MIT Libraries | DSpace@MIT

MIT Open Access Articles

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

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Neafsey, D. E. et al. "SNP Genotyping Defines Complex Gene-Flow Boundaries Among African Malaria Vector Mosquitoes." Science 330, 6003 (October 2010): 514–517 © 2010 American Association for the Advancement of Science

As Published: http://dx.doi.org/10.1126/science.1193036

Publisher: American Association for the Advancement of Science (AAAS)

Persistent URL: http://hdl.handle.net/1721.1/116704

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike

Massachusetts Institute of Technology

Europe PMC Funders Group
Author Manuscript
Science. Author manuscript; available in PMC 2016 March 29.

Published in final edited form as:
Science. 2010 October 22; 330(6003): 514–517. doi:10.1126/science.1193036.

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

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

[1]Broad Institute, Cambridge, MA 02142, USA.

[2]Imperial College London, London SW7 2AZ, UK.

[3]Malaria Research and Training Center, Bamako, Mali.

[4]Centre National de Recherche et Formation sur le Paludisme, Ouagadougou, Burkina Faso.

[5]Institut de Recherche pour le Développement, Unité de Recherche R016, Montpellier, France.

[6]Organisation de Coordination pour la Lutte contre les Endémies en Afrique Centrale, Yaounde, Cameroon.

[7]University of Notre Dame, Notre Dame, IN 46556, USA.

[8]Harvard School of Public Health, Boston, MA 02115, USA.

[9]Boston College, Chestnut Hill, MA 02467, USA.

[#] These authors contributed equally to this work.

Abstract

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.

[†]To whom correspondence should be addressed. marc.muskavitch@bc.edu.

Supporting Online Material
www.sciencemag.org/cgi/content/full/330/6003/514/DC1
Materials and Methods, Figs. S1 to S5, Tables S1 to S3, References, dbSNP Accession Numbers

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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 single-nucleotide 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

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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 nonpericentromeric 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 2L*a* 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

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

the large X*ag* 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 2R*j*, 2R*b*, 2R*c*, and 2R*u* inversions (table S3), the *b* arrangement of 2R*b* 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 2R*b*, as well as the uninverted arrangements of 2L*a* and 2R*u*, have differentiated between M and S (Fig. 2 and table S3). However, the inverted 2L*a* arrangement is an exception to this pattern: M, S, and Bamako mosquitoes homozygous for the 2L*a* inversion exhibit much less divergence between the 2L*a* 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 2L*a* arrangement) with individuals homozygous for 2L*a* 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 2L*a* 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 2R*j*, 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 2L*a* 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 ($F_{st} > 0.6$) between M and S exhibit significantly reduced polymorphism in one or both species (fig. S5) [M: one-tailed *t* test, $P = 1.14 \times 10^{-47}$ (1.14E-47); S: one-tailed *t* test, $P = 1.88E\text{-}120$], 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

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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, $P = 2.4E-104$ (14)] 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).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Acknowledgments

This work was supported by Burroughs Wellcome Fund Request 1008238, the Broad Institute Director's Fund, the Wellcome Trust Programme grant 077229/Z/05/Z to F.C.K. and G.K.C., National Human Genome Research Institute support to E.S.L., the Harvard School of Public Health Department of Immunology and Infectious Diseases, and the DeLuca Professorship from Boston College to M.A.T.M. In addition, M.K.N.L. was supported by a Biotechnology and Biological Sciences Research Council research grant BB/E002641/1 to G.K.C. Sequencing data for array validation have been deposited in the National Center for Biotechnology Information Short Read Archive with accession numbers SRX005397 to SRX005403. SNPs have been submitted to dbSNP Build B133 (ss/rs numbers in progress) and, meanwhile, are listed in the supporting online material. ArrayExpress accession number: A-AFFY-167.

References and Notes

1. della Torre A, et al. Insect Mol. Biol. 2001; 10:9. [PubMed: 11240632]
2. Lehmann T, Diabate A. Infect. Genet. Evol. 2008; 8:737. [PubMed: 18640289]
3. Manoukis NC, et al. Proc. Natl. Acad. Sci. U.S.A. 2008; 105:2940. [PubMed: 18287019]
4. Takken W. Trop. Med. Int. Health. 2002; 7:1022. [PubMed: 12460393]
5. Trung HD, et al. Trop. Med. Int. Health. 2005; 10:251. [PubMed: 15730510]
6. Fanello C, et al. Insect Mol. Biol. 2003; 12:241. [PubMed: 12752657]
7. Slotman MA, et al. Am. J. Trop. Med. Hyg. 2006; 74:641. [PubMed: 16606999]
8. Turner TL, Hahn MW, Nuzhdin SV. PLoS Biol. 2005; 3:e285. [PubMed: 16076241]
9. White BJ, Cheng C, Simard F, Costantini C, Besansky NJ. Mol. Ecol. 2010; 19:925. [PubMed: 20149091]
10. Enayati A, Hemingway J. Annu. Rev. Entomol. 2010; 55:569. [PubMed: 19754246]
11. Koufopanou V, Goddard MR, Burt A. Mol. Biol. Evol. 2002; 19:239. [PubMed: 11861883]
12. Lawniczak MKN, et al. Science. 2010; 330:512. [PubMed: 20966253]
13. Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V. Science. 2002; 298:1415. published online 3 October 2002. doi: 10.1126/science.1077769 [PubMed: 12364623]
14. della Torre A, Tu Z, Petrarca V. Insect Biochem. Mol. Biol. 2005; 35:755. [PubMed: 15894192]
15. Materials and methods are available as supporting material on *Science* Online.
16. Coluzzi M. Prog. Clin. Biol. Res. 1982; 96:143. [PubMed: 7178155]
17. Noor MA, Grams KL, Bertucci LA, Reiland J. Proc. Natl. Acad. Sci. U.S.A. 2001; 98:12084. [PubMed: 11593019]
18. Rieseberg LH. Trends Ecol. Evol. 2001; 16:351. [PubMed: 11403867]
19. Navarro A, Barton NH. Evolution. 2003; 57:447. [PubMed: 12703935]
20. Butlin RK. Mol. Ecol. 2005; 14:2621. [PubMed: 16029465]
21. Wu CI, Ting CT. Nat. Rev. Genet. 2004; 5:114. [PubMed: 14735122]
22. Machado CA, Kliman RM, Markert JA, Hey J. Mol. Biol. Evol. 2002; 19:472. [PubMed: 11919289]
23. Via S. Proc. Natl. Acad. Sci. U.S.A. 2009; 106(suppl. 1):9939. [PubMed: 19528641]
24. Stump AD, et al. Proc. Natl. Acad. Sci. U.S.A. 2005; 102:15930. [PubMed: 16247019]
25. Tripet F, et al. Mol. Ecol. 2001; 10:1725. [PubMed: 11472539]
26. Slotman MA, et al. Mol. Ecol. 2007; 16:639. [PubMed: 17257119]
27. Curtis, CJ. Recent Developments in the Genetics of Disease Vectors. Steiner, WM., editor. Stripes Publishing; Champaign, IL: 1982. p. 290-312.
28. Besansky NJ, et al. Proc. Natl. Acad. Sci. U.S.A. 2003; 100:10818. [PubMed: 12947038]
29. Slotman M, Della Torre A, Powell JR. Genetics. 2004; 167:275. [PubMed: 15166154]
30. White BJ, et al. PLoS Genet. 2007; 3:e217. [PubMed: 18069896]
31. Smith JM, Haigh J. Genet. Res. 1974; 23:23. [PubMed: 4407212]
32. Nielsen R, et al. Genome Res. 2005; 15:1566. [PubMed: 16251466]
33. Cassone BJ, et al. Mol. Ecol. 2008; 17:2491. [PubMed: 18430144]

34. Rogers DW, et al. Proc. Natl. Acad. Sci. U.S.A. 2008; 105:19390. [PubMed: 19036921]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

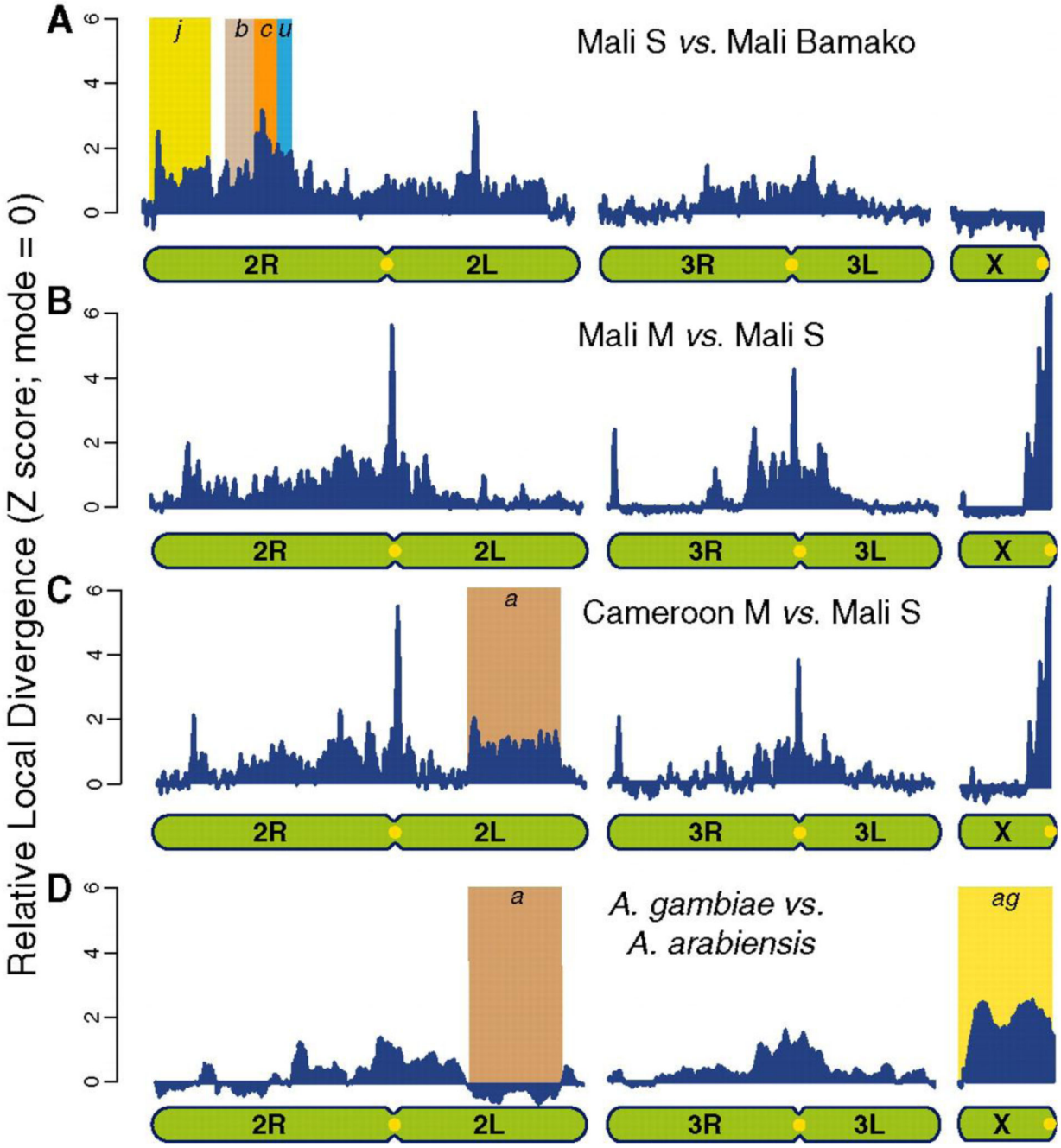

Fig. 1.

Relative local divergence profiles for pairwise comparisons of mosquito populations, represented by z scores (standard deviations) and scaled so that 0 reflects the modal divergence for each comparison. Plots represent average difference in allelic intensity ratios measured over adjacent 50 SNP stepping windows. The colored regions labeled with letters represent chromosomal inversion locations.(**A**) Divergence between S and Bamako forms of *A. gambiae* from Mali. (**B**) Divergence between *A. gambiae* M-form mosquitoes and S-form mosquitoes from Mali. (**C**) Divergence between M-form mosquitoes from Cameroon and S-

form mosquitoes from Mali. (**D**) Divergence between *A. arabiensis* from Burkina Faso and *A. gambiae* from Mali.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

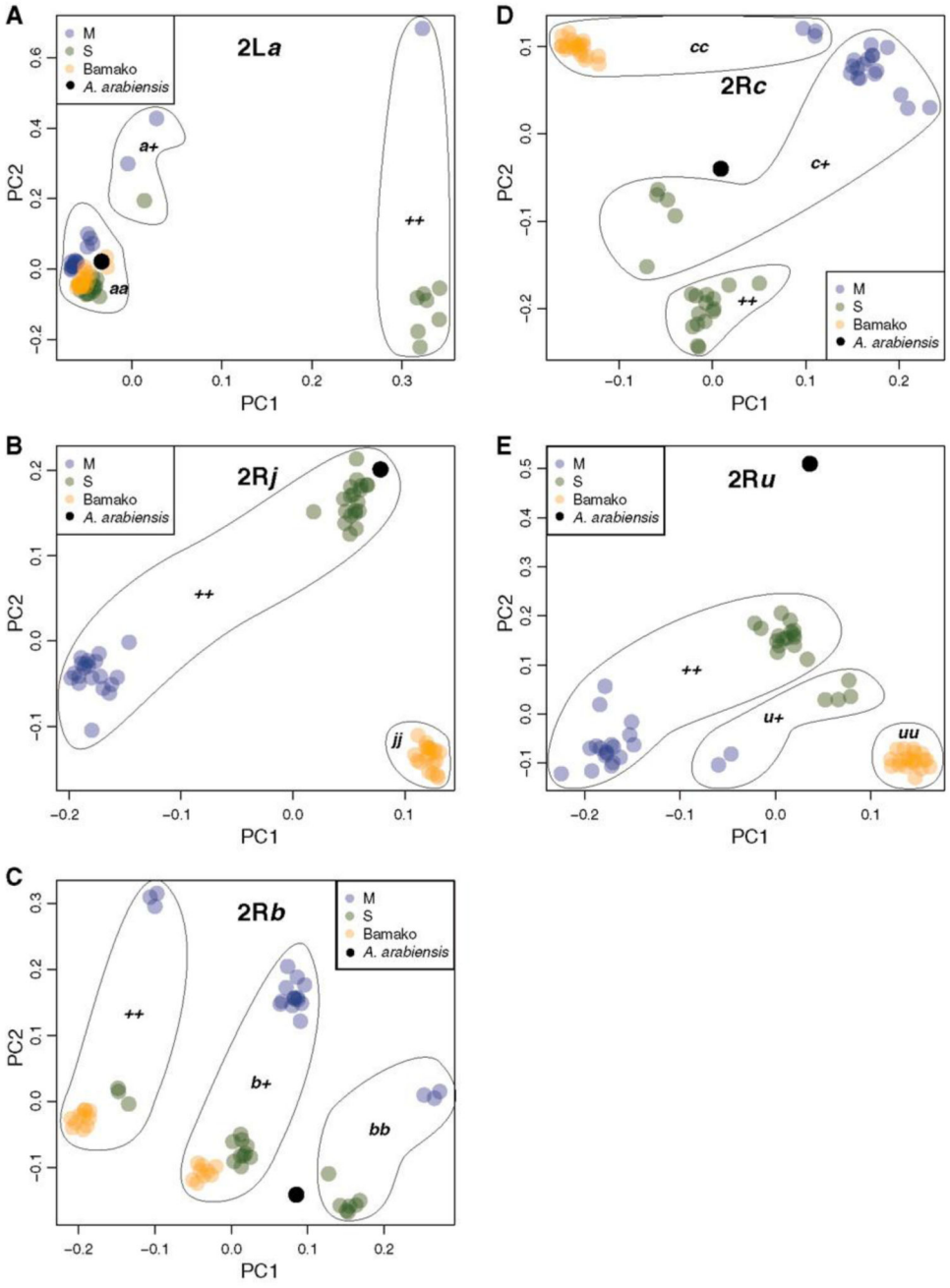

Fig. 2.

PCA plots of the 2L*a* (**A**), 2R*j* (**B**), 2R*b* (**C**), 2R*c* (**D**), and 2R*u* (**E**) inversion regions. Circled regions indicate groups of mosquitoes homozygous (*ii*) or heterozygous (*i*+) for the inversion or homozygous for the wild-type arrangement (++).

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

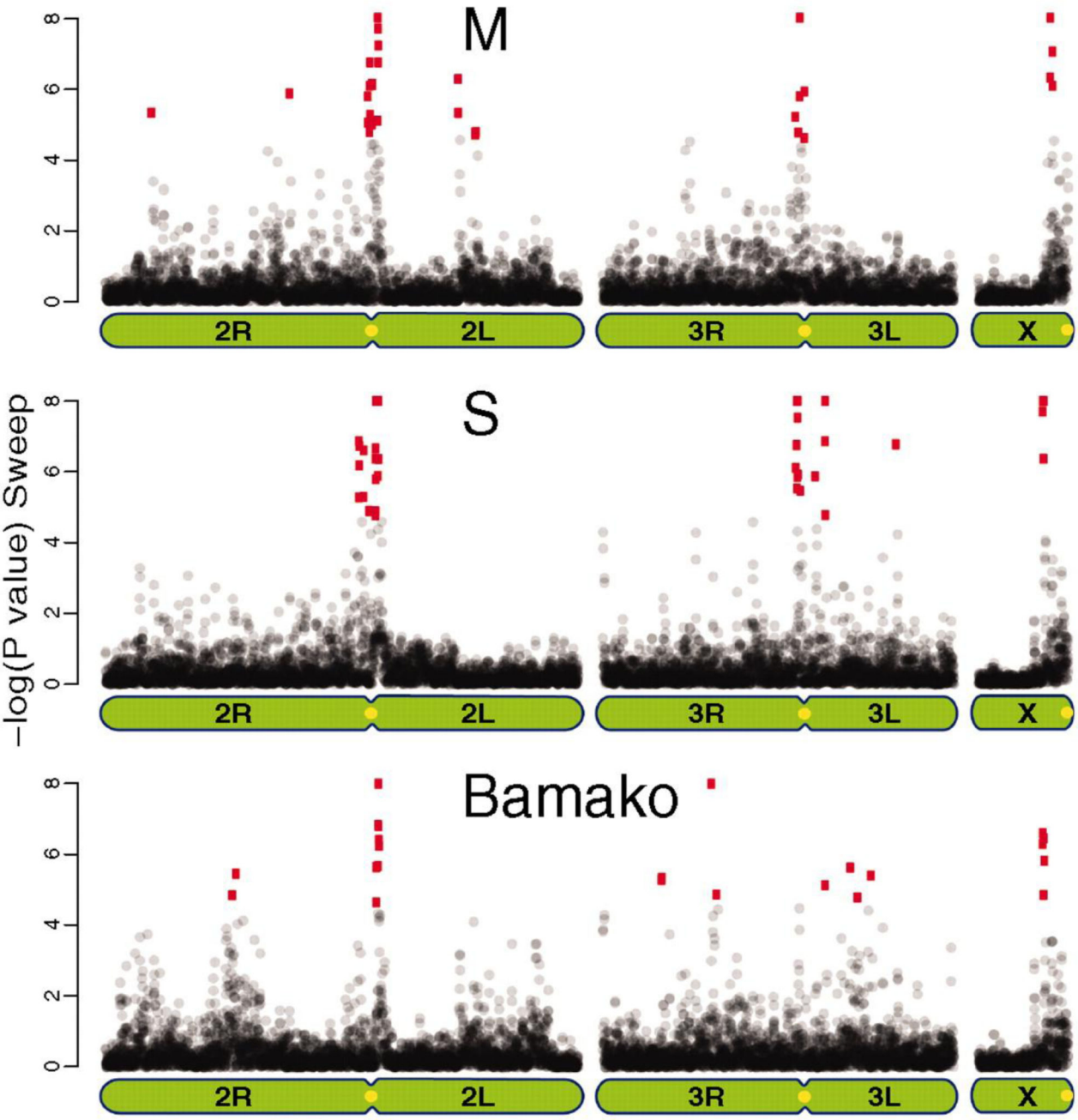

Fig. 3.

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 ~20 SNPs. Windows exhibiting significant signals of selection ($P < 0.05$ after Bonferroni correction) are indicated in red.