Adaptive divergence between incipient species of *Anopheles gambiae* increases resistance to *Plasmodium*

Bradley J. White[a], Mara K. N. Lawniczak[b], Changde Cheng[a], Mamadou B. Coulibaly[c], Michael D. Wilson[d], N'Fale Sagnon[e], Carlo Costantini[f,g], Frederic Simard[f,h], George K. Christophides[b], and Nora J. Besansky[a,1]

[a]Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556; [b]Division of Cell and Molecular Biology, Department of Life Sciences, Imperial College London, London SW7 2AZ, United Kingdom; [c]Malaria Research and Training Center, 1805 Bamako, Mali; [d]Noguchi Memorial Institute for Medical Research, University of Ghana, LG 581 Legon, Ghana; [e]Centre National de Recherche et Formation sur le Paludisme, 01 BP 2208 Ouagadougou 01, Burkina Faso; [f]Institut de Recherche pour le Développement, Unité de Recherche R016, BP 64501 Montpellier, France; [g]Organisation de Coordination pour la Lutte contre les Endémies en Afrique Centrale, BP 288 Yaounde, Cameroon; and [h]Institut de Recherche en Sciences de la Santé, 01 BP 545, Bobo-Dioulasso, Burkina Faso

Edited* by Francisco J. Ayala, University of California, Irvine, CA, and approved November 30, 2010 (received for review September 14, 2010)

The African malaria mosquito *Anopheles gambiae* is diversifying into ecotypes known as M and S forms. This process is thought to be promoted by adaptation to different larval habitats, but its genetic underpinnings remain elusive. To identify candidate targets of divergent natural selection in M and S, we performed genome-wide scanning in paired population samples from Mali, followed by resequencing and genotyping from five locations in West, Central, and East Africa. Genome scans revealed a significant peak of M-S divergence on chromosome 3L, overlapping five known or suspected immune response genes. Resequencing implicated a selective target at or near the *TEP1* gene, whose complement C3-like product has antiparasitic and antibacterial activity. Sequencing and allele-specific genotyping showed that an allelic variant of *TEP1* has been swept to fixation in M samples from Mali and Burkina Faso and is spreading into neighboring Ghana, but is absent from M sampled in Cameroon, and from all sampled S populations. Sequence comparison demonstrates that this allele is related to, but distinct from, *TEP1* alleles of known resistance phenotype. Experimental parasite infections of advanced mosquito intercrosses demonstrated a strong association between this *TEP1* variant and resistance to both rodent malaria and the native human malaria parasite *Plasmodium falciparum*. Although malaria parasites may not be direct agents of pathogen-mediated selection at *TEP1* in nature—where larvae may be the more vulnerable life stage—the process of adaptive divergence between M and S has potential consequences for malaria transmission.

ecological speciation | malaria vector | population genomics | thioester immune gene

Ecologically based divergent selection drives adaptive phenotypic variation and biological diversification (1, 2). However, the underlying genetic differences and often the specific adaptive traits they control are little known (1, 2). In cases where candidate gene and quantitative trait mapping approaches are unfeasible, population genomics provides a powerful tool to identify the genetic targets of divergent selection pressure, and thereby the adaptive traits themselves (2, 3). Population genomic surveys are expected to reveal elevated differentiation between populations and reduced sequence diversity within populations, around genes involved in local adaptation (4, 5).

Anopheles gambiae represents a particularly relevant study system, given its medical importance as a main vector of human malaria in Africa and its subdivision into sympatric ecotypes (termed M and S forms) (6, 7). The M and S forms cannot be distinguished morphologically or by other specific phenotypic traits. However, general ecological differences are known (8–11). The S form breeds in small ephemeral pools and puddles across sub-Saharan Africa, whereas the M form exploits larger, more stable breeding sites closely associated with agricultural or urban activity in West and Central Africa. Moreover, recent field studies have suggested fitness tradeoffs apparently related to the stability of their respective breeding sites (12, 13). The S form develops more rapidly and outcompetes M in predator-free space, but the M form is superior at predator avoidance, as expected in an ecotype adapted to longer-lived habitats.

Here, we have used a high-resolution population genomics approach to survey potential genic targets of divergent natural selection between M and S forms in Mali. By targeted resequencing and allele-specific genotyping of a candidate region on chromosome 3L from multiple geographic locations in Africa, we reveal a variable spatial pattern of adaptive divergence within and between these emerging species. Candidate gene analysis and subsequent phenotypic association implicate pathogen-mediated selection acting on the innate immunity gene, *TEP1*, as a factor contributing to adaptive divergence of M and S.

Results and Discussion

We used a gene-based oligonucleotide microarray with a density of approximately one marker every 2 kb to map divergence between M and S. Using 10 sympatric M- and S-form mosquitoes sampled from Mali, we identified four regions of significantly elevated differentiation on homosequential chromosome arms, three of which were adjacent to centromeres on chromosomes X, 2, and 3 (14). The imputation of pericentromeric genes in reproductive or ecological isolation between M and S is complicated by low sequence diversity and reduced recombination characteristic of centromere-proximal regions (14, 15), prompting us to concentrate follow-up effort on the fourth region, encompassing ~100 kb on chromosome arm 3L, ~11 Mb from the centromere (Fig. 1). Of the five genes predicted within this region, four (*TEP1*, *TEP3*, *TEP10*, and *TEP11*) belong to the thioester-containing protein family implicated in *An. gambiae* innate immunity (16, 17). The fifth, AGAP010817, encodes an unknown protein, potentially also playing a role in immunity on the basis of

Author contributions: B.J.W. and N.J.B. designed research; B.J.W. and M.K.N.L. performed research; M.B.C., M.D.W., N.S., C. Costantini, and F.S. contributed new reagents/analytic tools; B.J.W., M.K.N.L., C. Cheng, G.K.C., and N.J.B. analyzed data; B.J.W. and N.J.B. wrote the paper; and M.B.C., M.D.W., N.S., C. Costantini, and F.S. supervised field collections.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. GU394083–GU394938).

[1]To whom correspondence should be addressed. E-mail: nbesansk@nd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1013648108/-/DCSupplemental.

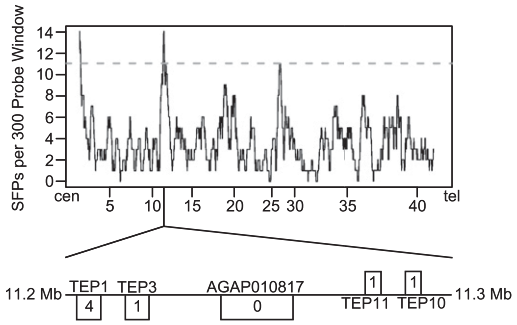

Fig. 1. Genomic microarray analysis of M-S divergence on chromosome 3L in Mali. A peak of significant differentiation spanning ~100 kb was detected at ~11 Mb from the centromere (cen), below which five genes are predicted, four belonging to the *TEP* immune gene family. Orientation of the coding strand (5′–3′) is indicated by boxes, above or below the horizontal line, which enclose the number of SFPs detected per gene.

its predicted peroxidase, CUB (complement C1r/C1s, Uegf, Bmp1), and low-density lipoprotein receptor domains.

To further explore this pattern of genetic variation on 3L in a larger geographic framework, we performed targeted resequencing of additional M and S sampled from the same Mali population and three other locations: Burkina Faso, Ghana, and Cameroon (Table S1). Portions of nine genes were sequenced from 32 to 73 mosquitoes per form: three candidate genes (*TEP1*, *TEP3*, and AGAP010817) from within the region of elevated divergence, and six reference (control) genes outside and flanking it (three on each side) (Table S2).

Concordant with the microarray results, sequence divergence between M and S was significantly higher in Mali (and Burkina Faso) in candidate genes than in flanking reference genes (Fig. 2*A* and Tables S3, S4, and S5). By contrast, interform divergence was indistinguishable between these gene classes in Ghana and Cameroon samples, suggesting that the geographic scope of sequence divergence in this genomic region is limited. Population-pairwise comparisons between geographic sites reinforced this observation. Only between Mali or Burkina Faso versus other sites, were candidate genes strongly differentiated relative to reference genes, and only in M form samples. In S, differentiation was uniformly low for both gene classes, regardless of geography (Fig. 2*B*).

Consistent with a selective sweep (the rapid fixation of a beneficial mutation and linked variants) in M populations from Mali and Burkina Faso, sequence polymorphism was reduced 5–16-fold in candidate versus reference genes from these samples (Table S6). Reduced diversity was not observed in S. A commonly applied statistical test of significant departure from neutral evolution (Hudson–Kreitman–Aguadé, HKA; ref. 18) evaluates within-species diversity to between-species divergence from an outgroup, with the null expectation that these levels should be correlated and their ratio equivalent across loci. This test can be difficult to apply in *An. gambiae*, due to the dearth of outgroups that are sufficiently closely related yet without extensive shared polymorphism (19). The best available option was the sibling species *Anopheles melas*; gene trees incorporating sequences from *An. gambiae* and *An. melas* are reciprocally monophyletic, suggesting that shared polymorphism is minimal (19, 20). A multilocus maximum-likelihood implementation of the HKA test (21) was used to compare the likelihood of two alternative models, in which (*i*) reference and candidate genes are neutral or (*ii*) candidate genes are subject to selection. By likelihood ratio test, reduced diversity relative to divergence was highly significant in Mali and Burkina Faso for M candidate genes (Table S7). Notably, among the three candidate genes surveyed, maximum-likelihood estimates of the selection

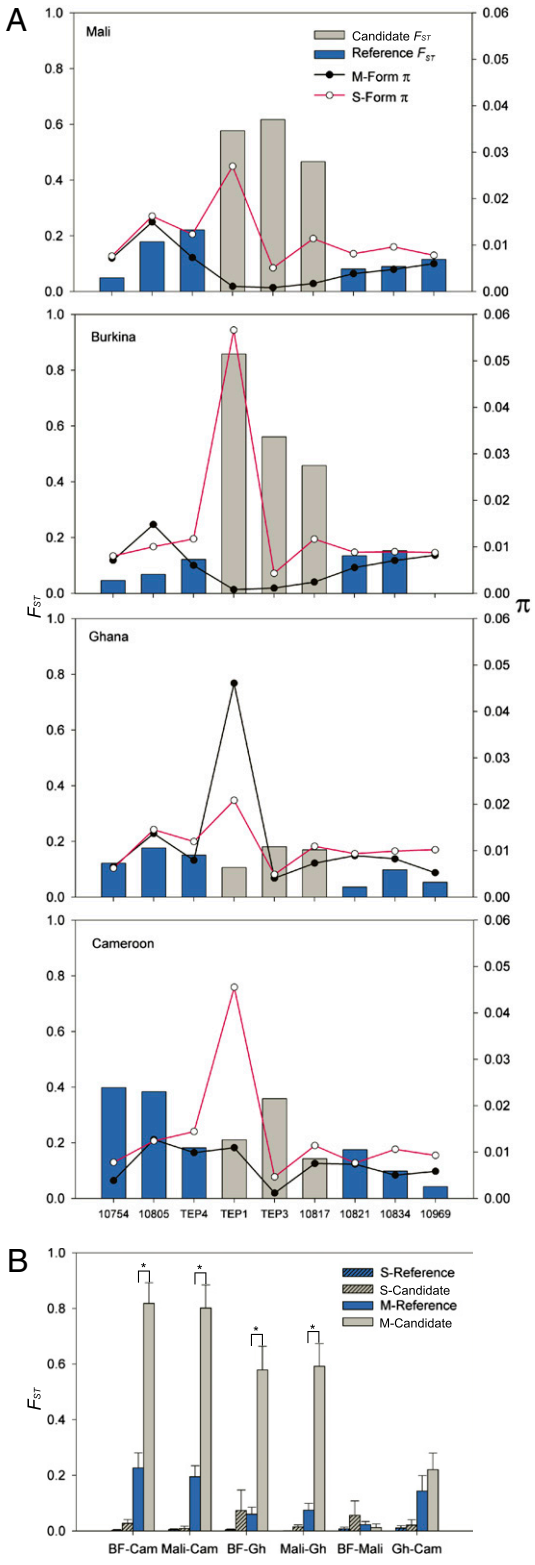

Fig. 2. Geographically restricted natural selection in the M form. (*A*) Country-specific plots of within-form nucleotide diversity (π, *Right y*-axis) and between-form divergence (F_{ST}, *Left y*-axis), at surveyed candidate and reference genes. (*B*) Within-form divergence between population pairs, at candidate and reference genes. Brackets and asterisks denote statistically significant contrasts between candidate and reference genes in the M form.

parameter (k) were most extreme at the *TEP1* gene, prompting further investigation of this gene.

Like its C3 vertebrate homolog, TEP1 is a complement-like opsonin whose binding triggers killing, via phagocytosis of bacteria or lysis/melanization of the malaria parasites *Plasmodium berghei* and *Plasmodium falciparum* (17, 22–24). Recently, allelic variants of *TEP1* from *An. gambiae* laboratory strains refractory to *P. berghei* were directly implicated in enhanced lysis and melanization of parasites (22), owing to faster binding kinetics relative to alleles in susceptible strains (25). Corresponding *TEP1* alleles (*TEP1r* and *TEP1s*) also segregate in natural populations of *An. gambiae* and its sibling species *Anopheles arabiensis* (26, 27). The majority of substitutions between these highly divergent alleles are found in the thioester binding domain (TED) where our *TEP1* sequence survey was focused. To compare the M- and S-form *TEP1* sequences obtained in this study with previously reported *TEP1r* and *TEP1s* alleles from natural populations and laboratory colonies (22, 26, 27), we performed a multiple sequence alignment on the basis of amino acid sequences deduced from the TED region (abridged in Fig. 3, for clarity). Inspection of the alignment revealed the coexistence of both resistant (*TEP1r*) and susceptible (*TEP1s*) allelic classes in all four of our S populations and in our M populations from Ghana and Cameroon, a factor contributing to high overall diversity at this gene in pooled (*TEP1r* and *TEP1s*) samples (26, 27) (Fig. 2*A* and Table S2). Although the co-occurrence of *TEP1r* and *TEP1s* alleles was expected (e.g., ref. 27), a distance-based gene tree showed that within the larger cluster of *TEP1r*-like resistant alleles, were two distinctive subclusters, referred to here as $TEP1r^A$ and $TEP1r^B$ (Fig. 3 and Fig. S1). The $TEP1r^A$ and $TEP1r^B$ subclusters differ by five fixed (or nearly fixed) amino acid replacements caused by six nonsynonymous nucleotide substitutions. Alleles within both subclusters have been previously recorded (22, 27), but neither the partitioning of sequence variation between subclusters nor their differential distribution between molecular forms of *An. gambiae* was appreciated. The $TEP1r^A$ subcluster contains the phenotypically resistant allele from the 4Arr laboratory strain (named *Tep1*R2*; ref. 22), as well as all *TEP1r* alleles sampled by Obbard et al. (27) from *An. gambiae* natural populations in Cameroon and Kenya. This $TEP1r^A$ subcluster also contains all *TEP1r* alleles sampled from the S form in the present study and a subset of those sampled from the M form, from Ghana and Cameroon. The $TEP1r^B$ subcluster contains only M-form alleles, found only in M sampled from Mali, Burkina Faso, and Ghana. This $TEP1r^B$ subcluster also contains the allele segregating in the M-form Mali-NIH strain and alleles sampled (exclusively) from Burkina Faso by Obbard et al. (27), but that study did not distinguish between M and S forms. Obbard et al. (27) emphasized recombination between *TEP1* and other *TEP* loci, as well as between *TEP1r* and *TEP1s*, as a likely force contributing to the evolution of *TEP1*. Consistent with this idea, possible recombination tracts are evident between $TEP1r^A$ and *TEP1s*, and between $TEP1r^A$ and $TEP1r^B$ (the L3-5 *TEP1*R1* allele) in Fig. 3. The same $TEP1r^A$*/s* recombinant alleles reported by Obbard et al. (27) from Mt. Cameroon also were sampled broadly in the present study, albeit at low frequency.

Although not definitive, two lines of evidence point to $TEP1r^B$ as a target of positive natural selection, as proposed previously for $TEP1r^A$. Whereas nucleotide diversity in *TEP1s* matches the genomewide average ($\pi \approx 1\%$; ref 26), it is sharply lower for both *TEP1r* subclasses ($TEP1r^A$, $\pi \approx 0.068\%$; $TEP1r^B$, $\pi \approx 0.095\%$) (Fig. S2). Additionally, nonneutral protein evolution between $TEP1r^A$ and $TEP1r^B$ allelic subclasses is suggested by the results of a test analogous to that of McDonald and Kreitman (ref. 27; polymorphic sites: 2 nonsynonymous (ns), 7 synonymous (s); fixed sites: 6 ns, 0 s; $P = 0.007$).

To examine the geographic distribution of the *TEP1* alleles in more detail, allele-specific primers were used to genotype additional *An. gambiae* and *An. arabiensis* from the same four countries and Tanzania (Fig. 4 and Fig. S3). On the basis of combined sequence and genotype evidence, the $TEP1r^B$ class is almost exclusive to the M form (one $TEP1r^B$ allele was found among 918 S-form chromosomes sampled; Table S8). Apparently $TEP1r^B$ has been swept to fixation in Mali and Burkina Faso in M, but is absent outside of West Africa. Very recent gene flow of $TEP1r^B$ into Ghana populations of M is likely: all three

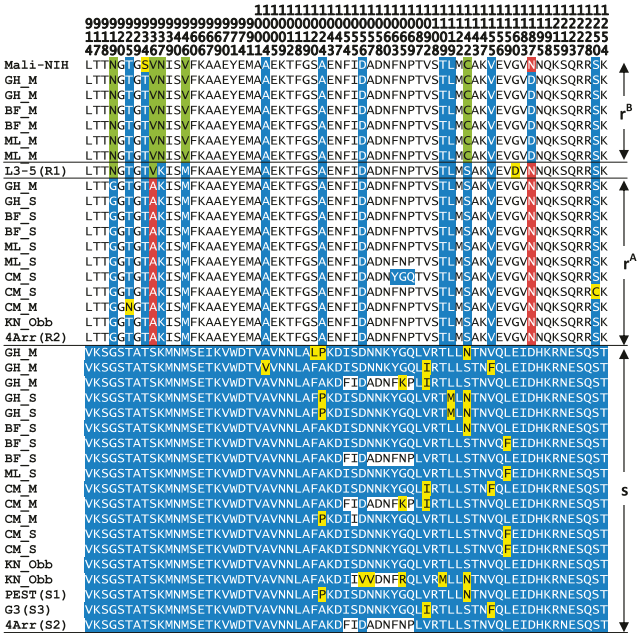

Fig. 3. Three allelic classes at the *TEP1* locus. Amino acid alignment based on the TED region of *TEP1* shows only variable positions, numbered following ref. 22 for a subset of available sequences. Shading highlights amino acid differences within and between allelic classes, which are separated by a horizontal black line and labeled at *Right* (s, *TEP1s*; r^A, $TEP1r^A$; and r^B, $TEP1r^B$). Allele designations at *Left* refer to sequences from laboratory strains (L3-5, Mali-NIH, 4Arr, PEST, and G3) or from the field. Alleles from the field are labeled by geographic origin (GH, Ghana; BF, Burkina Faso; ML, Mali; CM, Cameroon; and KN, Kenya) followed by molecular form (M, M form and S, S form) or by author (Obb, Obbard et al.; ref. 27).

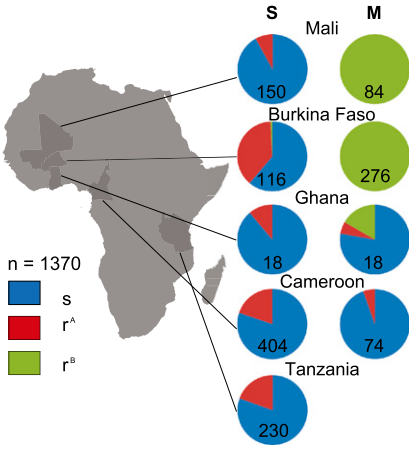

Fig. 4. Geographic structure of allelic variation at *TEP1*. Frequency of *TEP1s* (s, blue), $TEP1r^A$ (r^A, red), and $TEP1r^B$ (r^B, green) alleles of 1,370 chromosomes sampled in M and S from the indicated countries in Africa. Numbers within pie graphs represent chromosomes sampled per form in each country.

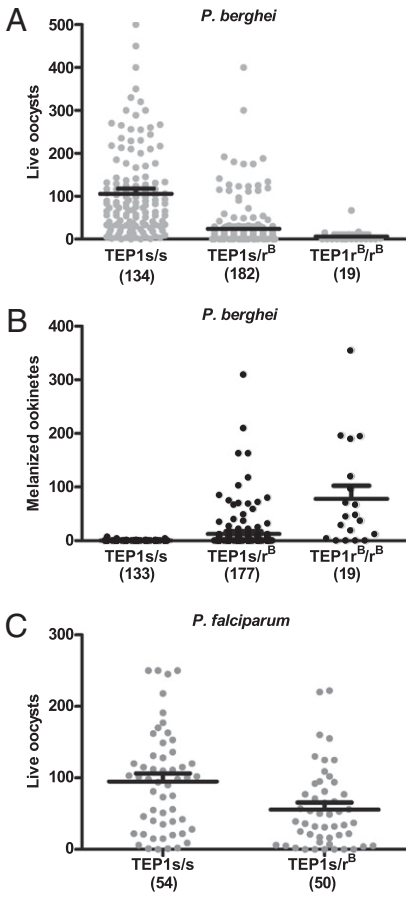

Fig. 5. The *TEP1r*^{B} allele increases mosquito resistance to *Plasmodium*. (*A*) *An. gambiae–P. berghei*: Number of live oocysts per midgut and mean (central bar) ± SE, $P < 0.0001$. (*B*) *An. gambiae–P. berghei*: Number of melanized ookinetes per midgut and mean (central bar) ± SE, $P < 0.0001$. (*C*) *An. gambiae–P. falciparum*: Number of live oocysts per midgut and mean (central bar) ± SE, $P < 0.01$.

TEP1r^B*/s* mosquitoes were heterozygous for a larger haplotype block (≥13 kb) in which *TEP1r*^{B} was in complete linkage disequilibrium with a CR1 retrotransposon inserted between *TEP1* and *TEP3* (*Materials and Methods*), and SNPs in *TEP3*. The *TEP1s* class predominates, but co-occurs with *TEP1r*^{A} in all S populations sampled, in M from Ghana and Cameroon, and in *An. arabiensis* (Table S8). The relative contributions of balancing and directional selection in the maintenance of the *TEP1s/r* polymorphisms in these populations cannot be determined from our data, but is a prime topic for future studies.

The resistance phenotype of the *TEP1r*^{B} allele has not been studied. To associate a phenotype with the *TEP1r*^{B} allelic class, we crossed two *An. gambiae* M colonies fixed for *TEP1r*^{B} (Mali-NIH) or *TEP1s* (Yaoundé). Using a set of sixth-generation advanced intercross progeny from reciprocal F_1 crosses to ensure recombination between the different genetic backgrounds, we bloodfed mosquitoes on mice infected with *P. berghei.* Ten days after bloodfeeding, the number of live and melanized oocysts was counted and *TEP1* was genotyped in each mosquito. As no qualitative differences were observed between the reciprocal crosses, results were combined. *TEP1* genotype explained a large percentage of variance in infection intensity (21% in the case of live oocysts, Fig. 5*A*; 13% in the case of melanized ookinetes, Fig. 5*B*). Parasite burdens were lowest in *TEP1r*^{B}*/r*^{B} mosquitoes, intermediate in *TEP1r*^{B}*/s*, and heaviest in *TEP1s/s*. Importantly, 96% of mosquitoes with no live oocysts ($n = 81$), and 100% with exclusively melanized parasites ($n = 48$), had at least one *TEP1r*^{B} allele (Fig. S4).

As the *P. berghei–An. gambiae* laboratory model is a combination not found in nature, we extended our phenotypic analysis of the *TEP1r*^{B} genotype to *An. gambiae* infected with the human malaria parasite it normally transmits, *P. falciparum.* Our experimental design was identical, except that F_8 rather than F_6 mosquitoes were analyzed. During these crosses, *TEP1r*^{B} allele frequency decreased rapidly in successive generations (*Materials and Methods*), leaving only heterozygotes and *TEP1s* homozygotes for analysis by F_8. Concordant with results for *P. berghei*, *P. falciparum* oocyst burden was significantly lower in *TEP1r*^{B}*/s* heterozygotes versus *TEP1s* homozygotes (Fig. 5*C*). These results strongly suggest that *TEP1r*^{B} or something very closely linked to it, acts in a semidominant but nonspecific manner to mediate increased mosquito resistance to both native and nonnative *Plasmodium* species.

Considering the key role of TEP1 in a generalized antimicrobial and antiparasitic response (17), adaptation at *TEP1r* in *An. gambiae* is most likely driven by pathogen pressure. Although the resistance phenotypes of *TEP1r*^{B} and *TEP1r*^{A} have never been compared directly in the laboratory, the recent selective sweep of *TEP1r*^{B} into West African populations of M suggests an advantage for this derived allele over both *TEP1s* and *TEP1r*^{A} under field conditions experienced by the M form. The absence of *TEP1r*^{B} from Cameroon populations of M is plausibly due to restricted gene flow between West and Central African populations of this form (28, 29), as reinforced in this study by higher differentiation of reference genes in contrasts involving Cameroon (Fig. 2*B*), but also could entail spatially varying selection. One explanation for the near absence of *TEP1r*^{B} in the S form is restricted gene flow, although it may also be the case that the pathogen community in the more permanent bodies of water in which M larvae develop present different challenges than those faced by S-form larvae. Previous evidence indicates that *An. gambiae* larval fitness can be drastically reduced by naturally occurring pathogens (30, 31), suggesting that pathogen pressure at the larval stage may be an even more potent selective force than immune challenges faced by adults.

In mediating antiparasite activity in the mosquito hemolymph, TEP1 does not act alone. In fact, the mature TEP1 protein must physically interact with two leucine-rich repeat proteins (APL1 and LRIM1) encoded by genes on an independent chromosome, 2L (32, 33). Thus, efficiency of parasite binding and killing may be a function of polymorphisms in TEP1 itself, and/or in APL1, LRIM1, or other proteins that control its function. However, the gene array-based scans performed here did not reveal M-S divergence at LRIM1 and are not informative about divergence at *APL1*, as unique probes targeting the *APL1* genes are either absent or underrepresented on the microarray.

In the context of ecological speciation, our study offers an example of how a bottom-up genomics approach can be used to gain insight into the genetic basis of adaptive traits without prior knowledge about their phenotypic affects. Our results also highlight the public health implications of ecological speciation within *An. gambiae*. In the process of adaptive divergence, aspects of physiology and behavior may change in ways that potentially impact vectorial capacity—with consequences for malaria epidemiology and control.

Materials and Methods

Mosquito Collections. Indoor resting mosquitoes were sampled from various locations (Table S1) by insecticide spray sheet collection and morphologically identified as *An. gambiae sensu lato* (34).

DNA Methods and Analysis. DNA was isolated from individual carcasses using the DNeasy Extraction kit (Qiagen). Sibling species and molecular forms were identified using diagnostic rDNA–PCR assays (35, 36). For microarray hybrid-

izations, fluorescently labeled genomic DNA samples from 10 M-form and 5 S-form *An. gambiae* from southern Mali were individually hybridized to 15 Affymetrix *Anopheles/Plasmodium* GeneChips, as described (37) (*SI Materials and Methods*). To test for significant clustering of single feature polymorphisms (SFPs), a sliding window analysis was performed with window and step sizes of 300 and 20 probes, respectively. Separately by chromosome arm, each window was tested (χ^2) for an excess of SFPs compared with the number expected, on the basis of the arm-specific frequency of significant probes, evaluated after Bonferroni correction (dashed line in Fig. 1). For sequence determination, PCR products amplified from genomic DNA template of single mosquitoes (*SI Materials and Methods*) were directly sequenced on both strands, using an Applied Biosystems 3730xl DNA Analyzer and Big Dye Terminator v3.1 chemistry. Sequences of AGAP10764 and AGAP10969 sampled from Mali were previously determined on the basis of primers and protocol given in ref. 14. For other genes, primers targeting exons (Table S9) were designed for this study using Primer3 (38) and custom synthesized. Sequences have been deposited in GenBank (GU394083–GU394938). Summary statistics of diversity and divergence were measured using DnaSP 4.20.2 (39). Significance values for F_{ST} and Tajima's *D* were determined by conducting 10,000 coalescent simulations in Arlequin 3.1 (40) and DnaSP, respectively. HKA tests (18) were performed using divergence to *An. melas.* For heuristic purposes, two alternative outgroups also were used: a sympatric sample of the S form and *An. arabiensis*. A maximum-likelihood implementation of the HKA test was performed using maximum-likelihood (ML) HKA (21) as described (14). Amino acid alignment was performed using MUSCLE. Phylogenetic inference was based on residues 891–1258 of *TEP1*, performed within Mega4 (41) using the neighbor-joining method (42) and evolutionary distances inferred from the Jones, Taylor, Thornton (JTT) matrix-based method (43). A test of neutral protein evolution analogous to the McDonald and Kreitman test (44) was performed following ref. 27, after removing recombinant sequences.

Allele-Specific PCR Genotyping of *TEP1*. We designed primers to distinguish the *TEP1s*, *TEP1r*^{A}, and *TEP1r*^{B} alleles by exploiting fixed nucleotide differences between them (Fig. S3 and *SI Materials and Methods*), using the artificial mismatch technique (45) in which primers carry an intentional mismatch to the target, three nucleotides from the 3′ end. Primers were validated on previously sequenced specimens and then applied to additional field and laboratory specimens (Table S9).

Crosses. Reciprocal crosses, initiated with ∼150 virgins of each sex, were performed between M-form colonies homozygous for *TEP1s* (Yaoundé) or *TEP1r*^{B} (Mali-NIH), as determined by sequencing. F_1's were intercrossed, as were the next four generations. Three groups of at least 100 F_6 specimens from both crosses were fed on mice infected with the GFP-expressing *P. berghei* strain PbGFP$_{CON}$ (46). Nonbloodfed females were removed the next day and the number of live and melanized oocysts was counted from all remaining female mosquitoes after 10 d. Following dissection, DNA extraction from a leg of each specimen was performed following ref. 47. Allele-specific genotyping at *TEP1* was performed with 10 μL of template DNA. The percentage of variance in both infection intensity and melanization explained by *TEP* genotype was calculated using linear regression. *P. falciparum* infections with the ND37 clone of the NF54 isolate (likely of African origin) followed the above protocol except that infected human blood was introduced to 100 F_8 mosquitoes of each cross via membrane feeder.

The *TEP1r*^{B} allele decreased in frequency as both intercrosses progressed. Under laboratory conditions, the cost of maintaining the allele appears to outweigh any immune benefits. In nature, the immune benefits of carrying *TEP1r*^{B} in pathogenic larval environments may override any intrinsic costs. Alternatively, decline of the allele in our crosses may be due to negative epistasis between *TEP1r*^{B} (or tightly linked variants) and mutations in the Yaoundé colony.

Structural Differences. Mali-NIH contains only *TEP1r*^{B} and Pimperena S contains *TEP1s* and *TEP1r*^{A} alleles. Sequence alignments of the M, S, and PEST strain in the five-gene region of divergence on 3L revealed an insertion present in M and PEST, but not S. In PEST, which like M (Mali-NIH) contains the *TEP1r*^{B} allele, the insertion is annotated as a 2.7-kb CR1 retrotransposon. This transposable element (TE) is located in genomic coordinates 11,209,718–11,212,323 in the intergenic distance between *TEP1* and *TEP3*, ∼3 kb upstream of *TEP1* and ∼3 kb downstream of *TEP3*. Due to the length of the TE, we designed alternative sets of primers to genotype its presence/absence in two separate PCR reactions, following ref. 48 (*SI Materials and Methods*).

ACKNOWLEDGMENTS. We thank N. Lobo and F. Collins for specimens; the entomological teams of Malaria Research and Training Center (Mali), Noguchi Memorial Institute for Medical Research (Ghana), Centre National de Recherche et de Formation sur le Paludisme and Institut de Recherche en Sciences de la Santé (Burkina Faso), and Organisation de Coordination pour la lutte contre les Endémies en Afrique Centrale (Cameroon) for fieldwork; M. Kern for insectary and laboratory assistance; B. Lazzaro for review and discussion; and anonymous reviewers for their suggestions. Funding was provided by National Institutes of Health Grants R01 AI63508 and R01 AI76584 (to N.J.B.).

1. Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proc Natl Acad Sci USA* 106(Suppl 1):9955–9962.
2. Via S (2009) Natural selection in action during speciation. *Proc Natl Acad Sci USA* 106 (Suppl 1):9939–9946.
3. Butlin RK (2008) Population genomics and speciation. *Genetica* 138:409–418.
4. Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Mol Ecol* 18:375–402.
5. Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 14:671–688.
6. della Torre A, et al. (2001) Molecular evidence of incipient speciation within *Anopheles gambiae s.s.* in West Africa. *Insect Mol Biol* 10:9–18.
7. della Torre A, et al. (2002) Speciation within *Anopheles gambiae*—the glass is half full. *Science* 298:115–117.
8. Edillo FE, Touré YT, Lanzaro GC, Dolo G, Taylor CE (2002) Spatial and habitat distribution of *Anopheles gambiae* and *Anopheles arabiensis* (Diptera: Culicidae) in Banambani village, Mali. *J Med Entomol* 39:70–77.
9. Costantini C, et al. (2009) Living at the edge: Biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecol* 9:16.
10. della Torre A, Tu Z, Petrarca V (2005) On the distribution and genetic differentiation of *Anopheles gambiae s.s.* molecular forms. *Insect Biochem Mol Biol* 35:755–769.
11. Simard F, et al. (2009) Ecological niche partitioning between the M and S molecular forms of *Anopheles gambiae* in Cameroon: The ecological side of speciation. *BMC Ecol* 9:17.
12. Diabaté A, et al. (2008) Evidence for divergent selection between the molecular forms of *Anopheles gambiae*: Role of predation. *BMC Evol Biol* 8:5.
13. Lehmann T, Diabate A (2008) The molecular forms of *Anopheles gambiae*: A phenotypic perspective. *Infect Genet Evol* 8:737–746.
14. White BJ, Cheng C, Simard F, Costantini C, Besansky NJ (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Mol Ecol* 19:925–939.
15. Noor MA, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* 103:439–444.
16. Christophides GK, et al. (2002) Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 298:159–165.
17. Blandin SA, Marois E, Levashina EA (2008) Antimalarial responses in *Anopheles gambiae*: From a complement-like protein to a complement-like pathway. *Cell Host Microbe* 3:364–374.
18. Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159.
19. Obbard DJ, Welch JJ, Little TJ (2009) Inferring selection in the *Anopheles gambiae* species complex: An example from immune-related serine protease inhibitors. *Malar J* 8:117.
20. Besansky NJ, et al. (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: Evidence from multilocus DNA sequence variation. *Proc Natl Acad Sci USA* 100:10818–10823.
21. Wright SI, Charlesworth B (2004) The HKA test revisited: A maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168:1071–1076.
22. Blandin SA, et al. (2009) Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. *Science* 326:147–150.
23. Dong Y, et al. (2006) Anopheles gambiae immune responses to human and rodent Plasmodium parasite species. *PLoS Pathog* 2:e52.
24. Levashina EA, et al. (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, Anopheles gambiae. *Cell* 104:709–718.
25. Blandin S, et al. (2004) Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell* 116:661–670.
26. Cohuet A, et al. (2008) SNP discovery and molecular evolution in *Anopheles gambiae*, with special emphasis on innate immune system. *BMC Genomics* 9:227.
27. Obbard DJ, et al. (2008) The evolution of *TEP1*, an exceptionally polymorphic immunity gene in *Anopheles gambiae*. *BMC Evol Biol* 8:274.
28. Slotman MA, et al. (2007) Evidence for subdivision within the M molecular form of *Anopheles gambiae*. *Mol Ecol* 16:639–649.
29. Lee Y, et al. (2009) Ecological and genetic relationships of the Forest-M form among chromosomal and molecular forms of the malaria vector *Anopheles gambiae sensu stricto*. *Malar J* 8:75.
30. Muspratt J (1946) On *Coelomomyces* fungi causing high mortality of *Anopheles gambiae* larvae in Rhodesia. *Ann Trop Med Parasitol* 40:10–17.
31. Service MW (1973) Identification of predators of *Anopheles gambiae* resting in huts, by the precipitin test. *Trans R Soc Trop Med Hyg* 67:33–34.

32. Fraiture M, et al. (2009) Two mosquito LRR proteins function as complement control factors in the TEP1-mediated killing of Plasmodium. *Cell Host Microbe* 5:273–284.
33. Povelones M, Waterhouse RM, Kafatos FC, Christophides GK (2009) Leucine-rich repeat protein complex activates mosquito complement in defense against Plasmodium parasites. *Science* 324:258–261.
34. Gillies MT, De Meillon B (1968) *The Anophelinae of Africa South of the Sahara* (South African Institute for Medical Research, Johannesburg), 2nd Ed.
35. Scott JA, Brogdon WG, Collins FH (1993) Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am J Trop Med Hyg* 49:520–529.
36. Santolamazza F, Della Torre A, Caccone A (2004) Short report: A new polymerase chain reaction-restriction fragment length polymorphism method to identify *Anopheles arabiensis* from *An. gambiae* and its two molecular forms from degraded DNA templates or museum samples. *Am J Trop Med Hyg* 70:604–606.
37. White BJ, et al. (2009) The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. *Genetics* 183:275–288.
38. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, eds Krawetz S, Misener S (Humana Press, Totowa, NJ), pp 365–386.
39. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
40. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.
41. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
42. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
43. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275–282.
44. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351:652–654.
45. Guo Z, Liu Q, Smith LM (1997) Enhanced discrimination of single nucleotide polymorphisms by artificial mismatch hybridization. *Nat Biotechnol* 15:331–335.
46. Franke-Fayard B, et al. (2004) A *Plasmodium berghei* reference line that constitutively expresses GFP at a high level throughout the complete life cycle. *Mol Biochem Parasitol* 137:23–33.
47. Lee CE, Frost BW (2002) Morphological stasis in the *Eurytemora affinis* species complex (Copepoda: Temoridae). *Hydrobiologia* 480:111–128.
48. González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA (2008) High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol* 6:e251.