Trends in Parasitology

CellPress
REVIEWS

Feature Review

Of Genes and Genomes: Mosquito Evolution and Diversity

Livio Ruzzante [1,2] Maarten J.M.F. Reijnders [1,2] and Robert M. Waterhouse [1,2,*]

Mosquitoes are widely despised for their exasperating buzzing and irritating bites, and more poignantly because, during blood-feeding, females may transmit pathogens that cause devastating diseases. However, the ability to transmit such viruses, filarial worms, or malaria parasites varies greatly amongst the ~3500 recognised mosquito species. Applying omics technologies to sample this diversity and explore the biology underlying these variations is bringing increasingly greater resolution that enhances our understanding of mosquito evolution. Here we review the current status of mosquito omics, or 'mozomics', resources and recent advances in their applications to characterise mosquito biology and evolution, with a focus on the intersection of evolutionary and functional genomics to understand the putative links between gene and genome dynamism and mosquito diversity.

Highlights

Genomic sampling of mosquito diversity has greatly improved in recent years and looks set to take advantage of emerging technologies to explore even further.

Evolutionary genomics analyses have unveiled dynamic patterns of gene and genome evolution likely linked to mosquito adaptability that will guide future research and control efforts.

Functional genomics assays have helped to characterise biological roles of thousands of genes, albeit with condition-patchy and species-biased coverage that is now starting to be remedied.

Comparative genomics approaches are increasingly being applied to contextualise and enhance the interpretation of results from multispecies studies with an evolutionary perspective.

These trends mean that effective data sharing will be critical to facilitate future integrative meta-analyses and fully harness the benefits of combined evolutionary and functional analyses.

Mozomics

The application of omics technologies, including genomics, transcriptomics, and proteomics, to the characterisation of molecular, cellular, and organismal biology has blossomed over the last two decades. As vectors of a plethora of human pathogens, mosquitoes have been the focus of many pioneering efforts to take advantage of the scale and throughput of these new omics technologies. The resulting nascent sequenced mosquito genomes represent a logical framework for building comprehensive knowledge bases that support and drive biological research. This framework then provides opportunities to explore the vast variety of animal biology through comprehensive assessments of the roles of different functional genomic elements throughout the mosquito life cycle. For example, transcriptomics or proteomics studies to determine in which tissues, and in response to what stimuli, different genes are expressed and subsequently translated into proteins. These approaches have understandably often focused on key biological processes associated with the capacity to transmit pathogens or with thwarting control efforts.

Increasingly, such investigations have been enhanced with an evolutionary perspective through combining functional assays with population and comparative genomics approaches to advance both basic and translational understanding of mosquito biology and biodiversity. This evolutionary perspective has been made possible by the recent genomic sampling of multiple mosquito species, but this still represents only a small fraction of species and is strongly biased towards the genus *Anopheles*. Current and future efforts promise to radically increase species sampling and improve the taxonomic balance, making it timely to consider what we have learnt so far and where this might lead. Here we review the current status of available mosquito genomic and transcriptomic resources and the major advances in mosquito omics – mozomics – applications they have facilitated. We focus on evolutionary characterisations of genomes and their encoded gene repertoires, and on functional genomics explorations of the biological roles of both protein-coding and non-protein-coding genes. While these

[1]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland
[2]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

*Correspondence:
robert.waterhouse@unil.ch
(R.M. Waterhouse).

 https://doi.org/10.1016/j.pt.2018.10.003

mozomics resources may be currently rather species-imbalanced, integrative evolutionary and functional genomics studies are already successfully informing our understanding of the putative links between dynamic gene and genome evolution and mosquito diversity.

Evolutionary Genomics: Characterising Conserved and Divergent Features

Mosquito Diversity and Genomic Sampling

The initial sequencing, assembly, and annotation of the first mosquito genome, that of the primary African malaria vector, *Anopheles gambiae* [1], provided the very first opportunity to explore the use of comparative genomics approaches to investigate insect biology. These pioneering efforts were by necessity limited to comparisons with the fruit fly, *Drosophila melanogaster*, as the only other insect with a sequenced genome at the time. While broad chromosomal-arm-level homology was still evident, and about half of the genes appeared to be maintained as **single-copy orthologues** (see Glossary), extensive differences between these two dipterans indicated much more rapid genomic divergence in insects compared with vertebrates [2]. These first pairwise comparisons laid the foundations for many subsequent advances in applying evolutionary genomics approaches to multispecies analyses as genome sequencing improved the sampling of insect diversity. From the ~3500 species of mosquitoes [3], genomic sampling then focused on the Culicinae with the primary vector for yellow and dengue fevers, *Aedes aegypti* [4], and the vector of West Nile and Saint Louis encephalitis viruses and filarial worms, *Culex quinquefasciatus* [5]. The larger genomes of these culicine mosquitoes revealed high levels of **gene turnover** with notable expanded gene repertoires of zinc-finger and insect pheromone-binding proteins, olfactory and gustatory receptors, as well as genes involved in immune responses and resistance to insecticides.

Comparative evolutionary genomics across the three major mosquito genera therefore established the baseline expectation for conserved or divergent and common or unique features that characterise mosquito genome biology. Recent efforts have increased genomic sampling with a focus on the Anophelinae and in particular on those most closely related to *An. gambiae* (Figure 1, Key Figure). The National Centre for Biotechnology Information (NCBI) Genome database currently lists a total of 28 genome projects for mosquitoes, a mere 1.7% of the 1657 species reported in the NCBI's Taxonomy database or 0.8% of the 3556 species currently recognised by the Mosquito Taxonomic Inventory[i]. The published assemblies and their associated annotations are hosted by VectorBase[ii] (Table 1), a National Institute of Allergy and Infectious Diseases Bioinformatics Resource Centre that provides the scientific community with tools to access and analyse genomic and associated data for invertebrate vectors of human pathogens [6]. Small though the current sampling of mosquito diversity might be, these resources have facilitated detailed characterisations of the evolutionary dynamics of mosquito genomes and their encoded genes, which offer insights into their adaptability to changing environments as well as their coevolution with the pathogens they transmit [4,5,7–9].

Dynamic Mosquito Genome Evolution

Concerted efforts to develop draft genome assemblies for representative *Anopheles* mosquitoes [9–12] facilitated the first detailed characterisations of mosquito genome evolution at nucleotide-level resolution across the genus. The sequenced species span a range of evolutionary distances from *An. gambiae*, are native to Africa, Asia, Australasia, Europe, South and Central America, exploit different ecological conditions, and display varying degrees of **vectorial capacity**. Short-read sequencing technologies with libraries of different insert sizes for species from laboratory colonies and wild-caught specimens produced 170–290 megabase-pair (Mbp) assemblies with variable levels of contiguity. Maintained and made available through VectorBase, these mozomics resources enable multispecies comparative analyses, from

Glossary

Bootstrap: a technique in phylogenetics used to estimate the support for the branching relationships presented on a phylogenetic tree that is based on recomputing the phylogeny using random sampling with replacement from the full multiple sequence alignment.

Chromosome quotient analysis: an approach used to discover Y chromosome sequences, based on the ratio of the number of aligned reads from female and male samples to reference assembly sequences where autosomal sequences have quotients distributed around one, X sequences are around two, and most Y sequences are near zero.

Dosage compensation: processes that balance the expression of sex-linked genes in the different sexes, for example, in XX/XY systems by inactivation of one X chromosome in the homogametic sex or by doubling X chromosome transcription in the heterogametic sex.

Gene model: a representation of the linear structure of an RNA transcript of a gene in its genomic context that describes its features such as start and stop codons, exon–intron boundaries, splice sites, untranslated regions, etc. that may include alternative transcripts of a single gene.

Gene turnover: quantification of gene gain and loss (also called gene birth and death) rates through comparative analyses of homologous genes in extant species and estimations of ancestral gene contents since the last common ancestor.

Hi-C: a technique that employs high-throughput sequencing with chromosome conformation capture protocols to estimate physical interaction frequencies between pairs of genomic loci; as contacts are more frequent for closely linked loci, this information can be used to order and orient draft scaffolds along each chromosome.

Introgression: also known as gene flow or introgressive hybridization. Introgression is the incorporation of genomic loci carrying alleles from one species into the genome, and hence the gene pool, of a second

individual genes and gene families to whole genomes, which also greatly benefit from extensive integration with results from functional and population genomics studies, building knowledge bases to drive new discoveries.

The *Anopheles* 16 genomes project exemplifies the power of multispecies comparisons [9,13]. Characterising conservation across the phylogeny in terms of maintained orthologous genomic regions, or **synteny**, showed that at the level of whole chromosomal arms synteny was highly conserved, which contrasted with extensive within-arm genome shuffling. Such rearrangements were much more frequent on the X chromosome than the autosomes, a trend also observed in fruit flies [14], but that appears greatly elevated in mosquitoes. Where inter-arm translocations were observed, the X chromosome again stood out, revealing a notably greater exodus of sequences to the autosomes. This could be linked to the possible effects of **dosage compensation**, transcriptional suppression, or high repeat content of the X chromosome [15–18] that may make it an inhospitable habitat for some genes. The more elusive *Anopheles* Y chromosomes required the application of new sequencing technologies to tackle the challenges of sequencing through repeat-rich heterochromatic regions [19]. Comparative analyses of these regions revealed superdynamic remodelling of the Y chromosome amongst closely related members of the *An. gambiae* species complex, with only a single gene found to be exclusively Y-linked in all examined species. This male-determining factor (M-factor) gene, called *YG2* or *YOB*, regulates male-specific splicing of the *doublesex* gene in early embryos where microinjection of *YOB* transcripts resulted in male-only broods and *YOB* silencing was male-lethal [20]. Despite their striking differences, hybrids with an *An. gambiae* Y chromosome in an *Anopheles arabiensis* background showed no substantial fitness reductions, indicating that hybrid incompatibility is unlikely to be due to loci on the Y chromosome and thus gene flow may be possible, albeit rare [21]. Interestingly, a candidate M-factor gene in *Anopheles stephensi*, *GUY1* [22], is the same length and has a similar secondary structure as *YOB* but shows no clear sequence homology. The comparative characterizations of *Anopheles* Y chromosomes offer important insights into male mosquito genetics and biology, critical to the successful development of Y chromosome-based control strategies. Ongoing and future efforts to unlock the secrets of these recalcitrant Y chromosomes will extend the current view of the *An. gambiae* species complex to other members of the genus.

In *Aedes* and *Culex* mosquitoes, sex determination is instead controlled by a dominant M-factor from a region called the M-locus on the otherwise homomorphic chromosome 1. In the hunt for the M-factor in *Ae. aegypti*, separate sequencing of male and female genomic DNA identified *myo-sex* as a male-biased gene that is tightly linked to the M-locus [23]. **Chromosome quotient analysis** with reassembled male sequencing data subsequently identified the M-factor *Nix*, a gene homologous to the splicing factor *transformer-2* [24]. *Nix* is persistently linked to the M-locus and is expressed in early male embryos, and somatic knockout resulted in feminization of males while ectopic expression in females led to masculinizing effects. Genetic analysis of the region tightly linked to the M-locus showed that rare crossing-over events caused sex ratio distortions through sex-specific lethal effects [25]. This suggested that there are several factors within the neighbourhood of the M-locus that may be lost or gained through recombination, causing lethality of either males or females. Although the M-locus itself is relatively small, it sits in a much larger sex-differentiated genomic region likely driven by reduced male recombination [26]. Exome sequencing of different populations also identified elevated levels of male/female diversity in the neighbourhood of the M-locus for the type form *Ae. aegypti aegypti* [27]. However, this was spread across the entire chromosome 1 in a Senegalese population, possibly driven by chromosomal rearrangements in this proposed cryptic subspecies. The lag between the release in 2007 of the *Ae. aegypti* draft genome and these key

species usually via hybridization and backcrossing.
Neofunctionalization: the process of functional divergence after gene duplication where one copy acquires a new biological role and the other maintains the ancestral function.
Phylogenomic analysis: refers to analyses that employ large-scale genomics data for evolutionary reconstructions, often specifically the estimation of species phylogenies using sequence alignments of whole genomes or large sets of orthologues.
Reticulate evolution: refers to speciation with natural hybridization events that give rise to lineages through the partial combining of ancestral lineages, meaning that a phylogenetic network provides a better description of the resulting species relationships than does a bifurcating tree.
Single-copy orthologues: orthologues are genes in extant species that have arisen by vertical descent from a single gene of the last common ancestor; when their evolutionary histories do not involve any gene duplication events then they are referred to as single-copy orthologues.
Synteny: in comparative genomics, this refers to the maintained order of blocks of orthologous genes or genomic regions between or amongst genomes of different species (note that in classical genetics it refers to the physical colocalisation of genetic loci in an individual or species).
Vectorial capacity: in general terms, describes a vector's capability to transmit a given pathogen to a host, which depends on many behavioural, ecological, or physiological factors such as feeding habits and host preferences, or longevity, or the effectiveness of its immune system.

Table 1. Sequenced Mosquito Genomes Available from VectorBase (VB-2018-06)

Genus	Subgenus	Species	Notes	Refs
Aedes	*Stegomyia*	*aegypti*	Liverpool strain, now chromosomal-level assembly (Liverpool AGWG strain), and cell line Aag2 assembly	Nene *et al*. 2007 [4] Dudchenko *et al*. 2017 [28] Matthews *et al*. 2017 [29] Whitfield *et al*. 2017 [95]
	Stegomyia	*albopictus*	Foshan (Chinese) and Rimini (Italian) strains, and cell line C6/36 assembly	Chen *et al*. 2015 [40] Dritsou *et al*. 2015 [110] Miller *et al*. 2018 [102]
Culex	*Culex*	*quinquefasciatus*	Johannesburg strain, now chromosomal-level assembly	Arensburger *et al*. 2010 [5] Dudchenko *et al*. 2017 [28]
Anopheles	*Nyssorhynchus*	*albimanus*	STECLA strain (El Salvador), now chromosomal-level assembly	Neafsey *et al*. 2015 [9] Artemov *et al*. 2017 [99]
		darlingi	Coari strain (Amazonas, Brazil)	Marinotti *et al*. 2013 [10]
	Anopheles	*atroparvus*	EBRO strain (Spain) now chromosomal-level assembly	Neafsey *et al*. 2015 [9] Artemov *et al*. 2018 [100]
		sinensis	China and SINENSIS (Korea) strains	Zhou *et al*. 2014 [12] Neafsey *et al*. 2015 [9]
	Cellia (*Neomyzomyia*)[a]	*dirus*	WRAIR2 strain (Thailand)	Neafsey *et al*. 2015 [9]
		farauti	FAR1 strain (Papua New Guinea)	
		koliensis	AKwgs3 strain (Papua New Guinea), contig-level assembly	Logue *et al*. 2015 [111]
		punctulatus	APwgs2 strain (Papua New Guinea), contig-level assembly	
		nili	Dinderesso strain (Burkina Faso), contig-level assembly	Peery *et al*. 2011 [112]
	Cellia (*Myzomyia*)[a]	*culicifacies*	A-37 strain (Iran)	Neafsey *et al*. 2015 [9]
		funestus	FUMOZ strain (Mozambique)	
		minimus	MINIMUS1 strain (Thailand)	
	Cellia (*Neocellia*)[a]	*maculatus*	maculatus3 strain (Malaysia)	
		stephensi	Indian and SDA-500 (Pakistan) strains	Jiang *et al*. 2014 [11] Neafsey *et al*. 2015 [9]
	Cellia (*Pyretophorus*)[a]	*coluzzii*	Formerly gambiae M (Mali-NIH) strain	Lawniczak *et al*. 2010 [113]
		gambiae	PEST strain, chromosomal-level assembly Pimperena (Mali) strain	Holt *et al*. 2002 [1] Lawniczak *et al*. 2010 [113]
		arabiensis	Dongola (Sudan) strain	Neafsey *et al*. 2015 [9]
		christyi	ACHKN1017 (Kenya) strain	
		epiroticus	Epiroticus2 (Vietnam) strain	
		melas	CM1001059_A (Cameroon) strain	
		merus	MAF (South Africa) strain	
		quadriannulatus	SANGWE (South Africa) strain	

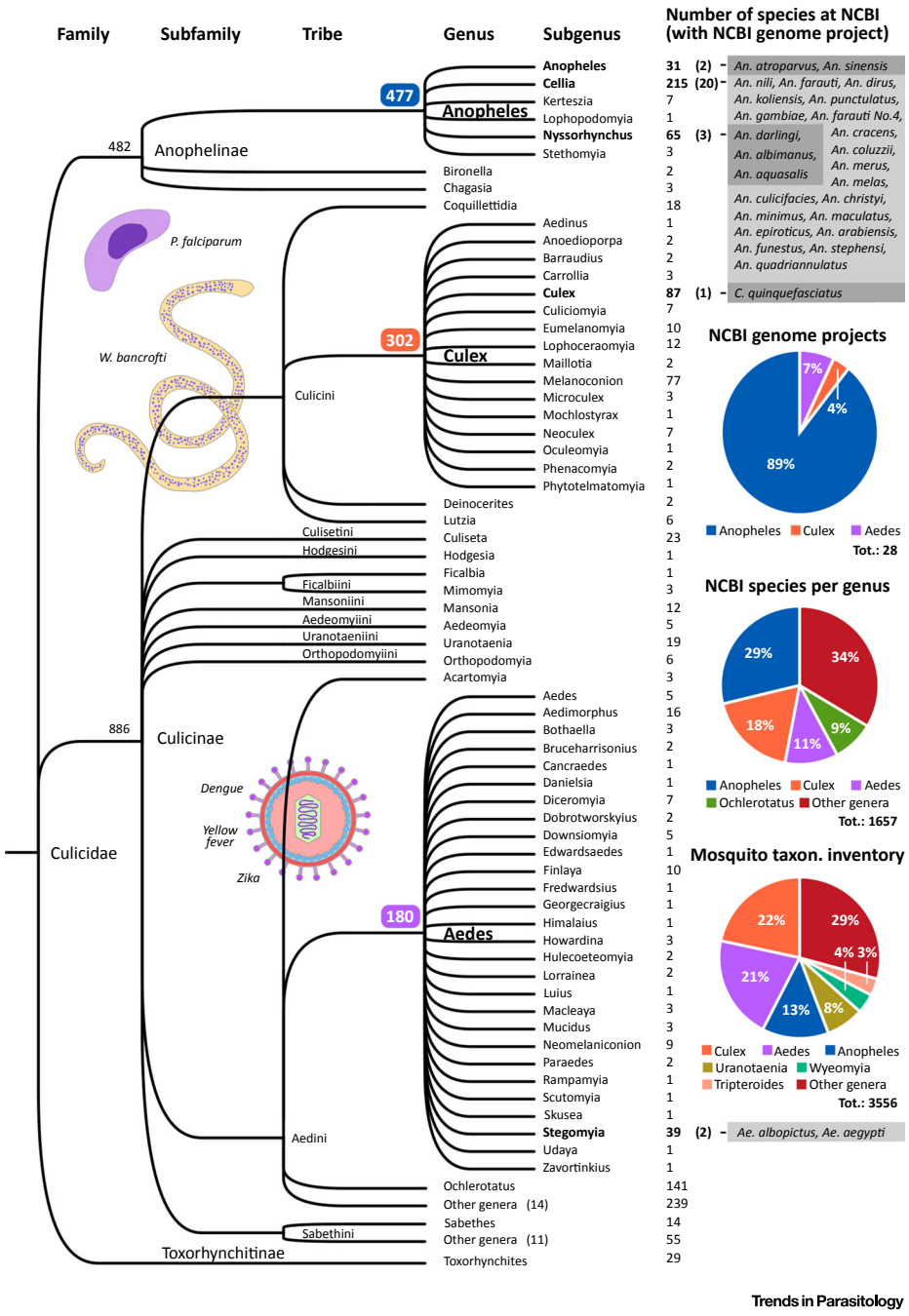
[a]Anopheline *Cellia* species are grouped by series.

discoveries is in part due to the fragmentation of the initial assembly but also because the M-locus was missing. Extensive efforts relying on several complementary approaches, including a **Hi-C** contact map-based chromosome-level assembly [28], have now successfully assembled the M-locus and the female m-locus [29]. The M-factor *Nix* with its 100 Kbp intron and

CellPress REVIEWS

Key Figure

Current View of Mosquito Taxonomic Diversity and Status of Genomic Sampling



(See figure legend on the bottom of the next page.)

neighbouring *myo-sex* gene sit within the highly repetitive ~1.5 Mbp M-locus, localised to the 1p pericentromeric region (1p11) in a ~100 Mbp region of excess sex-differentiation. This now makes it possible to begin to reliably assess the putative evolutionary conflicts that have prevented transformation into heteromorphic sex chromosomes.

Sex chromosomes cannot monopolise all the attention when it comes to evolutionary dynamism. Rather, in terms of gene flow within the *An. gambiae* species complex it is the autosomes that are the most dynamic: whole genome **phylogenomic analysis** uncovered pervasive **introgression** with extensive interspecies genetic exchanges [30]. These dynamics made the task of resolving the correct species branching order extremely challenging, as the extent of gene flow meant that taking a 'majority rule' approach using all alignable sites produced full **bootstrap** support for the wrong branching order. In contrast, using only X chromosome data identified the correct relationships, consistent with the hypothesis that it disproportionately harbours factors responsible for reproductive isolation (e.g., assortative mating genes [31]) and is thus largely resistant to introgression. These complex patterns of evolution, where species boundaries are or have been very fluid, prompt a rethink of the classic concept of a phylogenetic tree: in this case, a network view could be more useful to understand such 'rampant **reticulate evolution**' [32]. Indeed, subsequent application of phylogenetic network methods substantially refined the inference of the extent, direction, and distribution of gene flow across the chromosomes [33]. Furthermore, alternative approaches that take advantage of population resequencing data enabled the detection of putative introgressed regions between sister species [34]. Disentangling complex and sometimes conflicting evolutionary signals to understand the dynamic evolution of the *Anopheles* genomes has thus driven the development and refinement of several phylogenomic analysis methodologies, for example [33–35]. These will greatly facilitate the characterisation of potential gene flow in other species complexes such as that of the *Anopheles funestus* group, and inform control efforts. For example, in South Africa where *An. funestus sensu stricto* has been effectively eradicated but *An. vaneedeni* (genome yet to be sequenced) has recently been identified infected with *Plasmodium falciparum* [36].

Gene Repertoire Evolutionary Dynamics

Investigating mosquito diversity at the level of protein-coding genes has often focused on gene families or biological processes most closely linked to their ability to transmit pathogens. A key part of this is the response of the innate immune system, which must mobilise to clear infections if transmission is to be prevented. Comparative phylogenetic analyses of immune-related genes revealed both conserved and rapidly evolving features between *An. gambiae* and *Ae. aegypti* [7]. Combined functional and evolutionary genomics analyses of vector–pathogen interactions with arboviruses, filarial worms, bacteria, and malaria parasites, identified shared and unique transcriptional responses across three mosquito genera [8]. As infection and transmission occurs during blood-feeding, characterisation of chemosensory genes to identify

Figure 1. For a Figure360 author presentation of Figure 1, see the figure legend at https://doi.org/10.1016/j.pt.2018.10.003.
Species numbers and taxonomic groupings for mosquito families, subfamilies, tribes, genera, and subgenera as defined by the National Centre for Biotechnology Information (NCBI) Taxonomy database, with 28 listed species that have genome projects registered at the NCBI Genome database. Species with available genomic resources reflect their importance as vectors of malaria parasites (e.g., *Plasmodium falciparum*), filarial worms (e.g., *Wuchereria bancrofti*), and viruses (e.g., Dengue, Yellow Fever, and Zika viruses). These mosquito species are sampled from three major genera but represent a mere 1.7% of the 1657 NCBI Taxonomy species or 0.8% of the 3556 species currently recognised by the Mosquito Taxonomic Inventory (valid species list reconciled with the composite *Aedes* genus list), and are overwhelmingly dominated by the Anophelinae. The remaining major genera (shown in the pie charts for those with more than 100 recorded species) that genome sequencing is yet to sample include *Ochlerotatus*, *Uranotaenia*, *Wyeomyia*, and *Tripteroides* (depending on the taxonomy resource). Taxonomies, projects, and species counts retrieved from the NCBI and the Mosquito Taxonomic Inventory in April 2018.

putative molecular mechanisms underlying different host preferences has been another focus of comparative gene repertoire analyses. Here the radically different biology of non-blood-feeding mosquitoes (from the *Toxorhynchites*, *Malaya*, and *Topomya* genera) offers a useful contrast for understanding chemosensation and host-seeking [37], but reference genomes are still lacking (Figure 1). Detoxification gene families have also been intensively studied, principally with respect to their roles in resistance to insecticides but also in potentially enhancing adaptability to polluted urbanised environments. Here evolutionary characterisation is particularly useful, as cross-species analyses can help partition family members into widely maintained orthologues that are likely housekeeping enzymes versus those with more dynamic evolutionary histories possibly linked to species-specific traits [5,38,39].

Greater species sampling and improved reference genomes serve to enhance the resolution of traceable gene evolutionary histories as well as the confidence of their interpretations. For example, multispecies assessments of copy-number variation in homologous gene families across the genus *Anopheles* indicated a gene turnover rate at least five times that observed for drosophilids [9]. Sequence analyses further differentiated conserved genes involved in processes such as translation with the most rapidly diverging functional categories, including odorant receptors (ORs), gustatory receptors (GRs), salivary peptides, and male accessory gland (MAG) proteins. In contrast to their high levels of sequence divergence, chemosensory gene repertoires were found to have been relatively stably maintained across the genus, with a notable exception of a gain of about a dozen ORs in the common ancestor of the *An. gambiae* species complex. Thus, functional divergence of chemosensation in different anophelines appears to be generally better explained by gene sequence evolution (many also showed signatures of positive selection) than by dynamically changing gene repertoires. The culicine ORs also show high levels of sequence divergence, but numerous family expansions have greatly increased their gene repertoires: 117 genes in *Ae. aegypti* (about double that of the anophelines) [29] and possibly even more in the sequenced genomes of *Aedes albopictus* [40] and *C. quinquefasciatus* [5].

Unlike the stably maintained *Anopheles* ORs and GRs, families of salivary gland proteins did show relatively large numbers of gene gains and losses, where the resolution provided by multispecies comparisons, together with expert curation efforts and robust phylogenetic analyses, allowed for gene birth and death events to be pinpointed on the phylogeny [41]. Despite their generally low levels of sequence conservation, many of these salivary gland protein families have identifiable homologues in culicine mosquitoes [41–43]. However, the MAG genes present a more extreme example of evolutionary dynamism: the gene cluster on chromosome arm 3R in *An. gambiae* can be more or less completely identified in other members of the species complex but becomes largely untraceable in more distantly related species [9]. This mirrors dramatic differences in reproductive physiology, where MAG seminal secretions form a highly coagulated mating plug in *An. gambiae* that is less compact in more distantly related anophelines and cannot be detected in the New World species, *Anopheles albimanus* [44], a member of the divergent *Nyssorhynchus* subgenus. Tracing the origins of the gene encoding the *Ae. aegypti* MAG protein HP-I, a peptide that rapidly induces female refractoriness to subsequent mating, showed that it arose from the duplication of the *short neuropeptide F* gene in an *Aedes* ancestor [45]. This represents an intriguing case of putative **neofunctionalization**, with a neuropeptide that still targets its cognate receptor but now as part of the cocktail of seminal fluid proteins delivered during mating that induces many behavioural and physiological changes in females [46,47].

Contrasting patterns of gene repertoire evolution can even be found within the same gene family, where the increased species sampling allows for confident comparisons of trends

observed in different lineages. A striking example is that of the ancient family of *Argonaute* genes, whose protein products are effectors in RNA interference (RNAi)-related pathways. Detailed comparisons across numerous dipteran species revealed strict maintenance of *Ago1* with a single copy in each species; single copies of *Ago2* and *Ago3* in mosquitoes with numerous and rare duplications in other Diptera, respectively; several duplications of the *Piwi/Aubergine* ancestor, particularly in culicine mosquitoes, and a duplication early in the Brachycera radiation that established the separate *Piwi* and *Aubergine* subclades with few subsequent duplications [48]. Detoxification gene families such as cytochrome P450s, glutathione *S*-transferases, and carboxylesterases also exhibit contrasts between members that are almost always conserved as single-copy orthologues and others that duplicate frequently: where gene family expansions were particularly prominent amongst CYP3 and CYP4 clans of P450s in *Ae. aegypti* and *C. quinquefasciatus* [49].

Multispecies resolution also helps to guide hypotheses on the putative functional fates of gene duplicates, as the maintenance of a gene copy in all or most descendant species suggests the acquisition of some fitness advantage. These are exemplified by retrotransposed copies of single-copy progenitor genes, including the signal transducer and activator of transcription *STAT2* and its retrogene copy *STAT1* that emerged in a *Cellia* ancestor after the divergence of the *An. dirus-An. farauti* group [9], and the more recent retrotransposition event in the *Pyretophorus* ancestor that gave rise to the intronless copy of the gene encoding a SET-N chromatin protein [50]. Consistent with potential neofunctionalization of the *STAT1* retrogene, analyses of sequence divergence and polymorphism identified strong signatures of adaptive evolution that contrasted patterns of purifying selection for *STAT2* [51]. Tracing the evolutionary histories of mosquito long-wavelength-sensitive opsins identified at least six ancestral gene duplications that have been maintained in both the anophelines and the culicines, with several amino acid sites exhibiting evidence of adaptive evolution [52]. Population-level resolution offers insights into the evolutionary trajectories of more recent genomic duplications such as an insecticide resistance locus in *An. gambiae* that contains 11 genes, including the acetylcholinesterase *ace-1* gene. Here it appears that a variant of this duplicated locus, where all genes but *ace-1* have been deleted, is spreading through West Africa [53]: an elegant solution to balancing the benefits of the resistance-conferring *ace-1* and the presumed dosage-related fitness costs associated with the other ten duplicated genes.

Importantly, confident reconstructions of gene family evolutionary histories to characterise such events require close inspection of the sequence data to first correct any **gene model** errors and identify family members potentially missed by automated genome annotation pipelines. This is particularly relevant for dynamically evolving families where computational predictions are more likely to be inaccurate and incomplete. For example, curation of the large family of ionotropic receptors (IRs) in the new *Ae. aegypti* assembly corrected several existing models but most remarkably also identified 54 novel IRs, and prompted rescrutinising of the *An. gambiae* genome to find 64 previously overlooked IRs [29]. Such curation tasks are greatly facilitated by the tools and resources provided by VectorBase, including highly configurable genome browser views, interactive visualisations of precomputed comparative genomics analyses, and extensive sequence search options [6]. Furthermore, the Apollo [54] genomic annotation editors hosted by VectorBase provide a versatile platform enabling community curation of current genome annotations. This allows improved and/or novel gene models, as well as gene names, descriptions, or other metadata, to be continually incorporated into the official annotation sets for each species, thus enhancing the richness of these key mozomics knowledge bases.

Dynamic gene repertoire evolution need not be limited to gains and losses of entire genes. Instead, the insertions or deletions of introns or exons can alter transcript structures, possibly leading to new translated protein products – for example, the exon and intron losses that occurred during the evolution of the anopheline *salivary gene 7* family [41], or the exon gain in the *doublesex* gene of the culicines [55]. An intriguing and often overlooked phenomenon that can also lead to altered protein products occurs through the process of stop codon read-through, where instead of terminating upon encountering a stop codon the ribosome continues translation in the same frame until reaching a downstream stop codon, thereby producing a longer version of the protein [56]. Standard automated genome annotation tools normally ignore this possibility as confidently identifying putative readthrough regions requires additional supporting evidence. Here evolution comes to the rescue, as patterns of nucleotide substitution frequencies and insertions or deletions (indels) quantified from whole-genome alignments can distinguish protein-coding from non-coding regions, so the continuation of a strong protein-coding signal after a stop codon can be leveraged to locate candidate readthrough regions.

Aligning 20 *Anopheles* assemblies to the *An. gambiae* genome highlighted the dynamic landscape of conserved and divergent genomic regions, with lower alignability along most of the X chromosome and in autosomal centromeres (Figure 2). At basepair level, only ~13% of the non-repetitive genome was alignable to the most distantly related

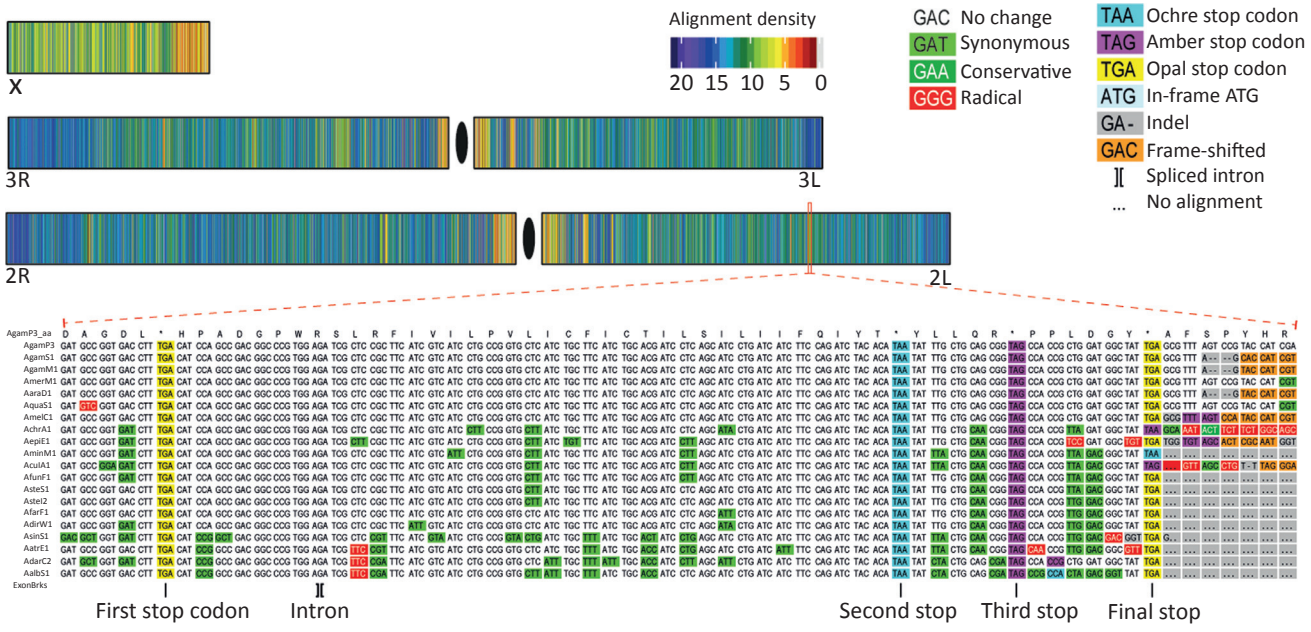


Figure 2. Whole-genome Alignments across the Anophelines Provides a Genome-wide Overview of Nucleotide-level Conservation and Enables the Identification of Stop-codon Readthrough Events. Multiple sequence alignments of 21 *Anopheles* genome assemblies allow for the quantification of conservation shown here as alignability averaged over 2000-basepair windows along each of the five chromosomal arms for the *Anopheles gambiae* genome (adapted from Figure 1D in Neafsey *et al*. 2015 [9]). This detailed view clearly shows the generally reduced alignability of repeat-rich centromeres and the fast-evolving X chromosome. Zooming in to nucleotide-level resolution uncovers numerous genes where in-frame conservation of codons beyond the annotated stop codon strongly supports the occurrence of translational stop-codon readthrough. This may occur when, instead of terminating upon encountering a stop codon, the ribosome continues translation in the same frame until reaching a downstream stop codon, thereby producing a longer version of the protein. The example of an *An. gambiae* rho-related BTB domain-containing protein (AGAP006474) shows a rare case of triple readthrough where in-frame conservation extends through three conserved stop codons before rapidly deteriorating after the final stop codon (adapted from Figure 1C in Jungreis *et al*. 2016 [56]).

Nyssorhynchus species, while 87–92% was alignable to members of the *An. gambiae* species complex [9]. These whole-genome alignments provided the basis for the phylogenomic analysis to characterise patterns of introgression discussed above, and they represent a rich source of comparative mozomics data for sequence-level analysis of the anopheline genomes. For example, examining conservation across the genus helped to select a region putatively under strong functional constraint of the *An. gambiae doublesex* gene for targeted disruption that produced sterile females [57]. With respect to the phenomenon of stop codon readthrough, the whole-genome alignments enabled the identification of evolutionary signatures of conserved, functional readthrough of 353 stop codons in *An. gambiae*, including several cases of double and triple readthrough (Figure 2), and an estimated total number of more than 600 readthrough regions [56]. Although the functional consequences of the alternative protein products that arise from readthrough are largely unknown, it could involve some 5% of all annotated *An. gambiae* protein-coding genes and it therefore certainly contributes to dynamic gene repertoire evolution in mosquitoes.

Functional Genomics: Large-scale Assays of Transcript Abundance

Building high-quality reference genome assemblies with comprehensive and accurate feature annotations facilitates genome-wide evolutionary analyses to build informed hypotheses on putative gene function. Nevertheless, without empirical evidence from detailed molecular biology studies or larger-scale assays of transcript and protein abundance, these hypotheses remain informed guesses. Early assessments of transcript abundance were performed using DNA microarray technologies that included several microarrays with increasingly comprehensive coverage of the gene space [58], as well as specialised resources such as the detoxification chips for *An. gambiae* [59] and *Ae. aegypti* [38]. Decreasing costs of newer technologies mean that RNA sequencing (RNA-seq) has, in many cases, superseded microarrays as the method of choice for transcriptional profiling [60]. As well as the cost incentives, RNA-seq does not rely on having an annotated genome and it can detect transcripts with even relatively low expression levels. As the functional products of expressed protein-coding genes are their correctly translated and folded proteins, proteomics techniques to quantify protein abundance levels can provide important complementary characterisations (reviewed in [61]). High-throughput transcriptomics and proteomics approaches therefore offer wide-ranging opportunities to characterise each gene's 'Where? When? Why?' profile, and thereby its key biological roles.

The current extent of transcriptomic sampling of mosquito species, anatomies, or life stages, and conditions or treatments, is difficult to assess because datasets are often not made publicly available until after publication. This is exacerbated when authors fail to submit their data to public repositories at all, or when metadata describing the species and experiments are incomplete or incorrect. Nevertheless, querying the NCBI's BioProject database provides an approximation ('Culicidae' [Organism] AND 'transcriptome gene expression' [Filter], 18 July 2018), with transcriptomes sampling almost 30 species from 330 projects, three quarters of which are associated with datasets deposited at the NCBI's Short Read Archive (SRA). A handful of species dominate these datasets, with a third of the projects involving *An. gambiae*, 27% *Ae. aegypti*, 11% *Ae. albopictus*, 5% *An. stephensi*, and 4% *C. quinquefasciatus*. While the SRA serves an important role as a public data repository, genomics data integration and interactive visualisation tools are required to build comprehensive knowledge bases that advance biological research. The development and implementation of the VectorBase Expression Browser exemplifies how this can be achieved, offering a queryable repository of transcriptomics data that are processed through a standard normalisation and analysis

pipeline to enable side-by-side visualisations of results from a variety of experimental designs [6]. This mozomics resource focuses on experimental comparisons that investigate patterns of differential gene expression, for example, infected versus non-infected [62–67], male versus female [15,18,24], chemosensory responses [68–70], blood-meal-induced changes [71–73], or effects of salinity stress [74] (Table 2). It currently (release VB-2018-06) integrates transcriptome data from 81 publications spanning the years 2005 to 2018 and including 12 mosquito species with 117 experimental comparisons. The majority of experiments involve *An. gambiae* or *Ae. aegypti*, and the VectorBase-curated experiment metadata offer a summarised view of the anatomies or life stages and conditions or treatments that have been examined (Figure 3). This overview also shows the general paucity of tissue-specific or non-adult life-stage sampling from other mosquitoes, and how broader species sampling for various conditions or treatments is dominated by blood-feeding and insecticide responses. Encouragingly, some of these imbalances will begin to be addressed as recently published transcriptomics data such as those summarised in Table 2 are incorporated into future VectorBase Expression Browser updates.

The numerous *An. gambiae* and *Ae. aegypti* studies represent a rich source of functional genomics data for the meta-analysis of differential gene expression patterns. At VectorBase this is implemented and visualised through dynamically queryable Expression Maps [75] that currently integrate expression data for 12 787 *An. gambiae* genes across 202 conditions and 17 090 *Ae. aegypti* genes from 92 conditions (Box 1, and Figure 4). These Expression Maps exemplify the importance of genomics data integration and interactive visualisation tools to support biological research by revealing coexpression patterns that inform the understanding of gene functions, for example, *LRIM9* and *vitellogenin* [76] or genes involved in melanisation responses [77]. Taking the example presented in Box 1 and Figure 4: the set of *An. gambiae* genes with expression patterns most similar to the known seminal fluid protein-encoding genes *transglutaminase 3* (*TG3*), *plugin*, and six of the MAG group genes includes 21 additional genes in the Expression Map. A preliminary hypothesis might simply assign a role in seminal fluid biology for at least some of these genes. Examining functional clues in more detail would show that at least three of them exhibit clear signatures of signal peptides (suggesting that they are secreted) and serine protease inhibitor-like protein domains. Interestingly these three genes produce proteins that are considerably shorter than typical serine proteases and they lack the normally conserved inhibitory site. A more nuanced hypothesis might therefore postulate that these three genes may have arisen from a partial duplication of a full-length serine protease gene and produce peptides in the seminal fluid that interfere with the activity of female proteases that might otherwise degrade the gelatinous mating plug. In this way, combining knowledge of coexpression patterns with clues from protein domain analyses can support novel hypotheses on gene function.

While transcriptional profiling has generally focused on protein-coding genes, several recent studies have investigated the expression of non-coding RNAs (ncRNAs), including micro-RNAs (miRNAs) [78–82] and long non-coding RNAs [83]. Such datasets enable more comprehensive and accurate cataloguing of these elements and provide insights into their biological roles, which have been generally poorly characterised in comparison with protein-coding genes, despite being key post-transcriptional regulators of protein-coding gene expression with roles in mosquito–pathogen interactions, reviewed in [84]. For example, microarray-based profiling of *An. gambiae* miRNAs characterised differentially expressed miRNAs from heads, fat bodies, midguts, and ovaries [78], and small RNA sequencing identified miRNAs that were modulated by Zika virus infection in *Ae. aegypti* [85]. The rewards from exploiting genomic and transcriptomic data and focusing on non-coding RNAs are

Table 2. Selected Recent (2015–2018) Mosquito Transcriptomics Studies Related to the Biological Themes of Immunity, Reproduction, Olfaction, Blood Feeding, and Salinity

Biological theme	Species	Notes	Refs
Immunity	*Aedes albopictus*	Dengue virus	Tsujimoto *et al*. 2017 [62][a]
	Ae. aegypti	Zika virus	Angleró-Rodríguez *et al*. 2017 [114]
		Microsporidia	Desjardins *et al*. 2015 [63][a]
		Chikungunya virus	Dong *et al*. 2017 [115]
		Zika virus	Etebari *et al*. 2017 [64][a]
		Brugia malayi	Juneja *et al*. 2015 [65][a]
		Dengue virus	Jupatanakul *et al*. 2017 [116]
	Anopheles coluzzii	O'nyong nyong virus	Carissimo *et al*. 2018 [117]
		Microbiota	Rodgers *et al*. 2017 [66][a]
	An. stephensi	Haemocytes	Thomas *et al*. 2016 [118]
		Plasmodium	Zhang *et al*. 2017 [67][a]
Reproduction/ sex-bias/ sex determination	*Ae. aegypti*	Mating-induced	Alfonso-Parra *et al*. 2016 [46]
		Male-determining factor	Hall *et al*. 2015 [24][a]
		Fat body	Roy *et al*. 2015 [119]
		Ovarian-biased genes	Whittle & Extavour 2017 [105]
	An. gambiae, An. arabiensis, An. minimus, An. albimanus	Sex-biased genes	Papa *et al*. 2017 [18][a]
	An. gambiae, An. merus	Testes	Cassone *et al*. 2017 [17]
	An. stephensi	Sex-biased genes	Biedler *et al*. 2015 [120]
		Sex-biased genes	Jiang *et al*. 2015 [15][a]
Olfaction/ chemosensation	*Ae. albopictus*	Olfactory repertoire	Lombardo *et al*. 2017 [68][a]
	An. coluzzii	Labella	Saveer *et al*. 2018 [121]
	An. sinensis	Antennae	Chen *et al*. 2017 [69][a]
	Culex quinquefasciatus	Chemosensory genes	Taparia *et al*. 2017 [70][a]
	Ae. aegypti	Female and male	Ribeiro *et al*. 2016 [42]
	C. tarsalis[b]	Female and male	Ribeiro *et al*. 2018 [43]
Blood feeding	*Ae. albopictus*	Malpighian tubules	Esquivel *et al*. 2016 [71][a]
		Diapause	Huang *et al*. 2015 [72][a]
	An. gambiae	Ivermectin	Seaman *et al*. 2015 [73][a]
	Wyeomyia smithii[b]	Non-biting mosquito	Bradshaw *et al*. 2018 [122]
Salinity	*An. coluzzii, An. merus*	Salinity stress	Uyhelji *et al*. 2016 [74][a]

[a]Included in VectorBase Expression Browser.
[b]No sequenced genome is available yet.

exemplified by the recent advances in understanding the activities of PIWI-interacting RNAs (piRNAs). The piRNA pathway is often recognised chiefly for its role in protecting the genome in germ cells by suppressing transposable element (TE) activity, while the small-interfering RNA (siRNA) pathway is usually considered the main RNAi mechanism responsible for suppressing viral replication. Indeed, in *An. gambiae* abundant piRNA expression was detected in reproductive tissues [86], and many piRNAs were found to target TEs [87].

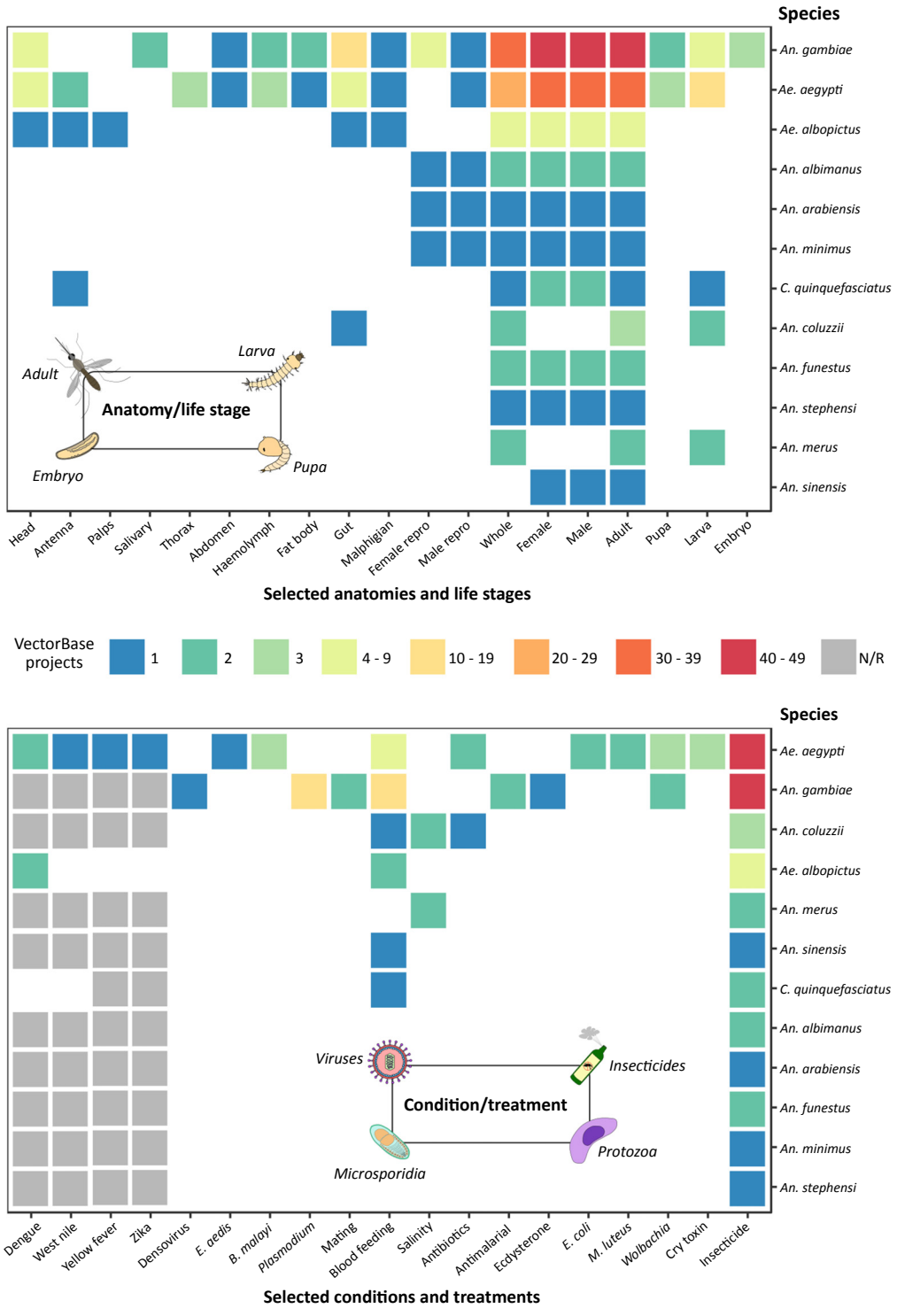


Figure 3. Summary of Transcriptomics Data Available through the VectorBase Expression Browser Highlighting the Most Sampled Species, Tissues, Life Stages, Conditions, or Treatments. Experimental factors for mosquito anatomies or life stages (top) and conditions or treatments (bottom) from

(Figure legend continued on the bottom of the next page.)

Box 1. VectorBase Expression Maps

VectorBase Expression Maps provide a systems-level view of gene expression by integrating normalised gene expression data from multiple experimental conditions to build clusters of genes with similar expression profiles [75]. Clustering employs a self-organising map algorithm with Pearson correlation coefficient-based distance measurements that define gene expression profile similarities. To facilitate data visualisation and browsing, the map is predefined to take the form of a 25 by 20 grid resulting in a total of 500 possible clusters (see Figure 4 in the main text). Self-organising map clustering then proceeds to add genes to the grid such that those with the most similar overall expression profiles will join the same cluster, and neighbouring clusters will usually show similar expression patterns. The resulting map therefore contains regions made up of several clusters that are characterised by significant differential gene expression observed from particular experimental conditions, life-stages, or tissues. Each cluster on the map is annotated with a list of member genes, the conditions for which these genes are most highly upregulated or downregulated, as well as any over-represented Gene Ontology terms. VectorBase currently (release VB-2018-06) provides Expression Maps for *Anopheles gambiae* (202 conditions) and *Aedes aegypti* (92 conditions) as this meta-analysis approach to clustering expression data requires integration across many experiments.

The Expression Maps can be queried with lists to locate clusters that contain matching genes. These lists may comprise gene identifiers or names, Interpro domain or Gene Ontology term identifiers, or annotation keywords or phrases. This allows users to locate clusters containing their gene(s) of interest and subsequently identify genes with the most similar expression profiles; and view the conditions that most strongly define their coexpression patterns. Alternatively; condition-defined queries can be used to identify sets of clusters on the map with genes showing significant upregulation or downregulation for a selected condition. These two types of query can be combined to interrogate the map with questions of biological interest; for example; to identify clusters containing (i) serine proteases that are upregulated in midgut tissues but not in ovaries after blood feeding; (ii) genes with homeobox domains that are upregulated mainly during early versus mid-versus late embryonic development; (iii) detoxification genes that are differentially upregulated or downregulated in insecticide-resistant versus susceptible mosquito strains. In this way; the Expression Maps facilitate exploration of the available data to build hypotheses on gene function informed by coexpression patterns across many different conditions.

However, deep sequencing of small RNAs from culicines discovered abundant viral piRNAs, 24–30 nucleotides in length and distinct from 21-nucleotide virus-derived siRNAs, revealing a key role for the piRNA pathway in the antiviral defences of *Ae. aegypti* [88–90], *Ae. albopictus* [91], and *C. pipiens* [92], reviewed in [93,94]. Furthermore, scanning mosquito genomes for nonretroviral integrated RNA virus sequences (NIRVSs) revealed an abundance of flavivirus and rhabdovirus integrations in *Aedes*, in stark contrast to almost no flavivirus and very few rhabdovirus sequences in *Anopheles* [91]. The genome assembly of the *Ae. aegypti* cell line, Aag2, was also found to be replete with such endogenous viral elements (EVEs), in loci enriched for both long terminal repeat retrotransposons and piRNA clusters [95]. Thus, analogous to the way that the piRNA machinery provides heritable immunity against transposable elements, EVE/NIRVS-derived piRNAs suggest a mechanism for heritable antiviral immune responses in some mosquitoes. Importantly, such a mechanism does not appear to operate in *Drosophila*, where the piRNA pathway does not play a major role in antiviral defence [96]. These recent extensions of mozomics to include ncRNAs have thus uncovered some intriguing lineage-specific novelties. Further functional and comparative genomics investigations of the activities of ncRNAs are thus clearly essential, going beyond the study of protein-coding genes to develop a comprehensive understanding of the roles of dynamic gene and genome evolution in the diversity and success of mosquitoes.

VectorBase-curated transcriptome metadata reconciled into 19 simplified factor-categories each (e.g., 'Head' combines 'head', 'adult head', 'pupal head', and 'head and thorax', and 'Insecticide' combines 23 insecticide-related factors). This expression-sampling overview highlights the historical focus on *Anopheles gambiae* and *Aedes aegypti*, the relative paucity of tissue-specific or non-adult life stage sampling from other mosquitoes, and how broader species sampling for various conditions or treatments is dominated by blood-feeding and insecticide responses. Recent additional studies, several of which are presented in Table 2, will address some of these imbalances when they are incorporated into future releases of the VectorBase Expression Browser. Species with the most-sampled-factor categories are at the top of each plot, and each species and factor-category pair is coloured according to the number of experimental comparisons (projects). Grey boxes indicate species factor-category pairs that are not relevant (N/R), that is, where some viruses are not thought to be able to naturally infect certain mosquito species. Data from VectorBase release VB-2018-06. Abbreviations: *B. malayi*, *Brugia malayi*; *E. aedis*, *Edhazardia aedis*; *E. coli*, *Escherichia coli*; *M. luteus*, *Micrococcus luteus*; Repro. reproductive tissues.

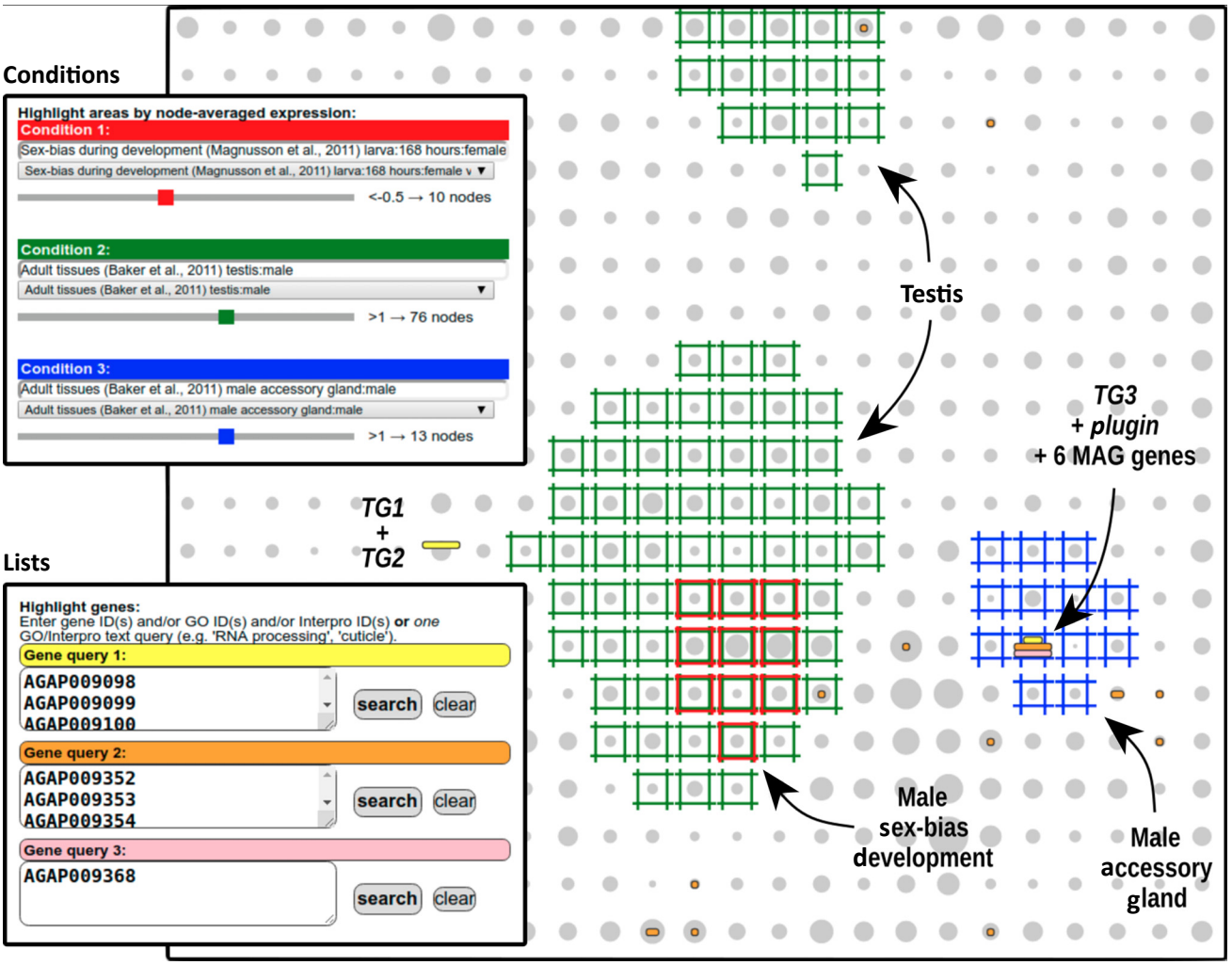


Figure 4. Example Queries Highlighting the Utility of the *Anopheles gambiae* Expression Map at VectorBase for Exploring Mosquito Gene Expression Data. The Expression Map (AgamP4.9 VB-2018-02) shows clusters of genes based on a meta-analysis of expression profiles across numerous experiments in a 25 by 20 grid self-organising map (Box 1). The map highlights regions for the results of three condition-based queries, Condition 1: red, male sex-biased expression during larval development; Condition 2: green, upregulated in testes; Condition 3: blue, upregulated in male accessory glands (MAGs). The results of three gene identifier list-based queries show the cluster-locations of the three *Anopheles gambiae* transglutaminase (TG) genes in yellow, 27 genes in orange that belong to the MAG gene group on chromosome 3R, and the *Plugin* gene in pink. The *TG3* gene is highly expressed in the MAGs, and its protein product is responsible for the formation of a gelatinous mating plug by acting upon the substrate protein, Plugin [103]. The combined queries show that *TG3* is found in an expression cluster of 29 genes characterised by upregulation in the MAGs that also includes six of the MAG group genes, as well as the *Plugin* gene, while *TG1* and *TG2* are clustered together elsewhere on the map. The clusters common to conditions one (red) and two (green) possibly contain genes important during larval development of the testes. Grey filled circles that make up the grid vary in size according to the numbers of genes within each cluster.

Concluding Remarks

Until recently, amongst the sequenced mosquitoes only the *An. gambiae* genome could be considered a chromosomal-level assembly, that is, a near-complete catalogue of the genomic content almost all of which is ordered and oriented along the chromosomal arms. This has certainly frustrated efforts to investigate the dynamics of mosquito genome evolution, something akin to being offered a tantalising glimpse of a beautiful vista but through a shattered window pane. Nevertheless, comprehensive efforts that combine data from complementary technologies offer new possibilities to substantially improve existing draft assemblies and their annotations, as exemplified by the new *Ae. aegypti* genome [29]. This showcases the concept of 'evolving' reference genomes, where assembly and annotation improvements are incorporated from manual curation efforts or from the large-scale integration of genomic, transcriptomic, and proteomic data, for example, for *An. stephensi* [97]. Physical mapping using fluorescent *in situ* hybridization (FISH) to chromosomally localise scaffolds on high-resolution cytogenetic maps [98] has played a key role in assembly improvements, for example, high-resolution cytogenetic photomaps with FISH-mapping together with computational synteny-based analyses have recently produced 98.2% and 89.6% chromosome-anchored *An. albimanus* and *An. atroparvus* reference genomes, respectively [99,100]. The recent application of synteny-based methods to the anopheline genome assemblies, in combination with physical mapping, transcriptomics, and other supporting evidence where available, further demonstrates the variety of approaches for improving these key resources [101]. Alternatively, Hi-C genomic proximity data promise fast, inexpensive, and accurate ordering and orienting of fragmented draft assemblies into chromosome-length scaffolds, as achieved recently for the *Ae. aegypti* and *C. quinquefasciatus* genomes [28]. Furthermore, long-read sequencing data are proving useful for assembling repeat-rich highly heterozygous genomes, for example, the 2.25 Gbp *Ae. albopictus* C6/36 cell line assembly presents both haplotypes (unphased) for most of the diploid genome [102]. While the fact that failed genome sequencing and assembly improvement attempts are not published may skew the outlook somewhat, the successful examples demonstrate what can be achieved. They foretell a future of substantial improvements to current draft assemblies as well as accelerated whole-genome sampling of mosquito diversity.

Though not without its challenges (see Outstanding Questions), this brings many exciting opportunities to characterise genome and gene evolutionary dynamics with increasing detail and confidence. Importantly, it will be imperative to concurrently expand functional genomics sampling of mosquito diversity, in terms of both species and conditions, to interrogate and begin to interpret emerging evolutionary patterns in the context of vector biology. The expression data already collated at VectorBase exemplify the potential benefits that data sharing for enhanced meta-analyses can bring, especially with the support of extensive comparative genomics data exploration and visualisation tools. The recent investigations of ncRNA biology extend these mozomics knowledge bases and reveal intriguing evolutionary twists where it is important not to take everything learnt from the fruit fly as automatically directly applicable to mosquito biology. Importantly, teasing apart conserved roles and lineage-specific innovations will require future interrogations that directly compare ncRNA biology across several mosquito species. The cointerrogation of evolutionary and functional genomics data applies to both focused studies on specific physiological or behavioural activities and larger-scale investigations of key biological processes. For example, *An. gambiae* reproductive biology studies have shown that TG3 enzymes act on Plugin proteins in the formation of mating plugs [103], and comparative evolutionary analyses of transglutaminases identified *TG3* as having arisen through a gene duplication event in an anopheline ancestor and subsequently having undergone rapid sequence divergence [9].

Outstanding Questions

Genomic sampling of mosquito species remains incomplete and biased; will this rectify itself through the 'sequence everything' movement, or is a more directed strategy required?

Rapidly growing amounts of genomic data and associated metadata pose challenges for quality assurance and long-term management; whose responsibility is it to tackle these challenges?

New approaches enable capturing of previously elusive sex-determining loci, so what led to heteromorphic sex chromosomes in anophelines and maintained homomorphic ones in culicines?

Chromosome-level assemblies and population genomics now allow for fine-scale mapping of introgression; can potential routes of gene-flow-mediated insecticide resistance be predicted?

Comparative evolutionary analyses reveal gene repertoire changes across phylogenies, but how can the all-important linking of such variation to variable organismal biology be robustly accelerated?

Gene repertoire changes cannot fully account for the observed variations, so will the delineation of detailed gene regulatory networks help to explain the missing links?

Gene families of particular interest to vector biology are often amongst the most dynamically evolving, so how can confidence of cross-species functional inferences be assured?

Stop-codon readthrough appears to be especially abundant in dipteran insects; what drives this phenomenon, and what are the functional implications of these alternative protein products?

Functional genomics sampling, mostly in the form of transcriptomics, remains very unevenly spread across species and conditions; how should the research community address this imbalance?

Comparative physiological analyses examined mating plugs and levels of 20-hydroxyecdysone (20E) hormone across nine anophelines, the amount of 20E transferred from males to females in five species, and the extent of post-20E-exposure oviposition induction and refractoriness to mating in three [44]. Together, these observations trace the emergence along the mosquito phylogeny and coevolutionary dynamics of important reproductive genes and traits that may be linked to differences in disease-vector capacity observed today. However, examining a larger set of mosquito species (three *Nyssorhynchus*, five *Anopheles*, and eight *Cellia*) suggested that sexually transferred steroids and their effects on female behaviour did not correlate with the transmission of malaria parasites [104]. Recent larger-scale mozomics functional assays are also starting to exploit the available genomes and their annotations to bring an evolutionary perspective to interpreting transcriptomics results from more than a single species, for example, in the context of sex- and tissue-biased gene expression in anophelines [17,18] and culicines [105]. Orthology delineation across mosquitoes allowed for genes with biased sex and tissue transcription patterns to be characterised with a suite of evolutionary metrics, including expression divergence, ratios of nonsynonymous-to-synonymous substitutions, protein sequence divergence levels, proportions of positively selected genes, the extent of gene turnover, and autosome versus X chromosome genomic locations. These implicate shared and distinct genes as well as species-reproducible patterns of evolutionary dynamism in the context of contrasting male and female mosquito biology.

Such integration of evolutionary and functional genomics analyses might appear a self-evident avenue of investigation when it comes to studying processes like sex and reproduction. However, with 24 sequenced and annotated mosquito genomes now available at VectorBase, similar integrative approaches can and should be applied to any focused studies or larger-scale investigations of gene function. Just as robust results rely on reproducibility across biological replicates, cross-species comparisons can provide 'evolutionary replicates' that enhance interpretations of the results by partitioning genes into evolutionarily stable or dynamic functional modules. Such stability or dynamism will be variable for different biological processes and across different evolutionary timescales, for example, even within the *An. gambiae* species complex immune resistance and tolerance can vary substantially [106], and even inversion polymorphisms can strongly influence organismal traits and transcription patterns [107]. Along the stable-to-dynamic spectrum, phenotypes are substantially more variable than proportions of species- or lineage-specific genes might suggest, clearly indicating the presence of additional tiers of complexity. One of these, which remains largely unexplored in mosquitoes to date, is variation in terms of the regulatory networks that govern each gene's 'Where? When? Why?' profile, delineation of which would greatly benefit from cataloguing transcription factor binding sites as for the fruit fly and the nematode worm [108]. Another, which genome sequencing is now beginning to explore in impressive detail, is variation in terms of the very high levels of genetic diversity observed in natural mosquito populations [109]. Mozomics assays like these, together with extended species sampling, will help to harness the power of combined evolutionary and functional genomics approaches to advance both basic and translational understanding of mosquito evolution, biology, and biodiversity.

The roles of non-coding RNA genes are deservedly receiving more attention; will these elements and their mechanisms of action produce more surprises that subvert dogma from *Drosophila*?

Accumulating genomics data offer new opportunities for integrative multi-species analyses, but how can we effectively cointerrogate function and evolution, and truly become better at meta?

Acknowledgments

The authors acknowledge support from the Swiss National Science Foundation (PP00P3_170664 to RMW), feedback on draft subsections of this manuscript from Mariangela Bonizzoni, Robert M. MacCallum, Daniel E. Neafsey, and Philippos A. Papathanos, feedback on the draft manuscript from Romain Feron, and constructive comments and critiques from the anonymous reviewers.

Resources

[i]http://mosquito-taxonomic-inventory.info

[ii]www.vectorbase.org

References

1. Holt, R.A. *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129–149
2. Zdobnov, E.M. *et al.* (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* 298, 149–159
3. Harbach, R.E. and Besansky, N.J. (2014) Mosquitoes. *Curr. Biol.* 24, R14–R15
4. Nene, V. *et al.* (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316, 1718–1723
5. Arensburger, P. *et al.* (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science* 330, 86–88
6. Giraldo-Calderón, G.I. *et al.* (2015) VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.* 43, D707–D713
7. Waterhouse, R.M. *et al.* (2007) Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* 316, 1738–1743
8. Bartholomay, L.C. *et al.* (2010) Pathogenomics of *Culex quinquefasciatus* and meta-analysis of infection responses to diverse pathogens. *Science* 330, 88–90
9. Neafsey, D.E. *et al.* (2015) Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science* 347, 1258522–1258522
10. Marinotti, O. *et al.* (2013) The genome of *Anopheles darlingi*, the main neotropical malaria vector. *Nucleic Acids Res.* 41, 7387–7400
11. Jiang, X. *et al.* (2014) Genome analysis of a major urban malaria vector mosquito, *Anopheles stephensi*. *Genome Biol.* 15, 459
12. Zhou, D. *et al.* (2014) Genome sequence of *Anopheles sinensis* provides insight into genetics basis of mosquito competence for malaria parasites. *BMC Genomics* 15, 42
13. Neafsey, D.E. *et al.* (2013) The evolution of the *Anopheles* 16 genomes project. *G3* 3, 1191–1194
14. Bhutkar, A. *et al.* (2008) Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics* 179, 1657–1680
15. Jiang, X. *et al.* (2015) Complete dosage compensation in *Anopheles stephensi* and the evolution of sex-biased genes in mosquitoes. *Genome Biol. Evol.* 7, 1914–1924
16. Rose, G. *et al.* (2016) Dosage compensation in the African malaria mosquito *Anopheles gambiae*. *Genome Biol. Evol.* 8, 411–425
17. Cassone, B.J. *et al.* (2017) Comparative transcriptomics of malaria mosquito testes: function, evolution, and linkage. *G3* 7, 1127–1136
18. Papa, F. *et al.* (2017) Rapid evolution of female-biased genes among four species of *Anopheles* malaria mosquitoes. *Genome Res.* 27, 1536–1548
19. Hall, A.B. *et al.* (2016) Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. *Proc. Natl. Acad. Sci. U. S. A.* 113, E2114–E2123
20. Krzywinska, E. *et al.* (2016) A maleness gene in the malaria mosquito *Anopheles gambiae*. *Science* 353, 67–69
21. Bernardini, F. *et al.* (2017) Cross-species Y chromosome function between malaria vectors of the *Anopheles gambiae* species complex. *Genetics* 207, 729–740
22. Criscione, F. *et al.* (2016) GUY1 confers complete female lethality and is a strong candidate for a male-determining factor in *Anopheles stephensi*. *eLife* 5, e19281
23. Hall, A.B. *et al.* (2014) Insights into the preservation of the homomorphic sex-determining chromosome of *Aedes aegypti* from the discovery of a male-biased gene tightly linked to the M-locus. *Genome Biol. Evol.* 6, 179–191
24. Hall, A.B. *et al.* (2015) A male-determining factor in the mosquito *Aedes aegypti*. *Science* 348, 1268–1270
25. Krzywinska, E. *et al.* (2016) The sex locus is tightly linked to factors conferring sex-specific lethal effects in the mosquito *Aedes aegypti*. *Heredity (Edinb)* 117, 408–416
26. Fontaine, A. *et al.* (2017) Extensive genetic differentiation between homomorphic sex chromosomes in the mosquito vector, *Aedes aegypti*. *Genome Biol. Evol.* 9, 2322–2335
27. Campbell, C.L. *et al.* (2017) Alternative patterns of sex chromosome differentiation in *Aedes aegypti* (L). *BMC Genomics* 18, 943
28. Dudchenko, O. *et al.* (2017) *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95
29. Mattews, B.J. *et al.* (2018) Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature* 563, 501–507 http://dx.doi.org/10.1038/s41586-018-0692-z
30. Fontaine, M.C. *et al.* (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347, 1258524–1258524
31. Aboagye-Antwi, F. *et al.* (2015) Experimental swap of *Anopheles gambiae*'s assortative mating preferences demonstrates key role of X-chromosome divergence island in incipient sympatric speciation. *PLoS Genet.* 11, e1005141
32. Clark, A.G. and Messer, P.W. (2015) Conundrum of jumbled mosquito genomes. *Science* 347, 27–28
33. Wen, D. *et al.* (2016) Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Mol. Ecol.* 25, 2361–2372
34. Rosenzweig, B.K. *et al.* (2016) Powerful methods for detecting introgressed regions from population genomic data. *Mol. Ecol.* 25, 2387–2397
35. Thawornwattana, Y. *et al.* (2018) Coalescent analysis of phylogenomic data confidently resolves the species relationships in the *Anopheles gambiae* species complex. *Mol. Biol. Evol.* Published online August 9, 2018. http://dx.doi.org/10.1093/molbev/msy158
36. Burke, A. *et al.* (2017) A new malaria vector mosquito in South Africa. *Sci. Rep.* 7, 43779
37. Zhou, X. *et al.* (2014) Divergent and conserved elements comprise the chemoreceptive repertoire of the nonblood-feeding mosquito *Toxorhynchites amboinensis*. *Genome Biol. Evol.* 6, 2883–2896
38. Strode, C. *et al.* (2008) Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. *Insect Biochem. Mol. Biol.* 38, 113–123
39. Waterhouse, R.M. *et al.* (2008) The *Aedes aegypti* genome: A comparative perspective. *Insect Mol. Biol.* 17, 1–8
40. Chen, X.-G. *et al.* (2015) Genome sequence of the Asian tiger mosquito, *Aedes albopictus*, reveals insights into its biology, genetics, and evolution. *Proc. Natl. Acad. Sci. U. S. A.* 112, E5907–E5915
41. Arcà, B. *et al.* (2017) Anopheline salivary protein genes and gene families: an evolutionary overview after the whole genome sequence of sixteen *Anopheles* species. *BMC Genomics* 18, 153
42. Ribeiro, J.M.C. *et al.* (2016) A deep insight into the sialome of male and female *Aedes aegypti* mosquitoes. *PLoS One* 11, e0151400
43. Ribeiro, J.M.C. *et al.* (2018) A deep insight into the male and female sialotranscriptome of adult *Culex tarsalis* mosquitoes. *Insect Biochem. Mol. Biol.* 95, 1–9

44. Mitchell, S.N. *et al.* (2015) Evolution of sexual traits influencing vectorial capacity in anopheline mosquitoes. *Science* 347, 985–988
45. Duvall, L.B. *et al.* (2017) A peptide signaling system that rapidly enforces paternity in the *Aedes aegypti* mosquito. *Curr. Biol.* 27, 3734–3742.e5
46. Alfonso-Parra, C. *et al.* (2016) Mating-induced transcriptome changes in the reproductive tract of female *Aedes aegypti*. *PLoS Negl. Trop. Dis.* 10, e0004451
47. Villarreal, S.M. *et al.* (2018) Male contributions during mating increase female survival in the disease vector mosquito *Aedes aegypti*. *J. Insect Physiol.* 108, 1–9
48. Lewis, S.H. *et al.* (2016) Duplication and diversification of dipteran argonaute genes, and the evolutionary divergence of Piwi and Aubergine. *Genome Biol. Evol.* 8, 507–518
49. Zhou, D. *et al.* (2015) Genomic analysis of detoxification supergene families in the mosquito *Anopheles sinensis*. *PLoS One* 10, e0143387
50. Jenkins, A.M. and Muskavitch, M.A. (2015) Evolution of an epigenetic gene ensemble within the genus *Anopheles*. *Genome Biol. Evol.* 7, 901–915
51. Rottschaefer, S.M. *et al.* (2015) Population genetics of *Anopheles coluzzii* immune pathways and genes. *G3* 5, 329–339
52. Giraldo-Calderón, G.I. *et al.* (2017) Retention of duplicated long-wavelength opsins in mosquito lineages by positive selection and differential expression. *BMC Evol. Biol.* 17, 84
53. Assogba, B.S. *et al.* (2018) Adaptive deletion in resistance gene duplications in the malaria vector *Anopheles gambiae*. *Evol. Appl.* 11, 1245–1256
54. Lee, E. *et al.* (2013) Web Apollo: a web-based genomic annotation editing platform. *Genome Biol.* 14, R93
55. Price, D.C. *et al.* (2015) Characterization of the doublesex gene within the *Culex pipiens* complex suggests regulatory plasticity at the base of the mosquito sex determination cascade. *BMC Evol. Biol.* 15, 108
56. Jungreis, I. *et al.* (2016) Evolutionary dynamics of abundant stop codon readthrough. *Mol. Biol. Evol.* 33, 3108–3132
57. Kyrou, K. *et al.* (2018) A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat. Biotechnol.* Published online September 24, 2018. http://dx.doi.org/10.1038/nbt.4245
58. Lawson, D. *et al.* (2009) VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res.* 37, D583–D587
59. David, J.-P. *et al.* (2005) The *Anopheles gambiae* detoxification chip: a highly specific microarray to study metabolic-based insecticide resistance in malaria vectors. *Proc. Natl. Acad. Sci. U. S. A.* 102, 4080–4084
60. Rinker, D.C. *et al.* (2016) Disease vectors in the era of next generation sequencing. *Genome Biol.* 17, 95
61. Hugo, R.L.E. and Birrell, G.W. (2018) Proteomics of *Anopheles* vectors of malaria. *Trends Parasitol.* Published online September 5, 2018. http://dx.doi.org/10.1016/j.pt.2018.08.009
62. Tsujimoto, H. *et al.* (2017) Dengue virus serotype 2 infection alters midgut and carcass gene expression in the Asian tiger mosquito, *Aedes albopictus*. *PLoS One* 12, e0171345
63. Desjardins, C.A. *et al.* (2015) Contrasting host–pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat. Commun.* 6, 7121
64. Etebari, K. *et al.* (2017) Global transcriptome analysis of *Aedes aegypti* mosquitoes in response to Zika virus infection. *mSphere* 2, e00456-17
65. Juneja, P. *et al.* (2015) Exome and transcriptome sequencing of *Aedes aegypti* identifies a locus that confers resistance to *Brugia malayi* and alters the immune response. *PLoS Pathog.* 11, 1–32
66. Rodgers, F.H. *et al.* (2017) Microbiota-induced peritrophic matrix regulates midgut homeostasis and prevents systemic infection of malaria vector mosquitoes. *PLoS Pathog.* 13, e1006391
67. Zhang, J. *et al.* (2017) Differential gene expression in *Anopheles stephensi* following infection with drug-resistant *Plasmodium yoelii*. *Parasit. Vectors* 10, 401
68. Lombardo, F. *et al.* (2017) Deciphering the olfactory repertoire of the tiger mosquito *Aedes albopictus*. *BMC Genomics* 18, 770
69. Chen, Q. *et al.* (2017) The antenna transcriptome changes in mosquito *Anopheles sinensis*, pre- and post- blood meal. *PLoS One* 12, e0181399
70. Taparia, T. *et al.* (2017) Blood meal induced regulation of the chemosensory gene repertoire in the southern house mosquito. *BMC Genomics* 18, 393
71. Esquivel, C.J. *et al.* (2016) A *de novo* transcriptome of the Malpighian tubules in non-blood-fed and blood-fed Asian tiger mosquitoes *Aedes albopictus*: insights into diuresis, detoxification, and blood meal processing. *PeerJ* 4, e1784
72. Huang, X. *et al.* (2015) Global transcriptional dynamics of diapause induction in non-blood-fed and blood-fed *Aedes albopictus*. *PLoS Negl. Trop. Dis.* 9, e0003724
73. Seaman, J.A. *et al.* (2015) Age and prior blood feeding of *Anopheles gambiae* influences their susceptibility and gene expression patterns to ivermectin-containing blood meals. *BMC Genomics* 16, 797
74. Uyhelji, H.A. *et al.* (2016) Transcriptomic differences between euryhaline and stenohaline malaria vector sibling species in response to salinity stress. *Mol. Ecol.* 25, 2210–2225
75. MacCallum, R.M. *et al.* (2011) An expression map for *Anopheles gambiae*. *BMC Genomics* 12, 620
76. Upton, L.M. *et al.* (2015) *Anopheles gambiae* blood feeding initiates an anticipatory defense response to *Plasmodium berghei*. *J. Innate Immun.* 7, 74–86
77. Zhang, X. *et al.* (2016) CLIPB8 is part of the prophenoloxidase activation system in *Anopheles gambiae* mosquitoes. *Insect Biochem. Mol. Biol.* 71, 106–115
78. Lampe, L. and Levashina, E.A. (2017) microRNA tissue atlas of the malaria mosquito *Anopheles gambiae*. *G3* 8, g3.300170.2017
79. Liu, W. *et al.* (2017) Comparative expression profile of microRNAs in *Anopheles anthropophagus* midgut after blood-feeding and *Plasmodium* infection. *Parasit. Vectors* 10, 86
80. Zhang, X. *et al.* (2017) Transcriptome-wide microRNA and target dynamics in the fat body during the gonadotrophic cycle of *Aedes aegypti*. *Proc. Natl. Acad. Sci. U. S. A.* 114, E1895–E1903
81. Feng, X. *et al.* (2018) Analysis of microRNA profile of *Anopheles sinensis* by deep sequencing and bioinformatic approaches. *Parasit. Vectors* 11, 172
82. Nouzova, M. *et al.* (2018) A comparative analysis of corpora allata-corpora cardiaca microRNA repertoires revealed significant changes during mosquito metamorphosis. *Insect Biochem. Mol. Biol.* 96, 10–18
83. Jenkins, A.M.A.M. *et al.* (2015) Long non-coding RNA discovery across the genus *Anopheles* reveals conserved secondary structures within and beyond the Gambiae complex. *BMC Genomics* 16, 337
84. Lampe, L. and Levashina, E.A. (2017) The role of microRNAs in *Anopheles* biology-an emerging research field. *Parasite Immunol.* 39, e12405
85. Saldaña, M.A. *et al.* (2017) Zika virus alters the microRNA expression profile and elicits an RNAi response in *Aedes aegypti* mosquitoes. *PLoS Negl. Trop. Dis.* 11, e0005760
86. Castellano, L. *et al.* (2015) The germline of the malaria mosquito produces abundant miRNAs, endo-siRNAs, piRNAs and 29-nt small RNAs. *BMC Genomics* 16, 100
87. George, P. *et al.* (2015) Increased production of piRNAs from euchromatic clusters and genes in *Anopheles gambiae* compared with *Drosophila melanogaster*. *Epigenet. Chromatin* 8, 50
88. Goic, B. *et al.* (2016) Virus-derived DNA drives mosquito vector tolerance to arboviral infection. *Nat. Commun.* 7, 12410

89. Miesen, P. *et al.* (2016) Small RNA profiling in Dengue virus 2-infected *Aedes* mosquito cells reveals viral piRNAs and novel host miRNAs. *PLoS Negl. Trop. Dis.* 10, e0004452
90. Varjak, M. *et al.* (2017) Characterization of the Zika virus induced small RNA response in *Aedes aegypti* cells. *PLoS Negl. Trop. Dis.* 11, e0006010
91. Palatini, U. *et al.* (2017) Comparative genomics shows that viral integrations are abundant and express piRNAs in the arboviral vectors *Aedes aegypti* and *Aedes albopictus*. *BMC Genomics* 18, 512
92. Göertz, G.P. *et al.* (2016) Noncoding subgenomic flavivirus RNA is processed by the mosquito RNA interference machinery and determines West Nile virus transmission by *Culex pipiens* mosquitoes. *J. Virol.* 90, 10145–10159
93. Miesen, P. *et al.* (2016) PIWIs go viral: arbovirus-derived piRNAs in vector mosquitoes. *PLoS Pathog.* 12, e1006017
94. Olson, K.E. and Bonizzoni, M. (2017) Nonretroviral integrated RNA viruses in arthropod vectors: an occasional event or something more? *Curr. Opin. Insect Sci.* 22, 45–53
95. Whitfield, Z.J. *et al.* (2017) The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. *Curr. Biol.* 27, 3511–3519.e7
96. Petit, M. *et al.* (2016) piRNA pathway is not required for antiviral defense in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 113, E4218–E4227
97. Prasad, T.S.K. *et al.* (2017) Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes. *Genome Res.* 27, 133–144
98. Artemov, G. *et al.* (2018) The development of cytogenetic maps for malaria mosquitoes. *Insects* 9, 121
99. Artemov, G.N. *et al.* (2017) The physical genome mapping of *Anopheles albimanus* corrected scaffold misassemblies and identified interarm rearrangements in genus *Anopheles*. *G3* 7, 155–164
100. Artemov, G.N. *et al.* (2018) Partial-arm translocations in evolution of malaria mosquitoes revealed by high-coverage physical mapping of the *Anopheles atroparvus* genome. *BMC Genomics* 19, 278
101. Waterhouse, R.M. *et al.* (2018) Leveraging evolutionary relationships to improve *Anopheles* genome assemblies. *bioRxiv* Published online October 4, 2018. http://dx.doi.org/10.1101/434670
102. Miller, J.R. *et al.* (2018) Analysis of the *Aedes albopictus* C6/36 genome provides insight into cell line utility for viral propagation. *Gigascience* 7, 1–13
103. Le, B.V. *et al.* (2013) Characterization of *Anopheles gambiae* transglutaminase 3 (AgTG3) and its native substrate plugin. *J. Biol. Chem.* 288, 4844–4853
104. Pondeville, E. *et al.* (2018) Evolution of sexually-transferred steroids in *Anopheles* mosquitoes. *bioRxiv* Published online January 15, 2018. http://dx.doi.org/10.1101/248112
105. Whittle, C.A. and Extavour, C.G. (2017) Rapid evolution of ovarian-biased genes in the yellow fever mosquito (*Aedes aegypti*). *Genetics* 206, 2119–2137
106. Habtewold, T. *et al.* (2017) Immune resistance and tolerance strategies in malaria vector and non-vector mosquitoes. *Parasit. Vectors* 10, 186
107. Cheng, C. *et al.* (2018) Systems genetic analysis of inversion polymorphisms in the malaria mosquito *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U. S. A.* Published online July 9, 2018. http://dx.doi.org/10.1073/pnas.1806760115
108. Kudron, M.M. *et al.* (2018) The modERN resource: genome-wide binding profiles for hundreds of *Drosophila* and *Caenorhabditis elegans* transcription factors. *Genetics* 208, 937–949
109. Miles, A. *et al.* (2017) Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature* 552, 96–100
110. Dritsou, V. *et al.* (2015) A draft genome sequence of an invasive mosquito: an Italian *Aedes albopictus*. *Pathog. Glob. Health* 109, 207–220
111. Logue, K. *et al.* (2015) Whole-genome sequencing reveals absence of recent gene flow and separate demographic histories for *Anopheles punctulatus* mosquitoes in Papua New Guinea. *Mol. Ecol.* 24, 1263–1274
112. Peery, A. *et al.* (2011) Improving the population genetics toolbox for the study of the African malaria vector *Anopheles nili*: microsatellite mapping to chromosomes. *Parasit. Vectors* 4, 202
113. Lawniczak, M.K. *et al.* (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* 330, 512–514
114. Angleró-Rodríguez, Y.I. *et al.* (2017) *Aedes aegypti* molecular responses to Zika virus: Modulation of infection by the toll and JAK/STAT immune pathways and virus host factors. *Front. Microbiol.* 8, 2050
115. Dong, S. *et al.* (2017) The midgut transcriptome of *Aedes aegypti* fed with saline or protein meals containing chikungunya virus reveals genes potentially involved in viral midgut escape. *BMC Genomics* 18, 382
116. Jupatanakul, N. *et al.* (2017) Engineered *Aedes aegypti* JAK/STAT pathway-mediated immunity to Dengue virus. *PLoS Negl. Trop. Dis.* 11, e0005187
117. Carissimo, G. *et al.* (2018) Highly focused transcriptional response of *Anopheles coluzzii* to O'nyong nyong arbovirus during the primary midgut infection. *BMC Genomics* 19, 526
118. Thomas, T. *et al.* (2016) Hemocytome: deep sequencing analysis of mosquito blood cells in Indian malarial vector *Anopheles stephensi*. *Gene* 585, 177–190
119. Roy, S. *et al.* (2015) Regulation of gene expression patterns in mosquito reproduction. *PLoS Genet.* 11, e1005450
120. Biedler, J.K. *et al.* (2015) Maternal germline-specific genes in the Asian malaria mosquito *Anopheles stephensi*: characterization and application for disease control. *G3* 5, 157–166
121. Saveer, A.M. *et al.* (2018) Characterization of chemosensory responses on the labellum of the malaria vector mosquito, *Anopheles coluzzii*. *Sci. Rep.* 8, 5656
122. Bradshaw, W.E. *et al.* (2018) Evolutionary transition from blood feeding to obligate nonbiting in a mosquito. *Proc. Natl. Acad. Sci. U. S. A.* 115, 1009–1014