Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes

Daniel E. Neafsey,‡ Robert M. Waterhouse,* et al.

INTRODUCTION: Control of mosquito vectors has historically proven to be an effective means of eliminating malaria. Human malaria is transmitted only by mosquitoes in the genus Anopheles, but not all species within the genus, or even all members of each vector species, are efficient malaria vectors. Variation in vectorial capacity for human malaria among Anopheles mosquito species is determined by many factors, including behavior, immunity, and life history.

RATIONALE: This variation in vectorial capacity suggests an underlying genetic/genomic plasticity that results in variation of key traits determining vectorial capacity within the genus. Sequencing the genome of Anopheles gambiae, the most important malaria vector in sub-Saharan Africa, has offered numerous insights into how that species became highly specialized to live among and feed upon humans and how susceptibility to mosquito control strategies is determined. Until very recently, similar genomic resources have not existed for other anophelines, limiting comparisons to individual genes or sets of genomic markers with no genome-wide data to investigate attributes associated with vectorial capacity across the genus.

RESULTS: We sequenced and assembled the genomes and transcriptomes of 16 anophelines from Africa, Asia, Europe, and Latin America, spanning ~100 million years of evolution and chosen to represent a range of evolutionary distances from An. gambiae, a variety of geographic locations and ecological conditions, and varying degrees of vectorial capacity. Genome assembly quality reflected DNA template quality and homozygosity. Despite variation in contiguity, the assemblies were remarkably complete and searches for arthropod-wide single-copy orthologs generally revealed few missing genes. Genome annotation supported with RNA sequencing transcriptomes yielded between 10,738 and 16,149 protein-coding genes for each species. Relative to Drosophila, the closest dipteran genus for which equivalent genomic resources exist, Anopheles exhibits a dynamic genomic evolutionary profile. Comparative analyses show a fivefold faster rate of gene gain and loss, elevated gene shuffling on the X chromosome, and more intron losses in Anopheles. Some determinants of vectorial capacity, such as chemosensory genes, do not show elevated turnover but instead diversify through protein-sequence changes. We also document evidence of variation in important reproductive phenotypes, genes controlling immunity to Plasmodium malaria parasites and other microbes, genes encoding cuticular and salivary proteins, and genes conferring metabolic insecticide resistance. This dynamism of anopheline genes and genomes may contribute to their flexible capacity to take advantage of new ecological niches, including adapting to humans as primary hosts.

CONCLUSIONS: Anopheles mosquitoes exhibit a molecular evolutionary profile very distinct from Drosophila, and their genomes harbor strong evidence of functional variation in traits that determine vectorial capacity. These 16 new reference genome assemblies provide a foundation for hypothesis generation and testing to further our understanding of the diverse biological traits that determine vectorial capacity.

Geography, vector status, and molecular phylogeny of the 16 newly sequenced anopheline mosquitoes and selected other dipterans.

The maximum likelihood molecular phylogeny of all sequenced anophelines and two mosquito outgroups was constructed from the aligned protein sequences of 1085 single-copy orthologs. Shapes between branch termini and species names indicate vector status and are colored according to geographic ranges depicted on the map. Ma, million years ago.

*These authors contributed equally to this work.
†Corresponding author. E-mail: neafsey@broadinstitute.org (D.E.N.); nbesansk@nd.edu (N.J.B.)

Cite this article as: D. E. Neafsey et al., Science 347, 1258522 (2015). DOI: 10.1126/science.1258522
Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes

Daniel E. Neafsey,1† Robert M. Waterhouse,2,3,4,5† Mohammad R. Abai,6 Sergey S. Aganezov,7 Max A. Alekseyev,7 James E. Allen,7 James Amon,9 Bruno Areia,10 Peter Arensburger,11 Gleb Artemov,12 Lauren A. Assour,13 Hamidreza Basseri,6 Aaron Berlin,1 Bruce W. Birren,1 Stephanie A. Blandin,14,15 Andrew I. Brockman,16 Thomas R. Burkot,17 Austin Burt,18 Clara S. Chan,2,3 Cedric Chauve,19 Joanna C. Chiu,20 Mikkel Christensen,8 Carlo Costantini,21 Victoria L. M. Davidson,22 Elena Deligianni,23 Tania Dottorini,16 Vicky Dritsou,24 Stacey C. Gabriel,25 Wamdaogo M. Guebo,26 Andrew B. Hall,27 Mira V. Han,28 Thaung Hlaing,29 Daniel S. T. Hughes,8,30 Mikkel Christensen,8 Carlo Costantini,21 Victoria L. M. Davidson,22 Elena Deligianni,23 Tania Dottorini,16 Vicky Dritsou,24 Stacey C. Gabriel,25 Wamdaogo M. Guebo,26 Andrew B. Hall,27 Mira V. Han,28 Thaung Hlaing,29 Daniel S. T. Hughes,8,30

Abstract

Malaria is a complex disease, mediated by obligate eukaryotic parasites with a life cycle requiring adaption to both vertebrate hosts and mosquito vectors. These relationships create a rich coevolutionary triangle. Just as Plasmodium parasites have adapted to their diverse hosts and vectors, infection by Plasmodium parasites has reciprocally induced adaptive evolutionary responses in humans and other vertebrates (1) and has also influenced mosquito evolution (2). Human malaria is transmitted only by mosquitoes in the genus Anopheles, but not all species within the genus, or even all members of each vector species, are efficient malaria vectors. This suggests an underlying genetic/genomic plasticity that results in variation of key traits determining vectorial capacity within the genus. In all, five species of Plasmodium have adapted to infect humans and are transmitted by ~60 of the 450 known species of anopheline mosquitoes (3). Sequencing the genome of Anopheles gambiae, the most important malaria vector in sub-Saharan Africa, has offered numerous insights into how that species became highly specialized to live among and feed upon humans and how susceptibility to mosquito control strategies is determined (4). Until very recently (5–7), similar genomic resources have not existed for other anophelines, limiting comparisons to individual genes or sets of genomic markers with no genome-wide data to investigate attributes associated with vectorial capacity across the genus. Thus, we sequenced and assembled the genomes and transcriptomes of 16 anophelines from Africa, Asia, Europe, and Latin America. We chose these 16 species to represent a range of evolutionary distances from An. gambiae, a variety of geographic locations and ecological conditions, and varying degrees of vectorial capacity (8) (Fig. 1A and B). For example, An. quadriannulatus, although extremely closely related to An. gambiae, feeds preferentially on bovines rather than humans, limiting its potential to transmit human malaria. An. merus, An. melas, An. farauti, and An. albimanus females can lay eggs in salty or brackish water, instead of the freshwater sites required by other species. With a focus on species most closely related to An. gambiae (9), the sampled anophelines span the three main subgenera that shared a common ancestor ~100 million years ago (Ma) (10).

Materials and methods summary

Genomic DNA and whole-body RNA were obtained from laboratory colonies and wild-caught specimens (tables S1 and S2), with samples for nine species procured from newly established isofemale colonies to reduce heterozygosity. Illumina sequencing libraries spanning a range of insert sizes were constructed, with ~100-fold paired-end 101–base pair (bp) coverage generated for small (180 bp) and medium (1.5 kb) insert libraries and lower coverage for large (38 kb) insert libraries (table S3). DNA template for the small and medium input libraries was sourced from single female mosquitoes from each species to further reduce heterozygosity. High-molecular-weight DNA template for each large insert library was derived from pooled DNA obtained from several hundred mosquitoes. ALLPATHS-LG (11) genome assemblies were produced using the three main subgenera that shared a common ancestor ~100 million years ago (Ma) (10) as an option to reduce haplotype assembly quality. Anopheles gambiae (12) genome assemblies were produced using the “haplotypes” option to reduce haplotype assembly caused by high heterozygosity. Assembly quality reflected DNA template quality and homology, with a mean scaffold N50 of 3.6 Mb, ranging to 18.1 Mb for An. albimanus (table S4). Despite variation in contiguity, the assemblies were remarkably complete and searches for arthropod-wide single-copy orthologs generally revealed few missing genes (fig. S1) (12).

Genome annotation with MAKER (13) supported with RNA sequencing (RNAseq) of transcriptomes (produced from pooled male and female larvae, pupae, and adults) (table S5) and comprehensive noncoding RNA gene prediction (fig. S2) yielded relatively complete gene sets
(fig. S3), with between 10,738 and 16,149 protein-coding genes identified for each species. Gene count was generally commensurate with assembly contiguity (table S6). Some of this variation in total gene counts may be attributed to the challenges of gene annotations with variable levels of assembly contiguity and supporting RNAseq data. To estimate the prevalence of erroneous gene model fusions and/or fragmentations, we compared the new gene annotations to An. gambiae gene models and found an average of 3.3% and 9.7% potentially fused and fragmented gene models, respectively. By analyzing the regions below that may be sensitive to variation in gene model accuracy or gene set completeness, we have conducted sensitivity analyses to rule out confounding results from these factors (12).

Rapidly evolving genes and genomes

Orthology delineation identified lineage-restricted and species-specific genes, as well as ancient genes found across insect taxa, of which universal single-copy orthologs were employed to estimate the molecular species phylogeny (fig. 1, B and C, and fig. S4). Analysis of codon frequencies in these orthologs revealed that anophelines, unlike drosophilids, exhibit relatively uniform codon usage preferences (fig. S5).

Polytene chromosomes have provided a glimpse into anopheline chromosome evolution (14). Our genome-sequence–based view confirmed the cytological observations and offers many new insights. At the base-pair level, ~90% of the non-gapped and nonmasked An. gambiae genome (i.e., excluding transposable elements, as detailed in table S7) is alignable to the most closely related species, whereas only ~13% aligns to the most distant (fig. 1D, fig. S6, and table S8), with reduced alignability in centromeres and on the X chromosome (fig. 1D). At chromosomal levels, mapping data anchored 35 to 76% of the An. stephensi, An. funestus, An. atroparvus, and An. albimanus genome assemblies to chromosomal arms (tables S12 to S14). These analyses of unanchored regions showed that synteny at the whole-arm level is highly conserved, despite several whole-arm translocations (fig. 2A and table S13). In contrast, small-scale rearrangements disrupt gene colinearity within arms over time, leading to extensive shuffling of gene order over a time scale of 29 million years or more (10, 15) (fig. 2B and fig. S7). As in Drosophila, rearrangement rates are higher on the X chromosome than on autosomes (fig. 2C and tables S14 to S16). However, the difference is significantly more pronounced in Anopheles, where X chromosome rearrangements are more frequent by a factor of 2.7 than autosomal rearrangements; in Drosophila, the corresponding ratio is only 1.2 (test, t0 = 7.3; P < 1 × 10−10) (fig. S8). The X chromosome is also notable for a significant degree of observed gene movement to other chromosomes relative to Drosophila (one sample proportion test, P = 2.2 × 10−5) (fig. 2D and tables S17 and S18), as was previously noted for Anopheles relative to Aedes (16), further underscoring its distinctive evolutionary profile in Anopheles compared with other dipteran genera.

Such dynamic gene shuffling and movement may be facilitated by the multiple families of DNA transposons and long terminal repeat (LTR) and non-LTR retroelements found in all genomes (table S7), as well as a weaker dosage compensation pattern in Anopheles compared with Drosophila (17). This shuffling, as well as paring genomic locations of orthologs can be successfully employed to reconstruct ancestral chromosomal arrangements (fig. S9) and to confidently improve assembly contiguity (tables S19 to S21).

Copy-number variation in homologous gene families also reveals striking evolutionary dynamics. Analysis of 11,636 gene families with CAFE 3 (18) indicates a rate of gene gain/loss higher by a factor of at least 5 than that observed for 12 Drosophila genomes (19). Overall, these Anopheles genomes exhibit a rate of gain or loss per gene per million years of 3.12 × 10−16 compared with 5.90 × 10−14 for Drosophila, suggesting substantially higher gene turnover within anophelines relative to fruit flies. This fivefold greater gain/loss rate in anophelines holds true under models that account for uncertainty in...
Fig. 1. Geography, vector status, molecular phylogeny, gene orthology, and genome alignability of the 16 newly sequenced anopheline mosquitoes and selected other dipterans. (A) Global geographic distributions of the 16 sampled anophelines and the previously sequenced *An. gambiae* and *An. darlingi*. Ranges are colored for each species or group of species as shown in (B), e.g., light blue for *An. farauti*. (B) The maximum likelihood molecular phylogeny of all sequenced anophelines and selected dipteran outgroups. Shapes between branch termini and species names indicate vector status (rectangles, major vectors; ellipses, minor vectors, triangles, nonvectors) and are colored according to geographic ranges shown in (A). (C) Bar plots show total gene counts for each species partitioned according to their orthology profiles, from ancient genes found across insects to lineage-restricted and species-specific genes. (D) Heat map illustrating the density (in 2-kb sliding windows) of whole-genome alignments along the lengths of *An. gambiae* chromosomal arms: from white where *An. gambiae* aligns to no other species to red where *An. gambiae* aligns to all the other anophelines.

gene family sizes at the tips of the species tree due to annotation or assembly errors and is not sensitive to inclusion or exclusion of taxa affecting the root age of the tree nor to the exclusion of taxa with the poorest assemblies and gene sets (fig. S10 and tables S22 and S23). Examples include expansions of cuticular proteins in *An. arabiensis* and neurotransmitter-gated ion channels in *An. albimanus* (table S24).

The evolutionary dynamics of *Anopheles* genes extend to their architecture. Comparisons of single-copy orthologs at deeper phylogenetic depths showed losses of introns at the root of the true fly order Diptera and revealed continued losses as the group diversified into the lineages leading to fruit flies and mosquitoes. However, anopheline orthologs have sustained greater intron loss than drosophilids, leading to a relative paucity of introns in the genes of extant anophelines (fig. S11 and table S25). Comparative analysis also revealed that gene fusion and fission played a substantial role in the evolution of mosquito genes, with apparent rearrangements affecting an average of 10.1% of all genes in the genomes of the 10 species with the most contiguous assemblies (fig. S12).

Furthermore, gene boundaries can be flexible; whole genome alignments identified 325 candidates for stop-codon readthrough (fig. S13 and table S26).

Because molecular evolution of protein-coding sequences is a well-known source of phenotypic change, we compared evolutionary rates among different functional categories of anopheline orthologs. We quantified evolutionary divergence in terms of protein sequence identity of aligned orthologs and the $d_{\mu}/d_{s}$ statistic (ratio of nonsynonymous to synonymous substitutions) computed using PAML (22, 20). Among curated sets of genes linked to vectorial capacity or species-specific traits against a background of functional categories defined by Gene Ontology or InterPro annotations, odorant and gustatory receptors show high evolutionary rates and male accessory gland proteins exhibit exceptionally high $d_{\mu}/d_{s}$ ratios (Fig. 3, figs. S14 and S15, and tables S27 to S29). Rapid divergence in functional categories related to malaria transmission and/or mosquito control strategies led us to examine the genomic basis of several facets of anopheline biology in closer detail.

**Insights into mosquito biology and vectorial capacity**

Mosquito reproductive biology evolves rapidly and presents a compelling target for vector control. This is exemplified by the *An. gambiae* male accessory gland protein (Acp) cluster on chromosomal arm 2R (fig. S17), where conservation is mostly lost outside the *An. gambiae* species complex (fig. S16). In *Drosophila*, male-biased genes such as Arps tend to evolve faster than loci without male-biased expression (23–25). We looked for a similar pattern in anophelines after assessing each gene for sex-biased expression using microarray and RNAseq data sets for *An. gambiae* (22). In contrast to *Drosophila*, female-biased genes show dramatically faster rates of evolution across the genus than male-biased genes (Wilcoxon rank sum test, $P = 5 \times 10^{-4}$) (fig. S17).

Differences in reproductive genes among anophelines may provide insight into the origin and function of sex-related traits. During
copulation, *An. gambiae* males transfer a gelatinous mating plug, a complex of seminal proteins, lipids, and hormones that are essential for successful sperm storage by females and for reproductive success (for successful sperm storage by females and for continuous mating plug, a complex of seminal proteins, lipids, and hormones that are essential for successful sperm storage by females and for reproductive success). The tree depicts the supported molecular topology for the tree describing the supported molecular topology for the species studied. (B) Conserved synteny blocks decay rapidly within chromosomal arms as the phylogenetic distance increases between species. Moving left to right, the dot-plot panels show gene-level synteny between chromosome 2R of *An. gambiae* (x axis) and inferred ancestral sequences (y axes; inferred using PATHGROUPS) at increasing evolutionary time scales (million years ago) estimated by an ultrametric phylogeny. Gray horizontal lines represent scaffold breaks. Discontinuity of the red lines/dots indicates rearrangement. (C) Anopheles X chromosomes exhibit higher rates of rearrangement ($P < 1 \times 10^{-7}$), measured as breaks per Mb per million years, compared with autosomes, despite a paucity of polymorphic inversions on the X. (D) The anopheline X chromosome also displays a higher rate of gene movement to other chromosomal arms than any of the autosomes. Chromosomal elements are labeled around the perimeter; internal bands are colored according to the chromosomal element source and match element colors in (A) and (C). Bands are sized to indicate the relative ratio of genes imported versus exported for each chromosomal element and the relative allocation of exported genes to other elements.

**Fig. 2. Patterns of anopheline chromosomal evolution.** (A) Anophele genomes have conserved gene membership on chromosome arms ("elements"; colored and labeled 1 to 5). Unlike *Drosophila*, chromosome elements reshuffle between chromosomes via translocations as intact elements and do not show fissions or fusions. The tree depicts the supported molecular topology for the species studied. (B) Conserved synteny blocks decay rapidly within chromosomal arms as the phylogenetic distance increases between species. Moving left to right, the dot-plot panels show gene-level synteny between chromosome 2R of *An. gambiae* (x axis) and inferred ancestral sequences (y axes; inferred using PATHGROUPS) at increasing evolutionary time scales (million years ago) estimated by an ultrametric phylogeny. Gray horizontal lines represent scaffold breaks. Discontinuity of the red lines/dots indicates rearrangement. (C) Anopheles X chromosomes exhibit higher rates of rearrangement ($P < 1 \times 10^{-7}$), measured as breaks per Mb per million years, compared with autosomes, despite a paucity of polymorphic inversions on the X. (D) The anopheline X chromosome also displays a higher rate of gene movement to other chromosomal arms than any of the autosomes. Chromosomal elements are labeled around the perimeter; internal bands are colored according to the chromosomal element source and match element colors in (A) and (C). Bands are sized to indicate the relative ratio of genes imported versus exported for each chromosomal element and the relative allocation of exported genes to other elements.

Proteins that constitute the mosquito cuticular exoskeleton play important roles in diverse aspects of anopheline biology, including development, ecology, and insecticide resistance, and constitute approximately 2% of all protein-coding genes (29). Comparisons among dipterans have revealed numerous amplifications of cuticular protein (CP) genes undergoing concerted evolution at physically clustered loci (30–32). We investigated the extent and time scale of gene cluster homogenization within anophelines by generating phylogenies of orthologous gene clusters (fig. S18 and table S31). Throughout the genus, these gene clusters often group phylogenetically by species rather than by position within tandem clusters, particularly in a subset of clusters. These include the 3RB and 3RC clusters of *An. gambiae* species complex (Fig. 4B), consistent with a relatively rapid (less than 20 million years) homogenization of sequences via concerted evolution. The emerging pattern of anopheline CP evolution is thus one of relative stasis for a majority of single-copy orthologs, juxtaposed with consistent concerted evolution of a subset of genes.

Anophelines identify hosts, oviposition sites, and other environmental cues through specialized chemosensory membrane-bound receptors. We examined three of the major gene families that encode these molecules: the odorant receptors (ORs), gustatory receptors (GRs), and variant ionotropic glutamate receptors (iGJs). Given the rapid chemosensory gene turnover observed in many other insects, we explored whether varying host preferences of anopheline mosquitoes could be attributed to chemosensory gene gains and losses. Unexpectedly in light of the elevated genome-wide rate of gene turnover, we found that the overall size and content of the chemosensory
Fig. 3. Contrasting evolutionary properties of selected gene functional categories. Examined evolutionary properties of orthologous groups of genes include a measure of amino acid conservation/divergence (evolutionary rate), a measure of selective pressure (dS/dN), a measure of gene duplication in terms of mean gene copy-number per species (number of genes), and a measure of ortholog universality in terms of number of species with orthologs (number of species). Notched box plots show medians and extend to the first and third quartiles; their widths are proportional to the number of orthologous groups in each functional category. Functional categories derive from curated lists associated with various functions/processes as well as annotated Gene Ontology or InterPro categories (denoted by asterisks).

Gene repertoire are relatively conserved across the genus. CAFE (38) analyses estimated that the most recent common ancestor of the anophelines had approximately 60 genes in each of the OR and GR families, similar to most extant anophelines (Fig. 4C and Fig. S20). Estimated gain/loss rates of OR and GR genes per million years (error-corrected λ = 1.5 × 10⁻⁵ for ORs and 2.0 × 10⁻⁴ for GRs) were much lower than the overall level of anopheline gene families. Similarly, we found almost the same number of antennae-expressed IRs (~20) in all anopheline genomes. Despite overall conservation in chemosensory gene numbers, we observed several examples of gene gain and loss in specific lineages. Notably, there was a net gain of at least 12 ORs in the common ancestor of the An. gambiae complex (Fig. 4C).

OR and GR gene repertoire stability may derive from their roles in several critical behaviors. Host preference differences are likely to be governed by a combination of functional divergence and transcriptional modulation of orthologs. This model is supported by studies of antennal transcriptomes in the major malaria vector An. gambiae (35) and comparisons between this vector and its morphologically identical sibling An. quadriannulatus (36), a very closely related species that plays no role in malaria transmission (despite vectorial competence) because it does not specialize on human hosts. Furthermore, we found that many subfamilies of ORs and GRs showed evidence of positive selection (19 of 53 ORs; 17 of 59 GRs) across the genus, suggesting potential functional divergence.

Several blood feeding–related behaviors in mosquitoes are also regulated by peptide hormones (37). These peptides are synthesized, processed, and released from nervous and endocrine systems and elicit their effects through binding appropriate receptors in target tissues (38). In total, 39 peptide hormones were identified from each of the sequenced anophelines (Fig. S21). Notably, no ortholog of the well-characterized head peptide (HP) hormone of the culicine mosquito Aedes aegypti was identified in any of the assemblies. In Ae. aegypti, HP is responsible for inhibiting host-seeking behavior after a blood meal (39). Because anophelines broadly exhibit similar behavior (40), the absence of HP from the entire clade suggests they may have evolved a novel mechanism to inhibit excess blood feeding. Similarly, no ortholog of insulin growth factor 1 (IGFI) was identified in any anophelines even though IGFI orthologs have been identified in other dipterans, including D. melanogaster (41) and Ae. aegypti (42). IGFI is a key component of the insulin/insulin growth factor 1 signaling (IIS) cascade, which regulates processes including innate immunity, reproduction, metabolism, and life span (43). Nevertheless, other members of the IIS cascade are present, and four insulin-like peptides are found in a compact cluster with gene arrangements conserved across anophelines (Fig. S22). This raises questions regarding the modification of IIS signaling in the absence of IGFI and the functional importance of this conserved genomic arrangement.

Epigenetic mechanisms affect many biological processes by modulation of chromatin structure, telomere remodeling, and transcriptional control. Of the 215 epigenetic regulatory genes in D. melanogaster (44), we identified 169 putative An. gambiae orthologs (table S32), which suggested the presence of mechanisms of epigenetic control in Anopheles and Drosophila. We find, however, that retrotransposition may have contributed to the functional divergence of at least one gene associated with epigenetic regulation. The ubiquitin-conjugating enzyme E2D (orthologous to effete (45) in D. melanogaster) duplicated via retrotransposition in an early anopheline ancestor, and the retrotransposed copy is maintained in a subset of anophelines. Although the entire amino acid sequence of E2D is perfectly conserved between An. gambiae and D. melanogaster, the retrogenes are highly divergent (Fig. 5A) and may contribute to functional diversification within the genus.

Saliva is integral to blood feeding; it impairs host hemostasis and also affects inflammation and immunity. In An. gambiae, the salivary proteome is estimated to contain the products of at least 75 genes, most being expressed solely in the adult female salivary glands. Comparative analyses indicate that anopheline salivary proteins are subject to strong evolutionary pressures, and these genes exhibit an accelerated pace of evolution, as well as a very high rate of gain/loss (Fig. 3 and Fig. S23). Polymorphisms within An. gambiae populations from limited sets of salivary genes were previously found to carry signatures of positive selection (46). Sequence analysis across the anophelines shows that salivary genes have the highest incidence of positively selected codons among the seven gene classes (fig. S24), indicating that coevolution with vertebrate hosts is a powerful driver of natural selection in salivary proteomes. Moreover, salivary proteins also exhibit functional diversification through new gene creation. Sequence similarity, intron-exon boundaries, and secondary structure prediction point to the birth of the SG7/SG7-2 inflammation-inhibiting (47) gene family from the genomic region encoding the C terminus of the 30-kD protein (Fig. 5B), a collagen-binding platelet inhibitor already present in the blood-feeding ancestor of mosquitoes and black flies (48). Based on phylogenetic representation, these events must have occurred before the radiation of anophelines but after separation from the culicines.

Resistance to insecticides and other xenobiotics has arisen independently in many anopheline species, fostered directly and indirectly by anthropogenic environmental modification. Metabolic resistance to insecticides is mediated by
multiple gene families, including cytochrome P450s and glutathione S-transferases (GSTs), which serve to generally protect against all environment stresses, both natural and anthropogenic. We manually characterized these gene families in seven anophelines spanning the genus. Despite their large size, gene numbers (87 to 104 P450 genes, 27 to 30 GST genes) within both gene families are highly conserved across all species, although lineage-specific gene duplications and losses are often seen (tables S33 and S34). As with the OR and GR olfaction-related gene families, P450 and GST repertoires may be relatively constant due to the large number of roles they play in anopheline biology. Orthologs of genes associated with insecticide resistance either via up-regulation or coding variation (e.g., Cyp6m2, Cyp6f2 [Cyp6f9 in An. funestus], Gste2, and Gste4) were found in all species, which suggested that virtually all anophelines likely have genes capable of conferring insecticide resistance through similar mechanisms. Unexpectedly, one member of the P450 family (Cyp18a1) with a conserved role in ecdysteroid catabolism [and consequently development and metamorphosis (49)] appears to have been lost from the ancestor of the An. gambiae species complex but is found in the genome and transcriptome assemblies of other species, which indicates that the An. gambiae complex may have recently evolved an alternate mechanism for catabolizing ecdysone.

Susceptibility to malaria parasites is a key determinant of vectorial capacity. Dissecting the immune repertoire (50, 51) (table S35) into its constituent phases reveals that classical recognition genes and genes encoding effector enzymes exhibit relatively low levels of sequence divergence. Signal transducers are more divergent in sequence but are conserved in gene set completeness. The OR observed gene counts and inferred ancestral gene counts on an ultrametric phylogeny. At least 10 OR genes were gained on the branch leading to the common ancestor of the An. gambiae species complex, although the overall number of OR genes does not vary dramatically across the genus.

Fig. 4. Phylogeny-based insights into anopheline biology. (A) Maximum-likelihood amino acid–based phylogenetic tree of three transglutaminase enzymes (TG1, green; TG2, yellow; and TG3, red) in 14 anopheline species with Culex quinquefasciatus (Cxqu), Ae. aegypti (Aeae), and D. melanogaster (Dmel) serving as outgroups. TG3 is the enzyme responsible for the formation of the male mating plug in An. gambiae, acting upon the substrate PlugIn, the most abundant mating plug protein. Higher rates of evolution for plug-forming TG3 are supported by elevated levels of dK. Mating plug phenotypes are noted where known within the TG3 clade. (B) Concerted evolution in CPFL cuticular proteins. Species symbols used are the same as in (A). In contrast to the TG1/TG2/TG3 phylogeny, CPFL paralogs cluster by subgeneric clades rather than individually recapitulating the species phylogeny. Gene family size variation among species may reflect both gene gain/loss and variation in gene set completeness. (C) OR observed gene counts and inferred ancestral gene counts on an ultrametric phylogeny. At least 10 OR genes were gained on the branch leading to the common ancestor of the An. gambiae species complex, although the overall number of OR genes does not vary dramatically across the genus.
Fig. 5. Genesis of novel anopheline genes. (A) Retrotransposition of the E2D/effete gene generated a ubiquitin-conjugating enzyme at the base of the genus, which exhibits much higher sequence divergence than the original multixenous gene. WebLogo plots contrast the amino acid conservation of the original effete gene with the diversification of the retrotransposed copy (residues 38 to 75; species represented are An. minimus, An. dirus, An. funestus, An. farauti, An. atroparvus, An. sinensis, An. darlingi, and An. albimanus). (B) The SG7 salivary protein—encoding gene was generated from the C-terminal half of the 30-kD gene. SG7 then underwent tandem duplication and intron loss to generate another salivary protein, SG7-2. Numerals indicate lengths of segments in base pairs. (C) The origin of STAT1, a signal transducer and activator of transcription gene involved in immunity, occurred through a retrotransposition event in the Cellia ancestor after divergence from An. dirus and An. farauti. The intronless STAT1 is much more divergent than its multixenous progenitor, STAT2, and has been maintained in all descendant species. An independent retrotransposition event created a retrogene copy in An. atroparvus, which is also more divergent than its progenitor.

REFERENCES AND NOTES
2. A. Cohuet, C. Harris, V. Robert, D. Fontenille, Evolutionary foundation for additional hypothesis generation and testing to further our understanding of the diverse biological traits that determine vectorial capacity.
6. Materials and methods are available as supplementary materials on Science Online.
20. F. Baldini et al., The interaction between a sexually transferred steroid hormone and a female protein regulates oogenesis in the malaria mosquito Anopheles gambiae. PLOS Biol. 11, e1001655 (2013). doi: 10.1371/journal.pbio.1001655; pmid: 24034223


52. C. Barillas-Mury, Y. S. Han, D. Seeley, F. C. Kafatos, Anopheles gambiae Ag-STAT, a new insect member of the STAT family, is activated in response to bacterial infection. EMBO J. 18, 999–967 (1999). doi: 10.1093/emboj/18.4.959; pmid: 10022838


ACKNOWLEDGMENTS

All sequencing reads and genome assemblies have been submitted to the National Center for Biotechnology Information (umbrella BioProject ID, PRJNA6751). Genome and transcriptome assemblies are also available from VectorBase (https://vectorbase.org) and the Broad Institute (https://olive.broadinstitute.org/collections/anopheles.4). The authors acknowledge the NIH Eukaryotic Pathogen and Disease Vector Sequencing Project Working Group for guidance and development of this project. Sequence data generation was supported at the Broad Institute by the National Human Genome Research Institute (U54 HG003067). We thank the many members of the Broad Institute Genomics Platform and Genome Sequencing and Analysis Program who contributed to sequencing data generation and analysis.

SUPPLEMENTAL MATERIALS

www.sciencemag.org/content/347/6217/1258522/suppl/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S25

Tables S1 to S36

References (55–248)

9 July 2014; accepted 14 November 2014

Published online 27 November 2014

10.1126/science.1258522
Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes


Science 347 (6217), 1258522
DOI: 10.1126/science.1258522originally 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

Use of this article is subject to the Terms of Service