression on the autosomes relative to the X chromosome was generally observed in all populations where M and S are sympatric (*SI Appendix, Table S4*). It has, however, been suggested that coastal populations represent unique environments and are therefore unlikely to be instructive about the genomics of RI in more “typical” areas (23). Results from the longitudinal survey conducted at Selinkenyi and the “snapshot” data for Goundry and Njigalap, however, demonstrate that hybridization and introgression do occur in typical *A. gambiae* habitats.

**Within and Between Divergence Island Recombination.** Recombination within islands was lowest ( $r^2 = 0.95$ ) for chromosome 3 and highest ( $r^2 = 0.90$ ) for chromosome 2 (*SI Appendix, Table S4*). Recombination within the X chromosome was similar ( $r^2 = 0.94$ ) to the level observed in chromosome 3, despite covering a larger region of the genome (37 times larger).

Recombination between islands was highest for the 2L island ( $r^2 = 0.87$  vs.  $r^2 = 0.94$  for X and  $r^2 = 0.93$  for 3L) with a mean proportion of mismatched islands = 20% (*SI Appendix, Table S4*). Collections that displayed high levels of 2L island mismatch ( $P > 10\%$ ) include Nathia, Bioko, and 2010–2012 Selinkenyi. The wide geographic range of this 2L island mismatched genome, as well as the loss of 2L divergence post-2006 in Selinkenyi, suggest that this particular genome combination is stable and likely to be present in other geographic regions.

## Discussion

The results presented here provide the basis for a critical reassessment of the relationship between the M and S forms of *A. gambiae*. A long-standing debate has centered on the level of gene flow between forms. On one side many authors consider that RI between the two is complete or nearly complete (7, 8, 45, 46) and that the two forms represent distinct species (29). On the other side are those who argued that RI, even if recent, cannot explain the “mosaic” pattern of genome divergence, and that it is selection operating on a few genomic regions in the face of gene flow that shapes this pattern (13, 16, 47). Here we have demonstrated that M/S hybridization occurs at frequencies well above the ~1% frequently cited and at locations throughout West and Central Africa (Figs. 2 and 3 and *SI Appendix, Figs. S3–S7*), outside the coastal areas where high levels of hybridization are well known, but have been dismissed as “anomalous” (23).

Genome-wide scans have identified genomic islands of divergence, consistent with the mosaic divergence identified in earlier work. These were initially interpreted as representing islands of speciation containing genes responsible, directly or

indirectly, for partial RI between forms and that the remaining, roughly 97% of the genome remains undifferentiated due to gene flow between the two forms (24). An opposing view recognizes islands of divergence as standing out, but argues this divergence is due to the location of the islands near centromeres where recombination is low and that the surrounding genome is relatively undifferentiated because complete RI evolved only recently (27). In support of this incidental islands model, it has been argued that the infrequent hybrids observed in nature are largely  $F_1$  hybrids and that these are effectively sterile, therefore their presence does not represent “effective” gene flow (26). The longitudinal study at Selinkenyei does suggest reduced fitness of hybrids; however, the loss of divergence on the chromosome 2L island after 2006 and the persistence of backcross genotypes in nearly all populations surveyed, often at high frequencies, reveals that hybridization does in fact represent gene flow and at levels far exceeding what has heretofore been recognized.

The existence of temporal variation in hybridization rates represents a unique and unanticipated feature in the relationship between M and S. The longitudinal study conducted over a period of 21 y at Selinkenyei reveals that RI is unstable (Fig. 3). Sustained maintenance of M and S populations, presumably by strong assortative mating, is periodically disrupted by episodes of hybridization, which in 2006 resulted in 12% of the population consisting of  $F_1$  hybrids. Remarkably this episode resulted in the complete loss of the 2L island of divergence present before 2006. In samples taken in 2010, 2011, and 2012, 2L SNP genotypes were at equilibrium frequencies.

This longitudinal survey also provided evidence for selection against hybrids. A smaller episode of hybridization occurred in 2002 with a frequency of  $F_1$  hybrids = 0.08, and backcrossed genotype frequencies = 0.11. Strong assortative mating appears to have been reestablished in 2004 with no  $F_1$ s present and a decline in backcross genotypes to a frequency of 0.05. Likewise, following the 2006 episode, the frequency of backcrossed genotypes for the X and 3L islands declined from 13.7% (X) and 2.4% (3L) in 2006 to 2.0% (X) and 0% (3L) in 2010. This decline in hybrids was presumably the result of selection. Loss of the 2L island after 2006 is likely the consequence of relatively weak selection being overwhelmed by gene flow that year [but see discussion by Weetman et al. concerning selection on the *kdr* locus on chromosome 2L (12)], whereas selection on the 3L and X islands appears to have been strong enough to maintain divergence, despite this relatively high level of gene flow.

Overall, it is the X chromosome island that remains the most highly diverged. This observation supports the suggestion that the X island contains genes responsible for RI (26, 48). This idea is further supported by the identification of high levels of divergence in the same region of the X between the Bamako and Savanna chromosomal forms of *A. gambiae*, which exhibit a high degree of RI, despite the fact that both are S form (49). Reduced introgression and recombination on the sex chromosome shown here accords well with other studies on flycatchers (50–52), *Drosophila simulans* group (53), and *Heliconius* butterflies (54) that demonstrated that sex chromosomes display a more advanced stage of speciation compared with autosomes.

Overall the picture that emerges is consistent with the islands of speciation model (24) as the best fit to describe the relationship between the M and S forms. However, in a larger sense these results challenge the more fundamental view of the relationship between M and S. The extent of gene flow between the two forms is far greater than presumed and the frequency of hybridization is dynamic, fluctuating in both time and space. This observation is consistent with a genic model of speciation where populations may diverge even in the face of considerable gene flow (4, 55). This process has been described as proceeding along a speciation continuum (56, 57) and it is possible for divergent populations to lie somewhere midway on this continuum (58,

59). Movement on the speciation continuum may be toward increasing divergence and speciation; however, the reverse is also possible (60, 61). Our data suggest the possibility that M and S forms represent diverged populations that sporadically move “forward” (diverging) and “backward” (introgressing).

The M form is generally believed to be the more recently evolved (62, 63), presumably via niche expansion into marginal habitats in parapatry with the S form (64). It is in these marginal habitats that they are believed to have diverged via adaptation to the peculiar ecological conditions that occur there and ultimately by the acquisition of premating mechanisms reducing the possibility of mating with the S form. Range expansion has increasingly brought M and S in contact, so that today they occur largely in sympatry throughout their range in West and Central Africa (1, 6). We provide evidence that pre- and postzygotic isolation between the two forms is relatively weak, such that varying degrees of introgression of the two forms has occurred throughout their range and that this is an ongoing phenomenon. The GOUNDRY population in Burkina Faso is the most extreme example where all form-specific SNP genotypes are in HWE (Fig. 2O and *SI Appendix*, Fig. S7E). Coastal populations in Guinea-Bissau (Abu and Prabis) have complete introgression of the autosomal islands of divergence; however, the X chromosome remains largely diverged (Fig. 2K–N and *SI Appendix*, Fig. S7A–D), presumably due to a strong asymmetry in assortative mating (22). Loss of the chromosome 2L island of divergence was observed in populations at Bioko and Selinkenyei (Fig. 2F and G and *SI Appendix*, Fig. S5). Evidence that this is a dynamic and ongoing process is provided by the longitudinal study conducted at Selinkenyei where periodic episodes of increased hybridization have recently culminated in the loss of the 2L island of divergence in this population (Fig. 3).

The results presented here confirm the importance of the M and S designation as a proxy for reproductively isolated populations (in most, but not all locations analyzed). However, it seems plausible that the M and S forms lie at an unstable position, midway on a species continuum, where forces promoting divergence (selection and assortative mating) are frequently overwhelmed by recurrent gene flow leading to introgression. It is likely that divergence occurs on a very long time scale, over hundreds to millions of generations (65). Our longitudinal study at Selinkenyei spans 21 y (~250 *A. gambiae* generations). In this regard, our data are still just a snapshot of a much longer process. In addition, our analysis covers a tiny fraction of the M and S genomes. Genome-wide surveys will provide a much better picture of the extent of M–S introgression and should provide insights into the importance of hybridization as an adaptive force in these mosquitoes. It is clear that *A. gambiae* remains an intriguing system for the exploration of mechanisms underlying the speciation process.

## Methods

**Mosquito Collections and Species Identification.** *A. gambiae* sensu lato were collected from inside human dwellings using mouth aspirators during the rainy season from various sites in Burkina Faso (ENDO), Cameroon (three sites), Equatorial Guinea (one site), Guinea-Bissau (three sites), Mali (seven sites) and Senegal (two sites) (*SI Appendix*, Table S3). The GOUNDRY, Burkina Faso collections were larvae from natural breeding sites. *A. gambiae* sensu stricto were identified using a diagnostic PCR developed by Scott et al. (66).

**SNP Genotyping.** We used the DIS method (35) to genotype *A. gambiae* individuals. The DIS method includes 15 SNPs occurring on all three chromosomes; 7 on the X, 5 on 2L, 3 on 3L. Full details of the DIS method are provided in *SI Appendix*. Genotyping was conducted using the Sequenom iPLEXGold assay platform at the Veterinary Genetics Laboratory at the University of California, Davis. Mass spectrogram visualization and genotype calls were done using TyperAnalyzer version 4.0 (Sequenom).

**Data Analysis.** We plotted DIS maps using R heatmap plotting. We calculated LD using the EMLD program that implements the maximum-likelihood method described in ref. 67. Arlequin version 3.5 (68) was used to test departure from HWE and to determine significance of LD.

**ACKNOWLEDGMENTS.** We thank Drs. K. Vernick (Pasteur Institute) and M. Riehle (University of Minnesota) for the GOUNDRY and ENDO specimens, Dr. B. Foy (Colorado State University) for the Senegal specimens, C. Neiman and A. Weakley (Vector Genetics Laboratory) for technical assistance, J. Malvick

(Veterinary Genetics Laboratory) for assistance in SNP genotyping, and Dr. M. Sanford (Harris County Institute of Forensic Sciences) for advice in developing this project. This research was supported by National Institutes of Health Grants R01AI 078183, R21AI062929, D43TW007390, and T32AI074550.

- Lanzaro GC, Lee Y (2013) Speciation in *Anopheles gambiae*—The distribution of genetic polymorphism and patterns of reproductive isolation among natural populations. *Anopheles Mosquitoes: New Insights into Malaria Vector*, ed Manguin S (InTech, Rijeka, Croatia).
- Pinho C, Hey J (2010) Divergence with gene flow: Models and data. *Annu Rev Ecol Syst* 41(41):215–230.
- Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends Genet* 28(7):342–350.
- Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Curr Opin Genet Dev* 16(6):592–596.
- Nosil P (2012) *Ecological Speciation* (Oxford Univ Press, Oxford), p xvii.
- della Torre A, Tu Z, Petrarca V (2005) On the distribution and genetic differentiation of *Anopheles gambiae* s.s. molecular forms. *Insect Biochem Mol Biol* 35(7):755–769.
- Favia G, Lanfrancotti A, Spanos L, Sidén-Kiamos I, Louis C (2001) Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s.s. *Insect Mol Biol* 10(1):19–23.
- della Torre A, et al. (2001) Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Mol Biol* 10(1):9–18.
- Tripet F, et al. (2001) DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Mol Ecol* 10(7):1725–1732.
- Diabaté A, Dabire RK, Millogo N, Lehmann T (2007) Evaluating the effect of post-mating isolation between molecular forms of *Anopheles gambiae* (Diptera: Culicidae). *J Med Entomol* 44(1):60–64.
- Turner TL, Hahn MW (2010) Genomic islands of speciation or genomic islands and speciation? *Mol Biol Evol* 27(5):848–850.
- Weetman D, Wilding CS, Steen K, Pinto J, Donnelly MJ (2012) Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S forms. *Mol Biol Evol* 29(1):279–291.
- Lanzaro GC, et al. (1998) Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proc Natl Acad Sci USA* 95(24):14260–14265.
- Wondji C, Simard F, Fontenille D (2002) Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. *Insect Mol Biol* 11(1):11–19.
- Wang R, Zheng L, Touré YT, Dandekar T, Kafatos FC (2001) When genetic distance matters: Measuring genetic differentiation at microsatellite loci in whole-genome scans of recent and incipient mosquito species. *Proc Natl Acad Sci USA* 98(19):10769–10774.
- Taylor C, et al. (2001) Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics* 157(2):743–750.
- Ndiath MO, 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.
- Caputo B, 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.
- Oliveira E, et al. (2008) High levels of hybridization between molecular forms of *Anopheles gambiae* from Guinea Bissau. *J Med Entomol* 45(6):1057–1063.
- Caputo B, et al. (2011) The “far-west” of *Anopheles gambiae* molecular forms. *PLoS ONE* 6(2):e16415.
- Nwakanma DC, et al. (2013) Breakdown in the process of incipient speciation in *Anopheles gambiae*. *Genetics* 193(4):1221–1231.
- Marsden CD, 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(23):4983–4994.
- Reidenbach KR, 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(12):1202–1212.
- Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* 3(9):e285.
- White BJ, et al. (2009) The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. *Genetics* 183(1):275–288.
- White BJ, Cheng C, Simard F, Costantini C, Besansky NJ (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Mol Ecol* 19(5):925–939.
- Neafsey DE, et al. (2010) SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science* 330(6003):514–517.
- Lawnciczak MK, et al. (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* 330(6003):512–514.
- Coetzee M, et al. (2013) *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa* 3619(2):246–274.
- Fanello C, Santolamazza F, della Torre A (2002) Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med Vet Entomol* 16(4):461–464.
- Santolamazza F, Della Torre A, Caccone A (2004) Short report: A new polymerase chain reaction-restriction fragment length polymorphism method to identify *Anopheles arabiensis* from *An. gambiae* and its two molecular forms from degraded DNA templates or museum samples. *Am J Trop Med Hyg* 70(6):604–606.
- Santolamazza F, et al. (2008) Insertion polymorphisms of SINE200 retrotransposons within speciation islands of *Anopheles gambiae* molecular forms. *Malar J* 7:163.
- Biswas S, Akey JM (2006) Genomic insights into positive selection. *Trends Genet* 22(8):437–446.
- Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nat Rev Genet* 14(6):404–414.
- Lee Y, Marsden CD, Nieman C, Lanzaro GC (2013) A new multiplex SNP genotyping assay for detecting hybridization and introgression between the M and S molecular forms of *Anopheles gambiae*. *Mol Ecol Resour*, 10.1111/1755-0998.12181.
- Hahn MW, White BJ, Muir CD, Besansky NJ (2012) No evidence for biased co-transmission of speciation islands in *Anopheles gambiae*. *Philos Trans R Soc Lond B Biol Sci* 367(1587):374–384.
- Riehle MM, et al. (2011) A cryptic subgroup of *Anopheles gambiae* is highly susceptible to human malaria parasites. *Science* 331(6017):596–598.
- Elseth GD, Baumgardner KD (1981) *Population Biology* (D. Van Nostrand, New York), p xvi.
- Aldridge G (2005) Variation in frequency of hybrids and spatial structure among *Ipomopsis* (Polemoniaceae) contact sites. *New Phytol* 167(1):279–288.
- Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152(2):713–727.
- Teeter KC, et al. (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* 64(2):472–485.
- McSweeney C, New M, Lizcano G, Lu X (2010) UNDP Climate Change Country Profiles (University of Oxford, Oxford). Available at <http://country-profiles.geog.ox.ac.uk/>. Accessed September 5, 2013.
- Lehmann T, et al. (2010) Aestivation of the African malaria mosquito, *Anopheles gambiae* in the Sahel. *Am J Trop Med Hyg* 83(3):601–606.
- Touré YT, et al. (1998) The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia* 40(4):477–511.
- della Torre A, et al. (2002) Speciation within *Anopheles gambiae*—the glass is half full. *Science* 298(5591):115–117.
- Mukabayire O, et al. (2001) Patterns of DNA sequence variation in chromosomally recognized taxa of *Anopheles gambiae*: Evidence from rDNA and single-copy loci. *Insect Mol Biol* 10(1):33–46.
- Lehmann T, et al. (2003) Population structure of *Anopheles gambiae* in Africa. *J Hered* 94(2):133–147.
- Slotman MA, et al. (2006) Reduced recombination rate and genetic differentiation between the M and S forms of *Anopheles gambiae* s.s. *Genetics* 174(4):2081–2093.
- Lee Y, et al. (2013) Chromosome inversions, genomic differentiation and speciation in the African malaria mosquito *Anopheles gambiae*. *PLoS ONE* 8(3):e57887.
- Ellegren H, et al. (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491(7426):756–760.
- Harr B, Price T (2012) Speciation: Clash of the genomes. *Curr Biol* 22(24):R1044–R1046.
- Saetre GP, et al. (2003) Sex chromosome evolution and speciation in *Ficedula* flycatchers. *Proc Biol Sci* 270(1510):53–59.
- Garrigan D, et al. (2012) Genome sequencing reveals complex speciation in the *Drosophila* simulans clade. *Genome Res* 22(8):1499–1511.
- Martin SH, et al. (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res* 23(11):1817–1828.
- Wu CI, Ting CT (2004) Genes and speciation. *Nat Rev Genet* 5(2):114–122.
- Mallet J, Beltrán M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: The species boundary as a continuum. *BMC Evol Biol* 7:28.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Mol Ecol* 18(3):375–402.
- Peccoud J, Ollivier A, Plantegenest M, Simon JC (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc Natl Acad Sci USA* 106(18):7495–7500.
- Merrill RM, et al. (2011) Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution* 65(5):1489–1500.
- Seehausen O, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol Ecol* 17(1):30–44.
- Taylor EB, et al. (2006) Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Mol Ecol* 15(2):343–355.
- Ayala FJ, Coluzzi M (2005) Chromosome speciation: Humans, *Drosophila*, and mosquitoes. *Proc Natl Acad Sci USA* 102(Suppl 1):6535–6542.
- Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V (2002) A polyploid chromosome analysis of the *Anopheles gambiae* species complex. *Science* 298(5597):1415–1418.
- Costantini C, et al. (2009) Living at the edge: Biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecol* 9:16.
- Abbott R, et al. (2013) Hybridization and speciation. *J Evol Biol* 26(2):229–246.
- Scott JA, Brogdon WG, Collins FH (1993) Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am J Trop Med Hyg* 49(4):520–529.
- Huang Q, Shete S, Swartz M, Amos CI (2005) Examining the effect of linkage disequilibrium on multipoint linkage analysis. *BMC Genet* 6(Suppl 1):S83.
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567.