MOSQUITO GENOMICS

Extensive introgression in a malaria vector species complex revealed by phylogenomics

Michael C. Fontaine,* James B. Pease,* Aaron Steele, Robert M. Waterhouse, Daniel E. Neafsey, Igor V. Sharakhov, Xiaofang Jiang, Andrew B. Hall, Flaminia Catteruccia, Evdoxia Kakani, Sara N. Mitchell, Yi-Chieh Wu, Hilary A. Smith, R. Rebecca Love, Mara K. Lawniczak, Michel A. Slotman, Scott J. Emrich, Matthew W. Hahn,† Nora J. Besansky†

INTRODUCTION: The notion that species boundaries can be porous to introgression is increasingly accepted. Yet the broader role of introgression in evolution remains contentious and poorly documented, partly because of the challenges involved in accurately identifying introgression in the very groups where it is most likely to occur. Recently diverged species often have incomplete reproductive barriers and may hybridize where they overlap. However, because of retention and stochastic sorting of ancestral polymorphisms, inference of the correct species branching order is notoriously challenging for recent speciation events, especially those closely spaced in time. Without knowledge of species relationships, it is impossible to identify instances of introgression.

RATIONALE: Since the discovery that the single mosquito taxon described in 1902 as *Anopheles gambiae* was actually a complex of several closely related and morphologically indistinguishable sibling species, the correct species branching order has remained controversial and unresolved. This Afrotropical complex contains the world’s most important vectors of human malaria, owing to their close association with humans, as well as minor vectors and species that do not bite humans. On the basis of phylogenetic clustering of the three major vectors of malaria, *An. gambiae*, *An. coluzzii*, and *An. arabiensis*. In stark contrast, the X chromosome strongly supported the grouping of *An. arabiensis* with a species that plays no role in malaria transmission, *An. quadriannulatus*. Although the autosomes, which make up the majority of the genome, overwhelmingly supported the grouping of the three major vectors of malaria, *An. gambiae*, *An. coluzzii*, and *An. arabiensis*. In stark contrast, the X chromosome strongly supports the grouping of *An. arabiensis* with a species that plays no role in malaria transmission, *An. quadriannulatus*. Although the whole-genome consensus phylogeny unequivocally agrees with the autosomal topology, we found that the topology most often located on the X chromosome follows the historical species branching order, with pervasive introgression on the autosomes producing relationships that group the three highly anthropophilic species together. With knowledge of the correct species branching order, we are further able to uncover introgression between another species pair, as well as a complex history of balancing selection, introgression, and local adaptation of a large autosomal inversion that confers aridity tolerance.

CONCLUSION: We identify the correct species branching order of the *An. gambiae* species complex, resolving a contentious phylogeny. Notably, lineages leading to the principal vectors of human malaria were among the first in the complex to radiate and are not most closely related to each other. Pervasive autosomal introgression between these human malaria vectors, including nonsister vector species, suggests that traits enhancing vectorial capacity can be acquired not only through de novo mutation but also through a more rapid process of interspecific genetic exchange.

RESULTS: We observed a large amount of introgression between trees generated from the autosomes and X chromosome. The autosomes, which make up the majority of the genome, overwhelmingly supported the grouping of the three major vectors of malaria, *An. gambiae*, *An. coluzzii*, and *An. arabiensis*. In stark contrast, the X chromosome strongly supported the grouping of *An. arabiensis* with a species that plays no role in malaria transmission, *An. quadriannulatus*. Although the whole-genome consensus phylogeny unequivocally agrees with the autosomal topology, we found that the topology most often located on the X chromosome follows the historical species branching order, with pervasive introgression on the autosomes producing relationships that group the three highly anthropophilic species together. With knowledge of the correct species branching order, we are further able to uncover introgression between another species pair, as well as a complex history of balancing selection, introgression, and local adaptation of a large autosomal inversion that confers aridity tolerance.

RELATED ITEMS IN SCIENCE

DOI: 10.1126/science.1258522

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. E-mail: mwh@indiana.edu (M.W.H.); nbesansk@ind.edu (N.J.B.)
Cite this article as M. C. Fontaine et al., Science 347, 1258524 (2015). DOI: 10.1126/science.1258524

Time-lapse photographs of an adult anopheline mosquito emerging from its pupal case.

PHOTO CREDIT: JAMES GATHANY, CENTERS FOR DISEASE CONTROL AND PREVENTION
Extensive introgression in a malaria vector species complex revealed by phylogenomics

Michael C. Fontaine,1,2,‡ James B. Pease,3‡ Aaron Steele,4 Robert M. Waterhouse,5,6,7,8 Daniel E. Neafsey,6 Igor V. Sharakhov,9,10 Xiaofang Jiang,10 Andrew B. Hall,10 Flaminia Catteruccia,11,12 Eudoxia Kanaki,11,12 Sara N. Mitchell,11 Yi-Chieh Wu,5 Hilary A. Smith,1,2 R. Rebecca Love,1,2 Mara K. Lawniczak,13‡ Michel A. Slotman,14 Scott J. Emrich,5,4 Matthew W. Hahn,5,15 Nora J. Besansky1,2

Introgressive hybridization is now recognized as a widespread phenomenon, but its role in evolution remains contested. Here, we use newly available reference genome assemblies to investigate phylogenetic relationships and introgression in a medically important group of Afrotropical mosquito sibling species. We have identified the correct species branching order to resolve a contentious phylogeny and show that lineages leading to the principal vectors of human malaria were among the first to split. Pervasive autosomal introgression between these malaria vectors means that only a small fraction of the genome, mainly on the X chromosome, has not crossed species boundaries. Our results suggest that traits enhancing vectorial capacity may be gained through interspecific gene flow, including between nonsister species.

The notion that species boundaries can be porous to introgression is increasingly accepted. Charismatic cases, such as gene flow between Neanderthals and anatomically modern humans (1) or between Heliconius butterflies (2, 3), show that introgression can transfer beneficial alleles between closely related species. Yet the broader role of introgression in evolution remains contentious and poorly documented, partly because of the challenges involved in accurately identifying introgression in the very groups where it is most likely to occur. Recently diverged species often have incomplete reproductive barriers, hence, may hybridize in sympathy. However, another feature of rapid radiations is that ancestral polymorphism predating lineage splitting may be sorted stochastically among descendant lineages in a process known as incomplete lineage sorting (ILS). Alleles shared through ILS can be difficult to distinguish from those shared through secondary contact and introgression. Newly developed methods can differentiate these two processes (1, 4) but only if the correct species branching order is known. Because both introgression and ILS cause discordance between gene trees and the species tree, inference of the correct species phylogeny (i.e., the historical branching order of the taxa) is notoriously challenging for recent radiations (5–7).

Since the discovery that the single mosquito taxon described in 1902 as Anopheles gambiae (8) was actually a complex of several closely related and morphologically indistinguishable sibling species (known as the An. gambiae complex) (9), the correct species branching order has retained controversial and unresolved. This Afro-tropical complex (10–13) contains three widely distributed and extensively sympatric species that rank among the world’s most important vectors of human malaria, owing to their association with humans (An. gambiae sensu stricto, its closest relative and sister species, Anopheles coluzzii, and Anopheles arabiensis) (Fig. 1A). Anopheles merus and Anopheles melas, salt-tolerant species that breed in brackish coastal waters of eastern and western Africa, respectively, are minor vectors. Anopheles quadrimaniatus plays no role in malaria transmission despite vector competence for Plasmodium falciparum, as it tends to bite animals other than humans. On the basis of ecology and behavior, one might predict phylogenetic clustering of the highly anthropophilic vector species. Yet such clustering has not been supported by chromosomal inversion polymorphisms and was instead attributed to introgression (15, 16). Introgression is plausible between any geographically overlapping pair of species in the complex, as reproductive isolation is incomplete: Adult female F1 hybrids—although uncommon in nature—are fertile and vigorous (only F2 hybrid males are sterile) (10). Nonetheless, comprehensive evidence for introgression between An. arabiensis and the other major vector lineage (An. gambiae + An. coluzzii) has been lacking, until now, because of insufficient genomic resources.

Important malaria vectors make up a small fraction of the genus Anopheles but almost invariably are embedded in species complexes similar to that of An. gambiae sensu lato (17). Thus, resolving the phylogeny of the An. gambiae complex has the potential to yield insights into the origin and evolution of traits that are associated with highly successful malaria vectors across the genus as a whole. Here, we have used newly available whole-genome reference assemblies (18) to infer the species phylogeny. We found that the two most important malaria vector lineages in the complex are not the most closely related, and we have uncovered pervasive introgression between them. The extent of introgression is such that the phylogenetic tree inferred from whole-genome alignments supports the incorrect species branching order with high confidence.

The X chromosome reflects the species branching order

As a first step in phylogenomic analysis, alignments were generated for existing (19, 20) and newly available genome assemblies, representing six species of the An. gambiae complex and two Pyretophorus out-group species (An. christyi and An. epiroticus) (supplementary text S1 and fig. S2). Among the in-group species, the alignment included ~60% of nongap and nonmasked base pairs of the An. gambiae PEST reference assembly. Across all assemblies, the proportion of aligned base pairs was lower (~53%) but, nevertheless, spanned more than 93 megabase pairs (Mb) (~40% of the euchromatic genome).

As the An. gambiae complex reference genomes were assembled from laboratory-colony-derived sequencing template (except for An. melas (18)), we also performed whole-genome shotgun sequencing of individual field-collected specimens sampled from at least one population of each of the six in-group species (supplementary text S2). Sequencing reads were aligned to the
Fig. 1. Distribution and phylogenetic relationships of sequenced members of the An. gambiae complex. (A) Schematic geographic distribution of An. gambiae (gam, formerly An. gambiae S form), An. coluzzii (col, formerly An. gambiae M form (13)), An. arabiensis (ara), An. quadriannulatus (qua), An. merus (mer), and An. melas (mel). (B) Species topology as estimated from a ML phylogeny of the X chromosome (see text) compared with the ML phylogeny estimated from the whole-genome sequence alignment. The scale-bar denotes nucleotide divergence as calculated by RAxML. Red branches of the chromosome (see text) compared with the ML phylogeny estimated from the whole-genome sequence. (C) Schematic of the species topology rooted by An. christyi (An. chr) and An. epiroticus (An. epi), showing the major introgression events (green arrows) and the approximate divergence and introgression times [Ma ± 1 standard deviation (SD)], if one assumes a substitution rate of $1.1 \times 10^{-9}$ per site, per generation, and 10 generations per year. The gray bar surrounding each node shows ±2 SD. Nodes marked by asterisks are not fully resolved owing to high levels of ILS. (D) Neighbor-joining tree displaying the Euclidian distance between individuals from population samples of each species, calculated using sequence data from the X chromosome. Each species is represented by a distinct symbol shape or shade.

species-appropriate reference assembly to avoid considerable interspecific mapping bias (supplementary text S2, figs. S3 to S4, and table S3), and single-nucleotide variant positions were converted to a common coordinate system relative to the reference genome alignment (supplementary texts S2 and S3 and fig. S2). The results presented below are based on the field-collected samples, but analyses were performed in parallel on the reference genome assemblies, and our findings were consistent in all cases.

To infer the correct species branching order in the face of anticipated ILS and introgression, maximum-likelihood (ML) phylogenies were constructed from 50-kilobase (kb) nonoverlapping windows across the alignments (referred to here as “gene trees” regardless of their protein-coding content), considering six in-group species rooted alternatively with An. christyi or An. epiroticus ($n = 4063$ windows) (supplementary text S3). As the choice of out-group did not materially alter our results, we present our findings based on the more closely related species, An. christyi. When the 85 tree topologies observed at least once across the genome were sorted by chromosome arm and relative phylogenetic positions of An. arabiensis and An. quadriannulatus, for simplicity, we grouped the most frequently observed topologies into three sets (Fig. 2 and fig. S16), (i) A+CG: those that supported An. arabiensis clustering with An. coluzzii + An. gambiae; (ii) A+CG, R+Q: those that supported both An. arabiensis (An. coluzzii + An. gambiae) and An. merus + An. quadriannulatus; and (iii) A+Q: those that supported the clustering of An. arabiensis and An. quadriannulatus. On the X chromosome, all three of the most frequently observed topologies (inferred from 64% of windows) strongly supported the relationship (melas (arabiensis, quadriannulatus)), which indicated a sister-taxon relationship between An. arabiensis and An. quadriannulatus (i.e., A+Q) (orange shades in Fig. 2). This relationship was shared among all field-collected samples (Fig. 1D). Notably, the X chromosomal windows supporting these topologies are concentrated distal to the centromere (Fig. 3D and supplementary text S3), in an ~15-Mb region corresponding to the Xag inversion whose orientation is ancestral to the An. gambiae complex and shared by An. gambiae, An. coluzzii, and An. merus (see supplementary text S5).

In stark contrast to the X chromosome, the overwhelming majority of window-based topologies across the autosomes supported An. arabiensis as sister to An. gambiae + An. coluzzii (green and purple shades in Fig. 2). On chromosomes 3 and 2R, a subset of these topologies also supported a sister-taxon relationship between An. quadriannulatus and An. merus, in further disagreement with the X chromosome (purple shades, Fig. 2).

Autosomal introgression between An. arabiensis and the ancestor of An. gambiae + An. coluzzii has long been postulated (10, 22) and could explain the strong discordance between the dominant tree topologies of the X and autosomes. However, before this study, the correct species branching order was unresolved, which precluded definitive interpretation of these conflicting signals. To infer the correct historical branching order, we applied a strategy based on sequence divergence (supplementary text S3 and fig. S16). Because introgression would reduce sequence divergence between the species exchanging genes, we expect that the correct species branching order revealed by gene trees constructed from nonintrogressed sequences will show deeper divergences than those constructed from introgressed sequences. If the hypothesis of autosomal introgression is correct, this implies that the topologies supported by the X chromosome should show significantly higher divergence times between An. arabiensis and either An. gambiae or An. coluzzii than topologies supported by the autosomes.

To test this hypothesis, we used An. arabiensis, An. gambiae, and An. melas, as these three species show strongly discordant trees on the X and autosomes. For this trio, there are three possible phylogenetic relationships and two divergence times for each, $T_1$ and $T_2$ (Fig. 3A). Using the metrics $T_1$ and $T_2$, we conducted two different tests. Initially,
Fig. 2. Phylogenies inferred from regions on the autosomes differ strongly from phylogenies inferred from regions on the X chromosome. ML-rooted phylogenies were inferred from $n = 4063$ 50-kb genomic windows for An. arabiensis (A), An. coluzzii (C), An. gambiae (G), An. melas (L), An. merus (R), and An. quadrinunulatus (Q) with An. christyi as out-group. The nine most commonly observed topologies across the genome (trees i to ix) are indicated on each of the five chromosome arms (if found) by correspondingly colored blocks whose length represents the proportion of all 50-kb windows on that arm that support the topology. Topologies specific to 2La are represented by the dark gray block on 2L: all other topologies observed on each chromosome arm were pooled, and their combined frequencies are indicated by the light gray blocks. The X chromosome most often indicates that A and Q are sister taxa (trees vii to ix), whereas the autosomes indicate that A and C+G are sister groups (trees i to vi). Large portions of the autosomes (particularly 3L and 3R) indicate that R and Q are sister taxa (trees iv to vi). Additional diversity in inferred trees is the result of rearrangement of three groups (R, C+G, and L+A+Q) because of ILS. The most common phylogeny on the X chromosome (vii) represents the most likely species branching order. The 2L arm has a markedly different distribution of phylogenies because of the unique history of the 2La inversion region (see Fig. 5), which creates unusual phylogenetic topologies found nowhere else in the genome. The autosome-like trees on the X chromosome (i and ii) are entirely found in the pericentromeric region (15 to 24 Mb) (see Fig. 3D), where introgression between An. arabiensis and An. gambiae + An. coluzzii has previously been implicated.

Fig. 3. Tree height reveals the true species branching order in the face of introgression. (A) Color-coded trees show the three possible rooted phylogenetic relationships for An. arabiensis (A), An. gambiae s.s. (G), and An. melas (L), with out-group An. christyi. For any group of three taxa, there are two species divergence times ($T_1$ and $T_2$). When introgression has occurred, a strong decrease is expected in these divergence times. (B) Trees on the X chromosome have significantly higher mean values of $T_1$ and $T_2$ than the autosomes, which indicates that introgression is more likely to have occurred on the autosomes than the X chromosome (diamond for mean, whiskers = SEM offset to the right for visual clarity; **$P < 1.0 \times 10^{-40}$). (C) Even among only autosomal loci, regions with the X majority relationship $G(LA)$ (orange) have significantly higher $T_1$ and $T_2$ (**$P < 1.0 \times 10^{-40}$), which indicates that the autosomal majority topology (green) is the result of widespread introgression between An. arabiensis and An. gambiae. The X majority relationship $G(LA)$ therefore represents the true species branching relationships. (D) The gene tree distributions or "chromoplots" for all five chromosomal arms show that the autosomal and X chromosome phylogenies strongly disagree. Three-taxon phylogenies were inferred from 10-kb chromosomal windows across the genome, and colored vertical bars represent the relative abundance of the three alternative topologies (A) in a 200-kb window. The proportions of the three 10-kb gene trees for each chromosome arm (left) indicate that the autosomes all strongly show a closer relationship between An. gambiae and An. arabiensis (green); the X indicates a closer relationship between An. arabiensis and An. melas (orange). Black semicircles indicate the location of centromeres.

we compared mean sequence divergence between X chromosome–based versus autosome-based topologies. As predicted, the estimated mean values of $T_1$ and $T_2$ for trees inferred from 10-kb windows on the X chromosome were significantly higher than their counterparts on the autosomes (both $P < 1.0 \times 10^{-40}$) (Fig. 3B). However, the possibility exists that this result was confounded or entirely driven by other factors that differ between the X and autosomes, including a faster rate of evolution on the X (29). Indeed, we found evidence supporting faster evolution of genes on the X chromosome (supplementary text S3 and fig. S23). To avoid this problem, we conducted a second test on the autosomes alone, focusing on trees inferred from 10-kb windows and the mean divergence levels among those supporting the three possible phylogenetic relationships illustrated in Fig. 3A. The result, congruent with the first test but providing unequivocal evidence for the correct species branching order, was that the set of autosomal trees supporting the major-
The region on 3L corresponding to the 3La inversion shows strong evidence of R-Q introgression and a larly notable is the separation of the sister taxa complex.

1258524-4

Fig. 5. Ancient trans-specific polymorphism of an inversion predates radiation of the An. gambiae complex. (A) All species in the complex are fixed for either the 2La (An. arabiensis, ara; An. merus, mer) or 2L+* (An. melas, mel) orientation of the 2La rearrangement except An. gambiae (gam) and An. coluzzii (col), which remain polymorphic for both orientations. The unique phylogenies observed in the 2La region (Fig. 2) are the result of differential loss (+) of the 2La and 2L+* orientations and the introgression of 2La from the ancestral population of An. gambiae + coluzzii into An. arabiensis (dotted arrow). The overall divergence between the 2La and 2L+* orientations inferred from sequences inside the inversion breakpoints is higher, on average, than the predicted divergence time of the species complex. (B) ML phylogeny inferred from sequences of the two different orientations of the 2La region (2La and 2L+*) shows that the divergence between opposite orientations is greater than the divergence between the same orientation present in different species (all nodes, 100% bootstrap support). Particularly notable is the separation of the sister taxa An. gambiae-2L+* and An. coluzzii-2La, as these are known sister taxa. The scale bar denotes nucleotide divergence as calculated by RAxML.

Fig. 4. Introgression between An. merus and An. quadriannulatus. Chromplots for all five chromosomal arms show a highly spatially heterogeneous distribution of phylogenies inferred from 50-kb genomic regions, particularly on 3L. The three possible rooted phylogenetic relationships for An. quadriannulatus (Q), An. melas (L), and An. merus (R), with out-group An. christyi are shown, colored according to the key in the lower right. The region on 3L corresponding to the 3La inversion shows strong evidence of R-Q introgression and a strong negative deviation of the D statistic. The region on 2L corresponding to the 2La inversion is highly enriched for the R(Q) relationship, as expected, given that L and Q both have the 2L+* orientation, whereas R has 2La (see Fig. 5). Across all the autosomes, the D statistic generally trends toward negative values, which indicates weak or ancient R-Q introgression may have been occurring across the autosomes (see supplementary text S3).

from character congruence when all available data are considered simultaneously. Alternative approaches seek taxonomic congruence between different data sets or data partitions, analyzed separately (25). Either way, standard phylogenomic practice entails some form of “majority rule,” based on the assumption that as the amount of data increases so will the probability of converging on the correct species branching order (apart from exceptional cases (26)). Because of both ILS and introgression, the historical species branching order for the An. gambiae complex is represented by only 1.9% of 50-kb windows across the entire genome (Fig. 2). As a result, when we inferred a ML tree for the An. gambiae complex on the basis of whole-genome alignments, we recovered the wrong species branching order supported by 100% of the bootstrap replicates at each node (Fig. 1B, supplementary text S3, and figs. S17A and S18A). The extent of autosomal introgression in the An. gambiae complex has the paradoxical effect that, as an increasing amount of the genome is sampled, support for the incorrect species branching order is maximized.

Autosomal permeability of species boundaries

Early cytogenetic evidence (10, 12, 27), as well as more recent ribosomal DNA-based evidence (28, 29), supports rare occurrences (<0.1%) of natural female F1 hybrids between An. arabiensis–An. gambiae + An. coluzzii, An. arabiensis–An. quadriannulatus, and An. melas–An. gambiae + An. coluzzii, although there are no reports of hybrids involving An. merus. The evolutionary importance of these rare hybrids as bridges to interspecific gene flow has remained controversial. Inference of the correct species branching order for the An. gambiae complex (Fig. 1B) allowed a systematic analysis of introgression across the genomes of six members of this complex using the D (1, 4) and Drefal (30) statistics (supplementary text S4).

As expected, such tests revealed pervasive introgression across all autosomes between An. arabiensis and the ancestor of An. gambiae + An. coluzzii (figs. S24 and S25). Although introgression was detected in both directions, the majority involved genetic transfer from An. arabiensis into the ancestor of An. gambiae + An. coluzzii. This recent and massive episode of introgression impeded our ability to detect older introgression events between these species. Unexpectedly, we also found evidence of extensive autosomal introgression between another species pair, An. merus and An. quadriannulatus (Fig. 4, supplementary text S3, and figs. S24 and S25). One of the most striking of the introgressed regions was a contiguous block of genes coincident with the 22-Mb 3La chromosomal inversion (S3). The corresponding sequence originally present in ancestral populations of An. quadriannulatus has been entirely replaced by its counterpart from An. merus, a conclusion supported by the clustering of An. merus with An. quadriannulatus in gene trees constructed from sequences in the 3La inversion (figs. S19, C and D, and S21B). Extant populations of both these species, and indeed all recognized species in the An. gambiae complex, are fixed for the standard (3La+) orientation except An. melas and its putative sister species An. brevicauda, both fixed for the 3La orientation (31). Considering that the exact ~22-Mb 3La region was replaced between species whose contemporary populations are collinear for 3L+, it is conceivable that ancestral An. quadriannulatus populations originally carried 3La before the 3La+ introgression. The expected reduced recombination between
inverted and standard arrangements in inversion heterozygotes may explain why the introgressed region is coincident with the entire 22-Mb 3La region.

The introgression profile was consistent between samples from natural populations and the reference genome assemblies, involving the same species pairs and the same chromosomal locations (supplementary text S4 and fig. S24). Moreover, based on population samples from multiple geographic locations in Africa, patterns of introgression between An. arabiensis and An. gambiae + An. coluzzii, and between An. merus and An. quadriannulatus, are similar across their geographic range (fig. S25 and table S10). These findings refute the possibility that introgression was an unnatural artifact of colonization and laboratory maintenance of multiple species. The lack of geographic variation in patterns of introgression occurred sufficiently long ago to have spread across subpopulations. By contrast, mitochondrial DNA (mtDNA) revealed patterns consistent with ongoing gene flow between An. arabiensis and An. gambiae or An. coluzzii (supplementary text 3.3 and fig. S22).

Transspecific inversion polymorphism

The unusually high sequence divergence between alternative orientations of a chromosomal inversion polymorphic within An. gambiae and An. coluzzii (2La and 2L+) has been noted previously (32, 33) but has not been adequately explained. Other species in the complex are fixed for either 2La (An. arabiensis and An. merus) or 2L+ (An. quadriannulatus and An. melas) (31) (Fig. 5A). In An. gambiae and An. coluzzii, the 2La arrangement has been shown to confer superior resistance to desiccating environments (34, 35) relative to 2L+, and its frequency correlates with environmental gradients of aridity (10, 32). It has been argued that 2La introgressed from An. arabiensis into the presumed 2L+ ancestor of An. gambiae + An. coluzzii (36), a crucial step facilitating the range expansion of a presumed forest-adapted species into drier savannas. Sequence divergence-based estimates of the age of 2La and 2L+ arrangements relative to the age of the species complex (Fig. 1C) suggest a radically different scenario in which 2La/2L+ is an ancient inversion polymorphism that predates the initial diversification of the entire complex but is maintained as polymorphic only in An. gambiae and An. coluzzii (Fig. 5). Consistent with this scenario, the topology of the gene tree built from sequences inside the inversion boundaries indicates that species are grouped by their 2La or 2L+ karyotype (Fig. 5B). Furthermore, contrary to the longstanding assumption (36), our data suggest that 2La introgressed from ancestral An. gambiae into An. arabiensis, not the other way around—eventually replacing the An. arabiensis 2L+ arrangement (Fig. 5A). Our inference about the direction of 2La introgression, as well as the underlying phylogenetic hypothesis for the An. gambiae complex, are consistent with the genome-enabled chromosomal inversion phylogeny of the An. gambiae complex recon-structed from ancestral and derived gene orders at the breakpoints of 10 fixed chromosomal inversions (supplementary text S5 and fig. S27).

Functional insights from differential genetic exchange

We found that introgression mainly involved the autosomes. Our data suggest that the X chromosome is largely resistant to introgression, consistent with studies in this and other systems indicating that the X (or Y) disproportionately harbors factors responsible for reproductive isolation (3, 6, 37-41). The nature, number, and chromosomal organization of these X-linked factors are unsolved puzzles for future research, but our data offer one tantalizing clue. An. gambiae males deliver to females large amounts of 20-hydroxyecdysone (20E) (42), a steroid hormone that increases female fertility (43) and fecundity (44) and regulates mating behavior and success (45). The cytochrome p450 gene CYP313A1 (AGAP000284) that synthesizes the 20E precursor ecdysone is located near the distal end of the X chromosome. Furthermore, this region is associated with male hybrid sterility between An. gambiae and An. arabiensis, with the An. gambiae X chromosome causing inviability in an An. arabiensis genetic background (40). Combined with our data showing that male An. gambiae produce significantly less 20E than An. gambiae (supplementary text S6 and fig. S28), these observations prompt the hypothesis that divergence in 20E function between the two species may have a role in speciation through possible effects on the reproductive fitness of hybrid males.

Pervasive autosomal introgression between An. arabiensis and the An. gambiae-clade ancestor is consistent with the paucity of sterility factors across much of the autosomes, although several autosomal quantitative trait loci have been mapped in both species (40). Accordingly, we explored the small subset of autosomal genes (n = 485) that showed no indication of introgression (supplementary text S6), as these are candidates contributing to reproductive isolation. We found a remarkable overrepresentation of genes encoding cyclic-nucleotide phosphodiesterases, enzymes that regulate the levels of the messengers cyclic adenosine monophosphate and cyclic guanosine monophosphate (46), which in turn control (among other processes) ecdysone synthesis (47, 48) (table S16).

Implications for the evolution of vectorial capacity

Initial radiation of the An. gambiae complex was both recent and rapid. Counter to the traditional view (49), it is now clear that the ancestor of the principal malaria vectors An. gambiae and An. coluzzii separated from other species in the group approximately 2 million years ago (Ma) and that An. gambiae and An. coluzzii are distantly related to the other primary vector in the group, An. arabiensis. Extant populations of An. gambiae and An. coluzzii are highly anthropophilic vectors, dependent upon humans for blood meals, adult shelter, and larval breeding sites, yet anthropogenic influences are unlikely to have triggered their cladogenesis an estimated 0.5 Ma. Instead, anthropophilic traits are likely to have developed in conjunction with the expansion of Neolithic human populations that occurred more recently. Despite a history of extensive introgression with An. arabiensis, An. gambiae and An. coluzzii are behaviorally, physiologically, ecologically, and epidemiologically distinct from An. arabiensis; the same is true for An. merus and An. quadriannulatus. Notably, experimental introgressions of certain autosomal inversions result in stable heterotic polymorphisms, whereas other introgressed autosomal and X chromosome inversions are rapidly eliminated (22), consistent with a role for natural selection in the fate of introgressed regions. Given evidence that the 2La inversion polymorphism is maintained by selection in An. gambiae and An. coluzzii (32, 50), it seems likely that its introgression into An. arabiensis was adaptive, and bidentational introgressions across the genome between these species probably contributed to their wide ecological flexibility and their vectorial capacity. Broad overlap exists between geographic ranges of these species, and the potential for ongoing hybridization and introgression remains, including the opportunity for introgression of insecticide resistance alleles (51). Our study establishes a foundation for further study of adaptive introgression in this species complex and its role in shaping vectorial capacity in this and other malaria vector species complexes.

REFERENCES AND NOTES

8. G. M. Giles, A handbook of the grats or mosquitoes giving the anatomy and life history of the Culicidae together with descriptions of all species noticed up to the present date. (John Bale, Sons & Danielsson, Ltd., London, 1902).
20. M. K. Lawniczak et al., Supplementary text, figures, and tables are available on Science Online.

ACKNOWLEDGMENTS


SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/347/6217/1258524/suppl/DC1

Downloaded from http://science.sciencemag.org/ on October 4, 2017
Extensive introgression in a malaria vector species complex revealed by phylogenomics

Michael C. Fontaine, James B. Pease, Aaron Steele, Robert M. Waterhouse, Daniel E. Neafsey, Igor V. Sharakhov, Xiaofang Jiang, Andrew B. Hall, Flaminia Catteruccia, Evdokia Kakani, Sara N. Mitchell, Yi-Chieh Wu, Hilary A. Smith, R. Rebecca Love, Mara K. Lawniczak, Michel A. Slotman, Scott J. Emrich, Matthew W. Hahn and Nora J. Besansky

Science 347 (6217), 1258524.
DOI: 10.1126/science.1258524 originally published online November 27, 2014

Mosquito adaptability across genomes

Virtually everyone has first-hand experience with mosquitoes. Few recognize the subtle biological distinctions among these bloodsucking flies that render some bites mere nuisances and others the initiation of a potentially life-threatening infection. By sequencing the genomes of several mosquitoes in depth, Neafsey et al. and Fontaine et al. reveal clues that explain the mystery of why only some species of one genus of mosquitoes are capable of transmitting human malaria (see the Perspective by Clark and Messer).

Science, this issue 10.1126/science.1258524 and 10.1126/science.1258522; see also p. 27