

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,⁸ James Amon,⁹ Bruno Arcà,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 B. Gabriel,²⁵ Wamdaogo M. Guelbeogo,²⁶ Andrew B. Hall,²⁷ Mira V. Han,²⁸ Thaung Hlaing,²⁹ Daniel S. T. Hughes,^{8,30} Adam M. Jenkins,³¹ Xiaofang Jiang,^{32,27} Irwin Jungreis,^{2,3} Evdoxia G. Kakani,^{33,34} Maryam Kamali,³⁵ Petri Kemppainen,³⁶ Ryan C. Kennedy,³⁷ Ioannis K. Kirmizoglou,^{16,38} Lizette L. Koekemoer,³⁹ Njoroge Laban,⁴⁰ Nicholas Langridge,⁸ Mara K. N. Lawniczak,¹⁶ Manolis Lirakis,⁴¹ Neil F. Lobo,⁴² Ernesto Lowy,⁸ Robert M. MacCallum,¹⁶ Chunhong Mao,⁴³ Gareth Maslen,⁸ Charles Mbogo,⁴⁴ Jenny McCarthy,¹¹ Kristin Michel,²² Sara N. Mitchell,³³ Wendy Moore,⁴⁵ Katherine A. Murphy,²⁰ Anastasia N. Naumenko,³⁵ Tony Nolan,¹⁶ Eva M. Novoa,^{2,3} Samantha O'Loughlin,¹⁸ Chioma Oringanje,⁴⁵ Mohammad A. Oshaghi,⁶ Nazzy Pakpour,⁴⁶ Philippos A. Papathanos,^{16,24} Ashley N. Peery,³⁵ Michael Povelones,⁴⁷ Anil Prakash,⁴⁸ David P. Price,^{49,50} Ashok Rajaraman,¹⁹ Lisa J. Reimer,⁵¹ David C. Rinker,⁵² Antonis Rokas,^{52,53} Tanya L. Russell,¹⁷ N'Fale Sagnon,²⁶ Maria V. Sharakhova,³⁵ Terrance Shea,¹ Felipe A. Simão,^{4,5} Frederic Simard,²¹ Michel A. Slotman,⁵⁴ Pradya Somboon,⁵⁵ Vladimir Stegnyy,¹² Claudio J. Struchiner,^{56,57} Gregg W. C. Thomas,⁵⁸ Marta Tojo,⁵⁹ Pantelis Topalis,²³ José M. C. Tubio,⁶⁰ Maria F. Unger,⁴² John Vontas,⁴¹ Catherine Walton,³⁶ Craig S. Wilding,⁶¹ Judith H. Willis,⁶² Yi-Chieh Wu,^{2,3,63} Guiyun Yan,⁶⁴ Evgeny M. Zdobnov,^{4,5} Xiaofan Zhou,⁵³ Flaminia Catteruccia,^{33,34} George K. Christophides,¹⁶ Frank H. Collins,⁴² Robert S. Cornman,⁶² Andrea Crisanti,^{16,24} Martin J. Donnelly,^{51,65} Scott J. Emrich,¹³ Michael C. Fontaine,^{42,66} William Gelbart,⁶⁷ Matthew W. Hahn,^{68,58} Immo A. Hansen,^{49,50} Paul I. Howell,⁶⁹ Fotis C. Kafatos,¹⁶ Manolis Kellis,^{2,3} Daniel Lawson,⁸ Christos Louis,^{41,23,24} Shirley Luckhart,⁴⁶ Marc A. T. Muskavitch,^{31,70} José M. Ribeiro,⁷¹ Michael A. Riehle,⁴⁵ Igor V. Sharakhov,^{35,27} Zhijian Tu,³² Laurence J. Zwiebel,⁷² Nora J. Besansky^{4,2†}

Author affiliations are shown at the end of the paper.

*These authors contributed equally to this work.

†Corresponding author. E-mail: neafsey@broadinstitute.org (D.E.N.); nbesansk@nd.edu (N.J.B.)

Variation in vectorial capacity for human malaria among *Anopheles* mosquito species is determined by many factors, including behavior, immunity, and life history. To investigate the genomic basis of vectorial capacity and explore new avenues for vector control, we sequenced the genomes of 16 anopheline mosquito species from diverse locations spanning ~100 million years of evolution. Comparative analyses show faster rates of gene gain and loss, elevated gene shuffling on the X chromosome, and more intron losses, relative to *Drosophila*. Some determinants of vectorial capacity, such as chemosensory genes, do not show elevated turnover, but instead diversify through protein-sequence changes. 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.

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 co-evolutionary 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 verte-

brates (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 approximately 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. 1, A and B). For example, *Anopheles quadriannulatus*, while extremely closely related to *An. gambiae*, feeds preferentially on bovines rather than humans, limiting its potential to transmit human malaria. *Anopheles merus*, *Anopheles melas*, *Anopheles farauti*, and *Anopheles albimanus* females can lay eggs in salty or brackish water, instead of the fresh-water 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 approximately 100 million years (MYr) ago (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 “haploidify” option to reduce haplotype assemblies caused by high heterozygosity. Assembly quality reflected DNA template quality and homozygosity, 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 RNAseq transcriptomes (produced from pooled male and female larvae, pupae, and adults) (table S5), and comprehensive non-coding 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. For analyses described 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 non-masked *An. gambiae* genome (i.e., excluding transposable elements, as detailed in table S7) is alignable to the most closely related species, while 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 35–76% of the *Anopheles stephensi*, *Anopheles funestus*, *Anopheles atroparvus*, and *An. albimanus* genome assemblies to chromosomal arms (tables S9 to S12). Analysis of genes in anchored 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 timescale of 29 MYA 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 2.7-fold more frequent than autosomal rearrangements; in *Drosophila*, the corresponding ratio is only 1.2 (*t* test, $t_{10} = 7.3$, $P < 1 \times 10^{-5}$) (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, $P < 2.2 \times 10^{-16}$) (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 to other dipteran genera.

Such dynamic gene shuffling and movement may be facilitated by the multiple families of DNA transposons and LTR and non-LTR retroelements found in all genomes (table S7), as well as a weaker dosage compensation phenotype in *Anopheles* compared to *Drosophila* (17). Despite such shuffling, comparing 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 dynamism. Analysis of 11,636 gene families with CAFE 3 (18) indicates a rate of gene gain/loss at least five-fold higher than that observed for 12 *Drosophila* genomes (19). Overall, these *Anopheles* genomes exhibit a rate of gain or loss/gene/million years of 3.12×10^{-3} compared to 5.90×10^{-4} for *Drosophila*, suggesting substantially higher gene turnover within anophelines relative to fruit flies. This five-fold greater gain/loss rate in anophelines holds true under models that account for uncertainty in gene family sizes at the tips of the species tree due to annotation/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 *Anopheles arabiensis* and neurotransmitter-gated ion channels in *An. albimanus* (table S24).

The evolutionary dynamism of *Anopheles* genes extends 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).

As 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_N/d_S statistic computed using PAML (12, 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_N/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 chromosome 3R (21, 22), where conservation is mostly lost outside the *An. gambiae* species complex (fig. S16). In *Drosophila*, male-biased genes such as Acps 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 datasets for *An. gambiae* (12). 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 (26–28). Coagulation of the plug is mediated by a seminal transglutaminase (*TG3*), which is found in anophelines but is absent in other mosquito genera that do not form a mating plug (26). We examined *TG3* and its two paralogs (*TG1* and *TG2*) in the sequenced anophelines, and investigated the rate of evolution of each gene (Fig. 4A). Silent sites were saturated at the whole-genus level, making d_S difficult to estimate reliably, but *TG1* (the gene presumed to be ancestral due to broadest taxonomic representation) exhibited the lowest rate of amino acid change ($d_N = 0.20$), *TG2* exhibited an intermediate rate ($d_N = 0.93$), and the anopheline-specific *TG3* has evolved even more rapidly ($d_N = 1.50$), perhaps due to male/male or male/female evolutionary conflict. Interestingly, plug formation appears to be a derived trait within anophelines, as it is not exhibited by *An. albimanus* and intermediate, poorly coagulated plugs were observed in taxa descending from early-branching lineages within the genus (table S30). Functional studies of mating plugs will be necessary to understand what drove the origin and rapid evolution of *TG3*.

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 (30–33). We investigated the extent and timescale 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 arrays, particularly in a subset of clusters. These include the 3RB and 3RC clusters of CP genes (30), the CPLCG group A and CPLCW clusters found elsewhere on 3R (32), and six tandemly arrayed genes on 3L designated *CPFL2* through *CPFL7* (34). CPLCW genes occur in a head-to-head arrangement with CPLCG group A genes, and exhibit highly conserved intergenic sequences (fig. S19). Furthermore, transcript localization studies using in situ hybridization revealed identical spatial expression patterns for CLPCW and CPLCG group A gene pairs suggestive of co-regulation (fig. S19). For these five gene clusters, complete grouping by organismal lineage was observed for most deep nodes as well as for many individual species outside the shallow *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 (*IRs*). Given 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 gene repertoire is relatively conserved across the genus. CAFE 3 (18) 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 $\lambda = 1.3 \times 10^{-3}$ 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). Interestingly, 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 following a blood meal (39). As 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 lifespan (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 impact many biological processes via 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), suggesting 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 co-evolution 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 kDa 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 (87-104 P450 genes, 27-30 GST genes) within both gene families are highly conserved across all species, though 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*, *Cyp6p3* [*Cyp6p9* in *An. funestus*], *Gste2*, *Gste4*) were found in all species, suggesting 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, indicating 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 representation across species and rarely duplicated. Cascade modulators, while also divergent, are more lineage-specific and generally have more gene duplications (Fig. 3 and fig. S25). A rare duplication of an immune signal transduction gene occurred through the retrotransposition of the signal transducer and activator of transcription, *STAT2*, to form the intronless *STAT1* after the divergence of *An. dirus* and *An. farauti* from the rest of the subgenus *Celia* (Fig. 5C and table S36). Interestingly, an independent retroposition event appears to have independently created another intronless STAT gene in the *An. atroparvus* lineage. In *An. gambiae*, *STAT1* controls the expression of *STAT2* and is activated in response to bacterial challenge (52, 53), and the STAT pathway has been demonstrated to mediate immunity to *Plasmodium* (53, 54), so the presence of these relatively new immune signal transducers may have allowed for rewiring of regulatory networks governing immune responses in this subset of anophelines.

Conclusion

Since the discovery over a century ago by Ronald Ross and Giovanni Battista Grassi that human malaria is transmitted by a narrow range of blood-feeding female mosquitoes, the biological basis of malarial vectorial capacity has been a matter of intense interest. Inasmuch as previous successes in the local elimination of malaria have always been accomplished wholly or in part through effective vector control, an increased understanding of vector biology is crucial for continued progress against malarial disease. These 16 new reference genome assemblies provide a foundation for additional hypothesis generation and testing to further our understanding of the diverse biological traits that determine vectorial capacity.

REFERENCES ANDNOTES

1. D. P. Kwiatkowski, How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet.* **77**, 171–192 (2005). [Medline doi:10.1086/432519](#)
2. A. Cohuet, C. Harris, V. Robert, D. Fontenille, Evolutionary forces on Anopheles: What makes a malaria vector? *Trends Parasitol.* **26**, 130–136 (2010). [Medline doi:10.1016/j.pt.2009.12.001](#)
3. S. Manguin, P. Carnivale, J. Mouchet, *Biodiversity of Malaria in the World* (John Libbey Eurotext, Paris, 2008).
4. R. A. Holt, G. M. Subramanian, A. Halpern, G. G. Sutton, R. Charlab, D. R. Nusskern, P. Wincker, A. G. Clark, J. M. Ribeiro, R. Wides, S. L. Salzberg, B. Loftus, M. Yandell, W. H. Majoros, D. B. Rusch, Z. Lai, C. L. Kraft, J. F. Abril, V. Anthouard, P. Arensburger, P. W. Atkinson, H. Baden, V. de Berardinis, D. Baldwin, V. Benes, J. Biedler, C. Blass, R. Bolanos, D. Boscus, M. Barnstead, S. Cai, A. Center, K.

- Chaturverdi, G. K. Christophides, M. A. Chrystal, M. Clamp, A. Cravchik, V. Curwen, A. Dana, A. Delcher, I. Dew, C. A. Evans, M. Flanigan, A. Grundschober-Freimoser, L. Friedli, Z. Gu, P. Guan, R. Guigo, M. E. Hillenmeyer, S. L. Hladun, J. R. Hogan, Y. S. Hong, J. Hoover, O. Jaillon, Z. Ke, C. Kodira, E. Kokoza, A. Koutsos, I. Letunic, A. Levitsky, Y. Liang, J. J. Lin, N. F. Lobo, J. R. Lopez, J. A. Malek, T. C. McIntosh, S. Meister, J. Miller, C. Mobarry, E. Mongin, S. D. Murphy, D. A. O'Brochta, C. Pfannkoch, R. Qi, M. A. Regier, K. Remington, H. Shao, M. V. Sharakhova, C. D. Sitter, J. Shetty, T. J. Smith, R. Strong, J. Sun, D. Thomasova, L. Q. Ton, P. Topalis, Z. Tu, M. F. Unger, B. Walenz, A. Wang, J. Wang, M. Wang, X. Wang, K. J. Woodford, J. R. Wortman, M. Wu, A. Yao, E. M. Zdobnov, H. Zhang, Q. Zhao, S. Zhao, S. C. Zhu, I. Zhimulev, M. Coluzzi, A. della Torre, C. W. Roth, C. Louis, F. Kalush, R. J. Mural, E. W. Myers, M. D. Adams, H. O. Smith, S. Broder, M. J. Gardner, C. M. Fraser, E. Birney, P. Bork, P. T. Brey, J. C. Venter, J. Weissenbach, F. C. Kafatos, F. H. Collins, S. L. Hoffman, The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129–149 (2002). [Medline doi:10.1126/science.1076181](#)
5. O. Marinotti, G. C. Cerqueira, L. G. de Almeida, M. I. Ferro, E. L. Loreto, A. Zaha, S. M. Teixeira, A. R. Wespiser, A. Almeida E Silva, A. D. Schindwein, A. C. Pacheco, A. L. Silva, B. R. Graveley, B. P. Walenz, B. A. Lima, C. A. Ribeiro, C. G. Nunes-Silva, C. R. de Carvalho, C. M. Soares, C. B. de Menezes, C. Matioli, D. Caffrey, D. A. Araújo, D. M. de Oliveira, D. Golenbock, E. C. Grisard, F. Fantinatti-Garbozzini, F. M. de Carvalho, F. G. Barcellos, F. Prosdocimi, G. May, G. M. Azevedo Junior, G. M. Guimarães, G. H. Goldman, I. Q. Padilha, J. S. Batista, J. A. Ferro, J. M. Ribeiro, J. L. Fietto, K. M. Dabbas, L. Cerdeira, L. F. Agnez-Lima, M. Brocchi, M. O. de Carvalho, M. M. Teixeira, M. M. Diniz Maia, M. H. Goldman, M. P. Cruz Schneider, M. S. Felipe, M. Hungria, M. F. Nicolás, M. Pereira, M. A. Montes, M. E. Cantão, M. Vincenz, M. S. Rafael, N. Silverman, P. H. Stoco, R. C. Souza, R. Vicentini, R. T. Gazzinelli, R. O. Neves, R. Silva, S. Astolfi-Filho, T. E. Maciel, T. P. Urményi, W. P. Tadei, E. P. Camargo, A. T. de Vasconcelos, The genome of *Anopheles darlingi*, the main neotropical malaria vector. *Nucleic Acids Res.* **41**, 7387–7400 (2013). [Medline doi:10.1093/nar/gkt484](#)
6. D. Zhou, D. Zhang, G. Ding, L. Shi, Q. Hou, Y. Ye, Y. Xu, H. Zhou, C. Xiong, S. Li, J. Yu, S. Hong, X. Yu, P. Zou, C. Chen, X. Chang, W. Wang, Y. Lv, Y. Sun, L. Ma, B. Shen, C. Zhu, Genome sequence of *Anopheles sinensis* provides insight into genetics basis of mosquito competence for malaria parasites. *BMC Genomics* **15**, 42 (2014). [Medline](#)
7. X. Jiang, A. Peery, A. Hall, A. Sharma, X. G. Chen, R. M. Waterhouse, A. Komissarov, M. M. Riehl, Y. Shouche, M. V. Sharakhova, D. Lawson, N. Pakpour, P. Arensburg, V. L. Davidson, K. Eiglmeier, S. Emrich, P. George, R. C. Kennedy, S. P. Mane, G. Maslen, C. Oringanje, Y. Qi, R. Settlege, M. Tojo, J. M. Tubio, M. F. Unger, B. Wang, K. D. Vernick, J. M. Ribeiro, A. A. James, K. Michel, M. A. Riehle, S. Luckhart, I. V. Sharakhov, Z. Tu, Genome analysis of a major urban malaria vector mosquito, *Anopheles stephensi*. *Genome Biol.* **15**, 459 (2014). [Medline doi:10.1186/s13059-014-0459-2](#)
8. D. E. Neafsey, G. K. Christophides, F. H. Collins, S. J. Emrich, M. C. Fontaine, W. Gelbart, M. W. Hahn, P. I. Howell, F. C. Kafatos, D. Lawson, M. A. Muskavitch, R. M. Waterhouse, L. J. Williams, N. J. Besansky, The evolution of the *Anopheles* 16 genomes project. *Genes, Genomes, Genetics* **3**, 1191–1194 (2013). [Medline doi:10.1534/g3.113.006247](#)
9. M. C. Fontaine, J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey, I. V. Sharakhov, X. Jiang, A. B. Hall, F. Catteruccia, E. Kakani, S. N. Mitchell, Y.-C. Wu, H. A. Smith, R. R. Love, M. K. Lawnczak, M. A. Slotman, S. J. Emrich, M. W. Hahn, N. J. Besansky, Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **346**, 1258524 (2014).
10. M. Moreno, O. Marinotti, J. Krzywinski, W. P. Tadei, A. A. James, N. L. Achee, J. E. Conn, Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time. *Malar. J.* **9**, 127 (2010). [Medline doi:10.1186/1475-2875-9-127](#)
11. S. Gnerre, I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, S. Sykes, A. M. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander, D. B. Jaffe, High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1513–1518 (2011). [Medline doi:10.1073/pnas.1017351108](#)
12. Materials and methods are available as supplementary materials on Science Online.
13. C. Holt, M. Yandell, MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011). [Medline doi:10.1186/1471-2105-12-491](#)
14. M. Coluzzi, A. Sabatini, A. della Torre, M. A. Di Deco, V. Petrarca, A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**, 1415–1418 (2002). [Medline doi:10.1126/science.1077669](#)
15. M. Kamali, P. E. Marek, A. Peery, C. Antonio-Nkondjio, C. Ndo, Z. Tu, F. Simard, I. V. Sharakhov, Multigene phylogenetics reveals temporal diversification of major African malaria vectors. *PLOS ONE* **9**, e93580 (2014). [Medline doi:10.1371/journal.pone.0093580](#)
16. M. A. Toups, M. W. Hahn, Retrogenes reveal the direction of sex-chromosome evolution in mosquitoes. *Genetics* **186**, 763–766 (2010). [Medline doi:10.1534/genetics.110.118794](#)
17. D. A. Baker, S. Russell, Role of testis-specific gene expression in sex-chromosome evolution of *Anopheles gambiae*. *Genetics* **189**, 1117–1120 (2011). [Medline doi:10.1534/genetics.111.133157](#)
18. M. V. Han, G. W. C. Thomas, J. Lugo-Martinez, M. W. Hahn, Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013). [Medline doi:10.1093/molbev/mst100](#)
19. M. W. Hahn, M. V. Han, S.-G. Han, Gene family evolution across 12 *Drosophila* genomes. *PLOS Genet.* **3**, e197 (2007). [Medline doi:10.1371/journal.pgen.0030197](#)
20. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007). [Medline doi:10.1093/molbev/msm088](#)
21. T. Dottorini, L. Nicolaidis, H. Ranson, D. W. Rogers, A. Crisanti, F. Catteruccia, A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16215–16220 (2007). [Medline doi:10.1073/pnas.0703904104](#)
22. F. Baldini, P. Gabrieli, D. W. Rogers, F. Catteruccia, Function and composition of male accessory gland secretions in *Anopheles gambiae*: A comparison with other insect vectors of infectious diseases. *Pathog. Glob. Health* **106**, 82–93 (2012). [Medline doi:10.1179/2047773212Y.0000000016](#)
23. R. Assis, Q. Zhou, D. Bachtrog, Sex-biased transcriptome evolution in *Drosophila*. *Genome Biol. Evol.* **4**, 1189–1200 (2012). [Medline doi:10.1093/gbe/evs093](#)
24. S. Grath, J. Parsch, Rate of amino acid substitution is influenced by the degree and conservation of male-biased transcription over 50 myr of *Drosophila* evolution. *Genome Biol. Evol.* **4**, 346–359 (2012). [Medline doi:10.1093/gbe/evs012](#)
25. J. C. Perry, P. W. Harrison, J. E. Mank, The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. *Mol. Biol. Evol.* **31**, 1206–1219 (2014). [Medline doi:10.1093/molbev/msu072](#)
26. D. W. Rogers, F. Baldini, F. Battaglia, M. Panico, A. Dell, H. R. Morris, F. Catteruccia, Transglutaminase-mediated semen coagulation controls sperm storage in the malaria mosquito. *PLOS Biol.* **7**, e1000272 (2009). [Medline doi:10.1371/journal.pbio.1000272](#)
27. F. Baldini, P. Gabrieli, A. South, C. Valim, F. Mancini, F. Catteruccia, The interaction between a sexually transferred steroid hormone and a female protein regulates oogenesis in the malaria mosquito *Anopheles gambiae*. *PLOS Biol.* **11**, e1001695 (2013). [Medline doi:10.1371/journal.pbio.1001695](#)
28. W. R. Shaw, E. Teodori, S. N. Mitchell, F. Baldini, P. Gabrieli, D. W. Rogers, F. Catteruccia, Mating activates the heme peroxidase HPX15 in the sperm storage organ to ensure fertility in *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5854–5859 (2014). [Medline doi:10.1073/pnas.1401715111](#)
29. J. H. Willis, Structural cuticular proteins from arthropods: Annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem. Mol. Biol.* **40**, 189–204 (2010). [Medline doi:10.1016/j.ibmb.2010.02.001](#)
30. R. S. Cornman, T. Togawa, W. A. Dunn, N. He, A. C. Emmons, J. H. Willis, Annotation and analysis of a large cuticular protein family with the R&R Consensus in *Anopheles gambiae*. *BMC Genomics* **9**, 22 (2008). [Medline doi:10.1186/1471-2164-9-22](#)
31. R. S. Cornman, J. H. Willis, Extensive gene amplification and concerted evolution within the CPR family of cuticular proteins in mosquitoes. *Insect Biochem. Mol. Biol.* **38**, 661–676 (2008). [Medline doi:10.1016/j.ibmb.2008.04.001](#)
32. R. S. Cornman, J. H. Willis, Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle. *Insect Mol. Biol.* **18**, 607–622 (2009). [Medline doi:10.1111/j.1365-2583.2009.00902.x](#)
33. R. S. Cornman, Molecular evolution of *Drosophila* cuticular protein genes. *PLOS*

- ONE 4, e8345 (2009). [Medline doi:10.1371/journal.pone.0008345](#)
34. T. Togawa, W. A. Dunn, A. C. Emmons, J. Nagao, J. H. Willis, Developmental expression patterns of cuticular protein genes with the R&R Consensus from *Anopheles gambiae*. *Insect Biochem. Mol. Biol.* **38**, 508–519 (2008). [Medline doi:10.1016/j.ibmb.2007.12.008](#)
 35. D. C. Rinker, R. J. Pitts, X. Zhou, E. Suh, A. Rokas, L. J. Zwiebel, Blood meal-induced changes to antennal transcriptome profiles reveal shifts in odor sensitivities in *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8260–8265 (2013). [Medline doi:10.1073/pnas.1302562110](#)
 36. D. C. Rinker, X. Zhou, R. J. Pitts, A. Rokas, L. J. Zwiebel; AGC Consortium, Antennal transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in *Anopheles gambiae*. *BMC Genomics* **14**, 749 (2013). [Medline doi:10.1186/1471-2164-14-749](#)
 37. M. Altstein, D. R. Nässel, Neuropeptide signaling in insects. *Adv. Exp. Med. Biol.* **692**, 155–165 (2010). [Medline doi:10.1007/978-1-4419-6902-6_8](#)
 38. J. P. Goetze, I. Hunter, S. K. Lippert, L. Bardram, J. F. Rehfeld, Processing-independent analysis of peptide hormones and prohormones in plasma. *Front. Biosci. (Landmark Ed.)* **17**, 1804–1815 (2012). [Medline doi:10.2741/4020](#)
 39. T. H. Stracker, S. Thompson, G. L. Grossman, M. A. Riehle, M. R. Brown, Characterization of the AeaHP gene and its expression in the mosquito *Aedes aegypti* (Diptera: Culicidae). *J. Med. Entomol.* **39**, 331–342 (2002). [Medline doi:10.1603/0022-2585.39.2.331](#)
 40. C. D. de Oliveira, W. P. Tadei, F. C. Abdalla, P. F. Paolucci Pimenta, O. Marinotti, Multiple blood meals in *Anopheles darlingi* (Diptera: Culicidae). *J. Vector Ecol.* **37**, 351–358 (2012). [Medline doi:10.1111/j.1948-7134.2012.00238.x](#)
 41. N. Okamoto, N. Yamanaka, Y. Yagi, Y. Nishida, H. Kataoka, M. B. O'Connor, A. Mizoguchi, A fat body-derived IGF-like peptide regulates postfeeding growth in *Drosophila*. *Dev. Cell* **17**, 885–891 (2009). [Medline doi:10.1016/j.devcel.2009.10.008](#)
 42. M. A. Riehle, Y. Fan, C. Cao, M. R. Brown, Molecular characterization of insulin-like peptides in the yellow fever mosquito, *Aedes aegypti*: Expression, cellular localization, and phylogeny. *Peptides* **27**, 2547–2560 (2006). [Medline doi:10.1016/j.peptides.2006.07.016](#)
 43. Y. Antonova, A. J. Arik, W. Moore, M. M. Riehle, M. R. Brown, in *Insect Endocrinology*, L. I. Gilbert, Ed. (Academic Press, 2012), pp. 63–92.
 44. A. Swaminathan, A. Gajan, L. A. Pile, Epigenetic regulation of transcription in *Drosophila*. *Front. Biosci. (Landmark Ed.)* **17**, 909–937 (2012). [Medline doi:10.2741/3964](#)
 45. F. Cipressa, S. Romano, S. Centonze, P. I. zur Lage, F. Verni, P. Dimitri, M. Gatti, G. Cenci, Effete, a *Drosophila* chromatin-associated ubiquitin-conjugating enzyme that affects telomeric and heterochromatic position effect variegation. *Genetics* **195**, 147–158 (2013). [Medline doi:10.1534/genetics.113.153320](#)
 46. B. Arcà, C. J. Struchiner, V. M. Pham, G. Sferra, F. Lombardo, M. Pombi, J. M. Ribeiro, Positive selection drives accelerated evolution of mosquito salivary genes associated with blood-feeding. *Insect Mol. Biol.* **23**, 122–131 (2014). [Medline doi:10.1111/imb.12068](#)
 47. H. Isawa, M. Yuda, Y. Orito, Y. Chinzei, A mosquito salivary protein inhibits activation of the plasma contact system by binding to factor XII and high molecular weight kininogen. *J. Biol. Chem.* **277**, 27651–27658 (2002). [Medline doi:10.1074/jbc.M203505200](#)
 48. E. Calvo, F. Tokumasu, O. Marinotti, J. L. Villeval, J. M. Ribeiro, I. M. Francischetti, Aegyptin, a novel mosquito salivary gland protein, specifically binds to collagen and prevents its interaction with platelet glycoprotein VI, integrin alpha2beta1, and von Willebrand factor. *J. Biol. Chem.* **282**, 26928–26938 (2007). [Medline doi:10.1074/jbc.M705669200](#)
 49. E. Guittard, C. Blais, A. Maria, J. P. Parvy, S. Pasricha, C. Lumb, R. Lafont, P. J. Daborn, C. Dauphin-Villemant, CYP18A1, a key enzyme of *Drosophila* steroid hormone inactivation, is essential for metamorphosis. *Dev. Biol.* **349**, 35–45 (2011). [Medline doi:10.1016/j.ydbio.2010.09.023](#)
 50. R. M. Waterhouse, E. V. Kriventseva, S. Meister, Z. Xi, K. S. Alvarez, L. C. Bartholomay, C. Barillas-Mury, G. Bian, S. Blandin, B. M. Christensen, Y. Dong, H. Jiang, M. R. Kanost, A. C. Koutsos, E. A. Levashina, J. Li, P. Ligoxygakis, R. M. Maccallum, G. F. Mayhew, A. Mendes, K. Michel, M. A. Osta, S. Paskewitz, S. W. Shin, D. Vlachou, L. Wang, W. Wei, L. Zheng, Z. Zou, D. W. Severson, A. S. Raikhel, F. C. Kafatos, G. Dimopoulos, E. M. Zdobnov, G. K. Christophides, Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* **316**, 1738–1743 (2007). [Medline doi:10.1126/science.1139862](#)
 51. L. C. Bartholomay, R. M. Waterhouse, G. F. Mayhew, C. L. Campbell, K. Michel, Z. Zou, J. L. Ramirez, S. Das, K. Alvarez, P. Arensburg, B. Bryant, S. B. Chapman, Y. Dong, S. M. Erickson, S. H. Karunaratne, V. Kokoza, C. D. Kodira, P. Pignatelli, S. W. Shin, D. L. Vanlandingham, P. W. Atkinson, B. Birren, G. K. Christophides, R. J. Clem, J. Hemingway, S. Higgs, K. Megy, H. Ranson, E. M. Zdobnov, A. S. Raikhel, B. M. Christensen, G. Dimopoulos, M. A. Muskavitch, Pathogenomics of *Culex quinquefasciatus* and meta-analysis of infection responses to diverse pathogens. *Science* **330**, 88–90 (2010). [Medline doi:10.1126/science.1193162](#)
 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**, 959–967 (1999). [Medline doi:10.1093/emboj/18.4.959](#)
 53. L. Gupta, A. Molina-Cruz, S. Kumar, J. Rodrigues, R. Dixit, R. E. Zamora, C. Barillas-Mury, The STAT pathway mediates late-phase immunity against Plasmodium in the mosquito *Anopheles gambiae*. *Cell Host Microbe* **5**, 498–507 (2009). [Medline doi:10.1016/j.chom.2009.04.003](#)
 54. A. C. Bahia, M. S. Kubota, A. J. Tempone, H. R. Araújo, B. A. Guedes, A. S. Orfanó, W. P. Tadei, C. M. Ríos-Velásquez, Y. S. Han, N. F. Secundino, C. Barillas-Mury, P. F. Pimenta, Y. M. Traub-Csekö, The JAK-STAT pathway controls Plasmodium vivax load in early stages of *Anopheles aquasalis* infection. *PLoS Negl. Trop. Dis.* **5**, e1317 (2011). [Medline doi:10.1371/journal.pntd.0001317](#)
 55. M. E. Sinka, M. J. Bangs, S. Manguin, Y. Rubio-Palis, T. Chareonviriyaphap, M. Coetzee, C. M. Mbogo, J. Hemingway, A. P. Patil, W. H. Temperley, P. W. Gething, C. W. Kabaria, T. R. Burkot, R. E. Harbach, S. I. Hay, A global map of dominant malaria vectors. *Parasit Vectors* **5**, 69 (2012). [Medline doi:10.1186/1756-3305-5-69](#)
 56. M. E. Sinka, M. J. Bangs, S. Manguin, T. Chareonviriyaphap, A. P. Patil, W. H. Temperley, P. W. Gething, I. R. Elyazar, C. W. Kabaria, R. E. Harbach, S. I. Hay, The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: Occurrence data, distribution maps and bionomic précis. *Parasit Vectors* **4**, 89 (2011). [Medline doi:10.1186/1756-3305-4-89](#)
 57. M. E. Sinka, M. J. Bangs, S. Manguin, M. Coetzee, C. M. Mbogo, J. Hemingway, A. P. Patil, W. H. Temperley, P. W. Gething, C. W. Kabaria, R. M. Okara, T. Van Boeckel, H. C. Godfray, R. E. Harbach, S. I. Hay, The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: Occurrence data, distribution maps and bionomic précis. *Parasit Vectors* **3**, 117 (2010). [Medline doi:10.1186/1756-3305-3-117](#)
 58. M. E. Sinka, Y. Rubio-Palis, S. Manguin, A. P. Patil, W. H. Temperley, P. W. Gething, T. Van Boeckel, C. W. Kabaria, R. E. Harbach, S. I. Hay, The dominant *Anopheles* vectors of human malaria in the Americas: Occurrence data, distribution maps and bionomic précis. *Parasit Vectors* **3**, 72 (2010). [Medline doi:10.1186/1756-3305-3-72](#)
 59. L. J. Williams, D. G. Tabbaa, N. Li, A. M. Berlin, T. P. Shea, I. Maccallum, M. S. Lawrence, Y. Drier, G. Getz, S. K. Young, D. B. Jaffe, C. Nusbaum, A. Gnirke, Paired-end sequencing of Fosmid libraries by Illumina. *Genome Res.* **22**, 2241–2249 (2012). [Medline doi:10.1101/gr.138925.112](#)
 60. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Muceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011). [Medline doi:10.1038/nbt.1883](#)
 61. M. K. Lawnczak, S. J. Emrich, A. K. Holloway, A. P. Regier, M. Olson, B. White, S. Redmond, L. Fulton, E. Appelbaum, J. Godfrey, C. Farmer, A. Chinwalla, S. P. Yang, P. Minx, J. Nelson, K. Kyung, B. P. Walenz, E. Garcia-Hernandez, M. Aguiar, L. D. Viswanathan, Y. H. Rogers, R. L. Strausberg, C. A. Sasaki, D. Lawson, F. H. Collins, F. C. Kafatos, G. K. Christophides, S. W. Clifton, E. F. Kirkness, N. J. Besansky, Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* **330**, 512–514 (2010). [Medline doi:10.1126/science.1195755](#)
 62. V. Nene, J. R. Wortman, D. Lawson, B. Haas, C. Kodira, Z. J. Tu, B. Loftus, Z. Xi, K. Megy, M. Grabherr, Q. Ren, E. M. Zdobnov, N. F. Lobo, K. S. Campbell, S. E. Brown, M. F. Bonaldo, J. Zhu, S. P. Sinkins, D. G. Hogenkamp, P. Amedeo, P. Arensburg, P. W. Atkinson, S. Bidwell, J. Biedler, E. Birney, R. V. Bruggner, J. Costas, M. R. Coy, J. Crabtree, M. Crawford, B. Debruyne, D. Decaprio, K. Eiglmeier, E. Eisenstadt, H. El-Dorry, W. M. Gelbart, S. L. Gomes, M. Hammond, L. I. Hannick, J. R. Hogan, M. H. Holmes, D. Jaffe, J. S. Johnston, R. C. Kennedy, H. Koo, S. Kravitz, E. V. Kriventseva, D. Kulp, K. Labutti, E. Lee, S. Li, D. D. Lovin, C. Mao, E. Muceli, C. F. Menck, J. R. Miller, P. Montgomery, A. Mori, A. L. Nascimento, H. F. Naveira, C. Nusbaum, S. O'leary, J. Orvis, M. Pertea, H.

- Quesneville, K. R. Reidenbach, Y. H. Rogers, C. W. Roth, J. R. Schneider, M. Schatz, M. Shumway, M. Stanke, E. O. Stinson, J. M. Tubio, J. P. Vanzee, S. Verjovski-Almeida, D. Werner, O. White, S. Wyder, Q. Zeng, Q. Zhao, Y. Zhao, C. A. Hill, A. S. Raikhel, M. B. Soares, D. L. Knudson, N. H. Lee, J. Galagan, S. L. Salzberg, I. T. Paulsen, G. Dimopoulos, F. H. Collins, B. Birren, C. M. Fraser-Liggett, D. W. Severson, Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* **316**, 1718–1723 (2007). [Medline doi:10.1126/science.1138878](#)
63. P. Arensburg, K. Megy, R. M. Waterhouse, J. Abrudan, P. Amedeo, B. Antelo, L. Bartholomay, S. Bidwell, E. Caler, F. Camara, C. L. Campbell, K. S. Campbell, C. Casola, M. T. Castro, I. Chandramouliswaran, S. B. Chapman, S. Christley, J. Costas, E. Eisenstadt, C. Feschotte, C. Fraser-Liggett, R. Guigo, B. Haas, M. Hammond, B. S. Hansson, J. Hemingway, S. R. Hill, C. Howarth, R. Ignell, R. C. Kennedy, C. D. Kodira, N. F. Lobo, C. Mao, G. Mayhew, K. Michel, A. Mori, N. Liu, H. Naveira, V. Nene, N. Nguyen, M. D. Pearson, E. J. Pritham, D. Puiu, Y. Qi, H. Ranson, J. M. Ribeiro, H. M. Roberston, D. W. Severson, M. Shumway, M. Stanke, R. L. Strausberg, C. Sun, G. Sutton, Z. J. Tu, J. M. Tubio, M. F. Unger, D. L. Vanlandingham, A. J. Vilella, O. White, J. R. White, C. S. Wondji, J. Wortman, E. M. Zdobnov, B. Birren, B. M. Christensen, F. H. Collins, A. Cornel, G. Dimopoulos, L. I. Hannick, S. Higgs, G. C. Lanzaro, D. Lawson, N. H. Lee, M. A. Muskavitch, A. S. Raikhel, P. W. Atkinson, Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science* **330**, 86–88 (2010). [Medline doi:10.1126/science.1191864](#)
64. M. D. Adams, S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. H. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. Gabor, J. F. Abril, A. Agbayani, H. J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M. H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G. H. Karpen, Z. Ke, J. A. Kennison, K. A. Ketchum, B. E. Kimmel, C. D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A. A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T. C. McIntosh, M. P. McLeod, D. McPherson, G. Merkulov, N. V. Milshina, C. Mobbary, J. Morris, A. Moshrefi, S. M. Mount, M. Moy, B. Murphy, L. Murphy, D. M. Muzny, D. L. Nelson, D. R. Nelson, K. A. Nelson, K. Nixon, D. R. Nusskern, J. M. Pacleb, M. Palazzolo, G. S. Pittman, S. Pan, J. Pollard, V. Puri, M. G. Reese, K. Reinert, K. Remington, R. D. Saunders, F. Scheeler, H. Shen, B. C. Shue, I. Sidén-Kiamos, M. Simpson, M. P. Skupski, T. Smith, E. Spier, A. C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A. H. Wang, X. Wang, Z. Y. Wang, D. A. Wassarman, G. M. Weinstock, J. Weissbach, S. M. Williams, K. C. Woodage, T. D. Worley, S. Wu, Q. A. Yang, J. Yao, R. F. Ye, J. S. Yeh, M. Zaveri, G. Zhan, Q. Zhang, L. Zhao, X. H. Zheng, F. N. Zheng, W. Zhong, X. Zhong, S. Zhou, X. Zhu, H. O. Zhu, R. A. Smith, E. W. Gibbs, G. M. Myers, J. C. Rubin, Venter, The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000). [Medline doi:10.1126/science.287.5461.2185](#)
65. G. S. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005). [Medline doi:10.1186/1471-2105-6-31](#)
66. R. M. Waterhouse, F. Tegenfeldt, J. Li, E. M. Zdobnov, E. V. Kriventseva, OrthoDB: A hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res.* **41** (D1), D358–D365 (2013). [Medline doi:10.1093/nar/gks1116](#)
67. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21** (Suppl 1), i351–i358 (2005). [Medline doi:10.1093/bioinformatics/bti1018](#)
68. Z. Bao, S. R. Eddy, Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269–1276 (2002). [Medline doi:10.1101/gr.88502](#)
69. A. Smit, R. Hubley, P. Green, RepeatMasker (1996-2010); www.repeatmasker.org
70. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004). [Medline doi:10.1186/1471-2105-5-59](#)
71. M. Stanke, B. Morgenstern, AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33** (Web Server), W465–W467 (2005). [Medline doi:10.1093/nar/gki458](#)
72. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009). [Medline doi:10.1093/bioinformatics/btp120](#)
73. UniProt Consortium, Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42** (D1), D191–D198 (2014). [Medline doi:10.1093/nar/gkt1140](#)
74. K. Megy, S. J. Emrich, D. Lawson, D. Campbell, E. Dyalynas, D. S. Hughes, G. Koscielny, C. Louis, R. M. Maccallum, S. N. Redmond, A. Sheehan, P. Topalis, D. Wilson; VectorBase Consortium, VectorBase: Improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Res.* **40** (D1), D729–D734 (2012). [Medline doi:10.1093/nar/gkr1089](#)
75. J. L. Jakubczak, W. D. Burke, T. H. Eickbush, Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3295–3299 (1991). [Medline doi:10.1073/pnas.88.8.3295](#)
76. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). [Medline doi:10.1093/nar/gkh340](#)
77. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007). [Medline doi:10.1093/bioinformatics/btm404](#)
78. T. M. Lowe, S. R. Eddy, tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997). [Medline doi:10.1093/nar/25.5.0955](#)
79. A. R. Gruber, S. Findeiß, S. Washietl, I. L. Hofacker, P. F. Stadler, RNAz 2.0: Improved noncoding RNA detection. *Pac. Symp. Biocomput.* **•••**, 69–79 (2010). [Medline doi:10.1186/1471-2105-10-S1-S35](#)
80. S. W. Burge, J. Daub, R. Eberhardt, J. Tate, L. Barquist, E. P. Nawrocki, S. R. Eddy, P. P. Gardner, A. Bateman, Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **41** (D1), D226–D232 (2013). [Medline doi:10.1093/nar/gks1005](#)
81. S. Kadri, V. Hinman, P. V. Benos, HHMMiR: Efficient de novo prediction of microRNAs using hierarchical hidden Markov models. *BMC Bioinformatics* **10** (Suppl 1), S35 (2009). [Medline doi:10.1186/1471-2105-10-S1-S35](#)
82. Y. Wu, B. Wei, H. Liu, T. Li, S. Rayner, MIRPara: A SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences. *BMC Bioinformatics* **12**, 107 (2011). [Medline doi:10.1186/1471-2105-12-107](#)
83. A. Kozomara, S. Griffiths-Jones, miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42** (D1), D68–D73 (2014). [Medline doi:10.1093/nar/gkt1181](#)
84. P. Schattner, A. N. Brooks, T. M. Lowe, The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33** (Web Server), W686–W689 (2005). [Medline doi:10.1093/nar/gki366](#)
85. J. Hertel, I. L. Hofacker, P. F. Stadler, SnoReport: Computational identification of snoRNAs with unknown targets. *Bioinformatics* **24**, 158–164 (2008). [Medline doi:10.1093/bioinformatics/btm464](#)
86. D. Tautz, J. M. Hancock, D. A. Webb, C. Tautz, G. A. Dover, Complete sequences of the rRNA genes of *Drosophila melanogaster*. *Mol. Biol. Evol.* **5**, 366–376 (1988). [Medline doi:10.1016/0092-8674\(77\)90169-6](#)
87. S. Artavanis-Tsakonas, P. Schedl, C. Tschudi, V. Pirrotta, R. Steward, W. J. Gehring, The 5S genes of *Drosophila melanogaster*. *Cell* **12**, 1057–1067 (1977). [Medline doi:10.1016/0092-8674\(77\)90169-6](#)
88. R. M. Waterhouse, E. M. Zdobnov, F. Tegenfeldt, J. Li, E. V. Kriventseva, OrthoDB: The hierarchical catalog of eukaryotic orthologs in 2011. *Nucleic Acids Res.* **39** (Database), D283–D288 (2011). [Medline doi:10.1093/nar/gkn930](#)
89. E. V. Kriventseva, N. Rahman, O. Espinosa, E. M. Zdobnov, OrthoDB: The hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res.* **36** (Database), D271–D275 (2008). [Medline doi:10.1093/nar/gkm845](#)
90. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009). [Medline doi:10.1093/bioinformatics/btp348](#)
91. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). [Medline doi:10.1093/bioinformatics/btu033](#)

92. J. O. McInerney, GCUA: General codon usage analysis. *Bioinformatics* **14**, 372–373 (1998). [Medline doi:10.1093/bioinformatics/14.4.372](#)
93. E. M. Novoa, L. Ribas de Pouplana, Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet.* **28**, 574–581 (2012). [Medline doi:10.1016/j.tig.2012.07.006](#)
94. S. Pechmann, J. Frydman, Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* **20**, 237–243 (2013). [Medline doi:10.1038/nmsb.2466](#)
95. A. Nakao, M. Yoshihama, N. Kenmochi, RPG: The Ribosomal Protein Gene database. *Nucleic Acids Res.* **32**, D168–D170 (2004). [Medline doi:10.1093/nar/gkh004](#)
96. S. Anders, P. Pyl, W. Huber, HTSeq: Analysing high-throughput sequencing data with Python. *bioRxiv preprint* (2014); [doi:10.1101/002824](#)
97. T. Ikemura, Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J. Mol. Biol.* **151**, 389–409 (1981). [Medline doi:10.1016/0022-2836\(81\)90003-6](#)
98. M. dos Reis, R. Savva, L. Wernisch, Solving the riddle of codon usage preferences: A test for translational selection. *Nucleic Acids Res.* **32**, 5036–5044 (2004). [Medline doi:10.1093/nar/gkh834](#)
99. S. Vicario, E. N. Moriyama, J. R. Powell, Codon usage in twelve species of *Drosophila*. *BMC Evol. Biol.* **7**, 226 (2007). [Medline doi:10.1186/1471-2148-7-226](#)
100. A. Smit, R. Hubble, P. Green, RepeatModeler Open-1.0 (2008-2010); [www.repeatmasker.org](#)
101. J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005). [Medline doi:10.1159/000084979](#)
102. R. C. Kennedy, M. F. Unger, S. Christley, F. H. Collins, G. R. Madey, An automated homology-based approach for identifying transposable elements. *BMC Bioinformatics* **12**, 130 (2011). [Medline doi:10.1186/1471-2105-12-130](#)
103. X. Huang, A. Madan, CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999). [Medline doi:10.1101/gr.9.9.868](#)
104. A. Stark, M. F. Lin, P. Kheradpour, J. S. Pedersen, L. Parts, J. W. Carlson, M. A. Crosby, M. D. Rasmussen, S. Roy, A. N. Deoras, J. G. Ruby, J. Brennecke, E. Hodges, A. S. Hinrichs, A. Caspi, B. Paten, S. W. Park, M. V. Han, M. L. Maeder, B. J. Polansky, B. E. Robson, S. Aerts, J. van Helden, B. Hassan, D. G. Gilbert, D. A. Eastman, M. Rice, M. Weir, M. W. Hahn, Y. Park, C. N. Dewey, L. Pachter, W. J. Kent, D. Haussler, E. C. Lai, D. P. Bartel, G. J. Hannon, T. C. Kaufman, M. B. Eisen, A. G. Clark, D. Smith, S. E. Celnik, W. M. Gelbart, M. Kellis; Harvard FlyBase curators; Berkeley *Drosophila* Genome Project, Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**, 219–232 (2007). [Medline doi:10.1038/nature06340](#)
105. K. Lindblad-Toh, M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. Di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz, D. B. Jaffe, I. Jungreis, W. J. Kent, D. Kostka, M. Lara, A. L. Martins, T. Massingham, I. Moltke, B. J. Raney, M. D. Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody, J. Baldwin, T. Bloom, C. W. Chin, D. Heiman, R. Nicol, C. Nusbaum, S. Young, J. Wilkinson, K. C. Worley, C. L. Kovar, D. M. Muzny, R. A. Gibbs, A. Cree, H. H. Dihn, G. Fowler, S. Jhangiani, V. Joshi, S. Lee, L. R. Lewis, L. V. Nazareth, G. Okwuonu, J. Santibanez, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson, K. Delehaunty, D. Dooling, C. Fronik, L. Fulton, B. Fulton, T. Graves, P. Minx, E. Sodergren, E. Birney, E. H. Margulies, J. Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, M. Kellis; Broad Institute Sequencing Platform and Whole Genome Assembly Team, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011). [Medline doi:10.1038/nature10530](#)
106. M. Blanchette, W. J. Kent, C. Riemer, L. Elnitski, A. F. Smit, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, W. Miller, Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004). [Medline doi:10.1101/gr.1933104](#)
107. E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004). [Medline doi:10.1101/gr.1865504](#)
108. K. P. Byrne, K. H. Wolfe, The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* **15**, 1456–1461 (2005). [Medline doi:10.1101/gr.3672305](#)
109. C. Zheng, Pathgroups, a dynamic data structure for genome reconstruction problems. *Bioinformatics* **26**, 1587–1594 (2010). [Medline doi:10.1093/bioinformatics/btq255](#)
110. M. V. Sharakhova, A. Xia, Z. Tu, Y. S. Shouche, M. F. Unger, I. V. Sharakhov, A physical map for an Asian malaria mosquito, *Anopheles stephensi*. *Am. J. Trop. Med. Hyg.* **83**, 1023–1027 (2010). [Medline doi:10.4269/ajtmh.2010.10-0366](#)
111. I. V. Sharakhov, A. C. Serazin, O. G. Grushko, A. Dana, N. Lobo, M. E. Hillenmeyer, R. Westerman, J. Romero-Severson, C. Costantini, N. Sagnon, F. H. Collins, N. J. Besansky, Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science* **298**, 182–185 (2002). [Medline doi:10.1126/science.1076803](#)
112. A. J. Cornel, F. H. Collins, Maintenance of chromosome arm integrity between two *Anopheles* mosquito subgenera. *J. Hered.* **91**, 364–370 (2000). [Medline doi:10.1093/jhered/91.5.364](#)
113. E. M. Zdobnov, C. von Mering, I. Letunic, D. Torrents, M. Suyama, R. R. Copley, G. K. Christophides, D. Thomasova, R. A. Holt, G. M. Subramanian, H. M. Mueller, G. Dimopoulos, J. H. Law, M. A. Wells, E. Birney, R. Charlab, A. L. Halpern, E. Kokoza, C. L. Kraft, Z. Lai, S. Lewis, C. Louis, C. Barillas-Mury, D. Nusskern, G. M. Rubin, S. L. Salzberg, G. G. Sutton, P. Topalis, R. Wides, P. Wincker, M. Yandell, F. H. Collins, J. Ribeiro, W. M. Gelbart, F. C. Kafatos, P. Bork, Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**, 149–159 (2002). [Medline doi:10.1126/science.1077061](#)
114. S. W. Schaeffer, A. Bhutkar, B. F. McAllister, M. Matsuda, L. M. Matzkin, P. M. O’Grady, C. Rohde, V. L. Valente, M. Aguadé, W. W. Anderson, K. Edwards, A. C. Garcia, J. Goodman, J. Hartigan, E. Kataoka, R. T. Lapoint, E. R. Lozovsky, C. A. Machado, M. A. Noor, M. Papaceit, L. K. Reed, S. Richards, T. T. Rieger, S. M. Russo, H. Sato, C. Segarra, D. R. Smith, T. F. Smith, V. Strelets, Y. N. Tobar, Y. Tomimura, M. Wasserman, T. Watts, R. Wilson, K. Yoshida, T. A. Markow, W. M. Gelbart, T. C. Kaufman, Polytene chromosomal maps of 11 *Drosophila* species: The order of genomic scaffolds inferred from genetic and physical maps. *Genetics* **179**, 1601–1655 (2008). [Medline doi:10.1534/genetics.107.086074](#)
115. L. Guy, J. R. Kultima, S. G. Andersson, genoPlotR: Comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334–2335 (2010). [Medline doi:10.1093/bioinformatics/btq413](#)
116. G. Tesler, GRIMM: Genome rearrangements web server. *Bioinformatics* **18**, 492–493 (2002). [Medline doi:10.1093/bioinformatics/18.3.492](#)
117. M. von Grotthuss, M. Ashburner, J. M. Ranz, Fragile regions and not functional constraints predominate in shaping gene organization in the genus *Drosophila*. *Genome Res.* **20**, 1084–1096 (2010). [Medline doi:10.1101/gr.103713.109](#)
118. G. Bourque, P. A. Pevzner, Genome-scale evolution: Reconstructing gene orders in the ancestral species. *Genome Res.* **12**, 26–36 (2002). [Medline doi:10.1101/gr.103713.109](#)
119. A. Bhutkar, S. W. Schaeffer, S. M. Russo, M. Xu, T. F. Smith, W. M. Gelbart, Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics* **179**, 1657–1680 (2008). [Medline doi:10.1534/genetics.107.086108](#)
120. M. Slotman, A. Della Torre, J. R. Powell, Female sterility in hybrids between *Anopheles gambiae* and *A. arabiensis*, and the causes of Haldane’s rule. *Evolution* **59**, 1016–1026 (2005). [Medline doi:10.1111/j.0014-3820.2005.tb01040.x](#)
121. M. Slotman, A. Della Torre, J. R. Powell, The genetics of inviability and male sterility in hybrids between *Anopheles gambiae* and *An. arabiensis*. *Genetics* **167**, 275–287 (2004). [Medline doi:10.1534/genetics.167.1.275](#)
122. V. A. Timoshevskiy, N. A. Kinney, B. S. deBruyn, C. Mao, Z. Tu, D. W. Severson, I. V. Sharakhov, M. V. Sharakhova, Genomic composition and evolution of *Aedes aegypti* chromosomes revealed by the analysis of physically mapped supercontigs. *BMC Biol.* **12**, 27 (2014). [Medline doi:10.1186/1741-7007-12-27](#)
123. M. V. Han, M. W. Hahn, Inferring the history of interchromosomal gene transposition in *Drosophila* using n-dimensional parsimony. *Genetics* **190**, 813–825 (2012). [Medline doi:10.1534/genetics.111.135947](#)
124. E. Betrán, K. Thornton, M. Long, Retroposed new genes out of the X in *Drosophila*. *Genome Res.* **12**, 1854–1859 (2002). [Medline doi:10.1101/gr.6049](#)
125. B. R. Jones, A. Rajaraman, E. Tannier, C. Chauve, ANGES: Reconstructing ANcestral GENomeS maps. *Bioinformatics* **28**, 2388–2390 (2012). [Medline doi:10.1093/bioinformatics/bts457](#)
126. C. Chauve, E. Tannier, A methodological framework for the reconstruction of contiguous regions of ancestral genomes and its application to mammalian genomes. *PLOS Comput. Biol.* **4**, e1000234 (2008). [Medline doi:10.1371/journal.pcbi.1000234](#)
127. A. Bergeron, J. Mixtacki, J. Stoye, HP distance via double cut and join distance. Proceedings of the 14th Annual Symposium on Combinatorial Pattern Matching

- (Springer Verlag, 2008), pp. 56–68. (2006).
128. R. Hilker, C. Sicking, C. N. Pedersen, J. Stoye, UniMoG—a unifying framework for genomic distance calculation and sorting based on DCJ. *Bioinformatics* **28**, 2509–2511 (2012). [Medline doi:10.1093/bioinformatics/bts440](#)
 129. S. Aganezov, N. Sydtnikova, A. G. C. Consortium, M. A. Alekseyev, Scaffold assembly based on genome rearrangement analysis. Proceedings of the 13th Asia-Pacific Bioinformatics Conference (2014).
 130. M. A. Alekseyev, P. A. Pevzner, Breakpoint graphs and ancestral genome reconstructions. *Genome Res.* **19**, 943–957 (2009). [Medline doi:10.1101/gr.082784.108](#)
 131. P. George, M. V. Sharakhova, I. V. Sharakhov, High-resolution cytogenetic map for the African malaria vector *Anopheles gambiae*. *Insect Mol. Biol.* **19**, 675–682 (2010). [Medline doi:10.1111/j.1365-2583.2010.01025.x](#)
 132. M. J. Sanderson, r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003). [Medline doi:10.1093/bioinformatics/19.2.301](#)
 133. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997). [Medline doi:10.1093/nar/25.17.3389](#)
 134. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002). [Medline doi:10.1093/nar/30.7.1575](#)
 135. M. Csurós, Malin: Maximum likelihood analysis of intron evolution in eukaryotes. *Bioinformatics* **24**, 1538–1539 (2008). [Medline doi:10.1093/bioinformatics/btn226](#)
 136. T. Rognes, Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. *BMC Bioinformatics* **12**, 221 (2011). [Medline doi:10.1186/1471-2105-12-221](#)
 137. Y. C. Wu, M. D. Rasmussen, M. Kellis, Evolution at the subgene level: Domain rearrangements in the *Drosophila* phylogeny. *Mol. Biol. Evol.* **29**, 689–705 (2012). [Medline doi:10.1093/molbev/msr222](#)
 138. M. F. Lin, I. Jungreis, M. Kellis, PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–i282 (2011). [Medline doi:10.1093/bioinformatics/btr209](#)
 139. M. F. Lin, J. W. Carlson, M. A. Crosby, B. B. Matthews, C. Yu, S. Park, K. H. Wan, A. J. Schroeder, L. S. Gramates, S. E. St Pierre, M. Roark, K. L. Wiley Jr., R. J. Kulathinal, P. Zhang, K. V. Myrick, J. V. Antone, S. E. Celniker, W. M. Gelbart, M. Kellis, Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res.* **17**, 1823–1836 (2007). [Medline doi:10.1101/gr.6679507](#)
 140. I. Jungreis, M. F. Lin, R. Spokony, C. S. Chan, N. Negre, A. Victorsen, K. P. White, M. Kellis, Evidence of abundant stop codon readthrough in *Drosophila* and other metazoa. *Genome Res.* **21**, 2096–2113 (2011). [Medline doi:10.1101/gr.119974.110](#)
 141. J. G. Dunn, C. K. Foo, N. G. Belletier, E. R. Gavis, J. S. Weissman, Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife* **2**, e01179 (2013). [Medline doi:10.7554/eLife.01179](#)
 142. C. S. Chan, I. Jungreis, M. Kellis, Heterologous stop codon readthrough of metazoan readthrough candidates in yeast. *PLOS ONE* **8**, e59450 (2013). [Medline doi:10.1371/journal.pone.0059450](#)
 143. P. Steneberg, C. Samakovlis, A novel stop codon readthrough mechanism produces functional Headcase protein in *Drosophila* trachea. *EMBO Rep.* **2**, 593–597 (2001). [Medline doi:10.1093/embo-reports/kve128](#)
 144. M. Suyama, D. Torrents, P. Bork, PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34** (Web Server), W609–W612 (2006). [Medline doi:10.1093/nar/gkl315](#)
 145. R. M. Waterhouse, E. M. Zdobnov, E. V. Kriventseva, Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. *Genome Biol. Evol.* **3**, 75–86 (2011). [Medline doi:10.1093/gbe/evq083](#)
 146. S. Falcon, R. Gentleman, Using GOSTats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007). [Medline doi:10.1093/bioinformatics/btl567](#)
 147. A. Alexa, J. Rahnenfuhrer, topGO: Enrichment analysis for Gene Ontology. R package version 2.18.0 (2010); www.bioconductor.org/packages/release/bioc/html/topGO.html
 148. S. B. Needleman, C. D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970). [Medline doi:10.1016/0022-2836\(70\)90057-4](#)
 149. T. F. Smith, M. S. Waterman, Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981). [Medline doi:10.1016/0022-2836\(81\)90087-5](#)
 150. C. Gillott, Male accessory gland secretions: Modulators of female reproductive physiology and behavior. *Annu. Rev. Entomol.* **48**, 163–184 (2003). [Medline doi:10.1146/annurev.ento.48.091801.112657](#)
 151. F. W. Avila, L. K. Sirot, B. A. LaFlamme, C. D. Rubinstein, M. F. Wolfner, Insect seminal fluid proteins: Identification and function. *Annu. Rev. Entomol.* **56**, 21–40 (2011). [Medline doi:10.1146/annurev-ento-120709-144823](#)
 152. A. G. Clark, M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, T. A. Markow, T. C. Kaufman, M. Kellis, W. Gelbart, V. N. Iyer, D. A. Pollard, T. B. Sackton, A. M. Larracuente, N. D. Singh, J. P. Abad, D. N. Abt, B. Adryan, M. Aguade, H. Akashi, W. W. Anderson, C. F. Aquadro, D. H. Ardell, R. Arguello, C. G. Artieri, D. A. Barbash, D. Barker, P. Barsanti, P. Batterham, S. Batzoglou, D. Begun, A. Bhutkar, E. Blanco, S. A. Bosak, R. K. Bradley, A. D. Brand, M. R. Brent, A. N. Brooks, R. H. Brown, R. K. Butlin, C. Caggese, B. R. Calvi, A. Bernardo de Carvalho, A. Caspi, S. Castrezana, S. E. Celniker, J. L. Chang, C. Chapple, S. Chatterji, A. Chinwalla, A. Civetta, S. W. Clifton, J. M. Comeron, J. C. Costello, J. A. Coyne, J. Daub, R. G. David, A. L. Delcher, K. Delehaunty, C. B. Do, H. Ebling, K. Edwards, T. Eickbush, J. D. Evans, A. Filipiński, S. Findeiss, E. Freyhult, L. Fulton, R. Fulton, A. C. Garcia, A. Gardiner, D. A. Garfield, B. E. Garvin, G. Gibson, D. Gilbert, S. Gnerre, J. Godfrey, R. Good, V. Gotea, B. Gravelly, A. J. Greenberg, S. Griffiths-Jones, S. Gross, R. Guigo, E. A. Gustafson, W. Haerty, M. W. Hahn, D. L. Halligan, A. L. Halpern, G. M. Halter, M. V. Han, A. Heger, L. Hillier, A. S. Hinrichs, I. Holmes, R. A. Hoskins, M. J. Hubisz, D. Hultmark, M. A. Huntley, D. B. Jaffe, S. Jagadeeshan, W. R. Jeck, J. Johnson, C. D. Jones, W. C. Jordan, G. H. Karpen, E. Kataoka, P. D. Keightley, P. Kheradpour, E. F. Kirkness, L. B. Koerich, K. Kristiansen, D. Kudrna, R. J. Kulathinal, S. Kumar, R. Kwok, E. Lander, C. H. Langley, R. Lapoint, B. P. Lazzaro, S. J. Lee, L. Levesque, R. Li, C. F. Lin, M. F. Lin, K. Lindblad-Toh, A. Llopart, M. Long, L. Low, E. Lozovsky, J. Lu, M. Luo, C. A. Machado, W. Makalowski, M. Marzo, M. Matsuda, L. Matzkin, B. McAllister, C. S. McBride, B. McKernan, K. McKernan, M. Mendez-Lago, P. Minx, M. U. Mollenhauer, K. Montooth, S. M. Mount, X. Mu, E. Myers, B. Negre, S. Newfeld, R. Nielsen, M. A. Noor, P. O'Grady, L. Pachter, M. Papacait, M. J. Parisi, M. Parisi, L. Parts, J. S. Pedersen, G. Pesole, A. M. Phillippy, C. P. Ponting, M. Pop, D. Porcelli, J. R. Powell, S. Prohaska, K. Pruitt, M. Puig, H. Quesneville, K. R. Ram, D. Rand, M. D. Rasmussen, L. K. Reed, R. Reenan, A. Reily, K. A. Remington, T. T. Rieger, M. G. Ritchie, C. Robin, Y. H. Rogers, C. Rohde, J. Rozas, M. J. Rubenfield, A. Ruiz, S. Russo, S. L. Salzberg, A. Sanchez-Gracia, D. J. Saranga, H. Sato, S. W. Schaeffer, M. C. Schatz, T. Schlenke, R. Schwartz, C. Segarra, R. S. Singh, L. Sirot, M. Sirota, N. B. Sisneros, C. D. Smith, T. F. Smith, J. Spieth, D. E. Stage, A. Stark, W. Stephan, R. L. Strausberg, S. Strepel, D. Sturgill, G. Sutton, G. R. Sutton, W. Tao, S. Teichmann, Y. N. Tobar, Y. Tomimura, J. M. Tsolas, V. L. Valente, E. Venter, J. C. Venter, S. Vicario, F. G. Vieira, A. J. Vilella, A. Villasante, B. Walenz, J. Wang, M. Wasserman, T. Watts, D. Wilson, R. K. Wilson, R. A. Wing, M. F. Wolfner, A. Wong, G. K. Wong, C. I. Wu, G. Wu, D. Yamamoto, H. P. Yang, S. P. Yang, J. A. Yorke, K. Yoshida, E. Zdobnov, P. Zhang, Y. Zhang, A. V. Zimin, J. R. Baldwin, A. Abdouelleil, J. Abdulkadir, A. Abebe, B. Abera, J. Abreu, S. C. Acer, L. Aftuck, A. Alexander, P. An, E. Anderson, S. Anderson, H. Arachi, M. Azer, P. Bachantsang, A. Barry, T. Bayul, A. Berlin, D. Bessette, T. Bloom, J. Blye, L. Boguslavskiy, C. Bonnet, B. Boukhgalter, I. Bourzgui, A. Brown, P. Cahill, S. Channer, Y. Cheshatsang, L. Chuda, M. Citroen, A. Collymore, P. Cooke, M. Costello, K. D'Acro, R. Daza, G. De Haan, S. DeGray, C. DeMaso, N. Dhargay, K. Dooley, E. Dooley, M. Doricent, P. Dorje, K. Dorjee, A. Dupes, R. Elong, J. Falk, A. Farina, S. Faro, D. Ferguson, S. Fisher, C. D. Foley, A. Franke, D. Friedrich, L. Gadbois, G. Gearin, C. R. Gearin, G. Giannoukos, T. Goode, J. Graham, E. Grandbois, S. Grewal, K. Gyaltsen, N. Hafez, B. Hagos, J. Hall, C. Henson, A. Hollinger, T. Honan, M. D. Huard, L. Hughes, B. Hurhula, M. E. Husby, A. Kamat, B. Kanga, S. Kashin, D. Khazanovich, P. Kisner, K. Lance, M. Lara, W. Lee, N. Lennon, F. Letendre, R. LeVine, A. Lipovsky, X. Liu, J. Liu, S. Liu, T. Lokyitsang, Y. Lokyitsang, R. Lubonja, A. Lui, P. MacDonald, V. Magnalis, K. Maru, C. Matthews, W. McCusker, S. McDonough, T. Mehta, J. Meldrim, L. Meneus, O. Mihai, A. Mihalev, T. Mihova, R. Mittelman, V. Mlenga, A. Montmayeur, L. Mulrain, A. Navidi, J. Naylor, T. Negash, T. Nguyen, N. Nguyen, R. Nicol, C. Norbu, N. Norbu, N. Novod, B. O'Neill, S. Osman, E. Markiewicz, O. L. Oyono, C. Patti, P. Phunkhang, F. Pierre, M. Priest, S. Raghuraman, F. Rege, R. Reyes, C. Rise, P. Rogov, K. Ross, E. Ryan, S. Settipalli, T. Shea, N. Sherpa, L. Shi, D. Shih, T.

- Sparrow, J. Spaulding, J. Stalker, N. Stange-Thomann, S. Stavropoulos, C. Stone, C. Strader, S. Tesfaye, T. Thomson, Y. Thouloutsang, D. Thouloutsang, K. Topham, I. Topping, T. Tsamla, H. Vassiliev, A. Vo, T. Wangchuk, T. Wangdi, M. Weiland, J. Wilkinson, A. Wilson, S. Yadav, G. Young, Q. Yu, L. Zembek, D. Zhong, A. Zimmer, Z. Zwirko, D. B. Jaffe, P. Alvarez, W. Brockman, J. Butler, C. Chin, S. Gnerre, M. Grabherr, M. Kleber, E. Mauceli, I. MacCallum; Drosophila 12 Genomes Consortium, Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**, 203–218 (2007). [Medline doi:10.1038/nature06341](#)
153. J. A. Andrés, L. S. Maroja, S. M. Bogdanowicz, W. J. Swanson, R. G. Harrison, Molecular evolution of seminal proteins in field crickets. *Mol. Biol. Evol.* **23**, 1574–1584 (2006). [Medline doi:10.1093/molbev/msl020](#)
154. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). [Medline doi:10.1093/bioinformatics/bts635](#)
155. D. A. Baker, T. Nolan, B. Fischer, A. Pinder, A. Crisanti, S. Russell, A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*. *BMC Genomics* **12**, 296 (2011). [Medline doi:10.1186/1471-2164-12-296](#)
156. J. Parsch, H. Ellegren, The evolutionary causes and consequences of sex-biased gene expression. *Nat. Rev. Genet.* **14**, 83–87 (2013). [Medline doi:10.1038/nrg3376](#)
157. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011). [Medline doi:10.1093/molbev/msr121](#)
158. L. Vannini, T. W. Reed, J. H. Willis, Temporal and spatial expression of cuticular proteins of *Anopheles gambiae* implicated in insecticide resistance or differentiation of M/S incipient species. *Parasit Vectors* **7**, 24 (2014). [Medline doi:10.1186/1756-3305-7-24](#)
159. J. Vontas, J. P. David, D. Nikou, J. Hemingway, G. K. Christophides, C. Louis, H. Ranson, Transcriptional analysis of insecticide resistance in *Anopheles stephensi* using cross-species microarray hybridization. *Insect Mol. Biol.* **16**, 315–324 (2007). [Medline doi:10.1111/j.1365-2583.2007.00728.x](#)
160. O. Wood, S. Hanrahan, M. Coetzee, L. Koekemoer, B. Brooke, Cuticle thickening associated with pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Parasit Vectors* **3**, 67 (2010). [Medline doi:10.1186/1756-3305-3-67](#)
161. Y. Goltsev, G. L. Rezende, K. Vranizan, G. Lanzaro, D. Valle, M. Levine, Developmental and evolutionary basis for drought tolerance of the *Anopheles gambiae* embryo. *Dev. Biol.* **330**, 462–470 (2009). [Medline doi:10.1016/j.ydbio.2009.02.038](#)
162. B. J. White, M. W. Hahn, M. Pombi, B. J. Cassone, N. F. Lobo, F. Simard, N. J. Besansky, Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. *PLOS Genet.* **3**, e217 (2007). [Medline doi:10.1371/journal.pgen.0030217](#)
163. T. Togawa, W. Augustine Dunn, A. C. Emmons, J. H. Willis, CPF and CPFL, two related gene families encoding cuticular proteins of *Anopheles gambiae* and other insects. *Insect Biochem. Mol. Biol.* **37**, 675–688 (2007). [Medline doi:10.1016/j.ibmb.2007.03.011](#)
164. M. Nei, A. P. Rooney, Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**, 121–152 (2005). [Medline doi:10.1146/annurev.genet.39.073003.112240](#)
165. J. Qiu, P. E. Hardin, Temporal and spatial expression of an adult cuticle protein gene from *Drosophila* suggests that its protein product may impart some specialized cuticle function. *Dev. Biol.* **167**, 416–425 (1995). [Medline doi:10.1006/dbio.1995.1038](#)
166. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). [Medline doi:10.1093/nar/gkf436](#)
167. K. Chen, D. Durand, M. Farach-Colton, NOTUNG: A program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429–447 (2000). [Medline doi:10.1089/106652700750050871](#)
168. W. Takken, N. O. Verhulst, Host preferences of blood-feeding mosquitoes. *Annu. Rev. Entomol.* **58**, 433–453 (2013). [Medline doi:10.1146/annurev-ento-120811-153618](#)
169. T. J. Wheeler, J. D. Kececioglu, Multiple alignment by aligning alignments. *Bioinformatics* **23**, i559–i568 (2007). [Medline doi:10.1093/bioinformatics/btm226](#)
170. W. P. Maddison, D. R. Maddison, Mesquite: A modular system for evolutionary analysis. Version 3.01 (2011); <http://mesquiteproject.org>
171. R. Lanfear, B. Calcott, S. Y. Ho, S. Guindon, Partitionfinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701 (2012). [Medline doi:10.1093/molbev/mss020](#)
172. A. Stamatakis, P. Hoover, J. Rougemont, A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* **57**, 758–771 (2008). [Medline doi:10.1080/10635150802429642](#)
173. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006). [Medline doi:10.1093/bioinformatics/btl446](#)
174. Y. Antonova, A. Arik, W. Moore, M. Riehle, M. Brown, Insulin-like peptides: Structure, signaling, and function. *Insect Endocrinology. Gilbert L.I. (Ed) Academic Press*, 63–92 (2012).
175. A. A. Horton, B. Wang, L. Camp, M. S. Price, A. Arshi, M. Nagy, S. A. Nadler, J. R. Faeder, S. Luckhart, The mitogen-activated protein kinase from *Anopheles gambiae*: Identification, phylogeny and functional characterization of the ERK, JNK and p38 MAP kinases. *BMC Genomics* **12**, 574 (2011). [Medline doi:10.1186/1471-2164-12-574](#)
176. N. Pakpour, L. Camp, H. M. Smithers, B. Wang, Z. Tu, S. A. Nadler, S. Luckhart, Protein kinase C-dependent signaling controls the midgut epithelial barrier to malaria parasite infection in anopheline mosquitoes. *PLOS ONE* **8**, e76535 (2013). [Medline doi:10.1371/journal.pone.0076535](#)
177. C. Rosse, M. Linch, S. Kermorgant, A. J. Cameron, K. Boeckeler, P. J. Parker, PKC and the control of localized signal dynamics. *Nat. Rev. Mol. Cell Biol.* **11**, 103–112 (2010). [Medline doi:10.1038/nrm2847](#)
178. S. L. Tan, P. J. Parker, Emerging and diverse roles of protein kinase C in immune cell signalling. *Biochem. J.* **376**, 545–552 (2003). [Medline doi:10.1042/B120031406](#)
179. J. Gong, B. Hoyos, R. Acin-Perez, V. Vinogradov, E. Shabrova, F. Zhao, M. Leitges, D. Fischman, G. Manfredi, U. Hammerling, Two protein kinase C isoforms, δ and ϵ , regulate energy homeostasis in mitochondria by transmitting opposing signals to the pyruvate dehydrogenase complex. *FASEB J.* **26**, 3537–3549 (2012). [Medline doi:10.1096/fj.11-197376](#)
180. B. H. Shieh, L. Parker, D. Popescu, Protein kinase C (PKC) isoforms in *Drosophila*. *J. Biochem.* **132**, 523–527 (2002). [Medline doi:10.1093/oxfordjournals.jbchem.a003252](#)
181. V. Sargsyan, M. N. Getahun, S. L. Llanos, S. B. Olsson, B. S. Hansson, D. Wicher, Phosphorylation via PKC Regulates the Function of the *Drosophila* Odorant Co-Receptor. *Front. Cell. Neurosci.* **5**, 5 (2011). [Medline doi:10.3389/fncel.2011.00005](#)
182. K. Liu, H. Tsujimoto, S. J. Cha, P. Agre, J. L. Rasgon, Aquaporin water channel AgAQP1 in the malaria vector mosquito *Anopheles gambiae* during blood feeding and humidity adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6062–6066 (2011). [Medline doi:10.1073/pnas.1102629108](#)
183. H. Tsujimoto, K. Liu, P. J. Linser, P. Agre, J. L. Rasgon, Organ-specific splice variants of aquaporin water channel AgAQP1 in the malaria vector *Anopheles gambiae*. *PLOS ONE* **8**, e75888 (2013). [Medline doi:10.1371/journal.pone.0075888](#)
184. L. L. Drake, D. Y. Boudko, O. Marinotti, V. K. Carpenter, A. L. Dawe, I. A. Hansen, The Aquaporin gene family of the yellow fever mosquito, *Aedes aegypti*. *PLOS ONE* **5**, e15578 (2010). [Medline doi:10.1371/journal.pone.0015578](#)
185. G. J. Filion, J. G. van Bommel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker, B. van Steensel, Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010). [Medline doi:10.1016/j.cell.2010.09.009](#)
186. E. L. Greer, Y. Shi, Histone methylation: A dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–357 (2012). [Medline doi:10.1038/nrg3173](#)
187. J. G. van Bommel, G. J. Filion, A. Rosado, W. Talhout, M. de Haas, T. van Welsom, F. van Leeuwen, B. van Steensel, A network model of the molecular organization of chromatin in *Drosophila*. *Mol. Cell* **49**, 759–771 (2013). [Medline doi:10.1016/j.molcel.2013.01.040](#)
188. C. H. Arrowsmith, C. Bountra, P. V. Fish, K. Lee, M. Schapira, Epigenetic protein families: A new frontier for drug discovery. *Nat. Rev. Drug Discov.* **11**, 384–400 (2012). [Medline doi:10.1038/nrd3674](#)
189. S. R. Schulze, L. L. Wallrath, Gene regulation by chromatin structure: Paradigms established in *Drosophila melanogaster*. *Annu. Rev. Entomol.* **52**, 171–192 (2007). [Medline doi:10.1146/annurev.ento.51.110104.151007](#)

190. S. Powell, K. Forslund, D. Szklarczyk, K. Trachana, A. Roth, J. Huerta-Cepas, T. Gabaldón, T. Rattei, C. Creevey, M. Kuhn, L. J. Jensen, C. von Mering, P. Bork, eggNOG v4.0: Nested orthology inference across 3686 organisms. *Nucleic Acids Res.* **42** (D1), D231–D239 (2014). [Medline doi:10.1093/nar/gkt1253](#)
191. E. Darbo, C. Herrmann, T. Lecuit, D. Thieffry, J. van Helden, Transcriptional and epigenetic signatures of zygotic genome activation during early *Drosophila* embryogenesis. *BMC Genomics* **14**, 226 (2013). [Medline doi:10.1186/1471-2164-14-226](#)
192. P. Dimitri, C. Pisano, Position effect variegation in *Drosophila melanogaster*: Relationship between suppression effect and the amount of Y chromosome. *Genetics* **122**, 793–800 (1989). [Medline](#)
193. E. Betrán, M. Long, Dntf-2r, a young *Drosophila* retroposed gene with specific male expression under positive Darwinian selection. *Genetics* **164**, 977–988 (2003). [Medline](#)
194. R. Kalamegham, D. Sturgill, E. Siegfried, B. Oliver, *Drosophila* mojeless, a retroposed GSK-3, has functionally diverged to acquire an essential role in male fertility. *Mol. Biol. Evol.* **24**, 732–742 (2007). [Medline doi:10.1093/molbev/msl201](#)
195. W. Wang, F. G. Brunet, E. Nevo, M. Long, Origin of sphinx, a young chimeric RNA gene in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4448–4453 (2002). [Medline doi:10.1073/pnas.072066399](#)
196. M. Berriman, K. Rutherford, Viewing and annotating sequence data with Artemis. *Brief. Bioinform.* **4**, 124–132 (2003). [Medline doi:10.1093/bib/4.2.124](#)
197. S. L. Pond, S. D. Frost, S. V. Muse, HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005). [Medline doi:10.1093/bioinformatics/bti079](#)
198. B. Murrell, J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, S. L. Kosakovsky Pond, Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012). [Medline doi:10.1371/journal.pgen.1002764](#)
199. R. L. Tatusov, N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, D. A. Natale, The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41 (2003). [Medline doi:10.1186/1471-2105-4-41](#)
200. S. Karim, P. Singh, J. M. Ribeiro, A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. *PLoS ONE* **6**, e28525 (2011). [Medline doi:10.1371/journal.pone.0028525](#)
201. J. M. Ribeiro, B. J. Mans, B. Arcà, An insight into the sialome of blood-feeding Nematocera. *Insect Biochem. Mol. Biol.* **40**, 767–784 (2010). [Medline doi:10.1016/j.ibmb.2010.08.002](#)
202. B. Arcà, F. Lombardo, J. G. Valenzuela, I. M. Francischetti, O. Marinotti, M. Coluzzi, J. M. Ribeiro, An updated catalogue of salivary gland transcripts in the adult female mosquito, *Anopheles gambiae*. *J. Exp. Biol.* **208**, 3971–3986 (2005). [Medline doi:10.1242/jeb.01849](#)
203. M. A. Okulate, D. E. Kalume, R. Reddy, T. Kristiansen, M. Bhattacharyya, R. Chaerkady, A. Pandey, N. Kumar, Identification and molecular characterization of a novel protein Saglin as a target of monoclonal antibodies affecting salivary gland infectivity of *Plasmodium* sporozoites. *Insect Mol. Biol.* **16**, 711–722 (2007). [Medline doi:10.1111/j.1365-2583.2007.00765.x](#)
204. C. Rizzo, R. Ronca, G. Fiorentino, V. D. Mangano, S. B. Sirima, I. Nèbiè, V. Petrarca, D. Modiano, B. Arcà, Wide cross-reactivity between *Anopheles gambiae* and *Anopheles funestus* SG6 salivary proteins supports exploitation of gSG6 as a marker of human exposure to major malaria vectors in tropical Africa. *Malar. J.* **10**, 206 (2011). [Medline doi:10.1186/1475-2875-10-206](#)
205. E. Calvo, B. J. Mans, J. F. Andersen, J. M. Ribeiro, Function and evolution of a mosquito salivary protein family. *J. Biol. Chem.* **281**, 1935–1942 (2006). [Medline doi:10.1074/jbc.M510359200](#)
206. H. Isawa, Y. Orito, S. Iwanaga, N. Jingushi, A. Morita, Y. Chinzei, M. Yuda, Identification and characterization of a new kallikrein-kinin system inhibitor from the salivary glands of the malaria vector mosquito *Anopheles stephensi*. *Insect Biochem. Mol. Biol.* **37**, 466–477 (2007). [Medline doi:10.1016/j.ibmb.2007.02.002](#)
207. H. Ranson, C. Claudianos, F. Orтели, C. Abgrall, J. Hemingway, M. V. Sharakhova, M. F. Unger, F. H. Collins, R. Feyereisen, Evolution of supergene families associated with insecticide resistance. *Science* **298**, 179–181 (2002). [Medline doi:10.1126/science.1076781](#)
208. C. V. Edi, B. G. Koudou, C. M. Jones, D. Weetman, H. Ranson, Multiple-insecticide resistance in *Anopheles gambiae* mosquitoes, Southern Côte d'Ivoire. *Emerg. Infect. Dis.* **18**, 1508–1511 (2012). [Medline doi:10.3201/eid1809.120262](#)
209. J. Pinto, A. Lynd, J. L. Vicente, F. Santolamazza, N. P. Randle, G. Gentile, M. Moreno, F. Simard, J. D. Charlwood, V. E. do Rosário, A. Caccone, A. Della Torre, M. J. Donnelly, Multiple origins of knockdown resistance mutations in the Afrotropical mosquito vector *Anopheles gambiae*. *PLoS ONE* **2**, e1243 (2007). [Medline doi:10.1371/journal.pone.0001243](#)
210. A. Lynd, D. Weetman, S. Barbosa, A. Egyir Yawson, S. Mitchell, J. Pinto, I. Hastings, M. J. Donnelly, Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in *Anopheles gambiae* s.s. *Mol. Biol. Evol.* **27**, 1117–1125 (2010). [Medline doi:10.1093/molbev/msq002](#)
211. D. K. Mathias, E. Ochomo, F. Atieli, M. Ombok, M. N. Bayoh, G. Olang, D. Muhia, L. Kamau, J. M. Vulule, M. J. Hamel, W. A. Hawley, E. D. Walker, J. E. Gimnig, Spatial and temporal variation in the kdr allele L1014S in *Anopheles gambiae* s.s. and phenotypic variability in susceptibility to insecticides in Western Kenya. *Malar. J.* **10**, 10 (2011). [Medline doi:10.1186/1475-2875-10-10](#)
212. C. V. Edi, L. Djogbénu, A. M. Jenkins, K. Regna, M. A. Muskavitch, R. Poupardin, C. M. Jones, J. Essandoh, G. K. Kéto, M. J. Paine, B. G. Koudou, M. J. Donnelly, H. Ranson, D. Weetman, CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito *Anopheles gambiae*. *PLoS Genet.* **10**, e1004236 (2014). [Medline doi:10.1371/journal.pgen.1004236](#)
213. S. N. Mitchell, B. J. Stevenson, P. Müller, C. S. Wilding, A. Egyir-Yawson, S. G. Field, J. Hemingway, M. J. Paine, H. Ranson, M. J. Donnelly, Identification and validation of a gene causing cross-resistance between insecticide classes in *Anopheles gambiae* from Ghana. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6147–6152 (2012). [Medline doi:10.1073/pnas.1203452109](#)
214. S. N. Mitchell, D. J. Rigden, A. J. Dowd, F. Lu, C. S. Wilding, D. Weetman, S. Dadzie, A. M. Jenkins, K. Regna, P. Boko, L. Djogbenou, M. A. Muskavitch, H. Ranson, M. J. Paine, O. Mayans, M. J. Donnelly, Metabolic and target-site mechanisms combine to confer strong DDT resistance in *Anopheles gambiae*. *PLoS ONE* **9**, e26662 (2014). [Medline doi:10.1371/journal.pone.0092662](#)
215. J. M. Riveron, C. Yunta, S. S. Ibrahim, R. Djouaka, H. Irving, B. D. Menze, H. M. Ismail, J. Hemingway, H. Ranson, A. Albert, C. S. Wondji, A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. *Genome Biol.* **15**, R27 (2014). [Medline doi:10.1186/gb-2014-15-2-r27](#)
216. C. F. Ayres, P. Müller, N. Dyer, C. S. Wilding, D. J. Rigden, M. J. Donnelly, Comparative genomics of the anopheline glutathione S-transferase epsilon cluster. *PLoS ONE* **6**, e29237 (2011). [Medline doi:10.1371/journal.pone.0029237](#)
217. R. Feyereisen, Evolution of insect P450. *Biochem. Soc. Trans.* **34**, 1252–1255 (2006). [Medline doi:10.1042/BST0341252](#)
218. R. Niwa, T. Matsuda, T. Yoshiyama, T. Namiki, K. Mita, Y. Fujimoto, H. Kataoka, CYP306A1, a cytochrome P450 enzyme, is essential for ecdysteroid biosynthesis in the prothoracic glands of *Bombyx* and *Drosophila*. *J. Biol. Chem.* **279**, 35942–35949 (2004). [Medline doi:10.1074/jbc.M404514200](#)
219. H. M. Ismail, P. M. O'Neill, D. W. Hong, R. D. Finn, C. J. Henderson, A. T. Wright, B. F. Cravatt, J. Hemingway, M. J. Paine, Pyrethroid activity-based probes for profiling cytochrome P450 activities associated with insecticide interactions. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19766–19771 (2013). [Medline doi:10.1073/pnas.1320185110](#)
220. G. K. Christophides, E. Zdobnov, C. Barillas-Mury, E. Birney, S. Blandin, C. Blass, P. T. Brey, F. H. Collins, A. Danielli, G. Dimopoulos, C. Hetru, N. T. Hoa, J. A. Hoffmann, S. M. Kanzok, I. Letunic, E. A. Levashina, T. G. Loukeris, G. Lycett, S. Meister, K. Michel, L. F. Moita, H. M. Müller, M. A. Osta, S. M. Paskewitz, J. M. Reichhart, A. Rzhetsky, L. Troxler, K. D. Vernick, D. Vlachou, J. Volz, C. von Mering, J. Xu, L. Zheng, P. Bork, F. C. Kafatos, Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298**, 159–165 (2002). [Medline doi:10.1126/science.1077136](#)
221. D. Vlachou, F. C. Kafatos, The complex interplay between mosquito positive and negative regulators of *Plasmodium* development. *Curr. Opin. Microbiol.* **8**, 415–421 (2005). [Medline doi:10.1016/j.mib.2005.06.013](#)
222. E. A. Levashina, L. F. Moita, S. Blandin, G. Vriend, M. Lagueux, F. C. Kafatos, 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 (2001). [Medline doi:10.1016/S0092-8674\(01\)00267-7](#)
223. S. Blandin, S. H. Shiao, L. F. Moita, C. J. Janse, A. P. Waters, F. C. Kafatos, E. A. Levashina, Complement-like protein TEPI is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell* **116**, 661–670 (2004). [Medline](#)

- [doi:10.1016/S0092-8674\(04\)00173-4](https://doi.org/10.1016/S0092-8674(04)00173-4)
224. S. Blandin, E. A. Levashina, Thioester-containing proteins and insect immunity. *Mol. Immunol.* **40**, 903–908 (2004). [Medline doi:10.1016/j.molimm.2003.10.010](https://doi.org/10.1016/j.molimm.2003.10.010)
225. R. H. Baxter, C. I. Chang, Y. Chelliah, S. Blandin, E. A. Levashina, J. Deisenhofer, Structural basis for conserved complement factor-like function in the antimalarial protein TEPI. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11615–11620 (2007). [Medline doi:10.1073/pnas.0704967104](https://doi.org/10.1073/pnas.0704967104)
226. M. A. Osta, G. K. Christophides, F. C. Kafatos, Effects of mosquito genes on *Plasmodium* development. *Science* **303**, 2030–2032 (2004). [Medline doi:10.1126/science.1091789](https://doi.org/10.1126/science.1091789)
227. M. M. Riehle, J. Xu, B. P. Lazzaro, S. M. Rottschaefer, B. Coulibaly, M. Sacko, O. Niare, I. Morlais, S. F. Traore, K. D. Vernick, *Anopheles gambiae* APL1 is a family of variable LRR proteins required for Rel1-mediated protection from the malaria parasite, *Plasmodium berghei*. *PLOS ONE* **3**, e3672 (2008). [Medline doi:10.1371/journal.pone.0003672](https://doi.org/10.1371/journal.pone.0003672)
228. M. Fraiture, R. H. Baxter, S. Steinert, Y. Chelliah, C. Frolet, W. Quispe-Tintaya, J. A. Hoffmann, S. A. Blandin, E. A. Levashina, Two mosquito LRR proteins function as complement control factors in the TEPI-mediated killing of *Plasmodium*. *Cell Host Microbe* **5**, 273–284 (2009). [Medline doi:10.1016/j.chom.2009.01.005](https://doi.org/10.1016/j.chom.2009.01.005)
229. C. Mitri, J. C. Jacques, I. Thiery, M. M. Riehle, J. Xu, E. Bischoff, I. Morlais, S. E. Nsango, K. D. Vernick, C. Bourgooin, Fine pathogen discrimination within the APL1 gene family protects *Anopheles gambiae* against human and rodent malaria species. *PLOS Pathog.* **5**, e1000576 (2009). [Medline doi:10.1371/journal.ppat.1000576](https://doi.org/10.1371/journal.ppat.1000576)
230. M. Povelones, L. M. Upton, K. A. Sala, G. K. Christophides, Structure-function analysis of the *Anopheles gambiae* LRIM1/APL1C complex and its interaction with complement C3-like protein TEPI. *PLOS Pathog.* **7**, e1002023 (2011). [Medline doi:10.1371/journal.ppat.1002023](https://doi.org/10.1371/journal.ppat.1002023)
231. R. M. Waterhouse, M. Povelones, G. K. Christophides, Sequence-structure-function relations of the mosquito leucine-rich repeat immune proteins. *BMC Genomics* **11**, 531 (2010). [Medline doi:10.1186/1471-2164-11-531](https://doi.org/10.1186/1471-2164-11-531)
232. M. Povelones, R. M. Waterhouse, F. C. Kafatos, G. K. Christophides, Leucine-rich repeat protein complex activates mosquito complement in defense against *Plasmodium* parasites. *Science* **324**, 258–261 (2009). [Medline doi:10.1126/science.1175000](https://doi.org/10.1126/science.1175000)
233. Y. Dong, G. Dimopoulos, *Anopheles* fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. *J. Biol. Chem.* **284**, 9835–9844 (2009). [Medline doi:10.1074/jbc.M807084200](https://doi.org/10.1074/jbc.M807084200)
234. Y. Dong, R. Aguilar, Z. Xi, E. Warr, E. Mongin, G. Dimopoulos, *Anopheles gambiae* immune responses to human and rodent *Plasmodium* parasite species. *PLOS Pathog.* **2**, e52 (2006). [Medline doi:10.1371/journal.ppat.0020052](https://doi.org/10.1371/journal.ppat.0020052)
235. M. Povelones, L. Bhagavatula, H. Yassine, L. A. Tan, L. M. Upton, M. A. Osta, G. K. Christophides, The CLIP-domain serine protease homolog SPCLIP1 regulates complement recruitment to microbial surfaces in the malaria mosquito *Anopheles gambiae*. *PLOS Pathog.* **9**, e1003623 (2013). [Medline doi:10.1371/journal.ppat.1003623](https://doi.org/10.1371/journal.ppat.1003623)
236. J. Volz, H. M. Müller, A. Zdanowicz, F. C. Kafatos, M. A. Osta, A genetic module regulates the melanization response of *Anopheles* to *Plasmodium*. *Cell. Microbiol.* **8**, 1392–1405 (2006). [Medline doi:10.1111/j.1462-5822.2006.00718.x](https://doi.org/10.1111/j.1462-5822.2006.00718.x)
237. J. Volz, M. A. Osta, F. C. Kafatos, H. M. Müller, The roles of two clip domain serine proteases in innate immune responses of the malaria vector *Anopheles gambiae*. *J. Biol. Chem.* **280**, 40161–40168 (2005). [Medline doi:10.1074/jbc.M506191200](https://doi.org/10.1074/jbc.M506191200)
238. C. Barillas-Mury, CLIP proteases and *Plasmodium* melanization in *Anopheles gambiae*. *Trends Parasitol.* **23**, 297–299 (2007). [Medline doi:10.1016/j.pt.2007.05.001](https://doi.org/10.1016/j.pt.2007.05.001)
239. E. G. Abraham, S. B. Pinto, A. Ghosh, D. L. Vanlandingham, A. Budd, S. Higgs, F. C. Kafatos, M. Jacobs-Lorena, K. Michel, An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16327–16332 (2005). [Medline doi:10.1073/pnas.0508335102](https://doi.org/10.1073/pnas.0508335102)
240. K. Michel, A. Budd, S. Pinto, T. J. Gibson, F. C. Kafatos, *Anopheles gambiae* SRPN2 facilitates midgut invasion by the malaria parasite *Plasmodium berghei*. *EMBO Rep.* **6**, 891–897 (2005). [Medline doi:10.1038/sj.embor.7400478](https://doi.org/10.1038/sj.embor.7400478)
241. A. K. Schnitger, H. Yassine, F. C. Kafatos, M. A. Osta, Two C-type lectins cooperate to defend *Anopheles gambiae* against Gram-negative bacteria. *J. Biol. Chem.* **284**, 17616–17624 (2009). [Medline doi:10.1074/jbc.M808298200](https://doi.org/10.1074/jbc.M808298200)
242. T. J. Little, N. Cobbe, The evolution of immune-related genes from disease carrying mosquitoes: Diversity in a peptidoglycan- and a thioester-recognizing protein. *Insect Mol. Biol.* **14**, 599–605 (2005). [Medline doi:10.1111/j.1365-2583.2005.00588.x](https://doi.org/10.1111/j.1365-2583.2005.00588.x)
243. A. Parmakelis, M. Moustaka, N. Poulakakis, C. Louis, M. A. Slotman, J. C. Marshall, P. H. Awono-Ambene, C. Antonio-Nkondjio, F. Simard, A. Caccone, J. R. Powell, *Anopheles* immune genes and amino acid sites evolving under the effect of positive selection. *PLOS ONE* **5**, e8885 (2010). [Medline doi:10.1371/journal.pone.0008885](https://doi.org/10.1371/journal.pone.0008885)
244. M. A. Slotman, A. Parmakelis, J. C. Marshall, P. H. Awono-Ambene, C. Antonio-Nkondjio, F. Simard, A. Caccone, J. R. Powell, Patterns of selection in anti-malarial immune genes in malaria vectors: Evidence for adaptive evolution in LRIM1 in *Anopheles arabiensis*. *PLOS ONE* **2** (e793), e793 (2007). [Medline doi:10.1371/journal.pone.0000793](https://doi.org/10.1371/journal.pone.0000793)
245. D. J. Obbard, J. J. Welch, T. J. Little, Inferring selection in the *Anopheles gambiae* species complex: An example from immune-related serine protease inhibitors. *Malar. J.* **8**, 117 (2009). [Medline doi:10.1186/1475-2875-8-117](https://doi.org/10.1186/1475-2875-8-117)
246. D. J. Obbard, D. M. Callister, F. M. Jiggins, D. C. Soares, G. Yan, T. J. Little, The evolution of TEPI, an exceptionally polymorphic immunity gene in *Anopheles gambiae*. *BMC Evol. Biol.* **8**, 274 (2008). [Medline doi:10.1186/1471-2148-8-274](https://doi.org/10.1186/1471-2148-8-274)
247. T. Lehmann, J. C. Hume, M. Licht, C. S. Burns, K. Wollenberg, F. Simard, J. M. Ribeiro, Molecular evolution of immune genes in the malaria mosquito *Anopheles gambiae*. *PLOS ONE* **4**, e4549 (2009). [Medline doi:10.1371/journal.pone.0004549](https://doi.org/10.1371/journal.pone.0004549)
248. C. Mendes, R. Felix, A. M. Sousa, J. Lamego, D. Charlwood, V. E. do Rosário, J. Pinto, H. Silveira, Molecular evolution of the three short PGRPs of the malaria vectors *Anopheles gambiae* and *Anopheles arabiensis* in East Africa. *BMC Evol. Biol.* **10**, 9 (2010). [Medline doi:10.1186/1471-2148-10-9](https://doi.org/10.1186/1471-2148-10-9)

ACKNOWLEDGMENTS

All sequencing reads and genome assemblies have been submitted to NCBI (umbrella BioProject ID, PRJNA67511). 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 wish to 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 would like to thank the many members of the Broad Institute Genomics Platform and Genome Sequencing and Analysis Program who contributed to sequencing data generation and analysis.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.1249770/DC1
www.sciencemag.org/cgi/content/full/science.1258522/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S25

Tables S1 to S36

References (55–248)

AUTHOR AFFILIATIONS

¹Genome Sequencing and Analysis Program, Broad Institute, 415 Main Street, Cambridge, MA 02142, USA. ²Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, MA 02139, USA. ³The Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, MA 02142, USA. ⁴Department of Genetic Medicine and Development, University of Geneva Medical School, rue Michel-Servet 1, 1211, Geneva, Switzerland. ⁵Swiss Institute of Bioinformatics, rue Michel-Servet 1, 1211 Geneva, Switzerland. ⁶Department of Medical Entomology and Vector Control, School of Public Health and Institute of Health Researches, Tehran University of Medical Sciences, Tehran, Iran. ⁷George Washington University, Department of Mathematics and Computational Biology Institute, 45085 University Drive, Ashburn, VA 20147, USA. ⁸European Molecular Biology Laboratory, European Bioinformatics Institute, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. ⁹National Vector Borne Disease Control Programme, Ministry of Health, Tafea Province, Vanuatu. ¹⁰Department of Public Health and Infectious Diseases, Division of

Parasitology, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy.

¹¹Department of Biological Sciences, Cal Poly Pomona, 3801 West Temple Avenue, Pomona, CA 91768, USA. ¹²Tomsk State University, 36 Lenina Avenue, Tomsk, Russia. ¹³Department of Computer Science and Engineering, Eck Institute for Global Health, 211B Cushing Hall, University of Notre Dame, Notre Dame, IN 46556, USA. ¹⁴Inserm, U963, Immune responses in the malaria vector *A. gambiae*, F-67084, Strasbourg, France. ¹⁵CNRS, UPR9022, Immune responses and development in insects, IBMC, F-67084, Strasbourg, France. ¹⁶Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. ¹⁷Faculty of Medicine, Health and Molecular Science, Australian Institute of Tropical Health Medicine, James Cook University, Cairns 4870, Australia. ¹⁸Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK. ¹⁹Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby BC, V5A1S6, Canada. ²⁰Department of Entomology and Nematology, One Shields Avenue, UC Davis, Davis, CA 95616, USA. ²¹Institut de Recherche pour le Développement, UMR MIVEGEC, 911, Avenue Agropolis, BP 64501 Montpellier, France. ²²Division of Biology, Kansas State University, 271 Chalmers Hall, Manhattan, KS 66506, USA. ²³Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, Hellas, Nikolaou Plastira 100 GR-70013, Heraklion, Crete, Greece. ²⁴Centre of Functional Genomics, University of Perugia, Perugia, Italy. ²⁵Genomics Platform, Broad Institute, 415 Main Street, Cambridge, MA 02142, USA. ²⁶Centre National de Recherche et de Formation sur le Paludisme, Ouagadougou 01 BP 2208, Burkina Faso. ²⁷Program of Genetics, Bioinformatics, and Computational Biology, Virginia Tech, Blacksburg, VA 24061, USA. ²⁸School of Life Sciences, University of Nevada, Las Vegas, NV 89154, USA. ²⁹Department of Medical Research, No. 5, Ziwaka Road, Dagon Township, Yangon 11191, Myanmar. ³⁰Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA. ³¹Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA. ³²Department of Biochemistry, Virginia Tech, Blacksburg, VA 24061, USA. ³³Harvard School of Public Health, Department of Immunology and Infectious Diseases, Boston, MA 02115, USA. ³⁴Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Università degli Studi di Perugia, Perugia, Italy. ³⁵Department of Entomology, Virginia Tech, Blacksburg, VA 24061, USA. ³⁶Computational Evolutionary Biology Group, Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester, M13 9PT, UK. ³⁷Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94143, USA. ³⁸Bioinformatics Research Laboratory, Department of Biological Sciences, New Campus, University of Cyprus, CY 1678, Nicosia, Cyprus. ³⁹Wits Research Institute for Malaria, Faculty of Health Sciences, and Vector Control Reference Unit, National Institute for Communicable Diseases of the National Health Laboratory Service, Sandringham 2131, Johannesburg, South Africa. ⁴⁰National Museums of Kenya, P.O. Box 40658-00100, Nairobi, Kenya. ⁴¹Department of Biology, University of Crete, 700 13 Heraklion, Greece. ⁴²Eck Institute for Global Health and Department of Biological Sciences, University of Notre Dame, 317 Galvin Life Sciences Building, Notre Dame, IN 46556, USA. ⁴³Virginia Bioinformatics Institute, 1015 Life Science Circle, Virginia Tech, Blacksburg, VA 24061, USA. ⁴⁴KEMRI-Wellcome Trust Research Programme, CGMRC, PO Box 230-80108, Kilifi, Kenya. ⁴⁵Department of Entomology, 1140 E. South Campus Drive, Forbes 410, University of Arizona, Tucson, AZ 85721, USA. ⁴⁶Department of Medical Microbiology and Immunology, School of Medicine, University of California Davis, One Shields Avenue, Davis, CA 95616, USA. ⁴⁷Department of Pathobiology, University of Pennsylvania School of Veterinary Medicine, 3800 Spruce Street, Philadelphia, PA 19104, USA. ⁴⁸Regional Medical Research Centre NE, Indian Council of Medical Research, PO Box 105, Dibrugarh-786001, Assam, India. ⁴⁹Department of Biology, New Mexico State University, NM 88003, USA. ⁵⁰Molecular Biology Program, New Mexico State University, Las Cruces, NM 88003, USA. ⁵¹Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L35QA, UK. ⁵²Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN 37235, USA. ⁵³Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA. ⁵⁴Department of Entomology, Texas A&M University, College Station, TX 77807, USA. ⁵⁵Department of Parasitology, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand. ⁵⁶Fundação Oswaldo Cruz, Av Brasil 4365, RJ Brazil. ⁵⁷Instituto de Medicina Social, UERJ, Rio de Janeiro, Brazil. ⁵⁸School of Informatics and Computing, Indiana University, Bloomington, IN 47405, USA. ⁵⁹Department of Physiology, School of Medicine, CIMUS, Instituto de Investigaciones, Sanitarias, University of Santiago de Compostela, Spain. ⁶⁰Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. ⁶¹School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, L3 3AF, UK. ⁶²Department of Cellular Biology, University of Georgia, Athens, GA 30602, USA. ⁶³Department of Computer Science, Harvey Mudd College, Claremont, CA 91711, USA. ⁶⁴Program in Public Health, College of Health Sciences, University of California, Irvine, Hewitt Hall, Irvine, CA 92697, USA. ⁶⁵Malaria Programme, Wellcome Trust Sanger Institute, Cambridge, CB10 1SJ, UK. ⁶⁶Centre of Evolutionary and Ecological Studies (CEES-MarECon group), University of Groningen, Nijenborgh 7, NL-9747 AG Groningen, Netherlands. ⁶⁷Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Ave, Cambridge, MA 02138, USA. ⁶⁸Department of Biology, Indiana University, Bloomington, IN 47405, USA. ⁶⁹Centers for Disease Control and Prevention, 1600 Clifton Road NE MSG49, Atlanta, GA 30329, USA. ⁷⁰Biogen Idec, 14 Cambridge Center, Cambridge, MA 02142, USA. ⁷¹Laboratory of Malaria and Vector Research, NIAID, 12735 Twinbrook Parkway, Rockville MD 20852, USA. ⁷²Departments of Biological Sciences and Pharmacology, Institutes for Chemical Biology, Genetics and Global Health, Vanderbilt University and Medical Center, Nashville, TN 37235, USA.

9 July 2014; accepted 14 November 2014

Published online 27 November 2014

10.1126/science.1258522

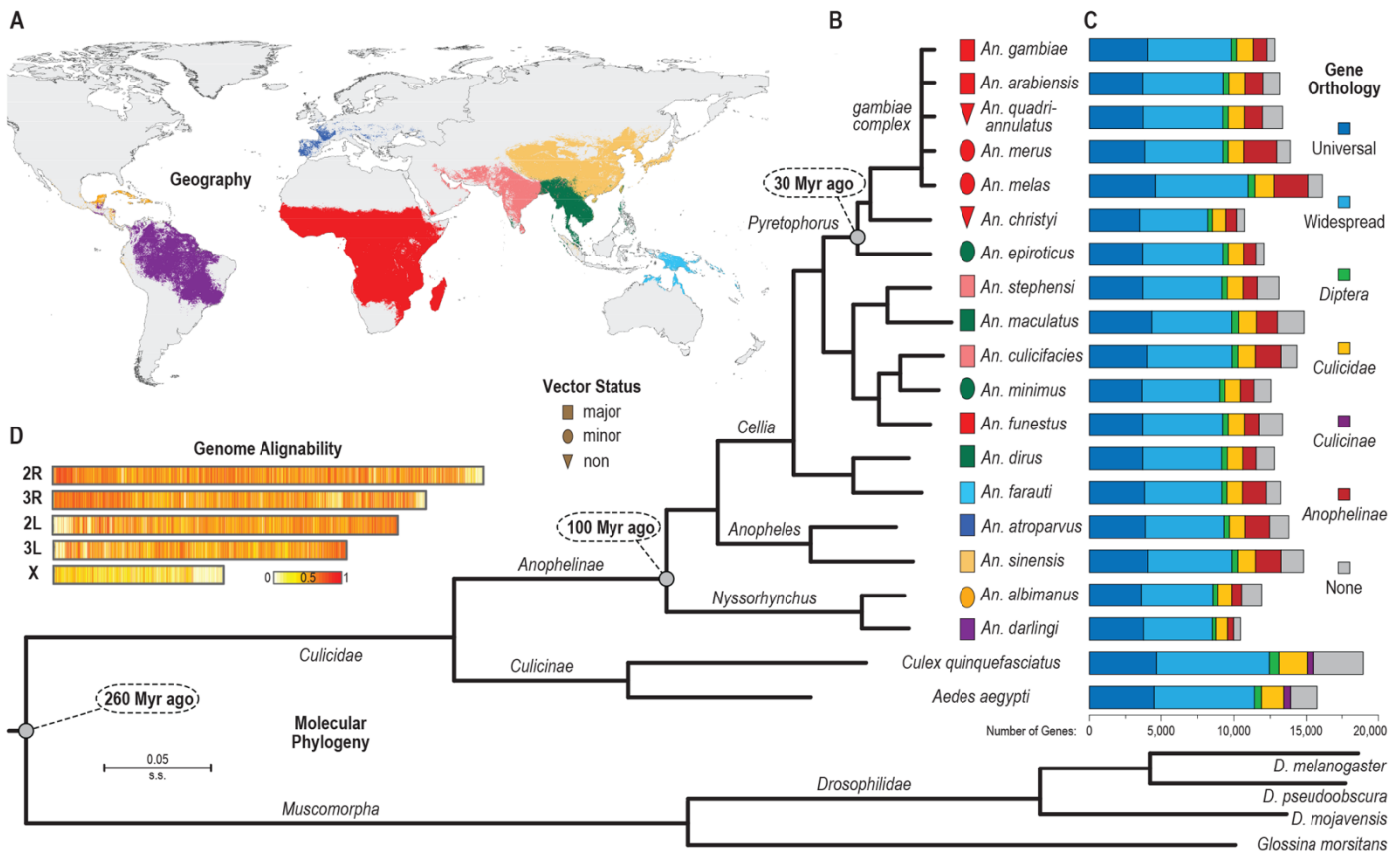


Fig. 1. Geography, vector status, molecular phylogeny, gene orthology, and genome alignability of the 16 newly sequenced anopheline mosquitoes and selected other dipterans. (A) Global geographic distributions of the 16 sampled anophelines and the previously sequenced *An. gambiae* and *An. darlingi*. Ranges are colored for each species or group of species as shown in panel 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, non-vectors) and are colored according to geographic ranges shown in panel A. (C) Barplots 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.

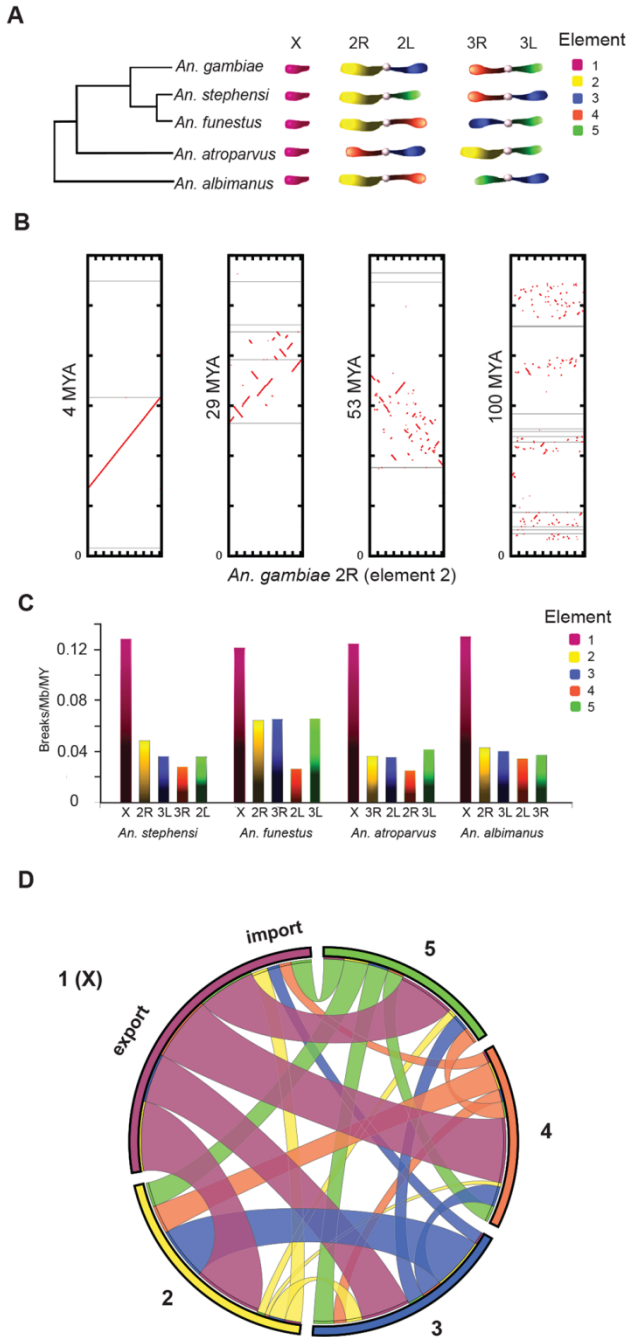


Fig. 2. Patterns of anopheline chromosomal evolution. (A) Anopheline genomes have conserved gene membership on chromosome arms (“elements”; colored and labeled 1-5). Unlike *Drosophila*, chromosome elements reshuffle between chromosomes via translocations as intact elements, and do not show fissions 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 dotplot 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 timescales (MYA, million years ago) estimated via an ultrametric phylogeny. Gray horizontal lines represent scaffold breaks. Discontinuity of the red lines/dots indicates rearrangement. (C) Anopheline X chromosomes exhibit higher rates of rearrangement ($P < 1 \times 10^{-5}$), measured as breaks per megabase (Mb) per million years (MY), 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 panels 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.

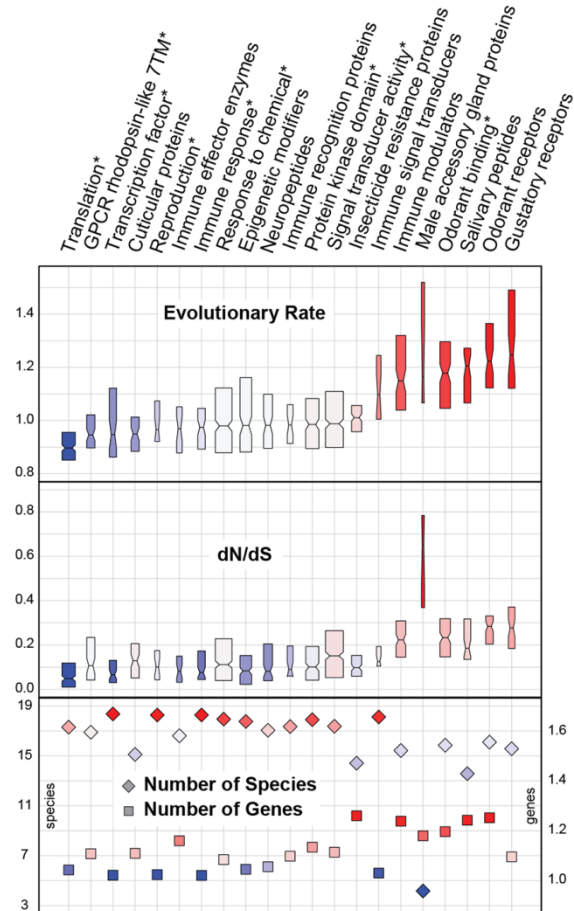


Fig. 3. Contrasting evolutionary properties of selected gene functional categories. Examined evolutionary properties of orthologous groups of genes include: a measure of amino acid conservation/divergence (evolutionary rate), a measure of selective pressure (d_N/d_S), a measure of gene duplication in 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 boxplots show medians, extend to the first and third quartiles, and 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).

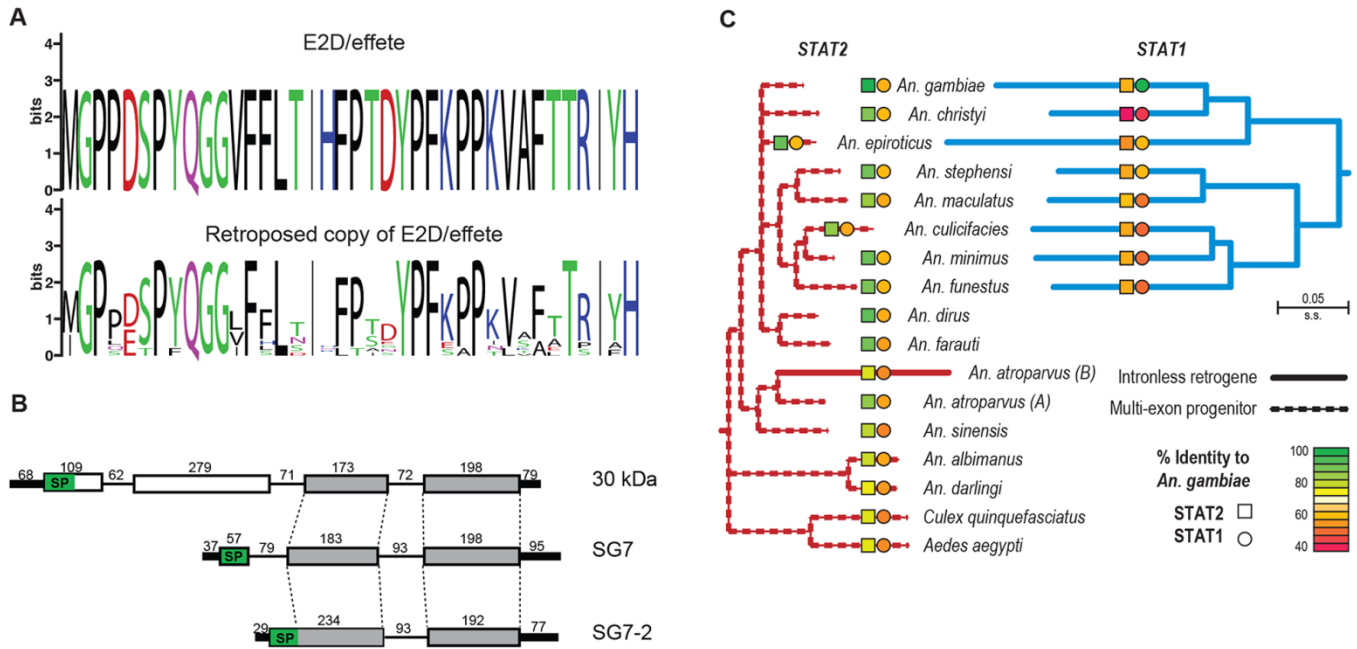


Fig. 5. Genesis of novel anopheline genes. (A) Retrotransposition of the *E2D/effete* gene generated a ubiquitin-conjugating enzyme at the base of the genus, which exhibits much higher sequence divergence than the original multi-exon gene. WebLogo plots contrast the amino acid conservation of the original *effete* gene with the diversification of the retrotransposed copy (residues 38-75; 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 kDa gene. SG7 then underwent tandem duplication and intron loss to generate another salivary protein, SG7-2. Numerals indicate length 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 multi-exon progenitor, *STAT2*, and has been maintained in all descendent species. An independent retrotransposition event created a retrogene copy in *An. atroparvus*, which is also more divergent than its progenitor.