N-terminal pore (as illustrated by the back arrow from the activated state to the histidine-locked state in Fig. 3) when the pH there is suddenly increased from 6 to 8. With these protons released, the histidine tetrad then becomes doubly protonated and the tryptophan gate becomes closed.
32. See supporting material on *Science* Online.
33. O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, M. S. Sansom, *J. Mol. Graph.* **14**, 354 (1996).

34. Supported by National Institute of Allergy and Infectious Diseases grant AI023007. The spectroscopy was conducted at the National High Magnetic Field Laboratory supported by Cooperative Agreement 0654118 between the NSF Division of Materials Research and the State of Florida. T.A.C., H.-X.Z., D.D.B., M.S., and M.Y. have applied for a patent on the mechanism reported here. The structure (an ensemble of eight models) has been deposited in the Protein Data Bank with accession code 2L0J.

Supporting Online Material
www.sciencemag.org/cgi/content/full/330/6003/509/DC1
Materials and Methods
Figs. S1 to S4
References

3 May 2010; accepted 26 August 2010
10.1126/science.1191750

Widespread Divergence Between Incipient *Anopheles gambiae* Species Revealed by Whole Genome Sequences

M. K. N. Lawniczak,[1]* S. J. Emrich,[2]* A. K. Holloway,[3] A. P. Regier,[2] M. Olson,[2] B. White,[4] S. Redmond,[1] L. Fulton,[5] E. Appelbaum,[5] J. Godfrey,[5] C. Farmer,[5] A. Chinwalla,[5] S.-P. Yang,[5] P. Minx,[5] J. Nelson,[5] K. Kyung,[5] B. P. Walenz,[6] E. Garcia-Hernandez,[6] M. Aguiar,[6] L. D. Viswanathan,[6] Y.-H. Rogers,[6] R. L. Strausberg,[6] C. A. Saski,[7] D. Lawson,[8] F. H. Collins,[4] F. C. Kafatos,[1] G. K. Christophides,[1] S. W. Clifton,[5] E. F. Kirkness,[6] N. J. Besansky[4]†

The Afrotropical mosquito *Anopheles gambiae* sensu stricto, a major vector of malaria, is currently undergoing speciation into the M and S molecular forms. These forms have diverged in larval ecology and reproductive behavior through unknown genetic mechanisms, despite considerable levels of hybridization. Previous genome-wide scans using gene-based microarrays uncovered divergence between M and S that was largely confined to gene-poor pericentromeric regions, prompting a speciation-with-ongoing-gene-flow model that implicated only about 3% of the genome near centromeres in the speciation process. Here, based on the complete M and S genome sequences, we report widespread and heterogeneous genomic divergence inconsistent with appreciable levels of interform gene flow, suggesting a more advanced speciation process and greater challenges to identify genes critical to initiating that process.

Population-based genome sequences provide a rich foundation for "reverse ecology" (*1*). By analogy to reverse genetics, reverse ecology uses population genomic data to infer the genetic basis of adaptive phenotypes, even if the relevant phenotypes are not yet known. This approach can be especially powerful for gaining insight into the genetic basis of ecological speciation, a process whereby barriers to gene flow evolve between populations as by-products of strong, ecologically based, divergent selection (*2*). Here, we apply reverse ecology to study incipient speciation within *Anopheles gambiae,* one of the most efficient vectors of human malaria. The complex population structure of *A. gambiae*, exemplified by the emergence of the M and S molecular forms (*3*), poses substantial challenges for malaria epidemiology and control, as underlying differences in behavior and physiology may affect disease transmission and compromise antivector measures. Genome-wide analysis of M and S can provide insight into the mechanisms promoting their divergence and open new avenues for malaria vector control.

Morphologically, M and S are indistinguishable at all life stages and can only be recognized by fixed differences in the ribosomal DNA genes (*3*). Geographically and microspatially, both forms co-occur across much of West and Central Africa (*4*), and in areas where they are sympatric, adults may be found resting in the same houses and even flying in the same mating swarms (*5*, *6*). Assortative mating limits gene flow between forms (*5*, *6*), but appreciable hybridization still occurs (*4*, *7*–*10*) without intrinsic hybrid inviability or sterility (*11*). Although the aquatic larvae of both forms also may be collected from the same breeding site, S-form larvae are associated with ephemeral and largely predator-free pools of rain water, whereas M-form larvae exploit longer-lived but predator-rich anthropogenic habitats (*12*). Thus, persistence of M and S despite hybridization may be driven by ecologically dependent fitness tradeoffs in the alternative larval habitats to which they are adapting (*12*).

Under a model of speciation in the presence of gene flow, genomic divergence between incipient species should be limited to regions containing the genes that confer differential adaptations or are involved in reproductive isolation (*13*). Consistent with this expectation, scans of genomic divergence between M and S at the resolution of gene-based microarrays revealed elevated divergence near the centromeres of all three independently assorting chromosomes, and almost nowhere else (*14*, *15*). Given the assumption of appreciable genetic exchange through hybridization, this pattern suggested that the genes causing ecological and behavioral isolation were located in the centromeric "speciation islands" (*14*). The small number, size, and gene content of these islands implied that speciation of M and S was very recent and involved only a few genes in a few isolated chromosomal regions—an influential model for speciation with gene flow (*13*, *16*, *17*). The complete genome sequences of *A. gambiae* M and S forms reported here provide much higher resolution than previous studies to address how genomes diverge during speciation.

Sequences were determined from colonies established in 2005 from Mali, where the rate of natural M-S hybridization (~1%) is theoretically high enough for introgression to homogenize neutral variation between genomes (*18*) in the absence of countervailing selection. Both colonies were homosequential and homozygous with respect to all known chromosomal inversions with the exception of 2L*a* and 2R*c* (*19*). Independent draft genome assemblies were generated based on ~2.7 million Sanger traces (*19*). Both assemblies were performed independently of the reference *A. gambiae* PEST genome (*20*), which is a chimera of the M and S forms. Genome assembly metrics were similar between M and S (table S1) (*19*). Lower coverage (~6-fold in M/S versus ~10-fold in PEST) contributed to assembly gaps, motivating alignment of the M and S scaffolds to the PEST assembly for transfer of genomic coordinates and gene annotations (www.vectorbase.org; table S2) (*19*). We confirmed the major trends of M-S divergence by direct alignment of M and S scaffolds to each other (fig. S1) (*19*).

More than two million single-nucleotide polymorphisms (SNPs) per form and more than 150,000 fixed differences between forms were identified in the sequence data using strict coverage and quality restrictions (table S3) (*19*). The chromosomes show significantly different patterns of divergence, with chromosome 2 showing proportionally more fixed differences than chromo-

[1]Division of Cell and Molecular Biology, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. [2]Department of Computer Science and Engineering and Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN 46556, USA. [3]J. David Gladstone Institutes, San Francisco, CA 94158, USA. [4]Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA. [5]The Genome Center at Washington University, St. Louis, MO 63108, USA. [6]The J. Craig Venter Institute, Rockville, MD 20850, USA. [7]Clemson University Genomics Institute, Clemson University, Clemson, SC 29634, USA. [8]The European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: nbesansk@nd.edu

some 3, and chromosome X showing the highest proportion of fixed differences [further explored in (*19*)] (table S3). The spatial distribution of polymorphism and divergence along chromosome arms also was investigated, using sliding window analyses to minimize noise from individual site-based divergence estimates (Fig. 1 and figs. S2 to S6) (*19*, *21*). Significant outlier divergence values falling in the top 1% of the empirical distribution (fig. S7) (*19*, *22*) are spread heterogeneously across the entire genome, not confined mostly to pericentromeric regions as observed in gene-based microarray studies (*14*, *15*, *19*).

The 436 genes overlapping with the top percentile of diverged 1-kb windows were tested for functional enrichment based on their gene ontology terms (database S1) (*19*). The 1-kb window size, smaller than the average gene size (~5.7 kb, including introns), mitigates the potentially confounding effect of physical clustering of functionally related members of gene families in *A. gambiae*. Functions related to G-protein–coupled receptor (GPCR) signaling, particularly neurohormone signaling, are significantly overrepresented in genomic regions of highest divergence (table S4). The neurohormone subfamily of GPCRs bind biogenic amines, neuropeptides, and protein hormone ligands, which in insects control development, feeding, reproduction, and complex behaviors (e.g., locomotion) that potentially bear on niche adaptation and mate recognition.

We also examined genes for evidence of divergence. Genes showing evidence of directional selection within forms, or amino acid fixations between forms (database S1 and figs. S2 to S6) (*19*), occur throughout the genome, suggesting that differential adaptation of M and S to their specific ecologies could involve an appreciable number of genes outside of pericentromeric regions. Some genomic regions appear to have experienced strong and recent selective sweeps, as illustrated by elevated divergence coupled with reductions in shared and private polymorphism (Fig. 1 and figs. S2 to 6). The most notable such region is on 2L (near Mb 25) centered on the *resistance to dieldrin* (*Rdl*) gene, which has been previously associated with insecticide resistance in *A. gambiae* and other insects (Fig. 1 and fig. S4) (*23*). In fact, M and S appear to carry different "resistant" substitutions (Ala296Ser in M, Ala296Gly in S) at *Rdl* (*23*) suggesting independent selective sweeps. Another notable region occurs on 3R near position ~40 Mb (Fig. 1 and fig. S4) and contains seven odorant receptors (ORs) whose closest match to the proteome of the fruit fly *Drosophila melanogaster* is OR67d. The single copy of this gene in *Drosophila* serves as the pheromone receptor for cis-vaccenyl acetate that mediates both social aggregation and female sexual receptivity (*24*), tempting speculation that these genes might play similar yet species-specific roles in M and S.

The pattern of genome-wide divergence inferred from colony-based genomic sequences is present in natural populations of M and S from the same region of Mali, based on a newly developed SNP genotyping array whose design included a subset of 400,000 SNPs derived from the M and S genome sequences (*25*). Indeed, visual and statistical concordance of patterns of divergence (fig. S8 and table S5) between the two data sets indicates that, at least in Mali, the widespread genomic divergence observed between M and S is not an artifact of laboratory culture (*19*). Future genome-wide studies spanning different geographic locations will be necessary to provide insight into whether and how this pattern varies spatially. Further population genomic sequencing by current short-read technologies will benefit from read-mapping to the independent M and S genome assemblies reported here.

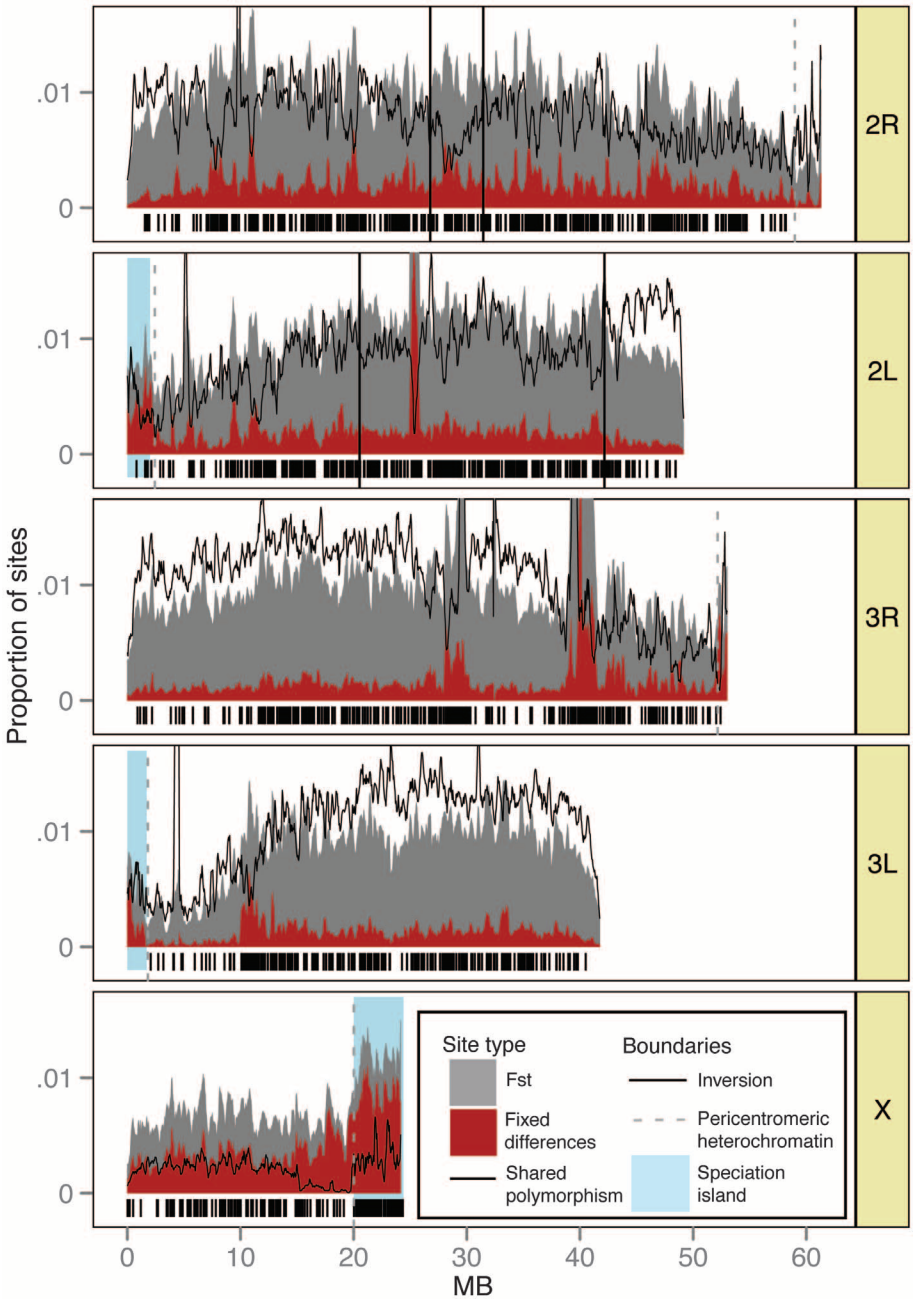

Fig. 1. Sliding window analysis of polymorphism and divergence in M and S based on 250-kb windows with 50-kb steps. Approximate boundaries of chromosomal rearrangements differing between M and S colonies (2R*c* and 2L*a*) are indicated by solid black vertical lines. Speciation islands sensu (*14*, *15*) are shaded in blue for reference. F_{ST} refers to the mean per-site estimate (*19*). Under the *x* axis, vertical black bars mark the approximate location of 1-kb windows whose divergence values fall in the top percentile of the distribution across autosomes (or the X chromosome, calculated separately). For both 250-kb and 1-kb windows, only windows meeting coverage and quality restrictions (*19*) are plotted.

The widely adopted model of ongoing speciation-with-gene-flow for M and S (*14*) posits that frequent hybridization leads to M-S genome homogenization in all except a few small regions near centromeres ("speciation islands"), which are barred from introgression because they contribute to differential fitness (i.e., ecological and reproductive isolation). Detection of much more widespread genomic divergence based on genotyping (*25*) and whole genome sequencing supports a very different model, in which realized gene flow between forms is currently much lower, and the process of speciation more advanced, than previously recognized, with the corollary that identification of genetic changes instrumental and not merely incidental to their ecological and behavioral divergence will be more difficult than initially hoped. However, powerful resources in the form of independently assembled M and S genomes and a SNP genotyping array (*25*) are now available for detecting morphologically cryptic vector subdivisions, probing their molecular basis, and ultimately developing innovative malaria interventions.

References and Notes

1. Y. F. Li, J. C. Costello, A. K. Holloway, M. W. Hahn, *Evolution* **62**, 2984 (2008).
2. H. D. Rundle, P. Nosil, *Ecol. Lett.* **8**, 336 (2005).
3. A. della Torre *et al.*, *Insect Mol. Biol.* **10**, 9 (2001).
4. A. della Torre, Z. Tu, V. Petrarca, *Insect Biochem. Mol. Biol.* **35**, 755 (2005).
5. A. Diabaté *et al.*, *Proc. Biol. Sci.* **276**, 4215 (2009).
6. A. Diabaté *et al.*, *J. Med. Entomol.* **43**, 480 (2006).
7. F. Tripet *et al.*, *Mol. Ecol.* **10**, 1725 (2001).
8. B. Caputo *et al.*, *Malar. J.* **7**, 182 (2008).
9. E. Oliveira *et al.*, *J. Med. Entomol.* **45**, 1057 (2008).
10. C. Costantini *et al.*, *BMC Ecol.* **9**, 16 (2009).
11. A. Diabaté, R. K. Dabire, N. Millogo, T. Lehmann, *J. Med. Entomol.* **44**, 60 (2007).
12. T. Lehmann, A. Diabaté, *Infect. Genet. Evol.* **8**, 737 (2008).
13. P. Nosil, D. J. Funk, D. Ortiz-Barrientos, *Mol. Ecol.* **18**, 375 (2009).
14. T. L. Turner, M. W. Hahn, S. V. Nuzhdin, *PLoS Biol.* **3**, e285 (2005).
15. B. J. White, C. Cheng, F. Simard, C. Costantini, N. J. Besansky, *Mol. Ecol.* **19**, 925 (2010).
16. M. Carneiro, N. Ferrand, M. W. Nachman, *Genetics* **181**, 593 (2009).
17. J. L. Feder, P. Nosil, *Evolution* **64**, 1729 (2010).
18. M. Slatkin, *Science* **236**, 787 (1987).
19. Materials and methods are available as supporting material on *Science* Online.
20. R. A. Holt *et al.*, *Science* **298**, 129 (2002).
21. B. S. Weir, L. R. Cardon, A. D. Anderson, D. M. Nielsen, W. G. Hill, *Genome Res.* **15**, 1468 (2005).
22. J. M. Akey *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1160 (2010).
23. W. Du *et al.*, *Insect Mol. Biol.* **14**, 179 (2005).
24. J. E. Mehren, *Curr. Biol.* **17**, R240 (2007).
25. D. E. Neafsey *et al.*, *Science* **330**, 514 (2010).

26. We thank W. M. Gelbart, an early advocate of this project, and J. L. Feder, P. Nosil, and M. W. Hahn for critical review. P. Howell of MR4 provided mosquitoes. Funding for genome sequencing of M (U54-HG00379) and S (U54-HG03068) was provided by the National Human Genome Research Institute. N.J.B. was supported by NIH (RO1 AI63508 and AI076584). M.K.N.L. was supported by Biotechnology and Biological Sciences Research Council research grant BB/E002641/1 to G.K.C. Sequence data are deposited with GenBank (accessions ABKP00000000 and ABKQ00000000).

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6003/512/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S5
References
Database S1

28 July 2010; accepted 9 September 2010
10.1126/science.1195755

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] M. A. T. Muskavitch[1,8,9]†

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 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

[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.
†To whom correspondence should be addressed. E-mail: marc.muskavitch@bc.edu