SNP Genotyping Defines Complex Gene-Flow Boundaries Among African Malaria Vector Mosquitoes

D. E. Neafsey,1,2 M. K. N. Lawniczak,2* (J. J. Park,3 S. N. Redmond,2 M. B. Coulibaly,2 S. F. Traoré,2 N. Sagnon,4 C. Costantini,5,6 C. Johnson,1 R. C. Wiegand,2 F. H. Collins,7 E. S. Lander,3 D. F. Wirth,3,8 F. C. Kafatos,2 N. J. Besansky,7 G. K. Christophides,2 M. A. T. Muskavitch1,8,9†

Mosquitoes in the Anopheles gambiae complex show rapid ecological and behavioral diversification, traits that promote malaria transmission and complicate vector control efforts. A high-density, genome-wide mosquito SNP-genotyping array allowed mapping of genomic differentiation between populations and species that exhibit varying levels of reproductive isolation. Regions near centromeres or within polymorphic inversions exhibited the greatest genetic divergence, but divergence was also observed elsewhere in the genomes. Signals of natural selection within populations were overrepresented among genomic regions that are differentiated between populations, implying that differentiation is often driven by population-specific selective events. Complex genomic differentiation among speciating vector mosquito populations implies that tools for genome-wide monitoring of population structure will prove useful for the advancement of malaria eradication.

Anopheles gambiae is the primary vector of human malaria in sub-Saharan Africa, where annual burdens of malaria-induced morbidity and mortality are greatest. Population subdivision within A. gambiae is pervasive but has been defined inconsistently and incompletely in the past. A. gambiae is composed of at least two morphologically identical incipient species known as the M and S molecular forms based on fixed ribosomal DNA sequence differences (1). The M and S forms are further divided by inversion karyotype into five distinct chromosomal forms, including Mopti (molecular form M), Savanna (molecular form S), and Bamako (molecular form S), each of which we examine here, and each of which has specialized for different breeding sites (2, 3). Furthermore, A. gambiae belongs to a species complex of seven recently diverged, morphologically identical sibling taxa, including another major malaria vector, A. arabiensis, which we also examine here. Population subdivision can increase disease transmission intensity and duration, as new mosquito populations evolve to exploit changing habitats and varied seasonal conditions. Vector control efforts can be complicated by population subdivision, because populations vary for traits on which interventions depend, such as indoor feeding behavior (4, 5) and insecticide susceptibility (6).

Genes underlying epidemiologically relevant phenotypic diversification among vector populations must reside within genomic regions that are differentiated among those populations. Most previous efforts to detect genetic differentiation between mosquito populations have been unable to localize differentiated regions, even when population divergence has been detected [for instance, between S and Bamako (7)] or lacked resolution to map all but the most highly differentiated regions [for example, between M and S (8, 9)]. High-resolution mapping of genomic regions differentiated between vector populations will advance our understanding of phenotypic diversification. Furthermore, ongoing assessment of gene flow among vector populations is essential for implementation of control measures designed for natural genetic variants [for instance, insecticide susceptibility alleles (10)] or introduced transgenic variants (11) within mosquito populations, as we strive yet again to eradicate malaria.

We used a customized Affymetrix singlenucleotide polymorphism (SNP) genotyping array to analyze 400,000 SNPs identified through sequencing of the M and S incipient species of A. gambiae (12). We hybridized individual arrays with genomic DNA from each of 20 field-collected females from the three known sympatric A. gambiae populations in Mali (M, S, and Bamako) that exhibit partial reproductive isolation (2, 13–15).

We then hybridized DNA pooled from the same 20 females from each population to determine the
degree to which quantitative differences in allele frequencies could be assessed with the use of pooled DNA. We also hybridized a pool of DNA from 20 field-collected individuals of the sister species *A. arabiensis*. Results obtained from pooled and individual hybridizations were highly correlated (Pearson’s correlation coefficient $r^2 = 0.96$ for M, S, and Bamako comparisons) (fig. S1), indicating that the majority of SNPs assayed on the array yield useful quantitative information regarding divergence in allele frequencies between pooled samples.

Pooled hybridization data revealed that the greatest differentiation between the recently subdivided S and Bamako populations maps within a cluster of inversions on chromosomal arm 2R (Fig. 1A). This pattern is concordant with models of speciation in the face of ongoing gene flow, which predict that early in the speciation process, divergence will be localized to regions of low recombination, such as inversions (16–20). In partially reproductively isolated populations like S and Bamako, these divergent genomic regions are most likely to contain genes (table S2) directly responsible for differential niche adaptation and reproductive isolation, whereas ongoing gene flow should homogenize the remainder of the genome (21–23).

The M and S mosquito populations in Mali exhibit divergence that is much greater and more heterogeneous overall than that observed between S and Bamako (Fig. 1B). This might be expected given the broader geographic ranges of M and S relative to Bamako and their presumed longer divergence time (2). We found that all pericentromeric regions exhibit high levels of differentiation between M and S (fig. S2), in accordance with previous observations (8, 9, 24). However, we unexpectedly detected shorter regions of substantial differentiation at various distances from centromeres along each chromosome. The existence of extensive divergence within pericentromeric regions suggests that realized gene flow between these two incipient species is low, despite the observation of hybrids between M and S at frequencies approaching 1% in Mali (25). These findings, which we obtained with the use of DNA isolated from field-collected mosquitoes, reinforce patterns observed in the sequencing-based SNP analysis of M and S mosquito colonies (12).

We also compared the Mali S pool to a pool of colony-derived M mosquitoes from Cameroon to address the possibility that differentiation observed between M and S is geographically restricted to Mali. Genetic differentiation is substantially greater between M populations from Mali and Cameroon than between S populations from these locations, and it has been speculated that another incipient speciation event may be occurring within M (26). However, with the exception of the 2La inversion, we find extremely similar patterns of differentiation between S and M, regardless of the geographic origin of the M population that was analyzed (Fig. 1C and fig. S3). This finding suggests that the genomic regions differentiated between M and S are probably similar throughout West and Central Africa and may harbor the genes facilitating niche differentiation as well as pre- and postmating isolation between these taxa. However, the great extent of genomic divergence also implies that identifying the genes involved in the earliest stages of the M and S speciation process will prove challenging.

Finally, we compared *A. gambiae* and its sister species *A. arabiensis*, between which hybridization can occur in nature, although it yields sterile males (27). Because SNPs assayed on the array are segregating in *A. gambiae* but may not be segregating in *A. arabiensis*, we could not compare the overall magnitude of genomic divergence between these taxa with the divergence between forms of *A. gambiae*. To avoid bias, we undertook this comparison with a subset (75,750) of the SNPs that were found to exhibit similar allelic intensity ratios in the M and S pools. This assay set was sufficient to indicate that the profile of relative differentiation between *A. gambiae* and *A. arabiensis* is less heterogeneous than that in the M versus S comparison (Fig. 1D), even as it echoes some of the same highly divergent regions. Chromosomes 2 and 3 exhibit slightly heightened pericentromeric differentiation, similar to the pattern we observed between the M and S forms of *A. gambiae*, with additional differentiation across the entire X chromosome, presumably due to the large Xag inversion fixed in *A. gambiae* and at least one additional inversion fixed in *A. arabiensis* relative to the ancestral X arrangement (28, 29).

Although particular inversion arrangements are not exclusive to any of the *A. gambiae* populations that we profiled, these genomic regions clearly harbor an excess of differentiation between populations compared with other regions of the genome (Figs. 1 and 2). Inversions may be hotspots for differentiation, even when maintained at similar frequencies in different populations, if recombination suppression facilitates functional divergence of the inverted and wild-type arrangements. Average linkage disequilibrium within all three forms of *A. gambiae* is extremely low, extending no more than a few thousand base pairs (fig. S4). Therefore, groups of loci that reside within regions of lower recombination in the *A. gambiae* genome would be more likely to establish consistent patterns of cosegregation.

It is important to distinguish a difference in inversion frequency between populations versus differentiation of alternative inversion arrangements between populations. Principal components analysis (PCA) of SNP genotypes within inversion boundaries indicates that, although the S and Bamako populations harbor different frequencies of the 2Rj, 2Rb, 2Rc, and 2Ru inversions (table S3), the b arrangement of 2Rb is divergent between S and Bamako (Fig. 2). This result indicates that, although this arrangement is frequent in both S and Bamako, it is differentiating independently within each population. Similarly, both arrangements of 2Ru, as well as the uninveterd arrangement of 2La and 2Ru, have differentiated between M and S (fig. 2 and table S3). However, the inverted 2La arrangement is an exception to this pattern: M, S, and Bamako mosquitoes homozygous for the 2La inversion exhibit much less divergence between the 2La breakpoints than is observed in the same three populations for all other inversions (Fig. 2A). The close clustering of
A. arabiensis (a species fixed for the inverted 2La arrangement) with individuals homozygous for 2La from each of the A. gambiae populations (M, S, and Bamako) supports earlier hypotheses regarding introgression between species within this region (30). Indeed, the region within the 2La inversion breakpoints shows divergence between A. gambiae and A. arabiensis that is lower than expected (Fig. 1D). Overall, these PCA plots highlight the degree of similarity within each of these partially isolated populations. With the exception of 2Rj, no inversion is diagnostic of a particular population in our sample. However, the consistently independent clustering of M, S, and Bamako mosquitoes by PCA across all inversions except 2La affirms the legitimacy and genetic distinctiveness of these groups.

We next examined the data for signals of natural selection. The genomic regions exhibiting greatest divergence (Fst > 0.6) between M and S exhibit significantly reduced polymorphism in one or both species (fig. S5) [M: one-tailed test, \( P = 1.14 \times 10^{-47} \) (1.14E-47); S: one-tailed test, \( P = 1.88 \times 10^{-120} \)], as might be expected if differences between populations were driven to fixation by polymorphism-eliminating selective sweeps (31). To explore selection more deeply, we analyzed SNP calls from individual hybridizations of M, S, and Bamako mosquitoes with the use of SweepFinder software (32), an approach that evaluates the likelihood of a sweep within a particular genomic region, given the allele frequency spectrum of local SNPs. Several genomic regions appear to have experienced recent sweeps within each of the three forms (Fig. 3). The pericentromeric regions of all three chromosomes exhibit the strongest signals of selective sweeps for M and S, suggesting that the extensive divergence observed in these regions has been driven by selection (Fig. 3). Indeed, the degree of concordance between the profiles of selection and differentiation for M and S [chi-squared test, \( P = 2.4E-104 \)] implies a causal role for selection within some genomic regions where differentiation is observed. Additionally, the fact that different regions of the genomes of M, S, and Bamako show evidence of selective sweeps suggests that these populations are experiencing different selective pressures that shape genetic variation independently (Fig. 3).

Analysis for functional enrichment (15) among 68 genes found in candidate sweep regions identified two interesting categories of genes significantly overrepresented after correction for multiple testing: (i) multicellular organismal development (\( P = 9.1E-4 \); five genes), and (ii) serine-type endopeptidase activity (\( P = 2.6E-2 \); nine genes, five of which occur in a pericentromeric cluster on 3L). Of the five genes annotated as being involved in development, three encode homeodomain-containing transcriptional regulators (AGAP004659, sweep in S; AGAP004660, sweep in S; AGAP004696, sweep in Bamako), one encodes a member of the Hedgehog signaling pathway (AGAP004637, sweep in S), and one encodes a member of the Wnt signaling pathway (AGAP010283, sweep in M), indicating that shifts in developmental regulatory programs may underlie ecological niche differentiation and/or reproductive isolation mechanisms that reinforce the ongoing process of speciation in these populations. The gene encoding CPF3, a cuticular protein speculated to bind sex pheromones (33), is also found in a pericentromeric sweep region in S on chromosome 2L. CPF3 is the gene exhibiting the most significant difference in expression between M and S (33), and it dramatically changes expression upon mating (34). These combined observations motivate further investigation of CPF3 and its potential relation to M and S mate discrimination. Table S2 presents a full list of the 536 genes found in differentiated and/or sweep regions. Among this set of 536 loci, genes from the X chromosome are significantly overrepresented (X total = 173; chi-squared test; \( P < 2.2E-16 \)).

Our findings demonstrate the power of high-resolution SNP arrays for mapping genetic divergence among vector mosquito taxa within the A. gambiae species complex. The ability to detect differentiation between distinct populations and selective sweeps within populations is valuable for identifying and monitoring alleles that mediate traits critical for malaria transmission and vector control. The differentiated genomic regions we have identified with these comparisons harbor genes (table S2) of epidemiological importance for disease transmission, including loci influencing reproduction, longevity, insecticide resistance, aridity tolerance, larval habitat, and other traits that differ among mosquito populations (2).
ATM Activation by Oxidative Stress

Zhi Guo,1 Sergei Kozlov,2 Martin F. Lavin,2 Maria D. Person,3 Tanya T. Paull1,*

The ataxia-telangiectasia mutated (ATM) protein kinase is activated by DNA double-strand breaks (DSBs) through the Mre11-Rad50-Nbs1 (MRN) DNA repair complex and orchestrates signaling cascades that initiate the DNA damage response. Cells lacking ATM are also hypersensitive to insults other than DSBs, particularly oxidative stress. We show that oxidation of ATM directly induces ATM activation in the absence of DNA DSBs and the MRN complex. The oxidized form of ATM is a disulfide–cross-linked dimer, and mutation of a critical cysteine residue involved in disulfide bond formation specifically blocked activation through the oxidation pathway. Identification of this pathway explains observations of ATM activation under conditions of oxidative stress and shows that ATM is an important sensor of reactive oxygen species in human cells.

Patients with ataxia-telangiectasia (A-T) lack functional A-T mutated (ATM) protein and exhibit a pleiotropic phenotype that includes cerebellar ataxia, immunodeficiency, premature aging, and a high incidence of lymphoma (1). These defects may be connected through dysregulation of reactive oxygen species (ROS), indicated by observations that mammalian cells lacking ATM exhibit high concentrations of ROS and hypersensitivity to agents that induce oxidative stress (2). Lymphoma incidence and the loss of hematopoietic stem cells that occurs in mice lacking ATM can be suppressed by antioxidants (3–5), indicating an important role for ATM in regulating cellular defenses against redox stress. ATM is activated in response to changes in intracellular redox status (6–10), but it has not been clear what the initiating event is in these cases or how this relates to Mre11-Rad50-Nbs1 (MRN)–mediated ATM activation that is dependent on DNA double-strand breaks (DSBs).

To address these questions, we used primary human fibroblasts and induced oxidative stress with H2O2 or DSBs with bleomycin (11). Auto-phosphorylation of ATM on Ser1981, phosphorylation of the tumor suppressor p53 on Ser15, and phosphorylation of the protein kinase Chk2 on Thr68 all occurred in response to H2O2 or to bleomycin treatment (Fig. 1, A to C). Phospho-

Fig. 3. Profiles of genomic regions subject to recent selective sweeps in M, S, and Bamako forms of *A. gambiae*. Each point represents the $-\log P$ value of a selective sweep for a window of $\pm 20$ SNPs. Windows exhibiting significant signals of selection ($P < 0.05$ after Bonferroni correction) are indicated in red.

References and Notes
1. A. della Torre et al., *Insect Mol. Biol.* 10, 9 (2001).
2. T. Lehmann, A. Diabate, *Infect. Genet. Evol.* 8, 737 (2008).
3. N. C. Manoukis et al., *Proc. Natl. Acad. Sci. U.S.A.* 105, 2940 (2008).
4. W. Takken, *Trop. Med. Int. Health* 7, 1022 (2002).
5. H. D. Trung et al., *Trop. Med. Int. Health* 10, 253 (2005).
6. C. Fanello et al., *Insect Mol. Biol.* 12, 241 (2003).
7. M. A. Slotman et al., *Am. J. Trop. Med. Hyg.* 74, 641 (2006).
8. T. L. Turner, M. W. Hahn, S. V. Nuzhdin, *PloS Biol.* 3, e285 (2005).
9. B. J. White, C. Cheng, F. Simard, C. Costantini, N. J. Besansky, *Mol. Ecol.* 19, 925 (2010).
10. A. Enayati, J. Hemingway, *Annu. Rev. Entomol.* 55, 569 (2010).
11. V. Koutoupanou, M. R. Goddard, A. Burt, *Mol. Biol. Evol.* 19, 239 (2002).
12. M. K. N. Lawrimcor et al., *Science* 330, 512 (2010).
13. M. Coluzzi, A. Sabatini, A. della Torre, M. A. Di Deco, V. Petrarca, *Science* 298, 1415 (2002); published online 3 October 2002 (10.1126/science.1077769).
14. A. della Torre, Z. Tu, V. Petrarca, *Insect Biochem. Mol. Biol.* 35, 755 (2005).
15. Materials and methods are available as supporting material on Science Online.
16. M. Coluzzi, *Prog. Clin. Biol. Res.* 96, 143 (1982).
17. M. A. Noor, K. L. Grams, L. A. Bertucci, J. Reiland, *Proc. Natl. Acad. Sci. U.S.A.* 98, 12084 (2001).
18. L. H. Rieseberg, *Trends Ecol. Evol.* 16, 351 (2001).
19. A. Navarro, N. H. Barton, *Evolution* 57, 447 (2003).
20. R. K. Butlin, *Mol. Ecol.* 14, 2621 (2005).
ERRATUM

Post date 10 December 2010

Reports: “SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes” by D. E. Neafsey et al. (22 October, p. 514). The authors mistakenly failed to include an acknowledgment in support of author G. K. Christophides: “the European Commission FP7 Collaborative project grant HEALTH-F3-2008-223736 to G.K.C.”