Insight into the Mechanism of the Influenza A Proton Channel from a Structure in a Lipid Bilayer

Mukesh Sharma,^{1,2} Myunggi Yi,^{3,4} Hao Dong,^{3,4} Huajun Qin,¹ Emily Peterson,⁵ David D. Busath,⁵ Huan-Xiang Zhou,^{3,4}* Timothy A. Cross^{1,2,4}*

The M2 protein from the influenza A virus, an acid-activated proton-selective channel, has been the subject of numerous conductance, structural, and computational studies. However, little is known at the atomic level about the heart of the functional mechanism for this tetrameric protein, a His³⁷-Trp⁴¹ cluster. We report the structure of the M2 conductance domain (residues 22 to 62) in a lipid bilayer, which displays the defining features of the native protein that have not been attainable from structures solubilized by detergents. We propose that the tetrameric His³⁷-Trp⁴¹ cluster guides protons through the channel by forming and breaking hydrogen bonds between adjacent pairs of histidines and through specific interactions of the histidines with the tryptophan gate. This mechanism explains the main observations on M2 proton conductance.

roton conductance by the M2 protein (with 97 residues per monomer) in influenza A is essential for viral replication (1). An M2 mutation, $Ser^{31} \rightarrow Asn$, in recent flu seasons and in the recent H1N1 swine flu pandemic renders the viruses resistant to antiviral drugs, amantadine and rimantadine (2). Previous structural determinations of M2 focused primarily on its transmembrane (TM) domain, residues 26 to 46 (3-8). Although the TM domain is capable of conducting protons, residues 47 to 62 following the TM domain are essential for the functional integrity of the channel. Oocyte assays showed that truncations of the post-TM sequence result in reduced conductance (9). Here, we report the structure of the "conductance" domain, consisting of residues 22 to 62, which in liposomes conducts protons at a rate comparable to that of the full-length protein in cell membranes (10, 11) and is amantadine-sensitive (fig. S1). This structure, solved in uniformly aligned 1,2-dioleoyl-sn-glycero-3-phosphatidylcholine:1,2dioleoyl-sn-glycero-3-phosphoethanolamine bilayers by solid-state nuclear magnetic resonance (NMR) at pH 7.5 and 30°C, shows striking differences from the structure of a similar construct (residues 18 to 60) solubilized in detergent micelles (12).

The structure is a tetramer (fig. S2), with each monomer comprising two helices (Fig. 1A). Residues 26 to 46 form a kinked TM helix, with the N-terminal and C-terminal halves having tilt angles of \sim 32° and \sim 22° from the bilayer normal, respec-

tively. The kink in the TM helix occurs around Gly³⁴, similar to the kinked TM-domain structure in the presence of amantadine (4). The amphipathic helix (residues 48 to 58) has a tilt angle of 105°, similar to that observed for the full-length protein (*13*) and resides in the lipid interfacial region (Fig. 1B). The turn between the TM and amphipathic helices is tight and rigid, as indicated by substantial anisotropic spin interactions for Leu⁴⁶ and Phe⁴⁷ (their resonances lie close to the TM and amphipathic helical resonance patterns, respectively; see fig. S3). The result is a structural base formed by the four amphipathic helices that stabilizes the tetramer.

The pore formed by the TM-helix bundle is lined by Val²⁷, Ser³¹, Gly³⁴, His³⁷, Trp⁴¹, Asp⁴⁴, and Arg⁴⁵, which include all of the polar residues of the TM sequence. The pore is sealed by the TM helices (Fig. 1B) and constricted by Val²⁷ at the N-terminal entrance and by Trp⁴¹ at the C-terminal exit (Fig. 1C). The gating role of Trp⁴¹ has long been recognized (*14*), and recently Val²⁷ was proposed to form a secondary gate (*15*). An important cavity between Val²⁷ and His³⁷ presents an amantadine-binding site (*3*, *7*, *15*). This drugbinding site is eliminated in a structure solubilized in detergent micelles (*12*) because of a much smaller tilt angle of the TM helices (*16*).

The linewidths of the NMR spectra (fig. S3) are much narrower for the conductance domain than for the TM domain (17), indicating substantially reduced conformational heterogeneity and higher stability (see also fig. S2). In the structure determined here, numerous nonpolar residues of the amphipathic helices extend the hydrophobic interactions interlinking the monomers, with their close approach facilitated by the small Gly⁵⁸ (Fig. 1D). In particular, Phe⁴⁸ interacts with Phe⁵⁵ and Leu⁵⁹ of an adjacent monomer and Phe⁵⁴ interacts with Leu⁴⁶ of another adjacent monomer.

The starting residues of the amphipathic helix, Phe⁴⁷ and Phe⁴⁸, are a sequence motif known to signal association of the helix with a lipid bilayer (18). The burial of the hydrophobic portion of the amphipathic helix in the tight intermonomer interface is consistent with hydrogen-deuterium exchange data showing it to be the slowestexchanging region for the full-length protein in a lipid bilayer (13). Furthermore, the Ser⁵⁰ hydroxyl here, which in the native protein is a palmitoylated Cys⁵⁰ residue, is located at an appropriate depth in the bilayer (at the level of the glycerol backbone; see Fig. 1B) for tethering the palmitic acid. A third native-like aspect of the amphipathic helix is the outward projection of the charged residues Lys⁴⁹, Arg⁵³, His⁵⁷, Lys⁶⁰, and Arg⁶¹ (Fig. 1B), which conforms to the "positive inside rule" such that M2 interacts favorably with negatively charged lipids in native membranes. The C termini of the amphipathic helices are situated to allow for the subsequent residues of the fulllength protein to form the tetrameric M1 binding domain. Contrary to the lipid interfacial location determined here, the detergent-solubilized structure has the four amphipathic helices forming a bundle in the bulk aqueous solution where the amides fully exchange with deuterium (12).

On the external surface at the C terminus of the TM helix, a hydrophobic pocket, with the Asp^{44} side chain at the bottom, has been described as a binding site for rimantadine (12). In the structure determined here, the large tilt of the TM helices widens the hydrophobic pocket, which is filled by the side chains of Ile⁵¹ and Phe⁵⁴ in the amphipathic helix, preventing accessibility to the Asp^{44} side chain from the exterior (Fig. 1D). Consequently, the formation of a rimantadine-binding site on the protein exterior is likely an artifact of the detergent environment used for that structural characterization.

The heart of acid activation and proton conductance in M2 is the tetrameric His³⁷-Trp⁴¹ cluster, referred to here as the HxxxW quartet (14, 19, 20). The pK_a values for the His³⁷ residues in the TM domain solubilized in a lipid bilayer were determined as 8.2, 8.2, 6.3, and <5.0 (20). At pH 7.5 used here, the histidine tetrad is doubly protonated; each of these two protons is shared between the $N_{\delta 1}$ of one histidine and the $N_{\epsilon 2}$ of an adjacent histidine, giving rise to substantial downfield 15N chemical shifts and resonance broadening for the protonated sites (20) indicative of a strong hydrogen bond (21, 22). The structure of the histidine tetrad as a pair of imidazoleimidazolium dimers is shown in Fig. 2A. In each dimer, the shared proton is collinear with the Ns1 and $N_{\epsilon 2}$ atoms (Fig. 2B); the two imidazole rings are within the confines of the backbones, to be nearly parallel to each other-a less energetically favorable situation than the perpendicular alignment of the rings in imidazole-imidazolium crystals and in computational studies (23–25). The two $N_{\delta 1}$ and two $N_{\epsilon 2}$ sites of the histidine tetrad not involved in the strong hydrogen bonds are protonated and project toward the C-terminal side (Fig. 2, A and B). The $N_{\epsilon 2}$ protons interact with the indoles of the Trp⁴¹ residues (Fig. 2A), and the two $N_{\delta 1}$ protons form hydrogen bonds with their

¹Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA. ²National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32310, USA. ³Department of Physics, Florida State University, Tallahassee, FL 32306, USA. ⁴Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA. ⁵Department of Physiology and Developmental Biology, Brigham Young University, Provo, UT 84602, USA.

^{*}To whom correspondence should be addressed. E-mail: hzhou4@fsu.edu (H.-X.Z.); cross@magnet.fsu.edu (T.A.C.)



Fig. 1. The tetrameric structure of the M2 conductance domain, solved by solid-state NMR spectroscopy and restrained molecular dynamics simulations, in liquid crystalline lipid bilayers. See (*32*) for details and fig. S3 for the NMR spectra. (**A**) Ribbon representation of the TM and amphipathic helices. One monomer is shown in red. The TM helix is kinked around the highly conserved Gly³⁴ (shown as C_α spheres). (**B**) Space-filling representation of the protein side chains in the lipid bilayer environment used for the NMR spectroscopy, structural refinement, and functional assay. C, O, N, and H atoms are colored green, red, blue, and white, respectively. The nonpolar residues of the TM and amphipathic helices form a continuous surface; the positively charged residues of the amphipathic helix are arrayed on the outer edge of the structure in optimal position to interact with charged lipids. The Ser⁵⁰ hydroxyl is also shown to be in an optimal position (as Cys⁵⁰) to accept a palmitoyl group in native membranes. (**C**) HOLE image (*33*) illustrating pore constriction at Val²⁷ and Trp⁴¹. (**D**) Several key residues at the junction between the TM and amphipathic helices, including Gly⁵⁸ (shown as C_α spheres), which facilitates the close approach of adjacent monomers, and lle⁵¹ and Phe⁵⁴, which fill a pocket previously described as a rimantadinebinding site (*12*).

respective histidine backbone carbonyl oxygens (Fig. 2B). Therefore, in this "histidine-locked" state of the HxxxW quartet, none of the imidazole N-H protons can be released to the C-terminal pore, resulting in a completely blocked channel. Furthermore, the only imidazole nitrogens available for additional protonation are the sites involved in the strong hydrogen bonds; acceptance and release of protons from the N-terminal pore by these sites, coupled with 90° side-chain χ_2 rotations, would allow the imidazole-imidazolium dimers to exchange partners (fig. S4). That the NMR data show a symmetric average structure suggests that such exchange occurs on a submillisecond time scale.

The structure of the HxxxW quartet at neutral pH suggests a detailed mechanism for acid activation and proton conductance (Fig. 3). Under acidic conditions in the viral exterior, a hydronium ion in the N-terminal pore attacks one of the imidazole-imidazolium dimers. In the resulting "activated" state, the triply protonated histidine tetrad is stabilized by a hydrogen bond with water at the newly exposed N₈₁ site on the N-terminal side and an additional cation- π interaction with a Trp⁴¹ residue at the N_{c2} site on the C-terminal side. Strong cation- π interactions between His³⁷ and Trp⁴¹ were observed by Raman spectroscopy in the TM domain under acidic conditions (*26*). These interactions protect the protons on the C-terminal



The histidine-locked state. (A) Top view of the tetrameric cluster of H³⁷xxxW⁴¹ (His³⁷ as sticks and Trp⁴¹ as spheres). Note the near-coplanar arrangement of each imidazole-imidazolium dimer that forms a strong hydrogen bond between N₀₁ and N_{e2}. In each dimer, the remaining N_{e2} interacts with the indole of a Trp⁴¹ residue through a cation- π interaction. The backbones have four-fold symmetry, as defined by the time-averaged NMR data. (B) Side view of one of the two imidazole-imidazolium dimers. Both the intraresidue N₆₁-H-O hydrogen bond and the interresidue N_{e2}-H-N₆₁ strong hydrogen bond can be seen. The near-linearity of the interresidue hydrogen bond is obtained at the expense of a strained C_α-C_β-C_γ angle (enlarged by ~10°) of the residue on the left.

side from water access. Conformational fluctuations of the helical backbones [in particular, a change in helix kink around $\text{Gly}^{34}(27)$] and motion of the Trp⁴¹ side chain could lead to occasional breaking of this cation- π interaction. In the resulting "conducting" state, the N_{e2} proton becomes exposed to water on the C-terminal side, allowing it to be released to the C-terminal pore. Upon proton release, the histidine-locked state is restored, ready for another round of proton uptake from the N-terminal pore and proton release to the C-terminal pore. In each round of proton conductance, the changes among the histidinelocked, activated, and conducting states of the HxxxW quartet can be accomplished by rotations (<45° change in χ_2 angle) of the His³⁷ and Trp⁴¹ side chains, which are much smaller than those envisioned previously (28) and even less than those needed for the imidazole-imidazolium dimers to exchange partners (fig. S4).

This proposed mechanism is consistent with many M2 proton conductance observations. The permeant proton is shuttled through the pore via



Fig. 3. Proposed mechanism of acid activation and proton conductance illustrated with half of the HxxxW quartet from a side view. The histidine-locked state (top) is shown with a hydronium ion waiting in the N-terminal pore. Acid activation is initiated with a proton transfer from the hydronium ion into the interresidue hydrogen bond between N_{$\delta 1$} and N_{$\epsilon 2$}. In the resulting activated state, the two imidazolium rings rotate so that the two nitrogens move toward the center of the pore; in addition, the protonated N_{$\delta 1$} forms a hydrogen bond with water in the N-terminal pore while the protonated N_{$\epsilon 2$} moves downward (via relaxing the C_{α}-C_{β}-C_{γ} angle) to form a cation- π interaction with an indole, thereby blocking water access from the C-terminal pore. The conducting state is obtained when this indole moves aside to expose the N_{$\epsilon 2$} proton to a water in the C-terminal pore. It was suggested previously (27) that the indole motion involves ring rotation coupled to backbone kinking. Once the N_{$\epsilon 2$} proton is released to C-terminal water, the HxxxW quartet returns to the histidine-locked state.

the histidine tetrad, at one point being shared between $N_{\delta 1}$ and $N_{\epsilon 2}$ of adjacent histidines. No other cations can make use of this mechanism, which explains why M2 is proton-selective. With the permeant proton obligatorily binding to and then unbinding from an internal site (i.e., the histidine tetrad), the proton flux is predicted to saturate at a moderate pH on the N-terminal side (27), which is consistent with conductance observations (29). Such a permeation model also predicts that the transition to saturation occurs at a pH close to the histidine-tetrad pK_a for binding or unbinding the permeant proton. Indeed, the transition is observed to occur around pH 6 (29), close to the third pK_a of the histidine tetrad. Without the histidines, as in the H37A mutant, the proton flux would lose pH dependence, as observed (19). Another distinguishing feature of the M2 proton channel is its low conductance, at ~100 protons per tetramer per second (fig. S1) (10, 11). Upon acid activation, the HxxxW quartet is primarily in the activated state. Only when the Trp⁴¹ gate opens occasionally to form

the conducting state is the proton able to be released to the C-terminal pore, thus explaining the low conductance.

In our mechanism, the histidine tetrad senses only acidification of the N-terminal side. This provides an explanation for an observation of Chizhmakov et al. (30) when they applied a positive voltage to drive protons outward in M2transformed MEL cells. A step increase in the bathing buffer pH from 6 to 8 (with the intracellular pH held constant at pH 6) produced a brief increase in outward current, as expected for the added driving force by the pH gradient (31); however, the outward current quickly decayed to a level lower even than that before the pH increase. In our structure, when the pH in the Nterminal side is 8, the HxxxW quartet is stabilized in the histidine-locked state and the Trp⁴¹ gate prevents excess protons in the C-terminal pore from activating the histidine tetrad. Upon removal of the Trp⁴¹ gate (e.g., by a W41A mutation), a substantial outward current would be produced, as was observed (14). A related observation is that

the M2 proton channel can be blocked by Cu²⁺ (through His³⁷ coordination) applied extracellularly, but not intracellularly (*14*). Again, the triply protonated histidine tetrad stays predominantly in the activated state, in which the Trp⁴¹ gate blocks Cu²⁺ access to His³⁷ from the C-terminal side. However, the W41A mutation opens that access (*14*).

The structure of the M2 conductance domain solved in liquid crystalline bilayers has led to a proposed mechanism for acid activation and proton conductance. The mechanism takes advantage of conformational flexibility, both in the backbones and in the side chains, which arises in part from the weak interactions that stabilize membrane proteins. M2 appears to use the unique chemistry of the HxxxW quartet to shepherd protons through the channel.

References and Notes

- 1. M. Takeda, A. Pekosz, K. Shuck, L. H. Pinto, R. A. Lamb, J. Virol. 76, 1391 (2002).
- L. Gubareva *et al.*; Centers for Disease Control and Prevention (CDC), *Morb. Mortal. Wkly. Rep.* 58, 1 (2009).
- K. Nishimura, S. Kim, L. Zhang, T. A. Cross, *Biochemistry* 41, 13170 (2002).
- 4. J. Hu et al., Biophys. J. 92, 4335 (2007).
- A. L. Stouffer *et al.*, *Nature* **451**, 596 (2008).
 S. D. Cady, M. Hong, *Proc. Natl. Acad. Sci. U.S.A.* **105**,
- 1483 (2008).
- 7. S. D. Cady et al., Nature 463, 689 (2010).
 - R. Acharya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15075 (2010).
 - C. Ma et al., Proc. Natl. Acad. Sci. U.S.A. 106, 12283 (2009).
- 10. J. A. Mould et al., J. Biol. Chem. 275, 8592 (2000).
- 11. T. I. Lin, C. Schroeder, J. Virol. 75, 3647 (2001).
- 12. J. R. Schnell, J. J. Chou, Nature 451, 591 (2008).
- C. Tian, P. F. Gao, L. H. Pinto, R. A. Lamb, T. A. Cross, *Protein Sci.* 12, 2597 (2003).
- 14. Y. Tang, F. Zaitseva, R. A. Lamb, L. H. Pinto, J. Biol. Chem. 277, 39880 (2002).
- M. Yi, T. A. Cross, H. X. Zhou, J. Phys. Chem. B 112, 7977 (2008).
- 16. T. A. Cross, M. Sharma, M. Yi, H. X. Zhou, *Trends Biochem. Sci.* 10.1016/j.tibs.2010.07.005 (2010).
- 17. C. Li, H. Qin, F. P. Gao, T. A. Cross, *Biochim. Biophys.* Acta **1768**, 3162 (2007).
- T. L. Lau, V. Dua, T. S. Ulmer, J. Biol. Chem. 283, 16162 (2008).
- P. Venkataraman, R. A. Lamb, L. H. Pinto, J. Biol. Chem. 280, 21463 (2005).
- 20. J. Hu et al., Proc. Natl. Acad. Sci. U.S.A. 103, 6865 (2006).
- X.-j. Song, A. E. McDermott, *Magn. Reson. Chem.* 39, S37 (2001).
- 22. X.-j. Song, C. M. Rienstra, A. E. McDermott, *Magn. Reson. Chem.* **39**, S30 (2001).
- 23. A. Quick, D. J. Williams, *Can. J. Chem.* **54**, 2465 (1976).
- J. A. Krause, P. W. Baures, D. S. Eggleston, Acta Crystallogr. B 47, 506 (1991).
- 25. W. Tatara, M. J. Wojcik, J. Lindgren, M. Probst, J. Phys. Chem. A 107, 7827 (2003).
- 26. A. Okada, T. Miura, H. Takeuchi, *Biochemistry* **40**, 6053 (2001).
- M. Yi, T. A. Cross, H. X. Zhou, Proc. Natl. Acad. Sci. U.S.A. 106, 13311 (2009).
- L. H. Pinto et al., Proc. Natl. Acad. Sci. U.S.A. 94, 11301 (1997).
- 29. I. V. Chizhmakov et al., J. Physiol. 494, 329 (1996).
- 30. I. V. Chizhmakov et al., J. Physiol. 546, 427 (2003).
- 31. The brief increase can be attributed to the release of protons from the triply protonated histidine tetrad to the

Downloaded from www.sciencemag.org on October 21, 2010

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.
- 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 Al023007. 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 2L0].

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. Besansky⁴†

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*,

*These authors contributed equally to this work. †To whom correspondence should be addressed. E-mail: nbesansk@nd.edu 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 2La and 2Rc (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-

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