DOI: 10.1534/genetics.109.105817

The Population Genomics of Trans-Specific Inversion Polymorphisms in *Anopheles gambiae*

Bradley J. White, Changde Cheng, Djibril Sangaré,[1] Neil F. Lobo, Frank H. Collins and Nora J. Besansky[2]

Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556

Manuscript received June 2, 2009
Accepted for publication July 1, 2009

ABSTRACT

In the malaria mosquito *Anopheles gambiae* polymorphic chromosomal inversions may play an important role in adaptation to environmental variation. Recently, we used microarray-based divergence mapping combined with targeted resequencing to map nucleotide differentiation between alternative arrangements of the 2L*a* inversion. Here, we applied the same technique to four different polymorphic inversions on the 2R chromosome of *An. gambiae*. Surprisingly, divergence was much lower between alternative arrangements for all 2R inversions when compared to the 2L*a* inversion. For one of the rearrangements, 2R*u*, we successfully mapped a very small region (∼100 kb) of elevated divergence. For the other three rearrangements, we did not identify any regions of significantly high divergence, despite ample independent evidence from natural populations of geographic clines and seasonal cycling, and stable heterotic polymorphisms in laboratory populations. If these inversions are the targets of selection as hypothesized, we suggest that divergence between rearrangements may have escaped detection due to retained ancestral polymorphism in the case of the youngest 2R rearrangements and to extensive gene flux in the older 2R inversion systems that segregate in both *An. gambiae* and its sibling species *An. arabiensis*.

MORE than 70 years ago Dobzhansky and Sturtevant (1938) first discovered polymorphic inversion arrangements carried by various *Drosophila pseudoobscura* populations. After observing correlations between environmental conditions and inversion frequencies, Dobzhansky proposed that inversions are under strong selection due to their role in promoting local adaptation to the heterogeneous conditions a species encounters both spatially and temporally (Dobzhansky 1944, 1948; Powell 1997). More recent studies have implicated chromosomal inversions in the adaptation of a diversity of eukaryotes including humans (Coluzzi *et al.* 1979; Feder *et al.* 2003; Hoffmann *et al.* 2004; Stefansson *et al.* 2005). Long known to be common in dipteran insects, more recent HapMap data suggest that polymorphic inversions may be numerous in human populations and by extension other mammals (Bansal *et al.* 2007). Given their potential importance in facilitating adaptation, surprisingly little is known about the mechanism(s) or the genes responsible for maintaining inversion polymorphisms in natural populations.

Gene exchange between inverted and standard arrangements, although reduced, can still occur through gene flux: the action of gene conversion and multiple crossovers in inversion heterozygotes (heterokaryotypes) (Chovnick 1973; Navarro *et al.* 1997; Schaeffer and Anderson 2005). Over time allelic variation unrelated to ecological adaptation should become homogenized between arrangements, while alleles which are under divergent selection pressures should remain in linkage disequilibrium with each other and with the inversion itself, leading to heightened differentiation between standard and inverted arrangements at and near the target loci. In principle, this process allows the identification of specific loci involved in adaptive divergence (Schaeffer *et al.* 2003; Schaeffer and Anderson 2005; Storz 2005). Consistent with this model, previous low-resolution studies of Drosophila inversions revealed heterogeneous patterns of nucleotide diversity relative to divergence, as well as the interspersion of regions of high and low genetic association potentially due to the interaction of selection and gene flux (Schaeffer *et al.* 2003; Kennington *et al.* 2006; but see Munte *et al.* 2005). The application of high-resolution tools flowing from completely sequenced genomes will facilitate the mapping of genes

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.109.105817/DC1.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. FJ891066–FJ892701.

[1] *Present address:* Malaria Research and Training Center, Bamako BP1805, Mali.

[2] *Corresponding author:* Eck Institute of Global Health, Department of Biological Sciences, 317 Galvin Life Sciences Bldg., Notre Dame, IN 46556. E-mail: nbesansk@nd.edu

that are the targets of divergent natural selection within gene arrangements.

Although Drosophila has been the favored model, the African malaria vector *Anopheles gambiae sensu stricto* also provides an excellent system for studying the maintenance of inversion polymorphisms, not only within a species but across speciation events of different ages in the *An. gambiae* sibling species complex. The nominal species *An. gambiae s.s.* (hereafter, *An. gambiae*) is synanthropic: almost exclusively biting humans, resting indoors, and exploiting anthropogenic larval habitats (Coluzzi 1999). This close association with humans, vital to making *An. gambiae* one of the most proficient vectors of malaria, is likely to have been facilitated by chromosomal inversions thought to confer adaptive benefits in heterogeneous climatic and ecological settings in Africa. Seven common polymorphic inversions exist on the second chromosome. Six of these are located on the right arm (2R): *j*, *b*, *c*, *u*, *d*, and *k*, while 2L*a* is the only inversion on the left arm (Coluzzi *et al.* 2002). Facilitated by the sequenced reference genome (Holt *et al.* 2002), some of the breakpoints for these polymorphic inversions have been localized to small genomic regions (Sharakhov *et al.* 2006; Coulibaly *et al.* 2007; Sangare 2007). Most of these inversions appear to be the targets of strong selection. Five of the inversions (2L*a* and 2R*b*, -*c*, -*d*, and -*u*) are nonrandomly associated with degree of aridity; each cycles seasonally with rainfall, and all except 2R*u* form stable geographic clines in frequency from mesic forest to xeric regions bordering the Sahara (Coluzzi *et al.* 1979; Toure *et al.* 1994, 1998; Powell *et al.* 1999). Inversion 2R*j* is not clinal, but its distribution in Mali is consistent with adaptation to novel rockpool niches (Coluzzi *et al.* 1985; Manoukis *et al.* 2008).

In the *An. gambiae* species complex, inversion polymorphisms can be maintained across the boundaries of emerging and even full species. *An. gambiae* and its sibling *An. arabiensis*, strictly sympatric throughout most of their extensive ranges in sub-Saharan Africa, differ by multiple fixed chromosomal rearrangements on the X but share three chromosome 2 inversions: 2L*a*, fixed in *An. arabiensis* and polymorphic in *An. gambiae*; and 2R*b* and -*c*, polymorphic in both species (Coluzzi *et al.* 1979, 2002). Moreover, these same inversions and all other common *An. gambiae* inversions with the exception of 2R*j* are shared and polymorphic in two lineages apparently undergoing ecological speciation within *An. gambiae*—the assortatively mating M and S molecular forms (della Torre *et al.* 2002, 2005). Inversion frequencies are correlated with climatic and ecological conditions in parallel in both lineages (Costantini *et al.* 2009; Simard *et al.* 2009). Unlike the full species, the M and S incipient species are not distinguished by any fixed inversion differences. Indeed, genomewide divergence mapping between the M and S forms revealed that significant differentiation was confined to two small low-recombination regions adjacent to the centromeres of 2L and X which are distant from any inversions (Turner *et al.* 2005). Thus, in distinction to models of speciation invoking inversions as facilitating the persistence of hybridizing species (Noor *et al.* 2001; Rieseberg 2001; Ortiz-Barrientos *et al.* 2002; Navarro and Barton 2003), the *An. gambiae* data suggest that chromosome 2 inversions are not directly responsible for reproductive isolation. Instead, the same chromosome 2 inversion polymorphisms appear to confer similar ecological benefits, within and across species boundaries. A long-term research goal is to identify the mechanisms and the genes controlling these processes.

Previously we conducted the first high-density genomic scan of divergence across a chromosomal inversion (2L*a*) in *An. gambiae* (White *et al.* 2007). By hybridizing genomic DNA from S form mosquitoes homokaryotypic for alternate gene arrangements on chromosome 2L (2L*a* or 2L+^{a}) to oligonucleotide microarrays we were able to measure divergence across the 22-Mb inversion at nearly 14,000 markers. Differentiation in the rearranged region was significantly higher than in collinear portions of chromosome 2L. Between breakpoints the pattern of differentiation was heterogeneous: two genomic clusters of significantly higher divergence were identified near but not adjacent to the breakpoints. Directed resequencing within the S form confirmed these results and suggested that both clusters contained genes targeted by selection. Observed levels of linkage disequilibrium between the 2L*a* breakpoints and markers in the clusters are highly unlikely under a neutral scenario, in light of known recombination rates and plausible estimates of the age of the inversion.

The present study characterizes the patterns of genetic variation in polymorphic rearrangements on the opposite (right) arm of chromosome 2: 2R*j*, -*b*, -*c*, and -*u*. With the goal of identifying candidate genes maintaining these inversions in natural populations, we applied microarray-based divergence mapping to measure differentiation between alternative 2R arrangements. Because three of four inversions have taxonomic distributions that span incipient and/or completed speciation events, we validated the microarray findings by targeted sequencing in multiple taxa: sympatric Malian populations of *An. gambiae* M and S forms, and the sibling species *An. arabiensis*.

MATERIALS AND METHODS

Mosquito collection, identification, and DNA isolation: Indoor resting mosquitoes were collected by spray catch in August through September 2004 from five villages in Mali: Banambani (12°48′N, 08°03′W), Bancoumana (12°20′N, 08°20′W), Fanzana (13°20′N, 06°13′W), Kela (11°52′N, 08°26′W), and Moribabougou (12°41′N, 07°57′W). *An. arabiensis* samples from this 2004 collection were augmented by collections made in 1995 from Banambani and Moribabougou, kindly provided by A. della Torre. Specimens were sorted morphologically to *An. gambiae s.l.* and by gonotrophic stage.

Ovaries of half-gravid specimens were dissected and placed in individual numbered microtubes containing Carnoy's solution; the remaining carcass was preserved over desiccant in individual microtubes with the corresponding number. Polytene chromosomes were scored for inversions with reference to the *An. gambiae* polytene chromosome map (published as a poster in Science 298, October 4, 2002 by M. Coluzzi and V. Petrarca; http://www.sciencemag.org/feature/data/mosquito/pdfs/poster.pdf) following della Torre (1997). Unused ovarian material was preserved for later validation, performed blind to the original scoring.

DNA was isolated from individual carcasses using the DNeasy Extraction kit (QIAGEN, Valencia, CA). *An. arabiensis* and *An. gambiae s.s.* molecular forms were identified using an rDNA-based PCR diagnostic (Fanello *et al.* 2002). Quantity and quality of eluted DNA for each specimen was checked by spectrophotometry using the Nanodrop-1000 (Nanodrop Technologies).

Microarray hybridization and analysis: Mosquito genomic DNA hybridized to Affymetrix Anopheles/Plasmodium GeneChip microarrays came from five villages in southern Mali, principally Kela. Arrays were hybridized with genomic DNA from individual mosquitoes, in sets of five biological replicates per each of four different 2R homokaryotypes (20 arrays, of 5 mosquitoes × 4 homokaryotypes: 2R+, 2R*b*, 2R*bc*, and 2R*jbcu*). Labeling and hybridization of genomic DNA followed White *et al.* (2007). Labeled DNA from each mosquito was hybridized to 20 individual microarrays using standard protocols for eukaryotic cRNA hybridization. Hybridization and scanning of the arrays was performed at the Indiana University School of Medicine. Ten arrays were processed on each of 2 consecutive days under otherwise identical conditions. Groups of mosquitoes with the same karyotype were split between the 2 days to avoid any bias introduced by the hybridization and scanning.

Affymetrix CEL files containing the raw probe intensities were imported into Bioconductor (http://www.bioconductor.org). Using the "affy" program, data quality was assessed by examining the distribution of raw probe intensities using both a histogram and box plot (Bolstad *et al.* 2005). Comparison between chips showed two with irregularly low hybridization intensities. Examination of chip images using raw and normalized probe intensities revealed that the same two aberrant chips had unusually low hybridization intensities across their surfaces (Irizarry *et al.* 2003). Both of these chips had been hybridized with specimens carrying the 2R*b* homokaryotype. After discarding these two chips, an insufficient number of biological replicates ($n = 3$) remained for this karyotype, leading us to eliminate it from further analysis.

Background adjustment and quantile normalization using the robust multiarray average (RMA) without summarization by probe set were performed (Irizarry *et al.* 2003). Probe level data were exported as comma separated value files for importation into Excel and are available from B.J.W. upon request. All Anopheles probes from the Anopheles/Plasmodium GeneChip were mapped against the AgamP3 assembly and filtered to remove those that exactly matched multiple genomic locations or had secondary one-off mismatches. For each of the 151,213 retained probes, a two-tailed *t*-test was performed to compare the background adjusted and normalized probe intensities obtained from the two alternative genetic classes being compared. Probes whose signal intensities differed at $P < 0.01$ between classes were considered to have single feature polymorphisms (SFPs) between the two groups (Turner *et al.* 2005; White *et al.* 2007).

The ~61.5-Mb chromosome arm 2R is interrogated by 43,363 probes. In order of size and distal-to-proximal location on 2R, inversion *j* spans ~12.5 Mb and 9,611 probes, inversion *b* ~8 Mb and 6,218 probes, inversion *c* ~4.5 Mb and 4,003 probes, and inversion *u* ~4 Mb and 3,013 probes (Coulibaly *et al.* 2007; Sangare 2007). To gauge whether SFPs were overrepresented on 2R, we used a χ^2 test to compare the observed and expected number of SFPs on 2R *vs.* all other chromosomes combined (2L+3R+3L+X; note that these other chromosome arms were devoid of inversion polymorphisms). We used the genomewide frequency of SFPs to determine the expected number for both categories. Similarly, overrepresentation of SFPs in rearranged *vs.* collinear regions of 2R was tested by comparing observed and expected numbers based on the 2R-specific proportion of significant probes. To test for significant clustering of SFPs across 2R we performed a sliding window analysis with a window size of 300 probes and a step size of 20 probes. Each window was tested (χ^2) for an excess of SFPs compared to the number expected on the basis of the 2R-specific frequency of significant probes. Significance was evaluated after Bonferroni correction for multiple tests (conservative because windows are correlated).

DNA sequencing and analysis: To shed light on the underlying evolutionary processes shaping the patterns observed from microarray divergence mapping, we performed targeted resequencing across the 2R chromosome. Chromosomal inversions have a complex history in *An. gambiae*, as they are shared across molecular forms and species boundaries (Della Torre *et al.* 1997; Coluzzi *et al.* 2002; Besansky *et al.* 2003). To account for this complexity, sequence data were obtained from Malian samples of both M and S molecular forms homokaryotypic for alternative arrangements of three 2R inversions: *b*, *c*, and *u*. As 2R*j* is exclusive to the S form (Della Torre *et al.* 2005), corresponding M sequences were obtained only for the standard arrangement. In addition, sequences were determined from Malian samples of the sibling species *An. arabiensis*. Subject to constraints of availability, average sample size per sequenced locus was 30 chromosomes for S form (range, 16–46), 17 for M form (6–24), and 16 for *An. arabiensis* (10–22), resulting in a total resequencing effort of ~1 Mb across 17 genes.

To avoid the complication of heterozygous insertion–deletions commonly encountered within introns of these species, we targeted exon segments of ~600 bp (range, 410–740 bp) from 17 genes on chromosome 2R predicted from the AgamP3.4 gene annotation (Figure 1). Among these were three genes near the estimated breakpoints of the *b*, *c*, and *u* inversions, for the purpose of dating their origin in *An. gambiae*. At least two additional genes distant from the breakpoints were sequenced for each inversion.

Primers targeting exons were designed using Primer3 software (Rozen and Skaletsky 2000) and custom synthesized (Invitrogen). Primer sequences and their corresponding VectorBase gene identifier are given in supporting information, Table S1. PCR was carried out in 25-μl reactions following the conditions of White *et al.* (2007). PCR products were purified using ExoSAP-IT (USB, Cleveland, OH) or the GeneClean Spin kit (MP Biomedicals) following excision of bands separated on a 1.5% agarose gel.

PCR products were directly sequenced on both strands using an Applied Biosystems 3730xl DNA Analyzer and Big Dye Terminator v3.1 chemistry. Electropherograms were trimmed and visually inspected for heterozygous SNPs and indels using SeqMan II (DNASTAR, Madison, WI). Because direct sequencing from diploid mosquitoes produced gene sequences whose polymorphic sites were not phased, haplotypes were inferred for each set of gene sequences using PHASE 2.1 software (Stephens *et al.* 2001; Stephens and Donnelly 2003). Sequences are available from GenBank under accession nos. FJ891066–FJ892701.

DnaSP version 4.2.02 (Rozas *et al.* 2003) and Arlequin version 3.1 (Excoffier *et al.* 2005) were used to calculate standard polymorphism and divergence statistics. Significance was determined from 10,000 permutations (Arlequin) or 10,000 coalescent simulations without recombination (DnaSP). Negative values for D_a and F_{ST} were reported as zero.

Hierarchical AMOVA were conducted in Arlequin. In the models, differentiation was represented by the mutational (pairwise nucleotide) distance between haplotypes for each gene in 2R*b*, -*c*, and *u* rearrangements. Two groups were defined: standard or inverted. Within groups, populations were defined by molecular form, M or S.

To evaluate whether Tajima's *D* values (Tajima 1989) for genes near inversion breakpoints were significantly lower than those of genes residing in the same inversion but more distant from the breakpoints, we implemented a modification of the heterogeneity test proposed by Hahn *et al.* (2002). Aimed at distinguishing demographic from selective events that may both produce an excess of rare mutations, the test compares the distribution of mutations at synonymous *vs.* nonsynonymous sites within a gene. However, the heterogeneity test also can be used to test for differences in Tajima's *D* between any two sets of nucleotide variation data (M. W. Hahn, personal communication). Here, the distribution of mutations was compared between gene sequences at breakpoints and concatenated sequences from nonbreakpoint genes, excluding mosquitoes for which the data set was incomplete.

To estimate the time to the most recent common ancestor for the 2R*b*, -*c*, and -*u* arrangements in *An. gambiae*, we used the expectation $E[T_{\mathrm{MRCA}}] = 4N_e f(1 - n_i^{-1})$, which is based on the number of segregating sites unique to each of the inverted and standard arrangements (Andolfatto *et al.* 1999); polymorphisms shared between arrangements and species were removed prior to calculation. This estimate assumes that each inversion instantly rose to its equilibrium (current) frequency immediately after entering the population. Violation of these assumptions makes $E[T_{\mathrm{MRCA}}]$ a minimum estimate of inversion age and, as noted by Andolfatto *et al.* (1999), the estimates have very large variances.

RESULTS

On chromosomes bearing inversion polymorphisms, higher divergence is expected between their rearranged *vs.* collinear regions, because of suppressed recombination between alternative gene arrangements. In our previous study of 2L in the S form of *An. gambiae*, divergence mapping by microarray revealed much higher differentiation between rearrangements of 2L*a* than collinear parts of that arm or elsewhere in the genome (White *et al.* 2007). In the present study, microarray hybridization of *An. gambiae* samples indicated that divergence on 2R also was consistently higher in rearranged compared to collinear regions ($P = 0.003$). However, rearranged 2R regions contained only 1.16 times more SFPs than collinear regions, while this ratio was >2 for the same comparison on 2L (Table 1). Among the four inversions we examined, 2R*b* and 2R*u* had marginally higher levels of differentiation between rearrangements than 2R*j* and 2R*c* although neither approached the level observed for alternative rearrangements of the 2L*a* inversion.

TABLE 1

Divergence in rearranged relative to collinear regions on chromosome 2 of *Anopheles gambiae*

Rearranged region	Divergence ratio[a]
2L*a*	2.475***
2R (all four)	1.204**
2R*j*	1.045 NS
2R*b*	1.351***
2R*c*	1.081 NS
2R*u*	1.340*

[a] Divergence (measured by proportion of probes with SFPs) between rearranged regions divided by divergence between collinear regions on the same chromosome arm. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, $P > 0.05$.

Apart from differences in overall divergence levels between rearranged and collinear regions, we sought to map locations with unusually high levels of divergence on 2R as a guide to identifying regions that may be responsible for the maintenance of inversions in populations. To test for regional clustering of SFPs, we performed a sliding window analysis of divergence on the basis of the comparative hybridization of 2R*jbcu* and $2R+^{j}+^{b}+^{c}+^{u}$ karyotypes of *An. gambiae* (Figure 1). This revealed a small region within the 2R*u* rearrangement that contained significantly more SFPs (14) than expected by chance, given the 2R-specific proportion of SFPs observed in this experiment and the number of windows tested ($P < 0.05$ after Bonferroni correction). If this region contains sequences that contribute to the maintenance of 2R*u*, we expect a peak of statistically significant divergence at the same position in comparisons between other *An. gambiae* genetic backgrounds, but only if they differ by the alternative arrangements 2R*u* and $2R+^{u}$. Comparison between genetic backgrounds that carry the same 2R*u* arrangement are not expected to show a peak of genetic divergence at this position. Our results confirmed these predictions for 2R*u* (Figure S1; Figure S2).

Inspection of the pattern of divergence across the 2R*u* rearrangement shows that the significant cluster of SFPs is near, yet distinct from, the distal breakpoint (Figure 1, Figure S1). Although the data do not allow for precise localization, the cluster spans a region of up to 250 kb between chromosomal coordinates ~32.25–32.50 Mb. From its midpoint, this cluster lies ~875 kb from the distal breakpoint. This 250-kb cluster contains 29 predicted genes whose location and inferred function in VectorBase (if any) is given in Table S2.

Unlike the results for 2R*u*, no significant regional clustering of SFPs was observed within the other three rearranged regions on 2R; the marginally significant peak within the 2R*b* rearrangement seen in Figure 1 was not validated in a separate comparison of alternative 2R*b* arrangements (Figure S2).

Overall, these results contrasted with our previous study of the 2L*a* rearrangement, where we observed

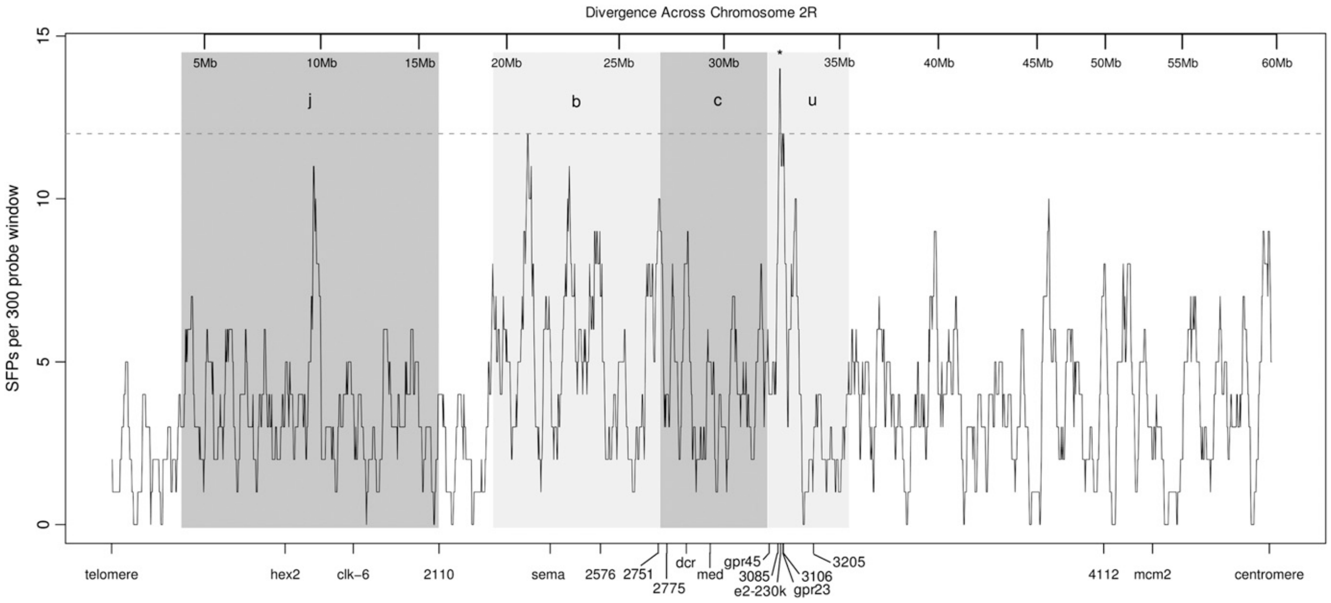

Figure 1.—Sliding window analysis of divergence across the 2R chromosome between standard and inverted karyotypes, 2R+ *vs.* 2R*jbcu*, measured in terms of proportion of SFPs per 300 probe window. Shaded areas represent chromosomal rearrangements. Horizontal dashed line is the significance threshold at 12 SFPs per window. A star denotes the significantly diverged genomic cluster in 2R*u*. Beneath the plot are indicated the sequenced loci and their relative positions along chromosome 2R.

substantially greater differentiation between rearranged and collinear regions, and greater heterogeneity in divergence across the inversion, allowing for the localization of two nonadjacent significantly differentiated regions putatively maintaining the polymorphism.

In an effort to validate and illuminate the microarray results, particularly in light of the contrast between 2R and 2L*a* rearrangements, we performed targeted sequencing of genes within and outside of rearranged regions. The expectation is that a newly arising inversion carries no variation, but its diversity recovers with age through mutation and gene flux (Chovnick 1973; Navarro *et al.* 2000; Andolfatto *et al.* 2001). Diversity is expected to recover more rapidly in the center of the inversion, away from the breakpoints where recovery is suppressed by low rates of gene flux (Chovnick 1973; Andolfatto *et al.* 2001; Schaeffer and Anderson 2005).

Age of rearrangements: To estimate age of the inversions, we used nucleotide variation data from genes near the breakpoints of *b*, *c*, and *u* (*2751*, *2775*, and *gpr45*, respectively) in the *An. gambiae* S form following Andolfatto *et al.* (1999) (Figure 1; Table 2). Taken together with a previous estimate of the age of 2R*j* on the basis of the same population sample (Coulibaly *et al.* 2007), the data agree with expectations on the basis of geographic and taxonomic distributions. They suggest that the 2R*j* and -*u* inversions are the youngest—at ~0.4 N_e and ~1.7 N_e generations, respectively—while the trans-specific inversions 2R*b* and -*c* both share a much older origin at ~2.6 N_e generations. It is remarkable that the other trans-specific inversion (2L*a*) dates to approximately the same time, ~2.7 N_e (White *et al.* 2007).

DNA sequence polymorphism: Concordant with their relative ages, nucleotide diversity was reduced by as much as twofold in the inverted relative to standard arrangements of the younger inversions 2R*j* and -*u*, but no pronounced or consistent reduction in diversity was observed between alternative arrangements for the older *b* and *c* inversions within M and S forms (Figure 2; Table 2). Similarly, the expected trend of lower diversity near inversion breakpoints relative to more central locations was observed most clearly for the youngest *j* inversion at locus *2110*, only 10 kb from the proximal breakpoint. For the other inversions, the trend (if present at all) was weaker, perhaps due not only to their greater age but also to the longer distance between the breakpoint and the gene that was surveyed (Table 2).

Genes in breakpoint regions where flux is lowest should show strong skews in the mutation frequency spectrum (Andolfatto *et al.* 2001). Skews toward rare mutations were apparent from the preponderance of negative values for Tajima's *D* statistic (Tajima 1989) in Table 2. However, as in previous studies on *An. gambiae*, the *D* values were negative regardless of molecular form, rearrangement, or chromosomal location, consistent with a proposed population expansion within the past ~10,000 years (Donnelly *et al.* 2001, 2002). Yet only four *D* values were significantly negative, three of which were associated with loci adjacent to the breakpoints of an inverted arrangement: *2110* (S form, *j* inversion), *2751* (S form, *b* inversion), and *gpr45* (M form, *u* inversion). To help distinguish heterogeneous selective

TABLE 2

Polymorphism statistics for 17 genes on chromosome 2R by arrangement class and molecular form of *An. gambiae*

Arrangement	Gene	Position (Mb)	Len	Form	*K*	*N*	*S*	π (%)	θ (%)	*D*
2R*j*	(distal BP)	3.26								
	hex2	7.55	731	S	*j*	38	19	0.496	0.619	−0.656
				S	$+^{j}$	16	21	0.578	0.866	−1.342
				M	$+^{j}$	16	12	0.459	0.495	−0.273
	clk-6	11.45	610	S	*j*	42	8	0.298	0.305	−0.065
				S	$+^{j}$	22	16	0.485	0.765	−1.182
				M	$+^{j}$	10	11	0.550	0.637	−0.614
	*2110**	15.74	410	S	*j*	22	2	0.040	0.130	−1.515*
				S	$+^{j}$	20	16	0.750	1.090	−1.075
	(proximal BP)	15.75								
2R*b*	(distal BP)	18.50								
	sema	21.20	610	S	*b*	28	47	1.870	2.022	−0.209
				S	$+^{b}$	24	35	1.747	1.668	0.523
				M	*b*	16	32	1.827	1.630	0.645
				M	$+^{b}$	16	40	1.975	1.976	−0.002
	2576	23.11	660	S	*b*	32	42	0.938	1.655	−1.484
				S	$+^{b}$	26	23	0.555	0.953	−1.426
				M	*b*	20	29	0.971	1.367	−0.844
				M	$+^{b}$	12	12	0.471	0.602	−0.917
	2751	26.66	633	S	*b*	38	26	0.401	0.978	*−2.022**
				S	$+^{b}$	26	15	0.369	0.621	−1.406
				M	*b*	14	11	0.372	0.546	−1.262
				M	$+^{b}$	16	12	0.465	0.571	−0.713
	(proximal BP)	26.74								
2R*c*	(distal BP)	26.78								
	2775	27.04	705	S	*c*	34	33	1.132	1.145	−0.039
				S	$+^{c}$	22	22	0.656	0.895	−0.874
				M	*c*	22	30	1.088	1.245	−0.261
				M	$+^{c}$	12	11	0.408	0.517	−0.872
	dcr	28.19	650	S	*c*	34	22	0.733	0.941	−0.392
				S	$+^{c}$	22	29	1.026	1.308	−0.620
				M	*c*	24	31	0.891	1.277	−1.143
				M	$+^{c}$	18	16	0.950	0.760	1.251
	med	29.11	655	S	*c*	28	14	0.371	0.549	−1.098
				S	$+^{c}$	22	27	0.961	1.131	−0.572
				M	*c*	16	18	0.613	0.828	−1.035
				M	$+^{c}$	6	11	0.743	0.735	0.062
	(proximal BP)	31.45								
2R*u*	(distal BP)	31.48								
	gpr45	31.52	671	S	*u*	42	21	0.451	0.727	−1.252
				S	$+^{u}$	24	23	0.614	0.918	−1.221
				M	*u*	18	8	0.147	0.347	−2.001**
				M	$+^{u}$	24	24	0.631	0.958	−1.264
	3085	32.33	740	S	*u*	38	24	1.011	0.772	1.052
				S	$+^{u}$	18	27	0.722	1.061	−1.272
				M	*u*	18	11	0.477	0.432	0.378
				M	$+^{u}$	22	25	0.834	0.917	−0.381
	e2-230k	32.39	590	S	*u*	38	16	0.922	0.645	1.393
				S	$+^{u}$	22	25	1.131	1.162	−0.101
				M	*u*	14	11	0.689	0.586	0.692
				M	$+^{u}$	20	30	1.255	1.433	−0.487
	gpr23	32.43	567	S	*u*	42	16	0.560	0.656	−0.467
				S	$+^{u}$	22	22	0.943	1.046	−0.428
				M	*u*	14	11	0.519	0.610	−0.586
				M	$+^{u}$	20	22	0.807	1.094	−1.006

(*continued*)

TABLE 2

(Continued)

Arrangement	Gene	Position (Mb)	Len	Form	*K*	*N*	*S*	π (%)	θ (%)	*D*
	3106	32.44	686	S	*u*	40	19	0.484	0.651	−0.844
				S	+^{u}	22	22	0.777	0.880	−0.436
				M	*u*	18	13	0.505	0.551	−0.311
				M	+^{u}	24	15	0.634	0.586	0.294
	3205	33.82	666	S	*u*	46	13	0.210	0.478	−1.602*
				S	+^{u}	20	30	0.985	1.322	−0.880
				M	*u*	12	16	0.571	0.796	−1.219
				M	+^{u}	24	18	0.781	0.724	0.286
	(proximal BP)	35.50								
Colinear	*4112*	50.09	625	S		44	27	0.977	1.030	−0.174
				M		18	20	0.799	0.930	−0.551
	mcm2	53.56	623	S		46	35	1.134	1.278	−0.385
				M		24	22	0.860	0.946	−0.333

Len, sequence length in bp; form, molecular form of *An. gambiae*; *K*, karyotype; *N*, number of chromosomes sampled; *S*, segregating sites; π, expected heterozygosity per site on the basis of average pairwise differences; θ, expected heterozygosity per site on the basis of the number of segregating sites; *D*, Tajima's *D* on the basis of total number of mutations; BP, breakpoint. *$P < 0.05$; **$P < 0.01$.

processes from the effects of more homogeneous demographic processes, heterogeneity among *D* estimates at different loci was assessed through two approaches. Where possible, we implemented a modified version of the test by Hahn *et al.* (2002) to demonstrate that *D* values for these breakpoint-adjacent genes were significantly different than those for genes located farther away inside the inversion (for *2751* in the S form, $P = 0.04$; for *gpr45* in the M form, $P = 0.045$; the test could not be performed for *2110* because sequence was derived from a different subsample of mosquitoes). Additionally, we explored the ratio of Tajima's *D* to its theoretical minimum value $D_{\min}$ as proposed by Schaeffer (2002). A plot of this statistic indicates that the same breakpoint-proximal loci gave the most extreme $D/D_{\min}$ values (near the theoretical minimum) (Figure S3).

Sequence divergence: We estimated sequence divergence in terms of numbers of shared polymorphisms *vs.* fixed differences between alternative arrangements, as well as two indices of differentiation, F_{ST} and D_{a}. As expected from the microarray results, the numbers of polymorphisms shared between collinear regions was higher than the numbers shared between rearranged regions, in all except three genes ("Inverted-Standard" columns in Table 3). Also expected, collinear regions

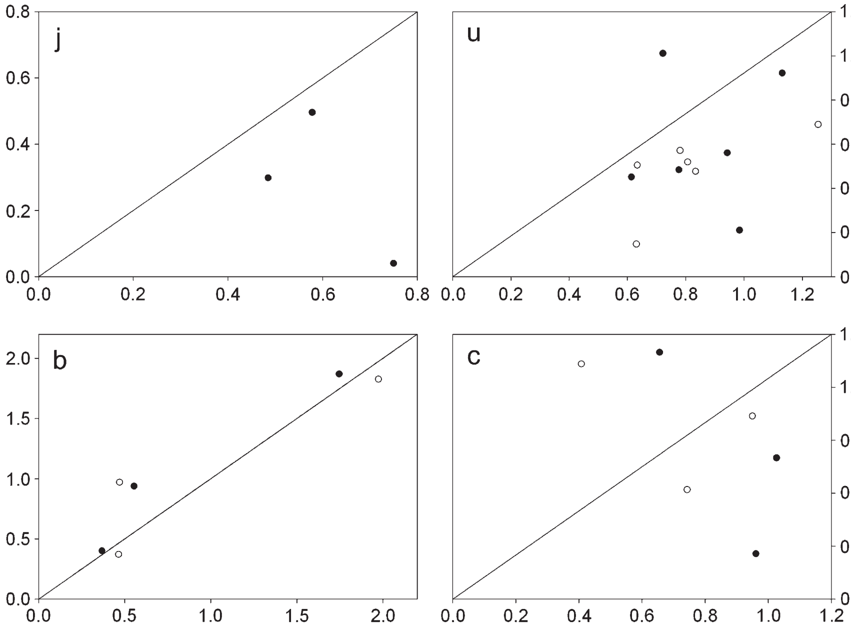

Figure 2.—Comparison of nucleotide diversity (%) between standard (*x*-axis) and inverted (*y*-axis) arrangements for each of four polymorphic 2R inversions in *An. gambiae*. Each circle represents a gene sequenced in either the M form (open) or the S form (solid). Circles falling below the *y*-*x* line indicate reduced diversity in the inverted arrangement.

TABLE 3

Divergence between alternative chromosomal arrangements within molecular forms (inverted standard) and between molecular forms within the same chromosomal arrangement (M-S)

		Position	S form, inverted standard				M form, inverted standard				Standard, M-S				Inverted: M-S			
Arrangement	Gene	(Mb)	*S*	*F*	D_a (%)	F_{ST}	*S*	*F*	D_a (%)	F_{ST}	*S*	*F*	D_a (%)	F_{ST}	*S*	*F*	D_a (%)	F_{ST}
2R*j*	(distal BP)	3.26																
	hex2	7.55	3	0	0.054	0.091*	—	—	—	—	5	0	0.062	0.106**	—	—	—	—
	clk-6	11.45	5	0	0.012	0.037	—	—	—	—	4	0	0.020	0.041	—	—	—	—
	2110	15.74	0	0	0.408	0.567***	—	—	—	—	—	—	—	—	—	—	—	—
	(proximal BP)	15.75																
2R*b*	(distal BP)	18.50																
	sema	21.20	28	0	0.426	0.190***	19	0	0.336	0.150***	18	0	0.258	0.124***	26	0	0.188	0.092**
	2576	23.11	7	0	0.255	0.249***	6	0	0.246	0.232***	2	0	0.271	0.337***	13	0	0.177	0.157***
	2751	26.66	5	0	0.009	0.021	1	0	0.167	0.284***	3	0	0.133	0.249***	3	0	0.050	0.112***
	(proximal BP)	26.74																
2R*c*	(distal BP)	26.78																
	2775	27.04	13	0	0.235	0.196***	8	0	0.208	0.186***	6	0	0.020	0.026	17	0	0.030	0.026
	dcr	28.19	13	0	0.208	0.199***	9	0	0.159	0.149***	12	0	0.124	0.110**	15	0	0.332	0.334***
	med	29.11	7	0	0.250	0.273***	7	0	0.095	0.123***	6	0	0.138	0.119*	6	0	0.313	0.409***
	(proximal BP)	31.45																
2R*u*	(distal BP)	31.48																
	gpr45	31.52	2	0	0.420	0.453***	4	1	0.557	0.574***	7	0	0.080	0.114***	3	0	0.411	0.529***
	3085	**32.33**	**6**	**0**	**0.108**	**0.102*****	**5**	**0**	**0.141**	**0.171*****	**10**	**0**	**0.030**	**0.037**	**6**	**0**	**0.109**	**0.107*****
	e2-230k	**32.39**	**6**	**2**	**1.00**	**0.502*****	**6**	**2**	**1.220**	**0.542*****	**13**	**0**	**0.214**	**0.152*****	**7**	**0**	**0.356**	**0.289*****
	gpr23	**32.43**	**7**	**0**	**0.273**	**0.287*****	**4**	**0**	**0.454**	**0.395*****	**14**	**0**	**0.084**	**0.087*****	**5**	**0**	**0.099**	**0.152*****
	3106	**32.44**	**12**	**0**	**0.086**	**0.131*****	**9**	**0**	**0.236**	**0.289*****	**10**	**0**	**0.131**	**0.157*****	**9**	**0**	**0.013**	**0.026**
	3205	33.82	6	0	0.203	0.327***	10	0	0.330	0.313***	14	0	0.043	0.048*	4	0	0.399	0.593***
	(proximal BP)	35.50																
Colinear	*4112*	50.09	13	0	0.001	0.008	9	0	0	0	8	0	0.126	0.137**	9	0	0	0
	mcm2	53.56	20	0	0	0	13	0	0	0	11	0	0.343	0.264***	11	0	0.148	0.107*

S, number of shared polymorphisms; *F*, number of fixed differences; D_a (%), average net nucleotide divergence per site (Nei 1987); F_{ST} (Hudson *et al.* 1992), estimate of differentiation; BP, breakpoint. Boldface type indicates loci in the significantly differentiated cluster within 2R*u*. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

were not differentiated within molecular forms: F_{ST} values were not significantly different from zero, and D_a values were at or near zero. By contrast, rearranged regions were characterized by highly significant F_{ST} values and elevated D_a values, particularly within 2R*u*. Moreover, genes flanking the breakpoints for the 2R*j*, -*b*, and -*u* inversions shared fewer polymorphisms between arrangements than genes more distant from the breakpoints, though this was not invariably reflected in higher F_{ST} and D_a values.

Four genes in 2R*u* were targeted for resequencing within the only significant cluster of SFPs identified by microarray. One of these, *e2-230k*, is represented by an expressed sequence tag in the AgamP3.4 assembly (ENSANGEST00000008826) that partially overlaps a predicted gene annotated as AGAP003090. At this locus, the level of net nucleotide divergence (D_a) between alternative arrangements was more than twice that seen at any other sequenced locus on 2R, and F_{ST} values also were much higher, the two exceptions being genes located close to the 2R*u* and -*j* inversion breakpoints. In addition, two of only three fixed nucleotide differences between alternative arrangements observed in our entire data set were identified in *e2-230k*, in both M and S population samples. These were in perfect linkage disequilibrium with the inversion despite a distance of nearly 1 Mb from the 2R*u* distal breakpoint. Such linkage is unlikely to have persisted as a remnant of complete inversionwide linkage disequilibrium established when the rearrangement first arose. On the basis of rates of recombination estimated in the 2L*a* rearrangement (Stump *et al.* 2007), the recombination fraction between *e2-230k* and the 2R*u* distal breakpoint is expected to be roughly $r = 0.007$. Given complete linkage disequilibrium upon origin of the inversion, in which fixed differences have linkage disequilibrium values of $D_0 = 0.25$, D_t should decay in the absence of selection to 0.001 in <800 generations (Lewontin and Kojima 1960). Assuming that 2R*u* arose ~1.7 N_e generations ago, and applying a conservative microsatellite-based estimate of long-term N_e of ~6500 (Pinto *et al.* 2003), the age of 2R*u* is at least 11,000 generations, a time-frame larger by an order of magnitude than the time required for linkage disequilibrium to decay. This result is consistent with natural selection maintaining the association between *e2-230k* (and presumably other genes not sequenced in the diverged cluster) with the distal breakpoint, and therefore the 2R*u* inversion itself.

Notwithstanding the expected overall pattern of significant differentiation between rearranged but not collinear regions on both arms of chromosome 2, we were struck by the much greater magnitude of differentiation between rearranged regions of 2L *vs.* 2R—a pattern consistent between microarray and sequencing results. On 2L, fixed nucleotide differences were found between arrangements at five of nine sequenced genes, and average F_{ST} and D_a values were relatively high (0.527 and 1.66%, respectively; White *et al.* 2007). On 2R, no fixed differences were found between arrangements except at two genes located within the 2R*u* inversion. Moreover, F_{ST} and D_a values also were much higher between 2L than between 2R rearrangements. In the S form, the average F_{ST} value for 2R rearranged regions was 0.242, more than twofold lower than the value for 2L*a* rearrangements. Even more striking, the corresponding average D_a value on 2R was approximately sixfold lower, 0.263%. (In the M form, 2R estimates were similar, but no 2L sequences from this form are available for comparison).

Species boundaries and chromosome rearrangements: Three of the four rearrangements considered on 2R are shared across taxonomic boundaries, a condition that we refer to here as "trans-specific" regardless of sharing between incipient or full species. In principle, the sharing of gene arrangements between taxa may occur due to multiple independent origins, or alternatively to a unique origin followed by retention of ancestral polymorphism and/or secondary introgression between taxa. Previous evidence suggests that 2L and 2R inversions likely share a common origin (Coluzzi *et al.* 1979; della Torre *et al.* 1997; Besansky *et al.* 2003). Our results (below) are consistent with this interpretation.

In the M and S forms, the standard arrangements from different molecular forms are more closely related to each other than to alternative arrangements within the same form; the same is true for inverted arrangements (Table 3). The most parsimonious explanation is that the rearrangements predate the origin of M and S forms, implying a unique origin of a shared polymorphism. However, it is notable that most comparisons between M and S within the same arrangement class revealed significant and moderately large values of F_{ST} (~0.1–0.5). Moreover, at genes within 2R*c* and -*u*, nucleotide divergence between M and S was almost invariably higher for inverted than for standard chromosomes (paired *t*-test, $P = 0.035$). This result is reinforced by hierarchical analyses of molecular variance (AMOVA, Table 4), which indicated a significant contribution of variance by M *vs.* S sequences within the same arrangement class, but not between arrangement classes. These results most likely reflect the fact that the combined forces of selection and drift (especially in the inverted arrangement) are stronger than migration due to interform gene flow, suggesting that rearrangements may be evolving largely independently in M and S forms.

Between the sibling species *An. gambiae* and *An. arabiensis*, a complete comparison of rearrangements analogous to that performed for the M and S forms was not possible, because our sample of *An. arabiensis* contained only one arrangement for each rearranged region considered. *An. arabiensis* sequences determined from seven 2R and six 2L genes (supplemented by existing 2L data; White *et al.* 2007), were compared to

TABLE 4

Hierarchical AMOVA of the contribution of chromosomal inversions and molecular forms to genetic structure in *An. gambiae*

Locus	Source of Variation	% variation
2R*b*	Between arrangement classes	0–7 NS
	Among forms within arrangement class	10–20***
	Within forms and arrangement class	73–84***
2R*c*	Between arrangement classes	0–18 NS
	Among forms within arrangement class	2 (NS)–31***
	Within forms and arrangement class	75–80***
2R*u*	Between arrangement classes	6–40 NS
	Among forms within arrangement class	7–26***
	Within forms and arrangement class	47–87***

Significance of variance component: NS, not significant; $***P < 0.001$.

corresponding gene sequences of the *An. gambiae* S form. Similar to the pattern observed for the incipient species, same-arrangement divergence was lower than alternative-arrangement divergence on both 2R and 2L (Table 5). However, the overall level of sequence differentiation was substantially larger for alternative arrangements of 2L*a* than for 2R (Figure 3). Whereas average D_a (F_{ST}) values between 2L arrangements were as high as 1.67 (0.589), those between 2R arrangements were only 0.265–0.574 (0.281–0.417). This result is even more striking because same-arrangement contrasts between sibling species did not reveal the same dichotomy between 2L and 2R: D_a and F_{ST} values for 2L*a* were within the range observed for the other inversions.

DISCUSSION

Our study was motivated by the goal of localizing regions within *An. gambiae* chromosomal inversions that are the targets of selection responsible for their maintenance and the promise of mapping those targets with oligonucleotide microarrays (Borevitz *et al.* 2003; Winzeler *et al.* 2003; Turner *et al.* 2005; White *et al.* 2007; Turner *et al.* 2008). Yet for the right arm of chromosome 2, significant divergence was detected for only one of four studied inversions (only 2R*u* and not 2R*j*, *-b*, or *-c*), and overall differentiation between 2R arrangements was sharply lower than between 2L arrangements studied previously (White *et al.* 2007). Possible reasons for these outcomes are discussed below.

TABLE 5

Divergence between chromosomal arrangements of *An. gambiae* (S form) and *An. arabiensis*

Arrangement	Gene	*S*	*F*	D_a (%)	F_{ST}	*S*	*F*	D_a (%)	F_{ST}
		Ara/2L*a*-Gam/2L*a*				Ara/2L*a*-Gam/2L+^{a}			
2L*a*	*asph*	9	0	0.362	0.274	1	7	2.223	0.716
	depcd5	9	0	0.145	0.161	7	0	0.302	0.360
	endp	17	0	0.768	0.300	3	10	3.258	0.675
	hdac	5	0	0.123	0.174	2	0	0.895	0.603
		40	0	0.350	0.227	13	17	1.670	0.589
		Ara/2R*b*-Gam/2R*b*				Ara/2R*b*-Gam/2R+^{b}			
2R*b*	*2576*	8	0	0.235	0.220	3	0	0.458	0.412
	2751	3	0	0.089	0.172	2	0	0.072	0.149
		11	0	0.162	0.196	5	0	0.265	0.281
		Ara/2R+^{c}-Gam/2R+^{c}				Ara/2R+^{c}-Gam/2R*c*			
2R*c*	*dcr*	5	0	0.562	0.375	5	0	0.723	0.478
	2775	1	0	0.200	0.275	1	0	0.424	0.356
		6	0	0.381	0.325	6	0	0.574	0.417
		Ara/2R+^{u}-Gam/2R+^{u}				Ara/2R+^{u}-Gam/2R*u*			
2R*u*	*gpr23*	4	0	0.149	0.174	3	0	0.355	0.408
	3085	1	0	0.124	0.181	1	0	0.174	0.294
	3205	7	0	0.223	0.186	4	0	0.417	0.413
		12	0	0.165	0.180	8	0	0.315	0.372
		Ara/2L*a*-Gam/2L*a*				Ara/2L*a*-Gam/2L+^{a}			
Collinear	*lys-c*	10	0	0.168	0.124	7	0	0.287	0.192
(2L)	*znf294*	12	0	0.749	0.350	10	0	0.815	0.365
		22	0	0.459	0.237	17	0	0.551	0.279

S, *F*, D_a, and F_{ST} as defined in Table 3; Ara, *An. arabiensis*; Gam, *An. gambiae*.

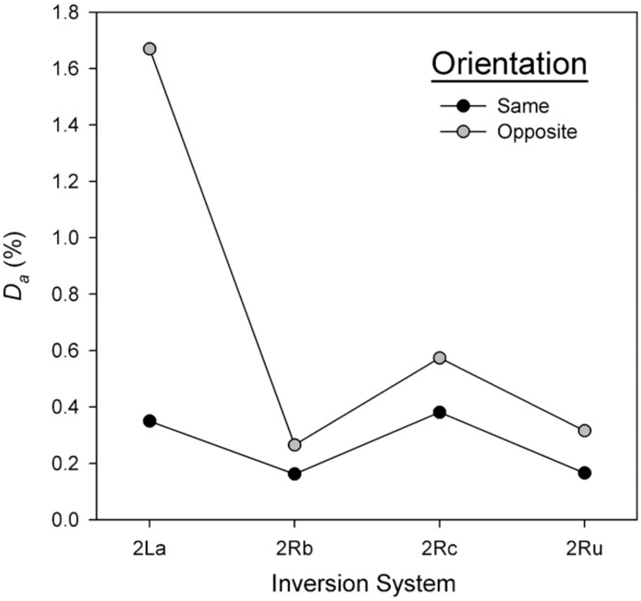

Figure 3.—Nucleotide divergence between *An. gambiae* and *An. arabiensis* in chromosome 2 inversion systems. Pairwise nucleotide divergence per site (D_a) is averaged across the genes sequenced within each inversion system.

One formal possibility is that *An. gambiae* 2R inversion polymorphisms *j*, *b*, and *c* are selectively neutral. Although this possibility cannot be dismissed on the basis of our molecular data, independent circumstantial evidence similar to that which influenced Dobzhansky (Krimbas and Powell 1992; Powell 1997) strongly favors an adaptive role for inversions in *An. gambiae*, including: (1) stable geographic clines that follow aridity gradients in distant parts of the African continent, such as Nigeria and Burkina Faso in West Africa (Coluzzi *et al.* 1979; Costantini *et al.* 2009), Cameroon in Central Africa (Simard *et al.* 2009), and the Comoros in East Africa (Petrarca *et al.* 1990); (2) microgeographic partitioning of alternative arrangements relative to degree of aridity, *e.g.*, differential indoor/outdoor resting behavior of mosquito carriers of alternative arrangements (Coluzzi *et al.* 1977); (3) seasonal cycling of inversions in association with rainfall (Coluzzi *et al.* 1979; Toure *et al.* 1998); and (4) stable heterozygote excess over Hardy-Weinberg expectations in some laboratory populations (della Torre *et al.* 1997). Schaeffer (2008) recently modeled the migration-selection balance required to maintain chromosomal polymorphisms across diverse geographic habitats in *D. pseudoobscura*, concluding that migration levels are too extensive in this species to explain observed clines in arrangement frequencies through neutral diffusion processes. Such modeling has yet to be attempted in *An. gambiae*, but the remarkable parallels to Drosophila—including high migration levels (Pinto *et al.* 2003)—suggest that a similar conclusion would not be surprising.

If selection is responsible for the maintenance of inversions 2R*j* and 2R*u* in natural populations, what can explain the failure to detect the anticipated signatures of selection by microarray divergence mapping? Under the model proposed by Kirkpatrick and Barton (2006), chromosomal inversions that capture multiple (at least two) locally adapted alleles spread because they suppress recombination with different genetic backgrounds disadvantageous under local conditions, in the face of migration or hybridization. However, these authors also suggest that the preconditions for the inversion to spread may still persist in the standard (uninverted) class of chromosomes. In other words, at least some of the locally adapted alleles captured by the inversion might continue to segregate among the standard chromosomes. Indeed, a proposed test of their hypothesis includes the expectation of linkage disequilibrium in standard chromosomes between the alleles carried by the inversion (Kirkpatrick and Barton 2006). Although such tests are beyond the scope of the present study, the possibility that beneficial alleles captured by the newly arisen inversion remain polymorphic in the parental (standard) population could explain the inability to measure divergence between arrangements of the relatively young 2R inversions using the microarray approach as applied here. Given samples of only 10 chromosomes per microarray hybridization, our array-based mapping technique lacks the sensitivity to uncover variants maintaining an inversion when those variants continue to segregate in the ancestral arrangement at frequencies much greater than 10% (*i.e.*, 1 chromosome of the 10 we hybridize). In the case of putatively recent 2R*j* and -*u* inversions, retention of beneficial alleles by the ancestral population represents a plausible explanation for the discovery of just a single candidate region within these two inversion systems—despite the theoretical prediction that inversions are maintained by adaptive benefits conferred by at least two loci.

The persistence of preconditions for the inversion to spread may be less plausible an explanation in the case of the apparently more ancient and trans-specific 2R*b* and -*c* rearrangements. Rather, multiple shared polymorphisms between alternative arrangements at every gene sequenced within the 2R*b* and -*c* rearrangements indicate extensive gene flux between both rearrangements, which could eventually winnow the region of divergence to a small size (Schaeffer *et al.* 2003; Schaeffer and Anderson 2005). The microarray that we utilized for divergence mapping contains only an average of eleven 25-mer probes per interrogated gene, and not all predicted genes are represented on the array. While this platform provides very high resolution compared to more traditional approaches with far fewer markers, genome coverage is neither complete nor uniform for the Affymetrix Anopheles chip. Furthermore, our technique for discovering highly diverged regions relied on a sliding 300-probe window, because adoption of smaller window sizes would increase the false positive discovery rate. Thus, regions of divergence that encompass only a single or even a few genes could escape detection with our method. Accordingly, we hypothesize that gene flux between alternative arrangements of 2R*b* and 2R*c* eroded divergence around adaptive variants below the resolution of the oligonucleotide array. In the future, overcoming this resolution threshold is possible with a higher density array, such as a tiling array (Turner *et al.* 2008), or by high throughput whole genome sequencing.

The extensive gene flux hypothesized between the 2R*b* and 2R*c* arrangements stands in contrast to our results for alternative 2L*a* arrangements, which exhibited only minimal genetic exchange and much higher levels of divergence (Figure 3). However, the estimated ages of these three inversions suggest that they have similar sojourn times in *An. gambiae*, which implies that gene flux between alternative arrangements should be the same for all three inversion systems. A possible explanation for this paradox rests on the fact that the sibling species *An. arabiensis* is monomorphic for 2L*a*, but polymorphic for both 2R*b* and 2R*c*. In Figure 4, we propose a speculative history for chromosomal evolution in the lineages leading to present-day *An. gambiae* and *An. arabiensis*. It reflects the fact that the lineage leading to *An. gambiae* was probably derived from a homokaryotypic standard ancestor (Ayala and Coluzzi 2005). The model posits that the sibling species were fixed for alternative 2R and 2L arrangements upon secondary contact, represented by horizontal arrows in Figure 4. The plausibility of introgressive hybridization upon secondary contact proposed in our model is supported by documented levels of female F_1 hybrids in nature between these strictly sympatric species (0.1–0.2%; White 1971; Temu *et al.* 1997) and their fertility (Davidson *et al.* 1967). Detailed arguments behind the interpretation that chromosome 2 inversions are shared between *An. arabiensis* and *An. gambiae* by introgression have been elaborated previously (see Garcia *et al.* 1996; Powell *et al.* 1999; Besansky *et al.* 2003). According to our speculative history, introgression from *An. arabiensis* brought the 2R*b*, 2R*c*, and 2L*a* arrangements into *An. gambiae*. Conversely, introgression in the opposite direction is proposed for the introduction of 2R$+^{b}$ and 2R$+^{c}$, but *not* 2L$+^{a}$ into *An. arabiensis*. The inability of the 2L$+^{a}$ to introgress from *An. gambiae* into *An. arabiensis*, in contrast to the bidirectional flow of 2R*b* and -*c* rearrangements, is supported by laboratory crossing experiments (della Torre *et al.* 1997). As a result of these semipermeable species boundaries (Besansky *et al.* 2003), both 2R rearrangements are polymorphic in both species, but the 2L rearrangement is polymorphic only in *An. gambiae*. Thus, the opportunity for gene flux is much greater between alternate arrangements of 2R than those on 2L. This process would be accelerated if multiple bidirectional introgressions of 2R arrangements had occurred. Because of low interspecific divergences relative to intraspecific diversity in the *An. gambiae* complex (*e.g.*, Obbard *et al.* 2009), the testing of this hypothesis must await solutions to complex, interrelated and as yet unsolved problems. These include the correct inference of phylogenetic relationships, the identification of an appropriate outgroup species, and the disentanglement of polymorphisms shared through recent common ancestry *vs.* introgression—problems that future genome sequencing projects could powerfully address.

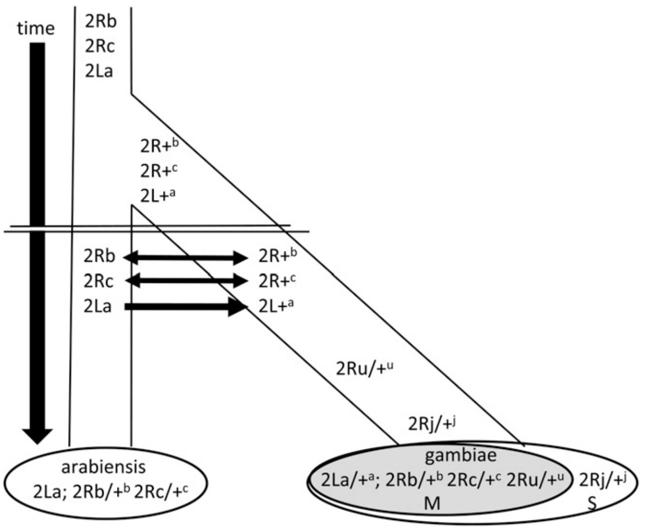

Figure 4.—Model of chromosomal inversion history in the lineages leading to present-day *An. arabiensis* and *An. gambiae* populations. Proposed bidirectional or unidirectional introgression of arrangements between species is indicated by horizontal arrows (double headed or single headed, respectively).

The single candidate region we mapped via microarray was further narrowed down to ~100 kb with targeted sequencing. This region contains only seven genes, which should make identifying candidate mutations practical in future studies. In fact, the ubiquitin-conjugating enzyme e2-230k that we sequenced is an interesting candidate not only due to the high differentiation in our sequenced samples, but also because this gene contained four SFPs in the microarray comparison—the most of any gene in the genome. Interestingly, the region of the gene we sequenced and the location of the SFPs are not the same, suggesting that this gene is highly differentiated between arrangements across its length. Unfortunately, nothing is known about the phenotypic differences between mosquitoes carrying different arrangements of 2R*u*. Unlike for 2L*a*, 2R*b*, and 2R*c*, no clines have been reported for 2R*u*; however, an ecological study in Mali revealed that the inversion cycles seasonally, increasing in frequency during the wet season and then decreasing during the dry season (Toure *et al.* 1994). Although finding candidate mutations is foreseeable, linking these mutations to phenotypic differences will be a challenging step in the path toward a more comprehensive understanding of the adaptive mechanisms underlying inversions in *An. gambiae*. Further knowledge of this fundamental aspect of *An. gambiae* biology may lead to more effective interventions against malaria.

We thank M. Kern for laboratory assistance, O. Grushko for karyotyping, and B. Cassone for R scripts. M. Coulibaly and S. Traoré managed the 2004 field collections in Mali and generously made these available to us; A. della Torre kindly provided karyotyped *An. arabiensis* in 1995. Z. Chunxiao and R. Jerome performed the microarray hybridizations under the direction of H. Edenberg at the Center for Medical Genomics, Indiana University School of Medicine. The authors are indebted to J. Feder, M. Hahn, two anonymous reviewers, and D. Begun, whose suggestions substantially improved the text. This work was funded by National Institutes of Health grant AI63508 (to N.J.B.). B.J.W. was supported by an Arthur J. Schmidt fellowship from the University of Notre Dame. The Center for Medical Genomics is funded in part by the Indiana Genomics Initiative of Indiana University (INGEN); INGEN is funded in part by The Lily Endowment.

LITERATURE CITED

Andolfatto, P., F. Depaulis and A. Navarro, 2001 Inversion polymorphisms and nucleotide variability in Drosophila. Genet. Res. **77:** 1–8.

Andolfatto, P., J. D. Wall and M. Kreitman, 1999 Unusual haplotype structure at the proximal breakpoint of In(2L)t in a natural population of *Drosophila melanogaster*. Genetics **153:** 1297–1311.

Ayala, F. J., and M. Coluzzi, 2005 Chromosome speciation: Humans, Drosophila, and mosquitoes. Proc. Natl. Acad. Sci. USA **102:** 6535–6542.

Bansal, V., A. Bashir and V. Bafna, 2007 Evidence for large inversion polymorphisms in the human genome from HapMap data. Genome Res. **17:** 219–230.

Besansky, N. J., J. Krzywinski, T. Lehmann, F. Simard, M. Kern *et al.*, 2003 Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. Proc. Natl. Acad. Sci. USA **100:** 10818–10823.

Bolstad, B. M., R. A. Irizarry, L. Gautier and Z. Wu, 2005 Preprocessing high-density oligonucleotide arrays, pp. 13–32 in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, edited by R. Gentleman, V. Carey, W. Huber, R. Irizarry and S. Dudoit. Springer-Verlag, New York.

Borevitz, J. O., D. Liang, D. Plouffe, H. S. Chang, T. Zhu *et al.*, 2003 Large-scale identification of single-feature polymorphisms in complex genomes. Genome Res. **13:** 513–523.

Chovnick, A., 1973 Gene conversion and transfer of genetic information within the inverted region of inversion heterozygotes. Genetics **75:** 123–131.

Coluzzi, M., 1999 The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of *Plasmodium falciparum*. Parassitologia **41:** 277–283.

Coluzzi, M., V. Petrarca and M. A. DiDeco, 1985 Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. Boll. Zool. **52:** 45–63.

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.

Coluzzi, M., A. Sabatini, V. Petrarca and M. A. Di Deco, 1977 Behavioural divergences between mosquitoes with different inversion karyotypes in polymorphic populations of the *Anopheles gambiae* complex. Nature **266:** 832–833.

Coluzzi, M., A. Sabatini, V. Petrarca and M. A. Di Deco, 1979 Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. Trans. R. Soc. Trop. Med. Hyg. **73:** 483–497.

Costantini, C., D. Ayala, W. M. Guelbeogo, M. Pombi, C. Y. Some *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.

Coulibaly, M. B., N. F. Lobo, M. C. Fitzpatrick, M. Kern, O. Grushko *et al.*, 2007 Segmental duplication implicated in the genesis of inversion 2Rj of *Anopheles gambiae*. PLoS ONE **2:** e849.

Davidson, G., H. E. Paterson, M. Coluzzi, G. F. Mason and D. W. Micks, 1967 The *Anopheles gambiae* complex in *Genetics of Insect Vectors of Disease*, edited by J. W. Wright and R. Pal. Elsevier, Amsterdam.

della Torre, A., 1997 Polytene chromosome preparation from anopheline mosquitoes, pp. 329–336 in *Molecular Biology of Disease Vectors: A Methods Manual*, edited by J. M. Crampton, C. B. Beard and C. Louis. Chapman & Hall, London.

della Torre, A., L. Merzagora, J. R. Powell and M. Coluzzi, 1997 Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. Genetics **146:** 239–244.

della Torre, A., C. Costantini, N. J. Besansky, A. Caccone, V. Petrarca *et al.*, 2002 Speciation within *Anopheles gambiae*—the glass is half full. Science **298:** 115–117.

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.

Dobzhansky, T., 1944 Chromosomal races in *Drosophila pseudoobscura* and *Drosophila persimilis*. Carnegie Inst. Washington Publ. **554:** 47–144.

Dobzhansky, T., 1948 Genetics of natural populations. XVIII. Experiments on chromosomes of *Drosophila pseudoobscura* from different geographic regions. Genetics **33:** 588–602.

Dobzhansky, T., and A. H. Sturtevant, 1938 Inversions in the chromosomes of *Drosophila pseudoobscura*. Genetics **23:** 28–64.

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

Donnelly, M. J., F. Simard and T. Lehmann, 2002 Evolutionary studies of malaria vectors. Trends Parasitol. **18:** 75–80.

Excoffier, L., G. Laval and S. Schneider, 2005 Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol. Bioinformatics Online **1:** 47–50.

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.

Feder, J. L., J. B. Roethele, K. Filchak, J. Niedbalski and J. Romero-Severson, 2003 Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. Genetics **163:** 939–953.

Garcia, B. A., A. Caccone, K. D. Mathiopoulos and J. R. Powell, 1996 Inversion monophyly in African anopheline malaria vectors. Genetics **143:** 1313–1320.

Hahn, M. W., M. D. Rausher and C. W. Cunningham, 2002 Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. Genetics **161:** 11–20.

Hoffmann, A. A., C. M. Sgro and A. R. Weeks, 2004 Chromosomal inversion polymorphisms and adaptation. Trends Ecol. Evol. **19:** 482–488.

Holt, R. A., G. M. Subramanian, A. Halpern, G. G. Sutton, R. Charlab *et al.*, 2002 The genome sequence of the malaria mosquito *Anopheles gambiae*. Science **298:** 129–149.

Hudson, R. R., D. D. Boos and N. L. Kaplan, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis *et al.*, 2003 Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4:** 249–264.

Kennington, W. J., L. Partridge and A. A. Hoffmann, 2006 Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion In(3R)Payne in *Drosophila melanogaster* are indicative of coadaptation. Genetics **172:** 1655–1663.

Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation and speciation. Genetics **173:** 419–434.

Krimbas, C. B., and J. R. Powell, 1992 Introduction, pp. 1–52 in *Drosophila Inversion Polymorphism*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Boca Raton.

Lewontin, R. C., and K. Kojima, 1960 The evolutionary dynamics of complex polymorphisms. Evolution **14:** 458–472.

Manoukis, N. C., J. R. Powell, M. B. Touré, A. Sacko, F. E. Edillo *et al.*, 2008 A test of the chromosomal theory of ecotypic speciation in *Anopheles gambiae*. Proc. Natl. Acad. Sci. USA **105:** 2940–2945.

Munte, A., J. Rozas, M. Aguade and C. Segarra, 2005 Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. Genetics **169:** 1573–1581.

Navarro, A., A. Barbadilla and A. Ruiz, 2000 Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. Genetics **155:** 685–698.

Navarro, A., and N. H. Barton, 2003 Chromosomal speciation and molecular divergence–accelerated evolution in rearranged chromosomes. Science **300:** 321–324.

Navarro, A., E. Betran, A. Barbadilla and A. Ruiz, 1997 Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics **146:** 695–709.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Noor, M. A., K. L. Grams, L. A. Bertucci and J. Reiland, 2001 Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA **98:** 12084–12088.

Obbard, D. J., J. J. Welch and T. J. Little, 2009 Inferring selection in the *Anopheles gambiae* species complex: an example from immune-related serine protease inhibitors. Malar. J. **8:** 117.

Ortiz-Barrientos, D., J. Reiland, J. Hey and M. A. Noor, 2002 Recombination and the divergence of hybridizing species. Genetica **116:** 167–178.

Petrarca, V., G. Sabatinelli, M. A. Di Deco and M. Papakay, 1990 The *Anopheles gambiae* complex in the Federal Islamic Republic of Comoros (Indian Ocean): some cytogenetic and biometric data. Parassitologia **32:** 371–380.

Pinto, J., M. J. Donnelly, C. A. Sousa, J. Malta-Vacas, V. Gil *et al.*, 2003 An island within an island: genetic differentiation of *Anopheles gambiae* in Sao Tome, West Africa, and its relevance to malaria vector control. Heredity **91:** 407–414.

Powell, J. R., 1997 *Progress and Prospects in Evolutionary Biology: The Drosophila Model.* Oxford University Press, Oxford.

Powell, J. R., V. Petrarca, A. della Torre, A. Caccone and M. Coluzzi, 1999 Population structure, speciation, and introgression in the *Anopheles gambiae* complex. Parassitologia **41:** 101–113.

Rieseberg, L. H., 2001 Chromosomal rearrangements and speciation. Trends Ecol. Evol. **16:** 351–358.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

Rozen, S., and H. J. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers, pp. 365–386 in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, edited by S. Krawetz and S. Misener. Humana Press, Totowa, NJ.

Sangare, D. M., 2007 Breakpoint analysis of the *Anopheles gambiae s.s.* chromosome 2Rb, 2Rc, and 2Ru inversions. Ph.D. Thesis, University of Notre Dame, Notre Dame.

Schaeffer, S. W., 2002 Molecular population genetics of sequence length diversity in the Adh region of *Drosophila pseudoobscura*. Genet. Res. **80:** 163–175.

Schaeffer, S. W., 2008 Selection in heterogeneous environments maintains the gene arrangement polymorphism of *Drosophila pseudoobscura*. Evolution **62:** 3082–3099.

Schaeffer, S. W., and W. W. Anderson, 2005 Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. Genetics **171:** 1729–1739.

Schaeffer, S. W., M. P. Goetting-Minesky, M. Kovacevic, J. R. Peoples, J. L. Graybill *et al.*, 2003 Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. Proc. Natl. Acad. Sci. USA **100:** 8319–8324.

Sharakhov, I. V., B. J. White, M. V. Sharakhova, J. Kayondo, N. F. Lobo *et al.*, 2006 Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the *Anopheles gambiae* complex. Proc. Natl. Acad. Sci. USA **103:** 6258–6262.

Simard, F., D. Ayala, G. C. Kamdem, J. Etouna, K. Ose *et al.*, 2009 Ecological niche partitioning between the M and S molecular forms of *Anopheles gambiae* in Cameroon: the ecological side of speciation. BMC Ecol. **9:** 17.

Stefansson, H., A. Helgason, G. Thorleifsson, V. Steinthorsdottir, G. Masson *et al.*, 2005 A common inversion under selection in Europeans. Nat. Genet. **37:** 129–137.

Stephens, M., and P. Donnelly, 2003 A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. **73:** 1162–1169.

Stephens, M., N. J. Smith and P. Donnelly, 2001 A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. **68:** 978–989.

Storz, J. F., 2005 Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. **14:** 671–688.

Stump, A. D., M. Pombi, L. Goeddel, J. M. C. Ribeiro, J. A. Wilder *et al.*, 2007 Genetic exchange in 2La inversion heterokaryotypes of *Anopheles gambiae*. Insect Mol. Biol. **16:** 703–709.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Temu, E. A., R. H. Hunt, M. Coetzee, J. N. Minjas and C. J. Shiff, 1997 Detection of hybrids in natural populations of the *Anopheles gambiae* complex by the rDNA-based, PCR method. Ann. Trop. Med. Parasitol. **91:** 963–965.

Toure, Y. T., V. Petrarca, S. F. Traore, A. Coulibaly, H. M. Maiga *et al.*, 1994 Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae s.str.* in Mali, West Africa. Genetica **94:** 213–223.

Toure, Y. T., V. Petrarca, S. F. Traore, A. Coulibaly, H. M. Maiga *et al.*, 1998 The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. Parassitologia **40:** 477–511.

Turner, T. L., M. W. Hahn and S. V. Nuzhdin, 2005 Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. **3:** e285.

Turner, T. L., M. T. Levine, M. L. Eckert and D. J. Begun, 2008 Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. Genetics **179:** 455–473.

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, G. B., 1971 Chromosomal evidence for natural interspecific hybridization by mosquitoes of the *Anopheles gambiae* complex. Nature **231:** 184–185.

Winzeler, E. A., C. I. Castillo-Davis, G. Oshiro, D. Liang, D. R. Richards *et al.*, 2003 Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics **163:** 79–89.

Communicating editor: D. Begun

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.109.105817/DC1

The Population Genomics of Trans-Specific Inversion Polymorphisms in *Anopheles gambiae*

Bradley J. White, Changde Cheng, Djibril Sangaré, Neil F. Lobo, Frank H. Collins and Nora J. Besansky

DOI: 10.1534/genetics.109.105817

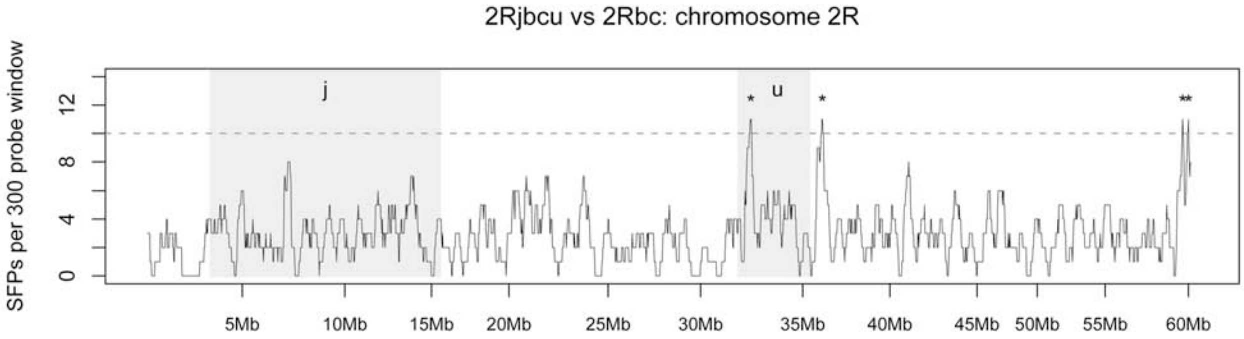

FIGURE S1.—Sliding window analysis of divergence between alternative arrangements across the 2R chromosome between karyotypes 2R*jbcu* and 2R*bc*, measured in terms of the proportion of SFPs per 300 probe window. Chromosome 2R is shown from telomere (left) to centromere (right). Shaded areas represent chromosomal inversions. Horizontal dashed line is the significance threshold at 12 SFPs per window. Asterisks denote significantly diverged regions.

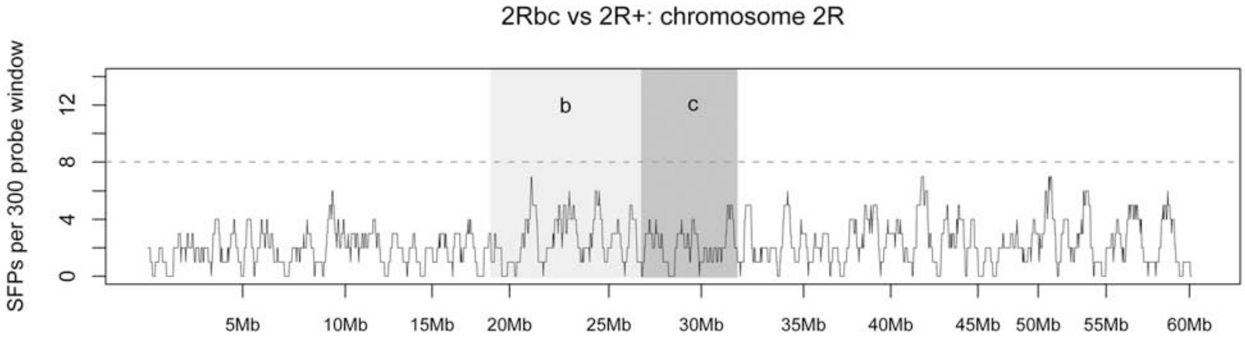

FIGURE S2.—Sliding window analysis of divergence between alternative arrangements across the 2R chromosome between karyotypes 2R*bc* and 2R+, measured in terms of the proportion of SFPs per 300 probe window. Chromosome 2R is shown from telomere (left) to centromere (right). Shaded areas represent chromosomal inversions. Horizontal dashed line is the significance threshold at 12 SFPs per window. In this contrast, no significant differentiation was detected.

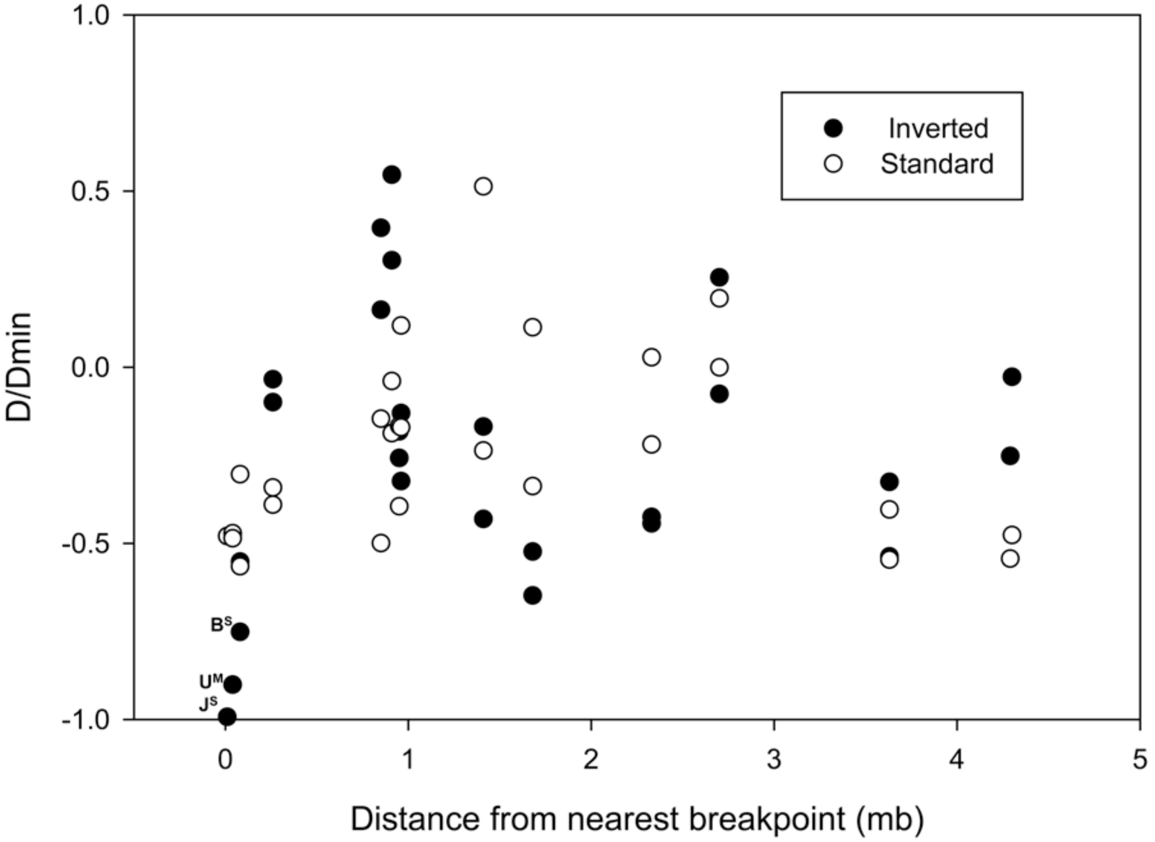

FIGURE S3.—Plot of the ratio of Tajima's *D* to the theoretical minimum of Tajima's *D* ($D_{\min}$) for all genes sequenced in each *Anopheles gambiae* molecular form and gene arrangement. Breakpoint loci are labeled by inversion and molecular form (in superscript).

TABLE S1

Genes and primers used for sequence determination on chromosome 2R in Malian populations of *An. gambiae*

Gene	Gene ID	Primer Sequences (5'-3')	Chromosomal Position (AgamP3.4)
hex2	AGAP001659	F: CGTTCCTGGAGAAGCAGAAG	7552268-7552287
		R: AGCCGTTGTGGTAGTTTTCG	7553116-7553097
clk-6	AGAP001856	F: CGTACCGGTTTCTGGTGAAC	11454087-11454068
		R: CGTTGCTTTCGGAGCTAAAA	11453386-11453405
Sema	AGAP002424	F: ATCGCCGTCACCAACTGTA	21200436-21200454
		R: ACTCGAGCTTCTCGCAGGT	21201153-21201135
2576	AGAP002576	F: TCGCTCGACATCGAGATACA	23117991-23117972
		R: GCCCCAGATGAGATTCGTTA	23117214-23117233
2751	AGAP002751	F: GAAGGTGCTCTGCCTCAAAG	26667813-26667832
		R: GTATGTTCGGGAACGAGTGC	26668540-26668521
2775	AGAP002775	F: TCACCAGAGGCTATGTGCTG	27041004-27040985
		R: CGAAAAACTGCTCCGACTTC	27040213-27040232
dcr	AGAP002836	F: GCGGAAATATGCAACCATCT	28196955-28196974
		R: TTTCGTTCGACCATGTACCA	28197702-28197683
med	AGAP002902	F: CAGCCTTCATCACAGTCCAA	29118353-29118372
		R: ACGGATCCACATACCCATC	29119145-29119127
gpr45	AGAP003053	F: GTGTACGGTGCTGATCGAAA	31519286-31519305
		R: TATAAACACCGCACCCATGA	31520044-31520025
3085	AGAP003085	F: AACAAGTTCGCCGACATACC	32330467-32330486
		R: CCTTCACCTTGTCCCACAGT	32331310-32331291
e2-230k	AGAP003090	F: AGGAAAACGACAATGCGAAC	32389332-32389351
		R: CATTACGCTCAGCAAGTCCA	32390014-32389995
gpr23	AGAP003098	F: AAGCTGCTGATCGTGTTCCT	32427925-32427906

		R: GATGTGAGCAGTTCCCGATT	32427264-32427283
3106	AGAP003106	F: CGACGAGAACATTGTGCAGT	32441688-32441669
		R: GCTCCGGATCGAGTATGAAG	32440901-32440920
3205	AGAP003205	F: GGGCTTTTGCTTCATCTACG	33827546-33827527
		R: GCCTAGAGCCGTGTCTTGAG	33826780-33826799
4112	AGAP004112	F: AATATCGGCCCCATACTTCC	50099114-50099095
		R: TCTCCATCCTCCACATCCTC	50098407-50098426
mcm2	AGAP004275	F: AACCGATATTGTCGCGTTTC	53568163-53568182
		R: AACAGTTCGCTTTCGAGGAA	53568871-53568852

TABLE S2

Predicted genes in a significantly diverged region of the *An. gambiae* 2R*u* rearrangement

Gene ID	Description	Affy Plasmodium Anopheles probe set	Gene Start (bp)	Gene End (bp)	Number of SFPs	
					2R*jbcu* vs 2R+	2R*jbcu* vs 2R*bc*
AGAP003079		Ag.2R.2722.0_CDS_at	32254248	32256017	2	0
AGAP003080			32273374	32288370	0	0
AGAP003081		Ag.2R.1721.0_CDS_at	32297951	32300052	1	0
AGAP003082		Ag.2R.3351.0_CDS_s_at	32301727	32302947	0	0
AGAP003083		Ag.2R.3351.0_CDS_s_at, Ag.2R.1722.0_CDS_at	32303704	32308399	0	0
AGAP003084		Ag.2R.1723.0_CDS_at, Ag.UNKN.568.0_CDS_s_at	32320490	32325718	0	0
AGAP003085		Ag.2R.1936.0_CDS_at	32329882	32331426	0	0
AGAP003086		Ag.2R.2723.0_CDS_at, Ag.2R.338.0_CDS_at	32378725	32379968	0	0
AGAP003086			32378725	32379968		
AGAP003087		Ag.2R.386.0_CDS_a_at, Ag.2R.719.0_CDS_s_at, Ag.2R.719.1_s_at	32380097	32381431	1	0
AGAP003088		Ag.2R.719.1_s_at, Ag.2R.386.0_CDS_a_at, Ag.2R.719.0_CDS_s_at	32381681	32383151	0	0
AGAP003089		Ag.2R.720.0_CDS_at	32383651	32385665	0	0
AGAP003090		Ag.2R.1114.0_CDS_a_at	32386739	32390959	0	1
ENSANGEST00000008826		Ag.2R.1114.1_at, Ag.2R.3684.0_at	32389514	32394038	4	4
AGAP003091		Ag.2R.1043.0_CDS_at	32394345	32395223	0	0
AGAP003092		Ag.2R.504.0_CDS_at, Ag.2R.889.0_CDS_a_at	32404897	32414429	0	0
AGAP003093		Ag.2R.504.0_CDS_at, Ag.2R.1415.0_CDS_at	32415560	32416806	0	0
AGAP003094		Ag.2R.504.0_CDS_at	32416873	32417819	1	2
AGAP003094		Ag.2R.1415.0_CDS_at	32416873	32417819		
AGAP003095	Dopachrome conversion enzyme	Ag.2R.25.1_CDS_a_at	32419121	32421544	0	0
AGAP003096		Ag.2R.989.0_CDS_at	32423179	32424670	0	0
AGAP003097		Ag.2R.989.0_CDS_at,	32425204	32426891	1	1

		Ag.2R.237.0_CDS_at				
AGAP003098	gustatory receptor gpr23	Ag.2R.989.0_CDS_at, Ag.2R.237.0_CDS_at	32427076	32428603	1	0
AGAP003099		Ag.2R.237.0_CDS_at, Ag.2R.844.0_CDS_at	32428675	32430131	1	1
AGAP003100	tRNA-Gly	Ag.2R.237.0_CDS_at, Ag.2R.844.0_CDS_at	32429430	32429500	0	0
AGAP003101	tRNA-Ile		32431955	32432028	0	0
AGAP003102	tRNA-Ile		32433892	32433965	0	0
AGAP003103	tRNA-Lys		32434922	32434994	0	0
AGAP003104	tRNA-Gly		32435182	32435252	0	0
AGAP003105	tRNA-Met		32435821	32435892	0	0
AGAP003106		Ag.2R.1724.0_CDS_at	32440110	32442980	1	0
AGAP003107	tRNA-Glu		32444715	32444786	0	0
AGAP003108		Ag.2R.1937.1_CDS_a_at	32497091	32500709	1	0

Underlined genes were re-sequenced in this study; see Table 2.