Supplementary Materials for

The genome sequence of the Asian tiger mosquito, *Aedes albopictus*, reveals insights into its biology, genetics and evolution

Xiao-Guang Chen*, Xuanting Jiang, Jinbao Gu, Meng Xu, Yang Wu, Yuhua Deng, Chi Zhang, Mariangela Bonizzoni, Wannes Dermauw, John Vontas, Peter Armbruster, Xin Huang, Yulan Yang, Zhang Hao, Weiming He, Hongjuan Peng, Yongfeng Liu, Kun Wu, Jiahua Chen, Manolis Lirakis, Pantelis Topalis, Thomas Van Leeuwen, Andrew Brantley Hall, Xiaofang Jiang, Chevon Thorpe, Rachel Lockridge Mueller, Cheng Sun, Robert M. Waterhouse, Guiyun Yan, Zhijian Tu, Xiaodong Fang*, Anthony A. James*

*Correspondence author. Email: 18664867266@qq.com (X.C.); Email: fangxd@genomics.cn (X.F.); aajames@uci.edu (A.A.J.)

This files includes:

Materials and Methods

Mosquitoes

The Foshan strain of *Ae. albopictus* was obtained from the Center for Disease Control and Prevention of Guangdong Province, China where it has been in culture since 1981. Mosquitoes were reared at 28 ℃ and 70-80% relative humidity with 14/10 hour light/dark cycles. Larvae were reared in pans and fed on finely ground fish food mixed 1:1 with yeast powder. Adults were kept in 30 cm^3 cages and allowed access to a cotton wick soaked in 20% sucrose as a carbohydrate source. Adult females were allowed to blood feed 3 and 4 days after eclosion on anesthetized mice.

Genome properties and evolution

DNA sequencing

Approximately 1.414 μg of genomic DNA were isolated from a single *Ae. albopictus* pupa of a 9^{th} generation iso-female line of the Foshan strain and subjected to whole-genome amplification (1) (WGA) to produce 243.2 μg of DNA. Amplified DNA was used to construct paired-end short-insert (170,500 and 800 base-pairs [bp] in length) and mate-paired long-insert (2 kilobase-pairs [kb], 5 kb, 10 kb and 20 kb) genomic libraries. These were sequenced using the Hiseq 2000 platform. A total of 943.59 giga base pairs (Gb) of genomic DNA sequencing data were obtained from all 23 libraries (Table S1.1).

Estimation of genome size using k-mer analysis

A k-mer refers to an artificial sequence division of K nucleotides in length generated iteratively from sequencing reads. A raw sequence read with L bp contains (L-K+1) k-mers if the length of each k-mer is K bp. The frequency of each k-mer can be calculated from the genome sequence reads. Typically, k-mer frequencies plotted against the sequence depth gradient follow a Poisson distribution in any given dataset; whereas sequencing errors may lead to a higher representation of low frequencies. The genome size, G, can be calculate from the formula G=K_num/K_depth, where the K_num is the total number of k-mers, and K_depth denotes the frequency occurring more frequently than the other frequencies (2). This analysis determined that K = 17, K_num = 69,801,535,654 and K_depth = 24, therefore the *Ae. albopictus* genome size was estimated to be 2.91 Gbp (Table S1.2; Figure S1.1).

Genome assembly

Data quality control and assembly

The raw sequence data were filtered before assembly by removing duplicated reads caused by gene amplification and reads contaminated by adapters, trimming continuous low-quality bases on 5'-ends according to quality graphs, and filtering reads with a significant excess of "N" and low-quality bases. A total of 689.59 Gb high-quality sequence data remained after removing the low-quality reads, and this corresponds to a 236.97-fold coverage of the genome based on our initial genome size estimate of 2.91 Gb (Table S1.3).

The assembler SOAPdenovo (3) (version 2.04), SSPACE (4) (version 2.0) and Gapcloser (3) (version 1.10) were used for genome assembly. Overlapped Pair-End reads from the 170 insert-size libraries were connected first to yield long sequences. A 97 bp sequence from the connected long reads was used next to construct contigs. All of the usable reads from different insert-size libraries then were realigned to the contigs using SSPACE (4). The resulting linking information was used to produce the final scaffold construction and this was followed by gap-filling of the scaffolds. The sequences of *Wolbachia pipientis* were aligned to the assembly and the scaffolds matching them were removed to avoid the contamination. Finally, the scaffold N50 length achieved 195.5 kb with a total length of 1.97 Gb. The assembly comprises 401,027 scaffolds with 147 Mb Ns, and 131,405 scaffolds with length ≥2 kb, which account for 1.83 Gb of the genome (Table S1.4).

Accuracy of genome assembly

The quality of the draft genome was evaluated by assessing the sequencing depth and coverage using available mRNA and fosmid sequences. All useable sequence reads were realigned to the draft

genome using SOAP2 (5). More than 85 % of the reads were mapped to the assembly allowing a maximum of three mismatches. Greater than 91.8% of the assembly sequences were covered by >3 genomic reads with a peak depth of 50-fold coverage, indicating that these regions had high single-base accuracy (Figure S1.2). The sequencing depth for individual nucleotides was unbiased, the GC content of non-overlapping sliding windows (1000 bp) was unbiased when compared with their average sequencing depth, and 99.6% of the regions with GC content between 30% and 50% had a sequencing depth of greater than 10-fold (Figure S1.3).

Nine Sanger-derived fosmid sequences were aligned to the scaffolds using blastn (E-value < 1^{-10}). The scaffolds (2,326 kb in length) were matched to the fosmid sequences completely with the coverage of 95 % (Table S1.5 and Figure S1.4). A total of 68,133 mRNA sequences downloaded from NCBI aligned to the assembly using BLAST (6) with default parameters. Among 56,807 mRNA sequences (>500 bp), 54,495 (95.87%) were mapped to scaffolds with ≥ 95% identity and ≥ 50% coverage (Table S1.6). Together, the two methods provided positive evidence that the draft genome has a reliable quality and coverage and that there is no evidence of WGA-induced artefacts.

Genomic features

Transposable elements in Aedes albopictus genome

Tandem repeat DNAs were searched across the *Ae. albopictus* genome using the software Tandem Repeats Finder (version 4.04, http://tandem.bu.edu/trf/trf.html; Benson, 1999). Transposable elements (TEs) were identified in the genome by a combination of homology-based and *de novo* approaches. Homology prediction used RepeatProteinMask and RepeatMasker (version 3.3.0, http://www.repeatmasker.org) (7) with default parameters against Repbase (release 16.03, http://www.girinst.org/repbase) (8). *De novo* prediction involved the *ab initio* prediction programs, RepeatModeler and LTR-FINDER (version 1.0.5, http://tlife.fudan.edu.cn/ltr_finde) (9) to build the *de novo* repeat library. Contaminating and multi-copy genes in the library were removed. RepeatMasker was run using the resulting library of sequences as a database to find and classified the TEs. *Ae. aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus* TEs were annotated with the same method (Table S1.7; Figure S1.5).

Transcriptome sequencing (RNA-seq)

Transcriptomes were derived from libraries comprising mRNA derived from seven developmental stages: mixed sex samples of embryos 0-24 hours post-deposition (hpd), embryos 24-48 hpd, a combined pool of 1^{st}-and 2^{nd}-instar larvae, a combined pool of 3^{rd}- and 4^{th}-instar larvae and pupae of all stages, and adult males and sugar-fed adult females (Table S1.8). The TRIzol reagent (Invitrogen) and RNase-free DNase I were used to extract and treat total RNA. Polyadenylated (polyA$^+$) mRNA was enriched using oligo-dT beads, fragmented and primed randomly during the first strand synthesis by reverse transcription. Second-strand was synthesized using RNaseH and DNA polymerase I to create double-stranded cDNA fragments. The double-stranded cDNA was applied to 200 bp paired-end RNA-seq libraries by Illumina's protocols and sequenced with 90bp at each end on the Illumina HiSeq 2000 platform. The cDNA library was normalized by the duplex-specific nuclease method (10) followed by cluster generation on the Illumina HiSeq 2000 platform. Transcript reads were mapped by Tophat and subsequently analyzed with in-house Perl scripts. Gene expression levels were calculated as RPKM (11). Differentially-expressed genes between two samples were detected using a method based on a Poisson distribution, and samples were normalized for differences in the RNA output size, sequencing depth, and gene length. Genes were identified in at least one experiment with a minimum two-fold difference (RPKM) in two experiments and an FDR of < 0.001 were defined as differentially expressed genes. Enrichment analysis was performed using EnrichPipeline (12)**.**

Gene annotation

We applied *de novo* gene prediction, homology-based methods and RNA-seq data to perform gene predictions. *Aedes aegypti*, *D. melanogaster*, *An. gambiae* and *Cx. quinquefasciatus* protein sequences were downloaded and aligned to *Ae. albopictus* genome using tblastn (13) to produce homology-based

predictions. Putatively homologous genome sequences then were aligned with the matching proteins using Genewise (14) to define gene models. Augustus (15) and Genscan (16) were employed using appropriate parameters for *de novo* prediction of coding genes. Homology-based and *de novo*-derived gene sets were merged to form a comprehensive and non-redundant reference gene set using GLEAN (http://sourceforge.net/projects/glean-gene). Transcriptome reads from the seven different samples were mapped to the genome assembly using TopHat (17) to give RNA-seq based predictions. TopHat mapping results were combined and applied Cufflinks (18) to predict transcript structures. A total of 1,000 intact genes also were selected from the homology-based prediction to pass a fifth-order Markov model, then to predict the ORF of RNA transcripts based on the Hidden Markov Model (HMM). Finally, the RNA transcripts were integrated with the GLEAN gene set to form the final non-redundant gene set. A total of 17,539 genes were identified in *Ae. albopictus* (Table S1.9).

Gene functional annotation

We aligned *Ae. albopictus* protein sequences to various protein databases, including InterPro (19), Swiss-Prot (20), KEGG (21) and Trembl (20) to infer their biological functions or their molecular pathways (Table S1.10 – S1.12). Gene Ontology (22) descriptions of gene products were retrieved from InterPro. The symbol of each gene was assigned based on the best match derived from the alignments with SwissProt databases using blastp. Motifs and domains were annotated by InterPro through searching against publicly-available databases, including Pfam, PRINTS, PANTHER, PROSITE, ProDom, and SMART. Genes also were mapped to KEGG pathway maps by searching KEGG databases and finding the best hit for each gene.

Gene evolution

Gene family clustering

A gene family denotes the set of orthologous and paralogous genes that descended from a single gene in the last common ancestor of the species. The TreeFam methodology (23) was used to define gene families using data from five mosquito species (*Ae. albopictus*, *An. gambiae*, *Ae. aegypti*, *Cx. quinquefasciatus* and *An. darlingi*) as references, and the fruit fly, *D. melanogaster*, as the out-group. Pipeline and parameters were as follows: blastp was used to find all the homologous relationship among protein sequences of the six species with E-value < 1e-10; Solar (in-house software, version 0.9.6) was used to conjoin high-scoring segment pairs (HSPs) between each pair of protein homologs. Protein sequence similarity then was assessed with bit-score, and protein genes are clustered into gene families by a hierarchical clustering algorithm (an implementation included in the Treefam pipeline, version 0.5.0) with an algorithm analogous to average-linkage clustering with the parameters set to be "-w 5 -s 0.33 -m 100000". A total of 6,787 orthologous gene groups could be partitioned among the mosquito and fruit fly species (Figure S1.6). We identified 1,564 genes that were in lineage-specific gene families within *Ae. albopictus*. GO enrichment and KEGG pathway enrichment were done for lineage-specific family genes (Tables S1.11 and S1.12).

Phylogenetic tree construction and divergence time estimate

A total of 2096 single-copy gene families were defined as orthologous genes according to Treefam pipelines and were chosen in this analysis. Proteins were assigned to a CDS based on the alignment results. All CDS and the 4d sites were extracted from each alignment and concatenated to one super gene for the six species. PhyMLv3.0 (Zang 1997) (parameters: -m HKY85, other default) was used to construct a phylogenetic tree for the six species. The chain length was set to 100,000 (1 sample/100 generations) and the first 1,000 samples were burned in. The transition/transversion ratio was estimated as a free parameter. Divergence time was estimated using the program MCMCTREE (version 4), which was part of the PAML package. "JC69" models in MCMCTREE program were used in our calculations (Figure 2).

Non-coding RNAs

Non-coding RNAs include a number of species-specific miRNAs identified in the diapause transcriptome analysis (Tables S1.13 and S4.5). These are discussed below with additional detail in the description of diapause related genes-small non-coding RNAs.

Expansion and contraction of gene families

Based on the phylogenetic tree topology, CAFÉ (24) (Computational Analysis of gene Family Evolution, version 2.1), a tool for the statistical analysis of the evolution of the size of gene families based on stochastic birth and death model, was used to detect gene family expansion and contraction in *Ae. albopictus*, *An. gambiae*, *Ae. aegypti*, *Cx. quinquefasciatus*, *An. darling* and *D. melanogaster* with the parameters "P-value threshold 0.05, number of random 10000 and search for the λ value". Gene families with P-values < 0.05 were analyzed manually. A total of 86 expansion gene families (773 genes) and 26 contraction gene families (108 genes) were identified in *Ae. albopictus* and function enrichment of expansion gene families determined (Tables S1.14 and S1.15).

Detection of positively selected genes

As described above, blastp and Treefam methodologies were used to define orthologs among in *Ae. albopictus*, *An. gambiae*, *Ae. aegypti*, *Cx. quinquefasciatus*, *An. darling and D. melanogaster*. In total, 2096 pairs of genes were identified as single copy orthologs. The coding sequences of the orthologs were aligned using Prank (25) (http://code.google.com/p/prank-msa/) software with default parameters. The genes were filtered even if the alignment rate of the gene was less than 80% in only one species. Ka and Ks were calculated for the aligned orthologs using KaKsCalculator software (26) (version 1.2, parameter "-m YN") with default parameters. Finally, we identified 239 positively-selected genes (P-value < 0.05) in *Ae. albopictus.* GO enrichment and KEGG pathway enrichment for the positively-selected genes identified some significant enrichment ($P < 0.05$, Table S1.1 – S1.16).

Repetitive DNA and TE

De novo and homology-based transposable element (TE) discovery

Transposable element discovery and classification were performed on the scaffold sequences of *Ae. albopictus*. RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html) was performed to identify transposable elements *de novo*. All repeats were compared to *Ae. aegypti* transposable elements deposited in TEfam (tefam.biochem.vt.edu) and known protein-encoding sequences to assist TE classification. RepeatModeler outputs < 500 base-pairs (bp) were used to search for MITEs and SINEs (27). Two methods were used to identify *Ae. albopictus* TEs that showed no homology to *Ae. aegypti*. The first, more-stringent, approach compared the *Ae. albopictus* TE library with the *Ae. aegypti* genome assembly by blastn (1e-5). Any *Ae. albopictus* TEs with similarity were removed from the library. The second approach used the *Ae. aegypti* repeatmodeller library to mask the *Ae. albopictus* TE library. The two methods produced similar results.

The *Ae. albopictus* genome harbors all major groups of TEs as shown by the analysis of RepeatModeller results (Table S2). Repetitive sequences comprise 71% of the *Ae. albopictus* genome, the highest of all sequenced genomes of mosquito species. This high repeat content is consistent with the large size of the genome, which is ~50% larger than that of *Ae. aegypti*, the only other sequenced mosquito with a genome >1 Gigabase. Notably, *Ae. albopictus* and *Ae. aegypti* belong to the same subgenus, *Stegomyia*. Non-LTR retrotransposons or long interspersed nuclear elements (LINEs) showed the highest genome occupancy in both species. A single LINE element (*Duo*, Fig S2.1) comprises 4.1% (82 Mb) of the entire genome. This element is homologous to TF000022 in *Ae. aegypti* (tefam.biochem.vt.edu), which occupies 3.17% of its genome. More than 20% of the *Ae. albopictus* genome is occupied by interspersed repeats that had no similarity (e-value cutoff = 1e-5) to *Ae. aegypti* sequences, indicating rapid repeat expansion after the divergence of the two species.

Recent TE insertion contributed to the expansion of the *Ae. albopictus* genome

We estimated the relative time of insertion of LINE and LTR retrotransposons, which account for 62.7% of genome (Figure S2.2; Table S2), by comparing sequence similarities between the best matching TE pairs within a cluster. LINE and LTR retrotransposons that are >100 bp were clustered using CD-HIT7 (28), with a threshold of 90% global sequence identity. The best mapping sequence within each cluster was chosen as the representative sequence. To date the insertion time of these

retrotransposons, we only considered clusters of more than 10 copies that cover at least 90% of the length of the representative sequence. For each cluster, we performed pairwise comparisons and selected the pairs that showed the highest percent identity. These pairs were aligned and the date of divergence (T) was calculated using Kimura's two-parameter method (29): T = K/2k, where k is 2.2×10^{-10} substitutions/site/year according to the previous report (30). K = -1/2 ×ln (1-2P-Q) ×sqrt (1-2Q), where P is the transition fraction in the aligned sequences, Q is the transversion fraction, and K is the evolutionary distance.

Shown in Figure S2.2 are insertion numbers of major clades of LINE and LTR retrotransposons over evolutionary time in *Ae. albopictus*. The same analysis were also performed on the same clades of LINE and LTR retrotransposons in *Ae. aegypti* and shown for comparison (Figure S2.2). Peak insertion activity occurred with 0–10 Myr ago in *Ae. albopictus* and is higher than *Ae. aegypti* for all but one of the clades (Figure S2.2).Thus it is clear that recent transposition of LINE and LTR retrotransposons contributes significantly to the expansion of the *Ae. albopictus* genome.

DNA loss analysis in mosquito genomes

Varied deletion rate also has been shown to drive genome size variations (31, 32). Thus we performed deletion-rate analysis using so-called "dead-on-arrival" non-LTR retrotransposons from *Ae. albopictus*, *Ae. aegypti*, and *Cx. quinquefasciatus*. The DNA loss rates for neutrally-evolved DNA sequences in mosquito genomes were estimated by using a previously described method (32). In brief, the consensus sequences of autonomous non-LTR retrotransposons in the focal mosquito genomes were collected. The consensus sequences for *Ae. aegypti* and *Cx. quinquefasciatus* were downloaded from TEfam (http://tefam.biochem.vt.edu/tefam/index.php). The consensus sequences for *Ae. albopictus* were generated in the present study using RepeatScout (33). Secondly, the obtained consensus sequences were trimmed to only keep protein-coding regions. Thirdly, the consensus sequences after trimming (Figure S2.3) were used as repeat library to mask their corresponding genomic sequences by RepeatMasker (http://www.repeatmasker.org/) to generate pairwise alignment files. We then used the obtained alignments to eliminate all non-LTR sequences with nonrandom distributions of substitutions across codon positions (χ^2 test; $P < 0.05$) to avoid counting substitutions that occurred along master element lineages. Finally, for each remaining non-LTR element copy, the numbers of insertions, deletions, and substitutions relative to the consensus sequence were obtained based on the RepeatMasker-generated alignment, and the sums of these values for every individual element copy were used to represent the total amounts of DNA gained and lost through small indels (≤ 30 bp) in the focal mosquito genome (bp deleted – bp inserted / substitution).

We observed more deletions than insertions in every mosquito genome (Table 1), which is the same as previously analyzed organisms (32, 34). *Ae. albopictus* has a slightly lower DNA loss rate compared to that of *Ae. aegypti* and *Cx. quinquefasciatus*, indicating that slower DNA loss may also contribute to the large genome size of *Ae. albopictus*.

Flavivirus-like sequences

BLAST analysis

Basic Local Alignment Search Tool (BLAST) was used to search the *Ae. albopictus* genome assembly using as query 261 sequences annotated as Cell Silent Agent (CSA) sequences (35), flavivirus-like DNA sequences from *Ae. albopictus*, *Ae. vexans*, *Ochlerotatus detritus*, *Oc. capius* and *Culiseta annulata* (36, 37), sequences corresponding to the whole genome or coding for the NS3, NS5 and the envelope (E) protein of representative members of insect specific flaviviruses (ISFs), mosquito-borne flaviviruses (MBVs), tick-born flaviviruses (TBVs) and flaviviruses with no known arthropod vector (NBVs). BLAST analysis was also extended to *Ae. aegypti*, *An. gambiae* and *Cx. quinquefasciatus* with the AaegL3, the AgamP4 and the CpipJ2 assemblies being downloaded from VectorBase (www.vectorbase.org), respectively. BLAST hits were retrieved from the *Ae. albopictus* genome using a custom script. Genome integrations from non-retroviral RNA viruses have been called NIRVs (non-retroviral integrated RNA viruses) (38, 39). We used this terminology throughout the text. Retrieved sequences were annotated using Argot2 (40). Argot2 was run with default parameters. Sequences associated with GO terms related to viral functions and/or DNA binding and integrations

were searched for Open Reading Frames (ORFs) using NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Sequences, in which partial or complete ORFs for flaviviral proteins were identified, were aligned to previously-characterized viral genomic integrations (35, 36) and sequences of the whole genome of representative ISFs, MBVs, NBVs and TBVs by ClustalW. Additionally, NIRVs were classified on the basis of the identified ORFs and aligned to representative E, NS1 and NS5 flaviviral genes to increase alignment accuracy. MEGA software version 5.2.2 (41) was used to search for the best-fit model of nucleotide substitution, which was then utilized to construct maximum-likelihood (ML) trees. The statistical robustness of the inferred nodes was assessed by 1000 bootstrapping. Trees were visualized by FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Flavivirus-like sequences in the genome of *Ae. albopictus*

DNA sequences similar to flaviviruses were recently detected in the genome of *Ae. albopictus* mosquitoes from Thailand, Italy, Cameroon, Madagascar and USA (Texas) (35, 36, 42, 43). Generally, host genome integrations from non-retroviral RNA viruses are called NIRVs (38, 39); the first integrations from flavivirus-like sequences in the genome of *Ae. albopictus* were called Cell Silent Agents (NIRV) to differentiate them from Cell Fusing Agent virus (CFAV) (35). Evidence for transcription was provided for one NIRV spanning the flaviviral NS1-NS4 genes (35).

We investigated the presence of NIRVs in the *Ae. albopictus* genome annotation of the Foshan strain by blast analyses using 261 sequences of previously characterized NIRVs, sequences of the complete genome or portions of the genome of representative ISF, MBV, TBV and NBV members as a query. blast analysis was extended to *Ae. aegypti*, *An. gambiae* and *Cx. quinquefasciatus* using the same queries as in *Ae. albopictus*. We identified hits (e-value < 2e-4) to ISFs and pathogenic flaviviruses in *Ae. aegypty* and *Ae. albopictus*, but none in *An. gambiae* and *Cx. quinquefasciatus* (Tables S3.2 and S3.3). In *Ae. albopictus*, some hits were duplicated as different parts of a scaffold sometimes gave separate hits; some hits were overlapping, flanked by another hit or located at the edge of a supercontig. The array of the identified hits is dependent on the assembly and may change with further assembly improvements. After elimination of redundancies, functional annotation of sequences corresponding to BLAST hits and scaffold analyses, 24 sequences with partial or complete ORFs for flaviviral proteins (44), primarily NS5 and NS1, were identified across ten scaffolds (Table S3.4). These NIRVs were embedded in regions rich with LTR retrotransposon sequences, primarily Ty1-copia and Ty3-gypsy (45, 46). There were instances of scaffolds harboring groups of contiguous or overlapping NIRVs, interspersed among defective retrotransposon-like sequences (i.e. scaffold15182, 14136, 6867, 5617 and 91) and instances of scaffolds with one viral integration (i.e. scaffold172623, 2646, 30874, 8815 and 4896) (Table S3.4). Contiguous NIRVs encompassed the same viral gene, either NS1 on scaffold 14636 or NS5 on scaffolds 15182, 6867, 5671 and 91, with the exception of Fo4904B and Fo7000 (scaffold15182) that included ORFs for the E, NS1, NS2A, NS2B and the NS4 proteins or the C and M proteins, respectively (Table S3.4). No duplications were observed at integration sites, suggesting viral integrations were derived from ectopic recombination with retrotransposons rather than being catalyzed by a classical retrotransposition activity (46, 47).

Besides sequences harboring complete or partial ORFs for flaviviral proteins, blast analyses in *Ae. albopictus* identified sequences supported by low e-values (<2e-4) that could not be functionally associated with viral functions probably due to their limited size, extensive sequence rearrangements, because they have homology with host proteins (48) or simply because they are sequence stretches with similarities to virus sequence stretches (45, 46, 49). Other sequences were associated with generic viral functions (i.e. Fo693 and Fo3594 in Table S3.5) or viruses other than flaviviruses such as the Negev virus (50) and the Wuhan Mosquito Virus 8 (51) (Table S3.5, Table S3.6).

Genetic relationship among the 24 sequences encompassing flaviviral ORFs, previously characterized NIRVs and the sequences of 92 complete genomes of 31 virus species belonging to the flavivirus lineage were investigated. We generated a ML tree including all the 24 identified NIRVs, previously characterized NIRVs and the complete genome of 31 viruses (Figure S3.1). Additionally, to increase the accuracy of the sequence alignment, we classified NIRVs on the basis of the identified ORF and compared them separately to the E, NS1 and NS5 genes of representative flaviviruses (Figures S3.2,

S3.3, S3.4). The derived trees confirm the close genetic relationships between NIRVs and ISFs as previously shown (35, 36, 42, 52), identify a potentially novel relationship between NIRVs and human-pathogenic flaviviruses and show that NIRVs from the Foshan strain do not form a unique cluster, suggesting multiple integrations. Overall BLAST and phylogeny results show that flaviviral integrations can occur in the germline and may be inherited in *Ae. albopictus* as mosquitoes of the Foshan strain have been excluded from contact with wild caught-mosquitoes and viruses for more than 30 years. At the same time, these data suggest that flaviviral integrations may be an on-going regional process because NIRVs recently reported for *Ae. albopictus* mosquitoes collected in northern Italy (36) formed a separate cluster from any of the NIRVs identified in the Foshan strain (Figure S3.1) and NIRVs with intact ORFs and high level of identity to circulating viruses were detected along with sequences harboring extensive rearrangements (52).

The larger number of NIRVs identified in the Foshan strain with respect to previous reports may be due to the fact that past characterizations were based on PCR analyses with flavivirus-specific primers (35-37, 42). Alternatively, the larger number of NIRVs in mosquitoes of the Foshan strain may indicate that these are the original integrations and that the *Ae. albopictus* invasion process out of its native range, which encompasses China, was associated with integration loss probably due to bottleneck events. The presence of a variable number of integrations across geographic populations may explain enticing reports of fluidity in the genome size of *Ae. albopictus* (53). The lower number of BLAST hits identified when using MBV, NBV and TBV sequences as query than ISFs and previously characterized NIRVs may be related to the low prevalence of MBV infection in natural mosquito populations (54-56). Alternatively, it may support ISF ancestral state with respect to other flaviviruses (57). The current variability in the viral integration sites, their sequence variability and their genetic relationships with respect to ISFs and human-pathogenic flaviviruses suggest that different regions of different length of the flavivirus genome can integrate.

Integrations of sequences from non-retroviral RNA viruses have been described in a number of eukaryotic organisms since the late ‘1970, but their geographic widespread within each species nor their biological relevance are completely understood yet (52, 58-68). In bees, shrimps, mice and plants, integrations from non-retroviral RNA viruses have been associated to subsequent immunity with respective viruses (48, 59, 62, 69). This phenomenon, called “viral accommodation” or endogenous viral elements (EVEs) derived immunity (70, 71), has already been exploited in plant biology to generate transgenic potato and tobacco plants resistant to Potato Virus Y (72). Additionally, genome integrations from non-retroviral RNA viruses could be associated with the emergence of viral DNA forms early in the infection cycle (73) and favor the establishment of persistent infections (45). Because *Ae. albopictus* can be chronically-infected with non-retroviral RNA viruses of different species for which genomic integrations have been detected, it represents an ideal system to study the origin and biological effects of integrations from non-retroviral RNA viruses.

Diapause-related genes

Gene annotation

Manual annotation of putative diapause-related genes was performed using WebApollo (74) to integrate the original GLEAN/Cuff annotations on the scaffolds with Maker annotations (75) based on a comprehensive diapause transcriptome (76-79). Annotated genes included those involved in chromatin remodeling, lipid metabolism, hormonal regulation, circadian rhythms and other functions. Final annotations were based on the presence of a start codon, stop codon, canonical splice sites, and extended 5' or 3' UTRs that were supported by Maker or exonerate (80) alignment of contigs from the transcriptome. A total of 71 genes with a putative diapause function were annotated (Table S4.1). Of these genes, 14 are duplicated uniquely in the *Ae. albopictus* genome, including several with known-diapause related functions such as lipid metabolism (81), elongation of long-chain hydrocarbons (82) and hormone signaling.

Gene expansion and diapause-associated differential expression

The list of genes from expanded gene families in the *Ae. albopictus* genome relative to *Ae. aegypti*, *Cx. quinquefasciatus*, and *An. gambiae* was searched against the *Ae. albopictus* diapause transcriptome

using blastn with an e-value cutoff of 1e-6. This search identified 211 genes (Table S4.2). Differential expression (DE) under diapause vs. non-diapause conditions of these 211 gene models was examined at seven physiological or developmental stages of the life cycle including two stages of embryonic development (77), three pharate larval stages (76) and two adult female stages (79) (Table S4.3). Overall 140 of the 211 gene models are DE during at least one of the seven stages. Results for each stage are indicated in Table S4.3 with the proportion of DE genes in the 211 gene expansion set vs. the proportion of DE genes in the overall transcriptome. An Exact Wilcoxon signed-rank test indicates that the percentage of DE genes in the expansion set is significantly higher than the percentage of DE genes in the overall transcriptome database across all seven stages ($P = 0.022$).

To further investigate diapause-associated differential expression of genes that show evidence of expansion in the *Ae. albopictus* genome, we further analyzed the 140 gene models from the expansion set that are expressed differentially during at least one stage of the life-cycle from the analysis described above. A total of 96 of these genes were classified into one of the following five protein super families (http://supfam.org/SUPERFAMILY/) with high confidence (e-value< 0.0001): stress response, lipid metabolism, gene expression regulation, serine protease related and others (Table S4.4). The differential expression patterns of the genes were compared between pre-adult (developing embryos and pharate larvae) and adult stages (blood fed and non-blood fed) to identify genes that had contrasting differential expression patterns. For example, genes that were up-regulated under diapause conditions during at least one time point within the pre-adult stage, and were down-regulated or not differentially expressed under diapause conditions during at least one time point within the adult stage were defined as having contrasting differential expression patterns between the pre-adult vs. adult stages. Similarly, genes that were down-regulated or not differentially expressed under diapause conditions at the pre-adult stage but were up-regulated under diapause conditions at the adult stage also were considered to have contrasting differential expression patterns. The expected proportion of genes with contrasting differential expression patterns for the complete *Ae. albopictus* diapause transcriptome was 0.55. Overall, the proportion of genes that exhibit contrasting patterns of differential expression across the life cycle was greater for genes from the expansion set with super-family annotations than for genes from the complete diapause transcriptome (Fisher's exact test, $p < 0.001$). The genes from the expansion set in all five super-family categories had a significantly greater proportion of contrasting patterns of differential expression than genes from the complete diapause transcriptome (Table S4.4). Lipid metabolism has been implicated previously as an important transcriptional component of the diapause program in *Ae. albopictus* (81), and gene expression regulation is implicated as an important component of diapause based on extensive differential gene expression under diapause vs. non-diapause conditions (76, 77, 79). The role of contrasting differential expression across the life-cycle for serine protease genes remains unclear. These results are consistent with the hypothesis that gene expansion can give rise to flexible gene expression across the life cycle and thereby contribute to the evolution of complex adaptive responses to environmental heterogeneity (83).

Small non-coding RNAs

To substantiate expression of putative miRNAs predicted from the *Ae. albopictus* genome, size-selected sequences (18-24bp) from six libraries of small, non-coding RNAs from mature oocytes of females maintained under diapause and non-diapause conditions (three libraries for each condition) were mapped using Bowtie 2 (84) to 1,548 genomic sequences annotated as non-coding RNAs. Using a mapping tolerance of 1bp mismatch, 57 genome annotated ncRNAs were identified with at least three reads per library in at least three libraries (Table S4.5). All of these predicted ncRNAs are putative miRNAs that were not previously described in *Ae. albopictus* by Gu et al. (85), and therefore represent novel and potentially *Ae. albopictus*-specific miRNAs with confirmed expression.

Detoxification

Phylogenetic analysis of the cytochrome P450 (CYP) gene family

Aedes albopictus CYP genes were identified using a tblastn approach (86) with default options and mosquito (*Ae. aegypti* and *An. gambiae*) CYP protein sequences as queries (87-89). Resulting high-scoring segment pairs (HSPs) were clustered based on their coordinates on the *Ae. albopictus* genome using in-house Perl scripts (53). Subsequently, HSP clusters and available RNA-Seq data were used to annotate CYP genes in the *Ae. albopictus* genome. CYP gene models were verified further by performing reciprocal blastn searches against the non-redundant nucleotide database of NCBI. Pseudogenes (gene models with one or two frame-shifts, late start codon or premature stop codon compared to the queries) and gene fragments were separated from putative full length CYP coding sequences. A total of 186 full length CYP genes (AalbCYPXXX) and 24 CYP pseudogenes (AalbCYPpseudoXX) were translated into protein sequences. These sequences were aligned with P450 protein sequences from *Ae. aegypti*, *An. gambiae* and *D. melanogaster* (87-90) and a set of P450 marker protein sequences of various organisms (91) using MUSCLE (92) with default settings, as it is incorporated in MEGA6 (93). Finally, a neighbor joining analysis was (poisson substitution model, uniform rates among sites, pairwise deletion of gaps) performed, bootstrapping with 1000 pseudoreplicates and the resulting tree was visualized and edited in MEGA6 (93). In addition, a separate neighbor joining phylogenetic analysis was performed with CYP4G protein sequences from representative insect species from the Diptera, Lepidoptera, Coleoptera, Hymenoptera and Phthiraptera and CYP4C protein sequences from *D. melanogaster* and *Blaberus discoidalis*. Finally, all *Ae. albopictus* CYPgene sequences were submitted to the P450 nomenclature committee (D. Nelson, Univ. Tennessee) for naming.

Phylogenetic analysis of the carboxyl/cholinesterase (CCE) gene family

Aedes albopictus CCE genes were identified using the previously-described CYP approach except that *Ae. aegypti* and *An. gambiae* CCE protein sequences were used as queries. Similarly, pseudogenes (gene models with one frameshift compared to the queries) and gene fragments were separated from putative full length CCE coding sequences. A total of 64 full length CCE genes (AalbCCEXXX) and 7 CCE pseudogenes (AalbCCEpseudoXX) were translated into protein sequences. These sequences were aligned with esterase protein sequences from *Ae. aegypti*, *An. gambiae* and *D. melanogaster* (87-90) using MUSCLE (92) with default settings, as it is incorporated in MEGA6 (93). The resulting alignment was trimmed at both ends according to Claudianos et al. 2006 (94). Model selection was done with ProtTest 2.4 (95) and according to the Akaike information criterion the LG+I+G+F model was optimum for phylogenetic analysis. Finally, a maximum likelihood analysis was performed with Treefinder (96) using the optimum model (LG+I+G+F), bootstrapping with 1000 pseudoreplicates. The resulting tree was visualized and edited in MEGA6 (93).

Phylogenetic analysis of the glutathione-S-transferase (GST) gene family

Aedes albopictus GST genes were identified using the previously-described CYP approach except that *Ae. aegypti* and *An. gambiae* GST protein sequences were used as queries. Similarly, pseudogenes (gene models with frameshifts, late or premature stop codon compared to the queries) and gene fragments were separated from putative full length GST coding sequences. A total of 32 full length cytosolic GST gene (AalbGSTXXX) and 5 GST pseudogenes (AalbGSTpseudoXX) were translated into protein sequences. These sequences were aligned with GST protein sequences from *Ae. aegypti*, *An. gambiae* and *D. melanogaster* (87-90) using MUSCLE (92) with default settings, as it is incorporated in MEGA6 (93). Model selection was done with ProtTest 2.4 (95) and according to the Akaike information criterion the LG+I+G model was optimum for phylogenetic analysis. Finally, a maximum likelihood analysis was performed with Treefinder (96), bootstrapping with 1000 pseudoreplicates. The resulting tree was visualized and edited in MEGA6 (93).

ATP-binding cassette transporter (ABC) gene family

Aedes albopictus ABC genes were identified in the *Ae. albopictus* genome using tblastn (E-value threshold < E-5) (86) and *D. melanogaster* and *An. gambiae* ABC protein sequences (97) as queries. *Aedes albopictus* ABC gene models were refined (CCGXXXXX IDs) or created (e.g. AalbABCxxx) on the basis of homology and available RNA-seq data. A similar approach was used to identify and annotate ABC genes in *Ae. aegypti* using *Ae. albopictus* ABC protein sequences as queries. *Aedes*

albopictus and *Ae. aegypti* full length ABC genes and incomplete ABC genes that 1) have a sequence length larger than 75% of the average sequence length of ABC full length genes per ABC subfamily per *Aedes* species and 2) that either contained a frame-shift or were located near a sequence gap or at the end/start of a scaffold, were considered as putative ABC genes and translated into protein sequences. Assignment of *Ae. albopictus* and *Ae. aegypti* sequences to the different ABC subfamilies (A-H) was assessed by a blastp search (86) against *D. melanogaster* ABC protein sequences. Except for the ABCE subfamily, *Ae. albopictus* and *Ae. aegypti* ABC protein sequences from each subfamily were aligned with those of *D. melanogaster* and *An. gambiae* using MUSCLE (92) and default settings. Model selection was done with ProtTest 2.4 (95) and according to the Akaike information criterion the LG+I+G+F, LG+G+F, LG+G+F, LG+I+G+F, LG+G, LG+G+F, LG+I+G+F, LG+G model was optimum for the phylogenetic analysis of the ABCA, B full transporter, B half transporter, C, D, F, G and H subfamily, respectively. Finally, a maximum likelihood analysis was performed for each subfamily with Treefinder (96) and the optimal model, bootstrapping with 1000 pseudoreplicates. The resulting tree was visualized and edited in MEGA6 (93).

The cytochrome P450 (CYP) gene family

Cytochrome P450, or CYP genes, constitute one of the largest gene families among all living organisms. They code for P450 enzymes that amongst a plethora of other functions have an important role in insecticide/xenobiotic metabolism (91). The insect CYP family can typically be divided into four major clans: CYP2, CYP3, CYP4 and mitochondrial CYPs. Based on a phylogenetic analysis with CYP sequences of *D. melanogaster, Ae. aegypti, An. gambiae* (Tables S5.1 and S5.2) and a set of CYP marker sequences from other organisms (91) the *Ae. albopictus* CYPs were assigned to appropriate CYP clans (Table S5.3, Figure S5.1). The *Ae. albopictus* genome contains a total of 186 full length CYP genes compared to 168, 104 and 87 in *Ae. aegypti, An. gambiae* and *D. melanogaster*, respectively (Table S5.3). Several large clusters of CYP genes are found in the *Ae. albopictus* genome (i.e. 13 on scaffold 64, CYP4 clan; 13 on scaffold 501, CYP3 clan; 10 on scaffold 4011, CYP3 clan). Similar to other Diptera, the majority of *Ae. albopictus* CYP genes belonged to the CYP3 and CYP4 clan (Table S5.3, Figure S5.2 to Figure S5.6) (89). These clans comprise insect specific families CYP4, CYP6, CYP9 and CYP325 that are well known for their involvement in environmental response/detoxification functions against xenobiotics and phytotoxins (91). Interestingly, it is exactly in two of these families that *Ae. albopictus* has a higher number of genes compared to the closely related *Ae. aegypti* (Table S5.4): 52 and 52 CYP genes were identified in the CYP6 and CYP325 clan of *Ae. albopictus* compared to 46 and 34 in *Ae. aegypti*, respectively. Our phylogenetic analyses confirm previous findings of limited orthologous relationship among members within the CYP3 and CYP4 clans (98). For instance, within the CYP4 clan, AalbCYP103 clustered with CYP4C50, CYP4C25 and Cyp4c3 of *Ae. aegypti*, *An. gambiae* and *D. melanogaster*, respectively. *Drosophila melanogaster* Cyp4c3 has not yet been characterized functionally, but an RNAi screen indicated that this P450 is vital (99). Remarkably, within the CYP4G clan we identified three CYP4G genes for each *Aedes* species while most insects only have one or two (100) (Figure S5.2, Figure S5.3). In *D. melanogaster Cyp4g1* is the most highly expressed of all *D. melanogaster* CYP genes and recently it was found that it encodes an enzyme that has a pivotal role in insect cuticular hydrocarbon synthesis (100). On the other hand, an ortholog of *D. melanogaster* Cyp4g15 in the wild silkmoth *Antheraea yamamai* (CYP4G25) was highly expressed during diapause in pharate first instar larvae (101).

In the CYP3 clan *D. melanogaster* Cyp308a1 clustered with AalbCYP001 and CYP6AH1 of *Ae. aegypti* and *An. gambiae* (Figure S5.4). *Cyp308a1*, however, seems not to be expressed in fruit flies (99), while its function in other dipteran species is unknown. Finally, all orthologues of the main pyrethroid metabolizers, members of the CYP9J family (within the CYP3 clan) in *Ae. aegypti* (102), are present in the *Ae. albopictus* genome with a 1:1 relationship (AalbCYP011 and CYP9J32; AalbCYP163/AalbCYPpseudo10 and CYP9J28/CYP9J24; AalbCYP164 and CYP9J26; Figure S5.4).

Contrary to the CYP3 and CYP4 clan, CYP genes in the mitochondrial and CYP2 clan are well conserved across Diptera. For the Halloween genes, of which gene products are involved in ecdysteroid biosynthesis (91, 103), clear orthologous relationships of shade (*D. melanogaster* Cyp314a1), disembodied (*D. melanogaster* Cyp302a1), spookiest (*Ae. aegypti* CYP307B1) could be

identified in all dipteran (shade, disembodied) and mosquito (spookiest) species included in our analysis (Figure S5.5 and Figure S5.6). Several *Ae. albopictus* CYP fragments showed tblastn hits with amino acid translations of other known Halloween genes [*phantom* (*D. melanogaster Cyp306a1*), *shadow* (*D. melanogaster Cyp315a1*) and *spook* (*D. melanogaster Cyp307a1*)] (Table S5.1) but their orthologous relationship with dipteran Halloween genes could not be derived from our phylogenetic analysis as only full length *Ae. albopictus* CYP genes and CYP pseudogenes were included (Table S5.1). A 1:1:1:1 relationship for *D. melanogaster* Cyp303a1 was found within the CYP2 clan in all Diptera and an *Ae. albopictus* orthologue (AalbCYP102) of the juvenile hormone epoxidase Cyp15b1 (104) could be identified.

The carboxyl/cholinesterase (CCE) gene family of *Ae. albopictus*

The carboxyl/cholinesterases (CCE) gene family catalyzes the hydrolysis of carboxylesters and displays a variety of physiological functions, such as neurone signalling, development and detoxification of xenobiotics (94, 105). Similar to GSTs and CYPs, CCEs have been shown to be involved in the detoxification of insecticides (106, 107). CCEs can be divided into 13 clades, which in turn can be organized into 3 classes: the dietary detoxification enzymes (clades A–C), the generally secreted enzymes (clades D–G) and the neurodevelopmental CCEs (clades I–M, mainly non catalytic esterases) (94, 105). We identified 64 full length CCE genes and 7 CCE pseudogenes in the *Ae. albopictus* genome (Table S5.4, Figure S5.7). The number of *Ae. albopictus* full length CCEs is similar to what is found in *Ae. aegypti* (59) but is higher than that in *D. melanogaster* (35) and *An. gambiae* (46) (Table S5.6). Based on a phylogenetic analysis with CCEs from *D. melanogaster, An. gambiae* and *Ae. aegypti*, the *Ae. albopictus* CCEs could be assigned to the different CCE clades (A-M) (94, 105). In nine cases [AalbCCE031/AalbCCE100 and AalbCCE101 (clade J), AalbCCE062 (clade K), AalbCCE061 (clade L), AalbCCEpseudo002 (clade M), AalbCCE005/050 (clade I), AalbCCE059/060 (clade M), AalbCCE034 (clade E) and AalbCCE011 (clade B)] clear orthologous relationships were identified between mosquito CCEs and those of *D. melanogaster* (Figure S5.7). Clade J is probably the best studied CCE clade, containing CCE genes coding for acetylcholinesterase (AChE) which is a key enzyme in the central nervous system and involved in organophosphate and carbamate resistance (89). Contrary to *D. melanogaster*, most insects have two AChEs and it was hypothesized that the two genes were derived from an old duplication before the split of the Arthropoda (108). Strikingly, three AChEs [AalbCCE031 and AalbCCE100, orthologs of Ace1, and AalbCCE101, ortholog of Ace2] could be identified in the *Ae. albopictus* genome compared to two in *Ae. aegypti* and *An. gambiae* (Figure S5.7). AalbCCE031 and AalbCCE100 are located on different scaffolds (329 and 13769), show 99.4 % nucleotide identity and have a nearly identical 4.5 kb downstream region. The possibility that AalbCCE031 and AalbCCE100 represent the same gene, due to a mis-assembly caused by allelic variants, cannot be excluded at this stage. AalBCCE013 and AalbCCE014 are both orthologs of CCEae3A, a gene that was recently strongly implicated in temephos resistance in *Ae. aegypti* (109), and are tandem duplicated on scaffold 178. Furthermore, within clade B multiple mosquito orthologues (10-18) were found of *D. melanogaster* cricklet, a CCE encoded by a gene located at a locus essential for mediating the response of adult tissues to juvenile hormone (110, 111) and of which allelic variants contribute to altitudinal variation in development time (112). Among the mosquito species, *Ae. albopictus* had the highest number of cricklet co-orthologues, with 5 cases where *Ae. albopictus* has 2-3 copies compared to one in *Ae. aegypti* (Figure S5.7). Similarly, an expansion of *Aedes* CCEs could be found in clade F with *Ae. albopictus* having the highest number of co-orthologues (9) of *D. melanogaster* juvenile hormone esterases FBpp0086362 and FBpp0086361. Finally, the number of glutactins (clade H) in both *Aedes* species is nearly double as high compared to those of *An. gambiae* and *D. melanogaster*. The function of glutactins however, is not well known (105, 113), but amongst others glutactins have been found to be a component of the diptera eggshell matrix (114, 115).

The glutathione-S-transferase (GST) gene family of *Ae. albopictus*

Glutathione S-transferases (GSTs) are a large family of multifunctional enzymes that mainly detoxify endogenous and xenobiotic electrophilic compounds through conjugation with reduced glutathione. Insect GSTs can be divided into two groups: cytosolic GSTs comprising seven known classes [delta

(δ), epsilon (ε), omega (ω), sigma (σ), theta (θ) and zeta (ζ)] and microsomal GSTs (88). We identified 36 full length GST genes and 5 GST pseudogenes in the *Ae. albopictus* genome, among them 32 cytosolic (Table S5.7, Figure S5.8), a number higher than those found in *Ae. aegypti* and *An. gambiae*, but lower than *D. melanongaster* (Table S5.9), which can be mainly attributed to the higher number of delta and epsilon class GSTs (Table S5.9, Figure S5.8), where all the majority of GSTs associated with resistance has been classified. Several tandem duplication events were suggested by the analysis in both delta and epsilon classes, with high nucleotide identity (> 90%) in some cases suggestive of a recent duplication event (Figure S5.8).

ABC transporter gene family in *Ae. albopictus*

ATP-binding cassette (ABC) proteins constitute one of the largest protein families and are present in all kingdoms of life. The majority of the ABC proteins function as primary active transporters, hydrolysing ATP to transport substrates across membranes. In addition to transporters, some ABC proteins function as receptor or are involved in translation. Metazoan ABC proteins are divided into eight (A to H) groups, of wich the ABCB full transporters (also named P-glycoproteins, P-gps, multidrug resistance proteins [MDRs]), ABCCs (also named multi drug resistance associated proteins [MRPs]) and ABCGs have been linked to xenobiotic resistance (97). We annotated 58 and 71 putative ABC genes in the genome of *Ae. aegypti* and *Ae. albopictus*, respectively (Table S5.10 and Table S5.11, Figure S5.9-S5.16). This number of *Aedes* ABC genes is higher than the 56 found in *D. melanogaster* and most other insect species of which the genome has been sequenced and the ABC gene family annotated (97) (Table S5.12). For both *Aedes* species we found clear orthologues of those ABC proteins that are considered as conserved in metazoan species [ABCB half transporters, ABCDs, ABCE, ABCFs, *D. melanogaster* CG7806 (ABCC), CG11069 and CG31121 (ABCG); (97); Figures S5.11-S5.15]. In the case of *Ae. albopictus* two copies of *D. melanogaster* CG2316 (ABCD) and CG9330 (ABCF) were found (Figure S5.13 and S5.14). In addition, other ABC proteins showing clear 1:1 orthology in most insects [*D. melanogaster* CG31731, CG34120 (ABCA), white, brown, scarlet, atet, CG3164, CG5853, CG17646 (ABCG), CG11147, CCG33970 and CG9990 (ABCH) (97)] also were detected in both *Aedes* species (Figure S5.9, S5.15 and S5.16). Similar to *Ae. aegypti* and *An. gambiae* only one member of the ABCB full transporter subfamily could be identified in *Ae. albopictus*, while four members are present in *D. melanogaster* (Figure S5.10). We also found clear *Aedes* orthologues for the *D. melanogaster* sulfonylurea receptor (SUR) within the ABCC subfamily, (Figure S5.12). However, only a fragment of the gene ~ 1.5 kb in length coding for this receptor was found in both *Aedes* species compared to full-length genes (~6kb) in *D. melanogaster* and *An. gambiae* (Table S5.12). This finding is consistent with recent data suggesting that *D. melanogaster sur* is not the true sulfonylurea-sensitive ABC transporter involved in chitin synthesis but that another sulfonylurea-sensitive ABC transporter must exist (97, 116). In our evolutionary analysis of the ABCC subfamily we also found 5 cases in which *Ae. aegypti* ABCC transporter genes are duplicated in *Ae. albopictus* (Figure S5.12). Interestingly, the majority of human ABCC transporters are well known for their role in multidrug resistance (117), nd as such these duplications might hint towards a possible higher "resistance potential" of *Ae. albopictus*. Similarly, we found six duplications of *Ae. aegypti* ABCG genes in the genome of *Ae. albopictus* (Figure S5.15). For most of these duplicated *Ae. albopictus* ABCG genes, the role of their counterparts in *D. melanogaster* is not known. In contrast, ABCG transporters in humans are well studied and are mainly involved in transport of endogenous and dietary lipids (118). In this light, the higher number of ABCG transporters in *Ae. albopictus* might be related to the complex regulation of increased lipid content in diapausing versus non-diapausing pharate larvae (81).

All *Ae. aegypti* ABC transporter gene sequences and *Ae. albopictus* sequences of genes involved in detoxification (CYPs, CCEs, GSTs, ABCs) that have been annotated in this study can be found in Table S5.13.

Odorant-binding and odorant receptor proteins

Bioinformatics analysis of odorant-binding proteins (OBPs) and odorant receptors (ORs)

Gene annotation, gene functional annotation and calculation of RPKM are described in the section "Genome properties and evolution". Verification of novel and other genes whose mRNA abundance could not be calculated with RPKM was performed using methods described in Deng et al. (2013) (119). Briefly, signal peptides (SP) present in the full-length conceptual translation products were predicted with SignalP4.0 Server (Version 4.0, http://www.cbs.dtu.dk/services/SignalP/) (120). The molecular weight (MW) was calculated using the ExPASy proteomics server (http://www.expasy.org/) (121), and PBP_GOBP motifs were identified using the NCBI conserved domains database with default setting (CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (122). Alignment of amino acid sequence to evaluate cysteine conservation was carried using Clustal W2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (123).

Predictions of olfactory proteins and construction of phylogenetic tree

OBP and OR genes of *Ae. aegypti*， *An. gambiae*, *Cx. quinquefasciatus* and *D. melanogaster* were downloaded from VectorBase. Blastp, Solar, GeneWise and InterProScan were used to identify *Ae. albopictus* sequences with 50% similarity and 50% coverage in protein domains, and these were interpreted to represent homologous relationships. Additionally, all of the predicted olfactory genes were used to construct a phylogenetic tree. The sensitivity of predicted genes and domains was calculated as: sensitivity = the number of predicted genes (or the number of predicted domain) / the number of genes derived from VectorBase.

Stage-specific gene expression profiles

OBP or OR genes were grouped arbitrarily based on relative transcriptional abundance into three expressional levels: 1 > 1, 2 = 0.1-1, and 3 < 0.1, corresponding to high, medium and low, respectively. The number of representative gene in each level of each specific stage was calculated, and the proportion was defined as that the number of gene in each level of each stage compared to the total number of gene in the same stage. Stacked bar chart was created using the R package (Version 3.1.1, http://www.r-project.org/).

The relative abundance of the transcriptional product of each OBP or OR gene was determined using mRNA samples derived from specific developmental stages: early (0-24 hours post-deposition [hpd]) and late (24-48 hpd) embryos, mixed 1^{st} and 2^{nd} instar larvae, mixed 3^{rd} and 4^{th} instar larvae, and adult males and females. The relative level is defined as the RPKM of individual OBP or OR gene in each specific stage/the total RPKM of individual OBP or OR gene in all stages, the result of which was presented as area chart using R-packages.

Bioinformatic predictions of olfactory genes and domains

Alignment of the olfactory genes derived from VectorBase with the *Ae. albopictus* genome using the criteria described above returned a sensitivity of OR prediction in *Ae. aegypti, An. gambiae, Cx. quinquefasciatus* and *D. melanogaster* of 93.5%, 100%, 100% and 96.7%, respectively, and that of 98.4%, 90.7%, 97.9% and 100%, respectively, for OBP prediction. Thus, the majority of genes can be found using the two parameters applied (Table S6.1). Additionally, protein domains were evaluated simultaneously using the same parameters, and not all of the predicted genes have protein function based on the predicted sensitivity level. However, most of the sensitivity in predicted genes and predicted domains exceeds 90%, thus, we set the parameter (50% sensitivity and 50% coverage) to further predict olfactory genes in *Ae.albopictus* using its homologs.

Comparison of OBP and OR between *Ae. albopictus* and mosquitoes and fruit fly

Comparisons of the numbers of OBP and OR genes in *D. melanogaster* with the four mosquito species, *An. gambiae*, *Ae. aegypti*, *Cx. quinquefasciatus* and *Ae. albopictus*, reveal that the fruit fly has fewer olfactory genes than the mosquitoes (Figure S6.1, Table S6.1). *Drosophila melanogaster* does not feed on blood and the larger number of OBP and OR genes may be an adaptation to hematophagy in the mosquitoes. Additionally, the differential blood feeding behavior among the mosquitoes (*Cx. quinquefasciatus* is a day-time, out-door feeder, *An. gambiae* is a night-time, indoor feeder, and *Ae. albopictus* is an aggressive outdoor, day-time feeder) may have influenced gene expansion in each group. *Aedes albopictus* has more annotated OBP and OR genes than other mosquitoes, and this may

contribute to its ability to adapt to varied and complex environments. Furthermore, although *Ae. aegypti* and *Ae. albopictus* share a number of genome features (124), they show different capabilities for large-scale geographical distribution.

Of the 158 OR and 86 OBP *Ae. albopictus* genes found in genome, only 112 OR and 83 OBP domains were found to have potential protein functions, which is higher than the other insects. Furthermore, phylogenetic analysis demonstrated that although the *D. melanogaster* olfactory genes showed some homology with mosquitoes, some of the OBP or OR were found only in the hematophagous-specific clades (Figure S6.2). However, some hematophagous-specific clades were dominated primarily by *Aedes* or *Culex*. Taken together, the bioinformatic analysis of olfactory genes derived from the genome analysis supports the conclusion that different life habits between *D. melanogaster* and the mosquitoes contribute to the differences in their respective olfactory genes.

Expressional profiles of AalbOBP and AalbOR

Most OBP or OR genes have no or low mRNA abundance during the embryonic developmental stages (Figures 3 and S6.3, Tables S6.2 and S6.3), but complexity and amount increase gradually following the exposure to the environment. Changing from the aquatic to the more complicated free-living environment is correlated with these increases, and is likely related to mating, oviposition and host-seeking behaviors.

Three AalbOBPs and 49 AalbORs with orthologs in *Ae. aegypti* had no transcriptional level in any tested developmental stages. The analysis of all predicted OBPs (containing the non-transcribed one) using CDD in NCBI supported their assignment to the PBP_GOBP family, and it is consistent with the GO and IPR prediction that they have odorant-binding activity and execute odorant-binding function (Table S6.4). Furthermore, CDD, GO and IPR analyses assigned all OR genes to the OR protein family with olfactory receptor function and sensory perception of smell and data published in Deng et al. (2013) (119). The non-expressed OBP or OR genes might have an undetectable transcriptional abundance or be not expressed in the sampled time at the developmental stages.

Identification of novel OBP and OR

A total of 43 putative OBP and two putative OR genes had no orthologs represented in PubMed and GenBank databases (Table S6.5). Most of these OBP genes have a predicted N-terminal signal peptide, a feature characteristic of these proteins (119, 125, 126). Molecular weights ranged from 14-41 kiloDaltons (kDa). CDD prediction showed that all of the novel OBPs belong to the PBP_GOBP family, and amino acid alignments confirmed the existence of six conservative cysteines, which is characteristic of OBPs (Figure S6.4). The putative OR genes were characterized as seven transmembrane domain proteins. These results support the conclusion that these are novel *Ae. albopictus* OBPs and ORs.

Sex-specific gene expression

cDNA library preparation and Illumina sequencing

Embryos were collected 0–24 (E1) and 24-48 (E2) hours after egg deposition by placing a damp collection cup within the cage. Larval samples were collected from combined 1st and 2nd instars (L1) and combined 3rd and 4th instar pools of mixed sex. Pupal samples (P) were collected from pools of pupae of varied ages. Male (M) and virgin female (F) adults were collected four days post-emergence. Total RNA was extracted with Trizol (Invitrogen Life Technologies, USA). The integrity of the RNA was verified by agarose gel electrophoresis and the RNA concentration was determined using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). The RNA quality for RNA-seq was verified further using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The cDNA library construction and sequencing were performed using Illumina protocols (Illumina, San Diego, CA, USA). The cDNA libraries were sequenced at BGI-Shenzhen, Shenzhen, PR China. The raw Illumina sequencing data generated are archived at the NCBI Sequence Read Archive database (SRA) (accession number SRA245721).

De novo assembly of sequencing reads

The raw reads of Illumina sequencing were preprocessed by removing adaptor sequences, low-quality reads (reads with ambiguous bases N), and duplicated sequences, and then assembled using SOAP2 denovo version 2 (5) with the default settings. Briefly, clean reads with a certain overlap length were combined to form longer fragments without Ns, which are called contigs, and then the reads were mapped back to contigs. Next, scaffolds were constructed using SOAPdenovo version 2 by connecting the contigs with Ns to represent unknown sequences between contigs in the same transcript. Gaps in the resulting scaffolds were filled by paired-end reads to obtain sequences with the least amount of Ns that cannot be extended on either end, which we defined as unigenes. The following analysis was performed on this set of unigenes (127, 128).

After removing short-length sequences and low-quality sequences containing more than 10% ambiguous 'N' nucleotides or 14 consecutive 'N' nucleotides, unigenes with a minimum length of 200 base-pairs (bp) were selected, and submitted to the NCBI Transcriptome Shotgun Assembly (TSA) database (http://www.ncbi.nlm.nih.gov/genbank/TSA.html, accession number GCLM000000000) and subjected to annotation analysis.

Statistical evaluation

RNA-seq expression profiles of 4 day sugar-fed virgin adult female and male mosquitoes were analyzed to identify differences in gene expression and corresponding transcript abundance between the sexes. The gene expression was normalized to Reads Per Kilobase per Million (RPKM) (11). For gene expression variance, the statistical *t*-test was used to identify genes expressed between libraries. *P* values were adjusted by the multiple testing procedures described by Benjamini & Yekutieli (2001) (129), by controlling the false discovery rate (FDR). In this study, we used a stringent value of FDR ≤ 0.001 and the absolute value of $|\log_2^{Ratio}| \leq 1$ as the threshold to judge a significant difference in gene expression. The correlation of the detected counts between parallel libraries was statistically assessed by the calculation of Pearson correlation coefficients.

Analysis of gene homologs involved in the *Ae. albopictus* sex determination cascade

Aedes albopictus doublesex, transformer 2, and *fruitless* genes were identified in the *Ae. albopictus* genome and transcripts, using a tblastn approach (86) with default options and *D. melanogaster* Transformer 2 (Swiss-Prot: P19018*)*, Doublesex (Swiss-Prot: P23023*)*, Fruitless (Swiss-Prot: Q8IN81*)* and Transformer (Swiss-Prot: P11596*)* sequences as queries.

Enrichment analysis of gene ontology (GO) functions

The DAVID functional annotation tool (http://david.abcc.ncifcrf.gov/) was used to perform GO classification and pathway annotation of genes with sex-biased expression (130). Functional annotation terms from the ontologies of "biological processes", "cellular compartments" and "molecular function" were recorded with an EASE threshold 0.1 and count threshold of 2. The enrichment score cutoff was set to 1.0.

Identification of genes exhibiting sex-biased expression

Analyses of the transcriptome profiles of adult sugar-fed female and male mosquitoes four days post-emergence revealed a total of 12,699 of 26,473 (47.97%) identified genes with sex-biased product abundance (FDR ≤ 0.001; $\log_2$ ratio ≥ 1). Of these, 8,559 and 4,140 had transcripts that were significantly higher in abundance in females and males, respectively (Table S7.1). Furthermore, genes with <2 RNA-seq alignments from females and males were selected and screened based on the log2 ratio of RPKM between female and male. Genes with RPKM $\log_2 \geq 10$ were considered as candidates for genes with sex-specific expression. In total, 268 and 246 sex-specifically expressed transcripts were identified in males (Table S7.2) and females (Table S7.3), respectively.

Genes in the sex-determination cascade

Sex-determination in *Aedes* species results from the activity of genes in the M-locus, which spans only a few megabases on one copy of what are otherwise homomorphic chromosomes in both sexes. Males contain this region whereas females do not and we were able to identify the *Ae. albopictus* ortholog of *Nix*, the Male-determining locus in *Ae. aegypti* (131). The sex-determination cascade in insects

represents an exquisite model of mechanisms of gene regulation that evolved hierarchically. A master gene at the top of the cascade determines the sex of each individual. In *Drosophila*, *Sex-lethal* (*Sxl*) is at the top of this regulatory cascade and its product controls the splicing of its own pre-mRNA as well as the splicing of the pre-mRNA from the downstream gene, *transformer* (*tra*) (132). The Tra product and the product of the constitutive gene *transformer-2* (*tra-2*) control the sex-specific splicing of the *doublesex* (*dsx*) pre-mRNA, which is transcribed in both sexes in *Drosophila* but gives rise to two different proteins, and these control aspects of sexual differentiation (133). We anticipate that there will be conservation of function in the mosquito orthologs of the *Drosophila* genes that are further downstream of the cascade. The orthologs were identified for a number of these genes in *Ae. albopictus*, including *dsx*, *tra-2*, and *fru* (Table S7.4). Remarkably, no *tra* ortholog was detected.

Gene ontology analysis of sex-specific genes

Gene ontology terms (GO) enriched significantly in the sets of sex-biased genes were identified and are listed in Tables S7.5-7.7. Genes with sex-biased expression are enriched for 26 biological process categories (BPGO), 11 Cellular Compartment categories (CCGO) and 34 molecular function categories (MFGO) ($P < 0.01$). The highest representations were in RNA metabolic processes, nucleus and ion binding (3.2%, $P = 5.99 \times 10^{-18}$; 9.5%, $P = 5.63 \times 10^{-49}$ and 34.4%, $P = 5.08 \times 10^{-37}$, respectively). The hierarchical network structure of the significant GO categories is visualized in Figure S7.1.

Expression profiles of sex-biased genes

Gene transcript abundance levels were expressed by RPKM in seven transcripts libraries representing early (0-24 hours post-deposition [hpd]) and late (24-48 hpd) embryos, mixed 1st and 2nd instar larvae, mixed 3rd and 4th instar larvae, pupae and adult males and sugar-fed virgin females. All transcript libraries, except those of the two adult stages, come from mixed-sex samples, so the mean value of expression of adult males and females were used in the adult stages. A total of 97 genes had coefficient of variation (CV) values stable throughout development (CV<0.2), and the remaining 30 were differentially-expressed (Table S7.8). Furthermore, we grouped sex-biased genes into 30 categories based on developmental expression profiles (Figure S7.2).

Immune-Related Genes

Gene predictions

The majority of genes with immune-related functions belong to families with multiple members that share high sequence similarity. This similarity makes it difficult to make predictions *de novo* on the total number of distinct genes based on primary sequence alone. A comparative approach was used to identify putative *Ae. albopictus* immune-related genes by looking at a number of other mosquito species, *Ae. aegypti*, *An. gambiae* and *Cx. quinquefasciatus*, and as an outgroup, the fruit fly, *D. melanogaster*. The immune protein sequences of *Ae. aegypti*, *An. gambiae*, *Cx. quinquefasciatus and D. melanogaster* were obtained from ImmunoDB (134) and the corresponding DNA aligned using GeneWise to the genomes of all of the species. We first filtered the alignments with an align rate lower than 50%. Next, best reciprocal matches (BRM) were identified first for the known immune genes. The high similarity complicates identifying all BRMs, and if an individual gene duplicated or if the family expanded after speciation, no BRM may exist for the new genes. Therefore, the predicted genes were aligned to self-species' predicted genes and we selected the identity based on the procedure where ≥BRM & identity≥ 80%. These combined analyses predicted 554, 476, 536, 400 and 345 immune-related genes for *Ae. albopictus*, *Ae. aegypti*, *An. gambiae*, *Cx. quinquefasciatus and D. melanogaster*, respectively (Table S8.1). *Aedes albopictus* has the largest number of predicted genes with members of the SPZ, BGBP, SRRP, GALE, TOLL, TOLLPATH, SOD, IMDPATH, PPO and CLIP families represented more in this species than any of the others.

Immune-related expansion genes and having species-specific polymorphisms

A maximum-likelihood phylogenetic tree was constructed using the amino acid sequences corresponding to each *Ae. albopictus* gene and this was used to identify branches containing large numbers of genes that we designate as "expansion genes" (Table S8.2, Figures S8.1, S8.2). Immune-related genes of the five dipteran species also were clustered by TreeFam and a multiple

sequence alignment for each cluster was generated by MUSCLE. Inspection of the alignments show a number of specific amino acid insertions found only in the *Ae. albopictus* genes in the APHAG16 (11 aa), CASPA1 (3 aa), CLIP39 (5 aa), SCR5 and SCR18 (3-4 aa) (Figure S8.2). A deletion (3 aa) is present in the CASPA3-D1 genes and SRPN22-D1 has an insertion of 4 aa and a deletion of 3 aa also found in *Ae. aegypti* and *Cx. quinquefasciatus*. The polymorphisms were verified by the transcripts which were assembled by Tophat and Cufflinks. Through predicting 3D model of the proteins by SWISS-MODEL (http://swissmodel.expasy.org/), we found that APHAG16, CASPA3 and CLIP39 of *Ae. albopictus* have changed the 3D model (Figures S8.3-S8.5).

Expression profiles of immune-related genes

Transcriptome analyses of mRNAs derived from seven samples, combined 1st- and 2nd-instar larvae, combined 3rd- and 4th-instar larvae, mixed sex pupae, adult males, adult females, and eggs 0-24 and 24-48 hours after deposition, were performed to determine immune-related gene activity during development. CLUSTER (version 3.0) (135) analyses distanced the two egg samples away from the larval, pupal and adult stages (Figure S8.6). A total of 468 immune-related transcripts were predicted the adult mosquitoes, including 166 related to immune recognition, 106 involved in gene modulation, 100 in signal transduction and 96 in effector molecule (Figure S8.7). The top three most abundant transcripts are effector AMP (56), recognition LRR (53) and modulation CLIP (51).

Figures and Tables

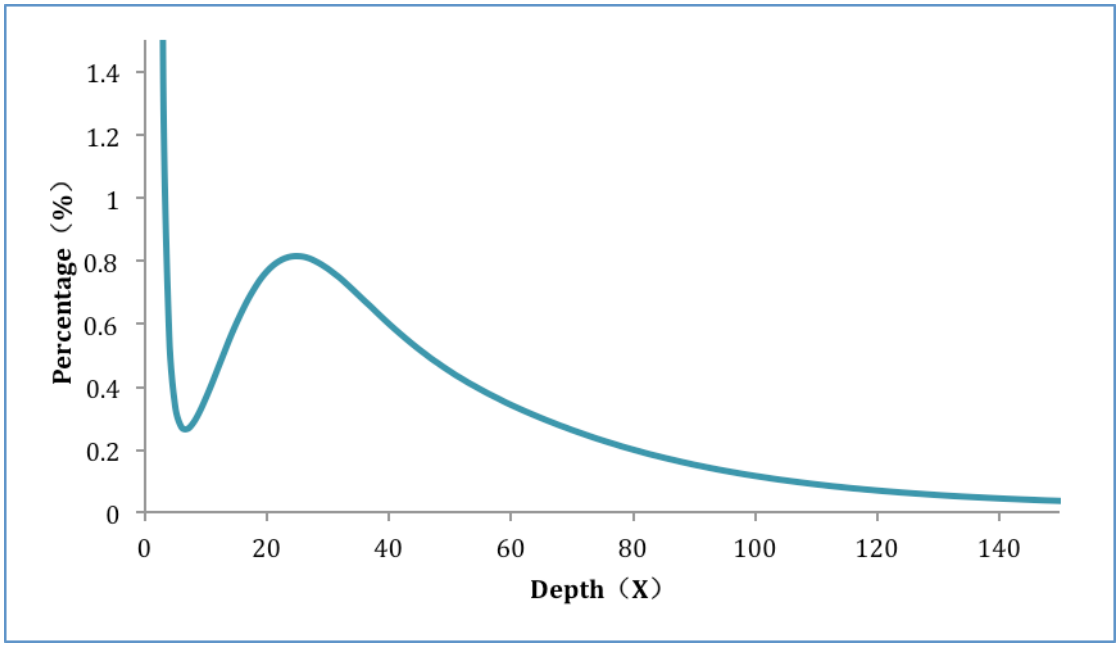

Figure S1.1.
17-kmer estimation of genome size. The genome size of *Aedes albopictus* was estimated to be 2.91 Gb based on reads from short insert size libraries.

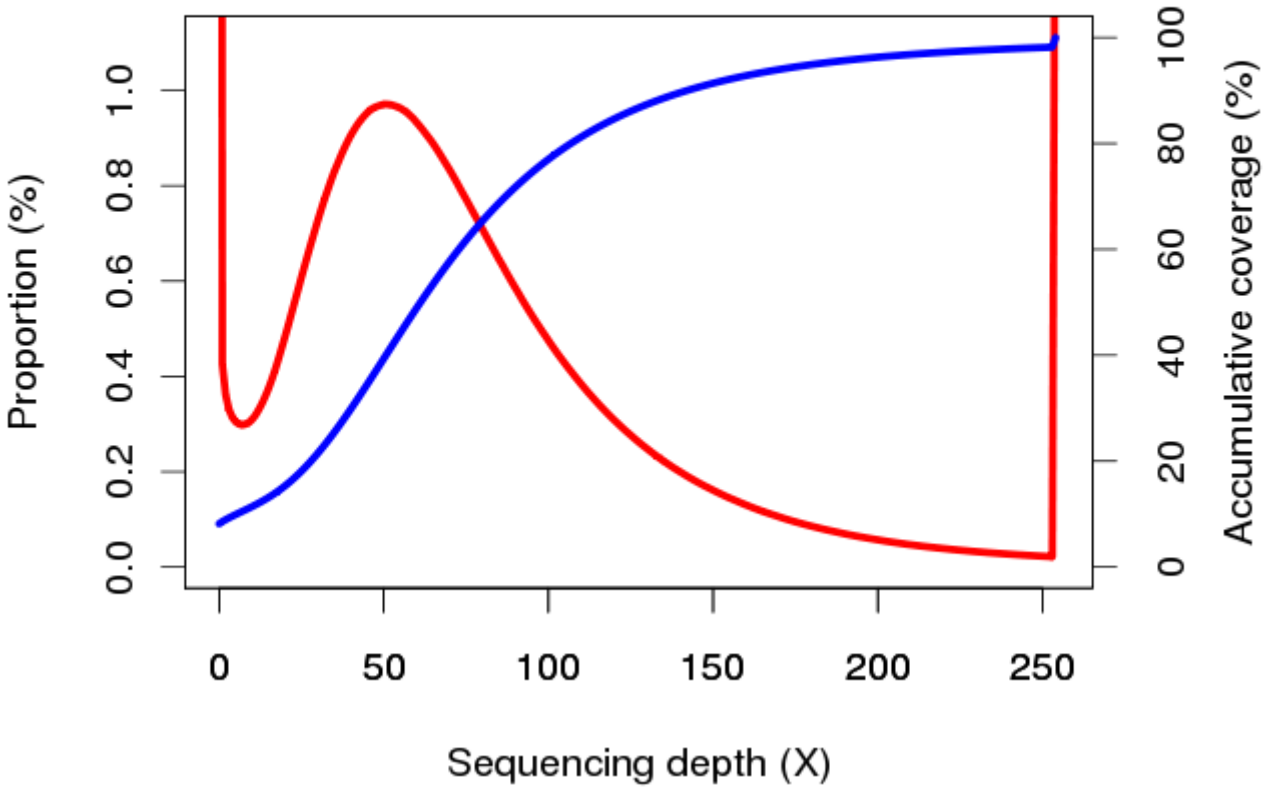

Figure S1.2.
Distribution of sequencing depth. The X-axis represents sequencing depth and the Y-axis represent the proportion of total bases at a given depth (left) and the frequency for each of the covered genome bases (right).

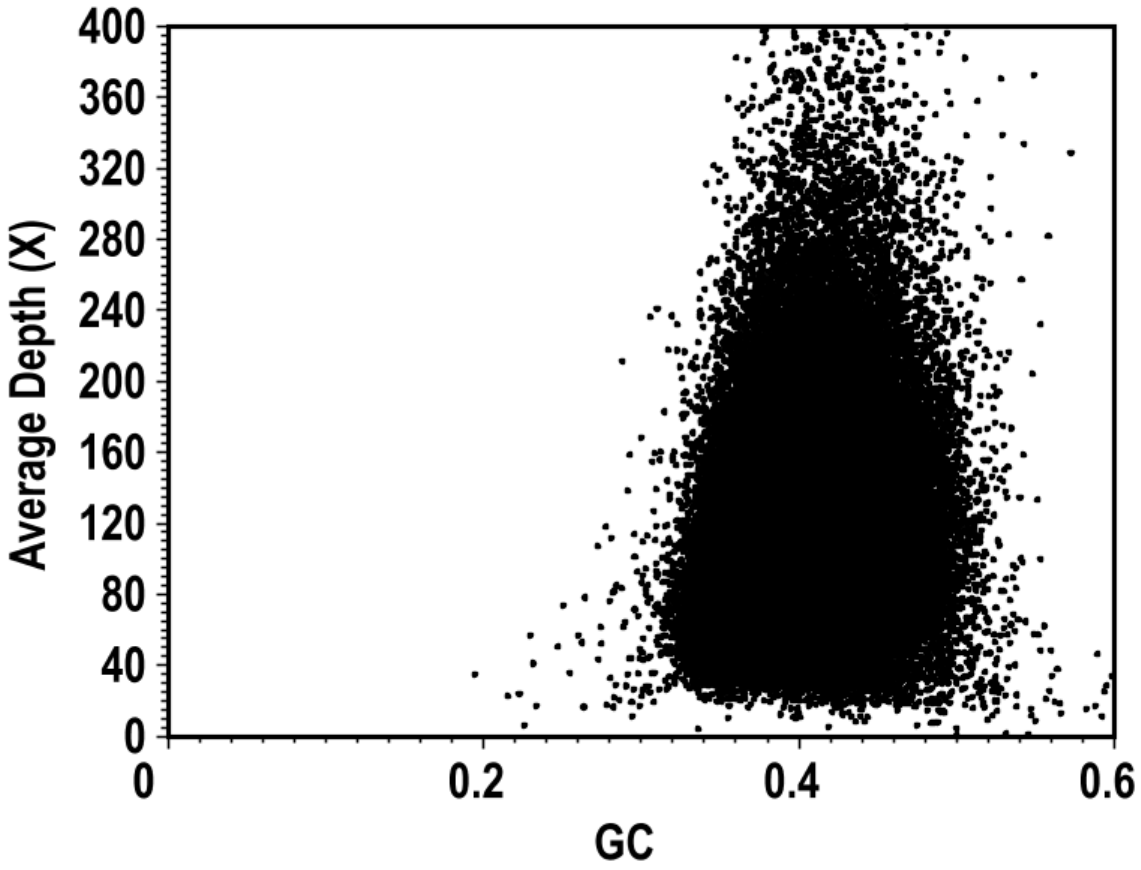

Figure S1.3.
Correlation between GC content and sequencing depth. The X-axis represents GC content; the Y-axis represents average sequencing depth. We used 10kb non-overlapping sliding windows and calculated the GC content and average depth among these windows.

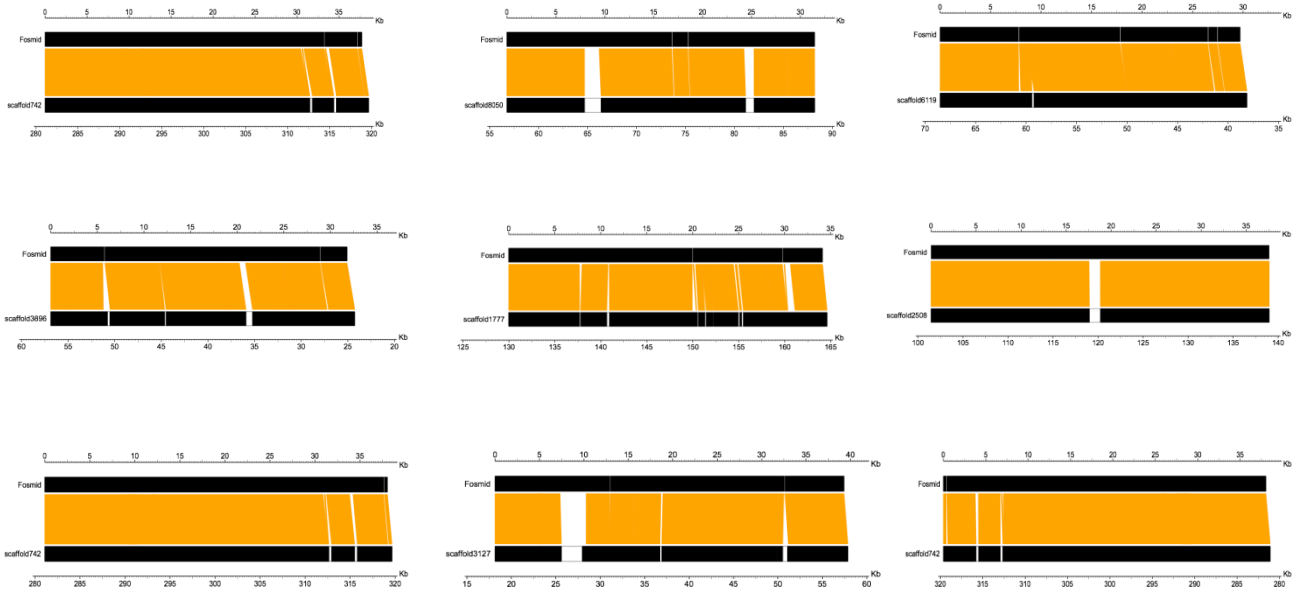

Figure S1.4.
Comparison of the assembled genome with Fosmid sequences. The orange blocks represent the alignments between fosmids and scaffolds/contigs and the blank regions are unmapped gaps.

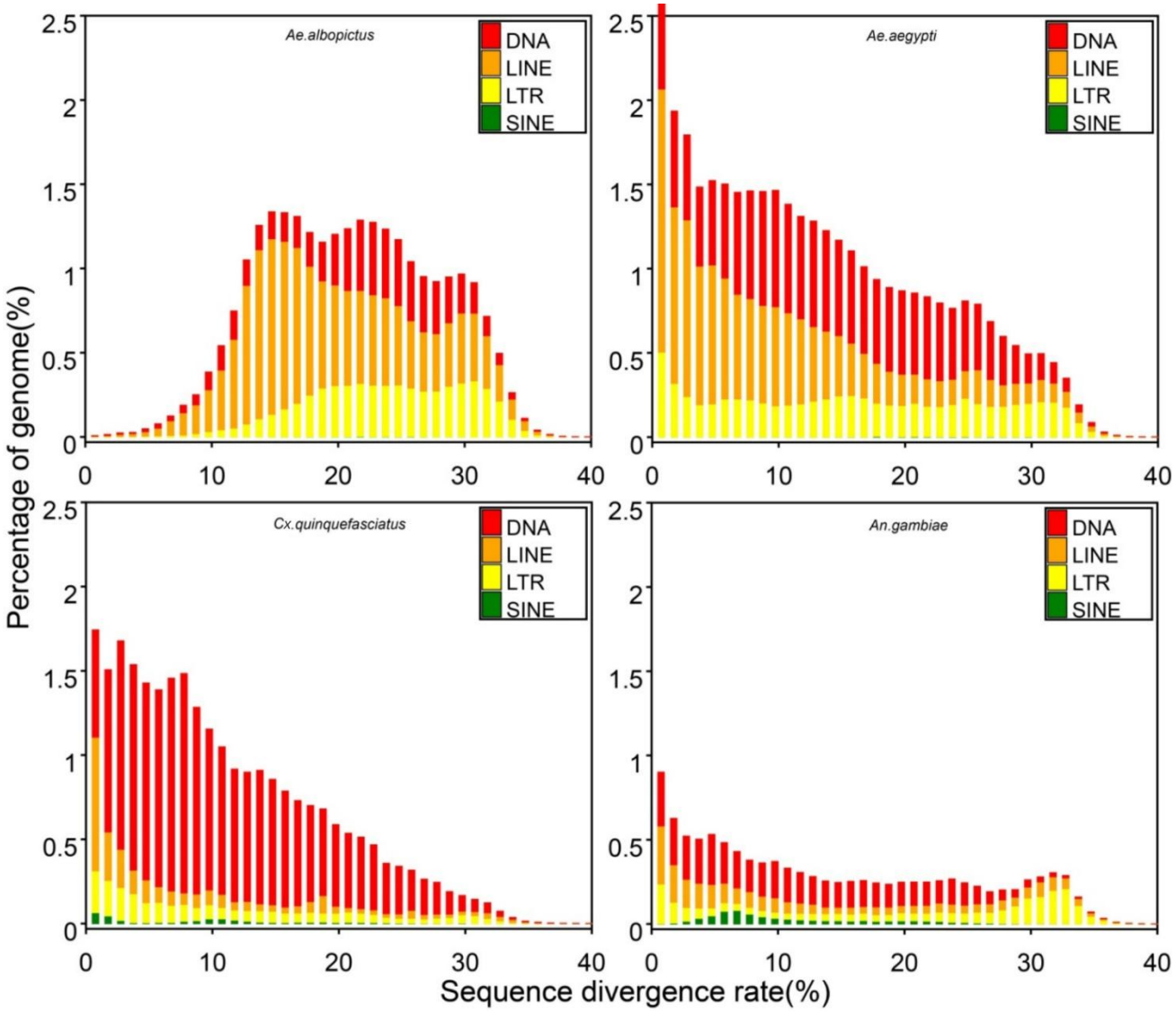

Figure S1.5.
Divergence distribution of classified TE families. Transposon types in *Ae. albopictus*, *Ae. aegypti*, *Cx. quinquefasciatus* and *An. gambiae* genomes aligned with consensus sequences in Repbase (8).

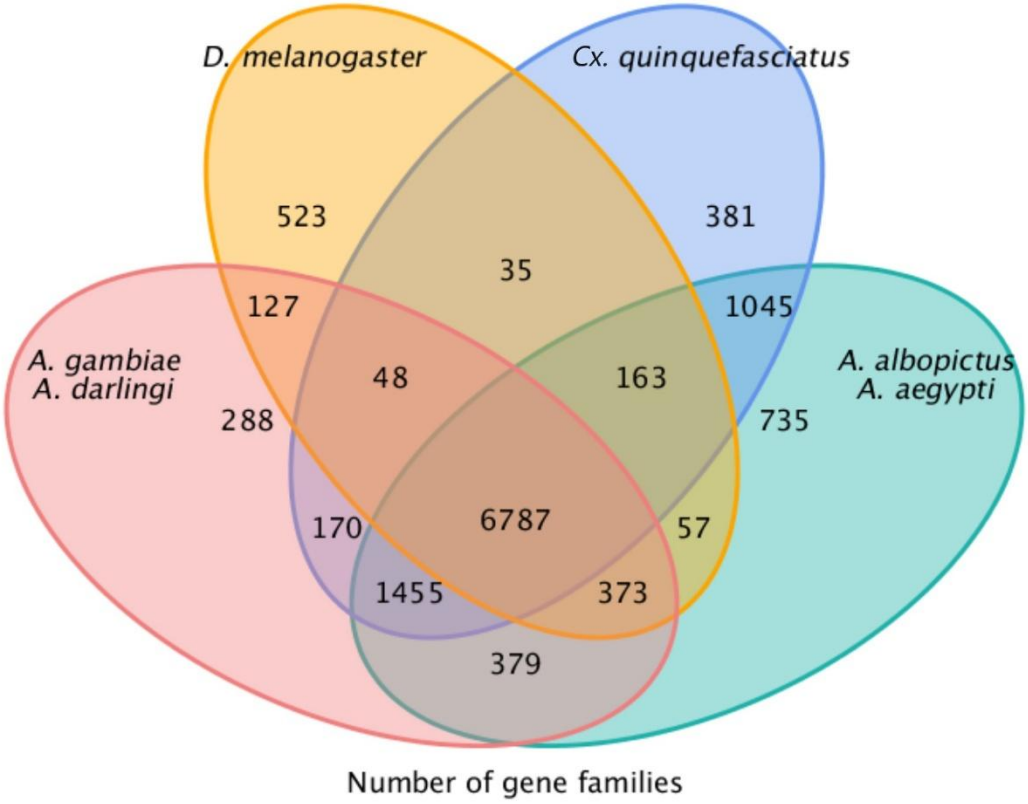

Figure S1.6.
Protein orthology comparisons among genomes of five mosquitoes, *Aedes aegypti, Aedes albopictus, Anopheles darlingi, Anopheles gambiae* and *Culex quinquefasciatus,* and *Drosophila melanogaster*. Predicted proteins were subjected to blastp analysis, and those with similarity of e=1e-5, and >30% of the aligned sequence with identity >30% in other species, were designated the ortholog for each *Ae. albopictus* protein. A total of 6,787 orthologous gene groups could be partitioned among the mosquito and fruit fly species.

Duo

GGCACGGTAAAGGCTGGGTATGCTGCGCAATTCAGATCCCATTGTGATCCACTAGCCTCT
GCCCAGCAACTCCTATCCCTACCTCCNCGCGGTACCGGCCGGAAACTACGAGCAACCTTA
GGGAAGATCGGGTAACCAACCCCGGTGGGAACTTTGGTCGTAGGCTGACAGGGAAGGGGG
GGGGTTTGCTTCGGCAAACCTGAGCGTCTGTTCTCCAGGAGGAGCGGCTCACAACAGCGT
CTGATCCCCATGTTAGGGGCGGCTGATCTACGTCCGAGTGCCAGGGAAGGACTCTAAGCT
CAACTGTGCACTATGGTCCTCCGGAAAGTAGGGGGTTGGTGTCAGGCCCTACGAGCCAGC
CGTAAAAAACCATTGTAACGGAAAATCAGCAACAGAATAATACGAACCGAGACCAACGGC
AACGACCCCAGCGAACAAAAAGGACTTGCGATTGGAAACTCGGTACGTGGAACTGCCGAT
CTCTCAACTTCATCGGGAGCACCCGCATACTCGCCGATCTACTGAAGGACCGCGGGTTCG
GCATCGTAGCGCTGCAGGAGGTGTGTTGGACAGGATCCATGGTGCGAACGTTTAGAGGTA
ATCATACCATCTACCAGAGCTGCGGCAACACACGCGAGCTGGGAACAGCTTTCATCGTGA
TGGGCGATATGCAGAGGCGCGTGATCGGTTGGTGGCCGATCGACGAAAGAATGTGCAGGT
TGAGGATCAAGGGCCGATTCTTCAACTTCAGCATAATAAACGTGCACAGCCCACACTCCG
GAAGCACTGATGATGACAAGGACGCATTTTACGCGCAGCTCGAACGCGAGTACGACCGCT
GCCCAAGCCACGACGTCAAGATCATCATAGGAGATTTGAACGCTCAGGTAGGCCAGGAGG
AGGAATTCAGACCGACGATTGGTAAGTTCAGCGCCCACCAGCAGACGAACGAAAACGGCC
TACGACTCATCGATTTCGCCGCCTCCAAAAATATGGCCATACGTAGCACCTTTTTCCAAC
ACAGCCTCCCTTATCGTTACACCTGGAGATCACCACAGCAGACGGAATCTCAAATCGACC
ACGTTCTGATTGACGGACGGCACTTCTCCGACATTATCGACGTCAGGACCTATCGTGGCG
CCAACATCGACTCCGACCACTATCTGGTGATGGTCAAACTGCGCCCAAAACTCTCCGTCA
TCAACAATGTACGGTACCGGCGACCGCCACGGTACAACCTAGAGCGACTGAAGCAACCGG
ATGTCGCCTCAGCATACGCGCAGAATCTCGAGGCCGCGTTGCCAGACGAGGGCGAGCTCG
ATGAGGCCCCTCTAGAGGACTGCTGGAGTACAGTGAAAGCAGCCATCAACGACGCAGCCG
AGAGCACCATCGGGTACGTGGAACGGAATCGACGGAACGAATGGTTCGACGAAGAGTGCA
GAACGGTTTTGGAGGAGAAGAACGCAGCGAGGGCGGTAATGCTGCAGCAAGGGACCCGAC
AGAACGTGGAACGTTACAAACAGAAGCGGAAACAGCAGACCCGCCTCTTTCGGGAGAAAA
AGCGCCGCCTGGAAGAAGCGGAGTGCGAAGAAATGGAACTGCTGTGCCGTTCCCAAGAAA
CACGGAAGTTCTATCAGAAGCTCAACGCATCCCGCAACGGCTTCGTGCCGCGAGCCGAAA
TATGCAGGGATAAAGACGGAGGCCTCTTGACGGACGGACGTGAGGTGATCGAAAGGTGGA
AGCAGCACTTCGATCAGCACCTGAACGGCGTGGAGAACGTAGGCACGGGAGACCACGGCA
ACGGAGGAAACGACGACGCCGGTGCAGCGGAGGACGGAAACGAACCAACTCCCACGCTGA
GGGAAGTTAAGGATGCCATTCACCAGCTCAAAACCAACAAAGCGGCTGGTAAGGACGGTA
TCGCAGCTGAACTCATCAAGATGGGCCCGGAAAAGTTGGCCACCTGTCTGCATCGGCTGA
TAGTCAGGATCTGGGAAACCGAACAGCTACCGGAGGAGTGGAAGGAAGGGGTAATCTGCC
CCATTCACAAGAAAGGCGACCATTTGGAATGTGAGAACTTCAGGGCGATCACTATTTTGA
ATGCCGCCTACAAAGTGCTATCCCAGATCATCTTCCGTCGTCTGTCACCTAAAACGAATG
AGTTCGTGGGAAGTTATCAAGCCGGCTTCATCGACGGCCGGTCGACAACGGACCAGATCT
TCACCGTACGGCAAATCCTCCAGAAATGCCGTGAATACCAGGTCCCAACGCATCACCTGT
TCATCGACTTCAAAGCGGCATACGACAGTATCGACCGCGCAGAGCTATGGAGAATCATGG
ACGAAAACGGCTTTCCCGGGAAGCTGACTAGACTGATTAAAGCAACGATGGACGGTGTGC
AAAACTGCGTAAGGGTTTCGGGTGAACTATCCAGTTCATTCGAATCTCGCCGGGGACTGC
GACAAGGTGACGGACTCTCATGCCTACTCTTCAACATCGCTCTGGAAGGTGTGATGCGAC
GAGCCGGGCTCAACAGCCGGGGAACGATTTTCACGAAATCCGGTCAATTTGTGTGCTTTG
CGGACGACATGGACATTATCGCCAGAACATTTGGAACGGTGGCAGAGCTGTACACCCGCC
TGAAACGCGAAGCAGCAAAGGTCGGACTGGTGGTGAATGCCTCAAAAACAAAGTACATGC
TGGTAGGCGGAACCGAACGCGACCGGATCCGTCTGGGTAGTAATGTTACGATAGACGGGG
ATACTTTCGAGGTGGTGGAGGAATTCGTCTACCTCGGATCCTTACTGACGGCTGACAACA
ACGTGAGCCGTGAAATTCGGAGGCGCATCATCAGCGGAAGTCGGGCCTACTACGGGCTCC
AGAAGAAACTGCGGTCGAAAAAGATTCACCCACGCACCAAATGCACCATGTACAAGACGC
TAATAAGACCGGTGGTCCTCTACGGACACGAGACATGGACCATGCTCGAGGAGGACCTGC
AAGCACTCGGAGTTTTCGAGCGACGCGTGCTAAGGACGATCTTCGGCGGCGTGCAGGAGA
ACGGTGTGTGGCGGAGAAGGATGAACCACGAGCTCGCTGCACTTTACGGCGAACCCAGCA
TCCAGAAGGTGGCCAAAGCCGGAAGGATACGGTGGGCAGGGCATGTTGCAAGAATGCCGG
ACAACAACCCTGCAAAGCTGGTGTTTGCAACCGATCCGGTTGGCACAAGAAGGCGTGGAG
CGCAGAGAGCACGATGGGCGGACCAGGTGGAGCGTGACCTGGCGAGCGTTGGGCGCGACC
GAGGATGGAGAGCGGCAGCCGCAAACCGAGTATTGTGGCGNANTATTGTTGATTCNGTNT
TGTCTTGAATTTGATGTTGAACAAATAAATGTATG

Figure S2.1.
The most abundant LINE element in *Ae. albopictus*.

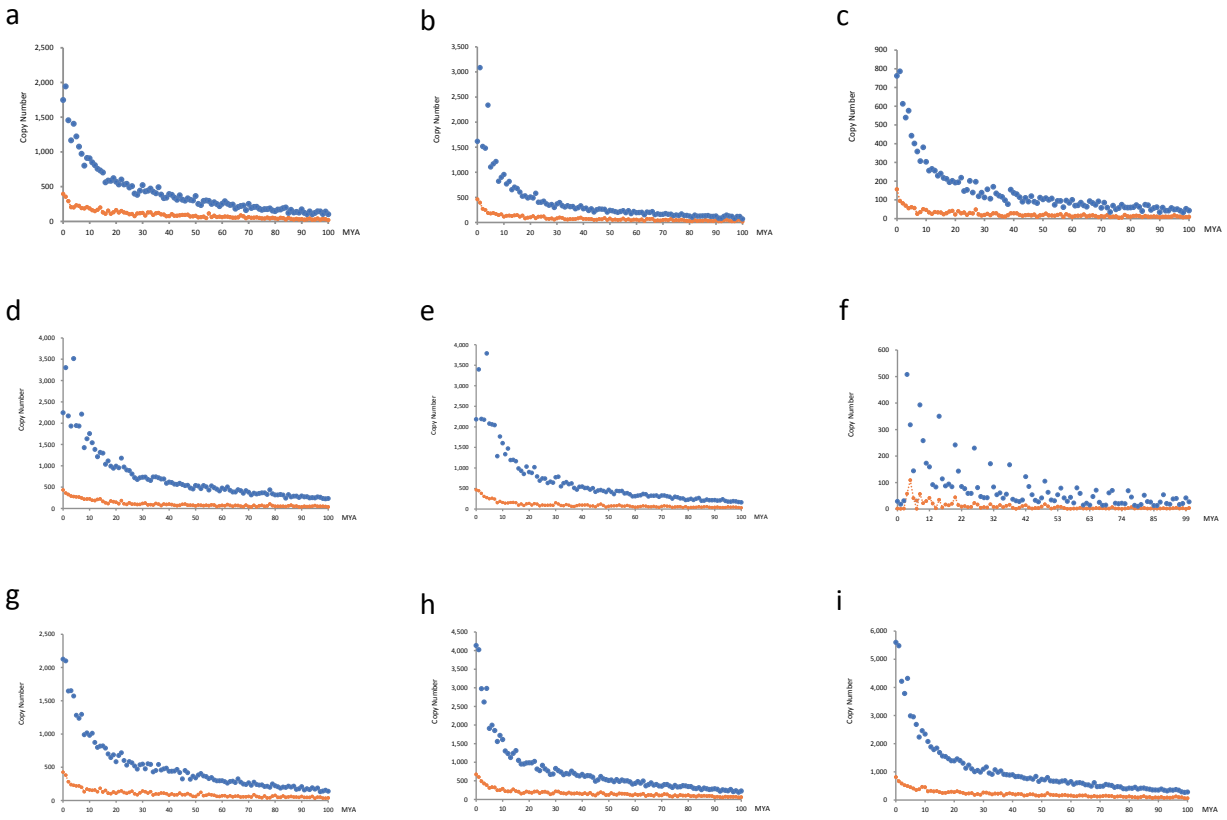

Figure S2.2.
Number of TE insertions relative to the estimated time of insertion. AEDALB (blue lines) indicate *Aedes albopictus* TEs while AEDAGE (orange lines) are *Ae. aegypti* TEs. Shown here are LINE/CR1(a), LINE/I(b), LINE/L2(c), LINE/LOA(d), LINE/R1(e), LINE/RTE(f), LTR/Copia(g), LTR/Pao(h) and LTR/Gypsy(i) elements.

Figure S2.3.
Non-LTR retrotransposon consensus sequences in *Aedes albopictus*. These sequences were used as repeat library to mask their corresponding genomic sequences by RepeatMakser (http://www.repeatmasker.org/) to generate pairwise alignment files for deletion rate analysis.
See additional data file S1 (separate file).

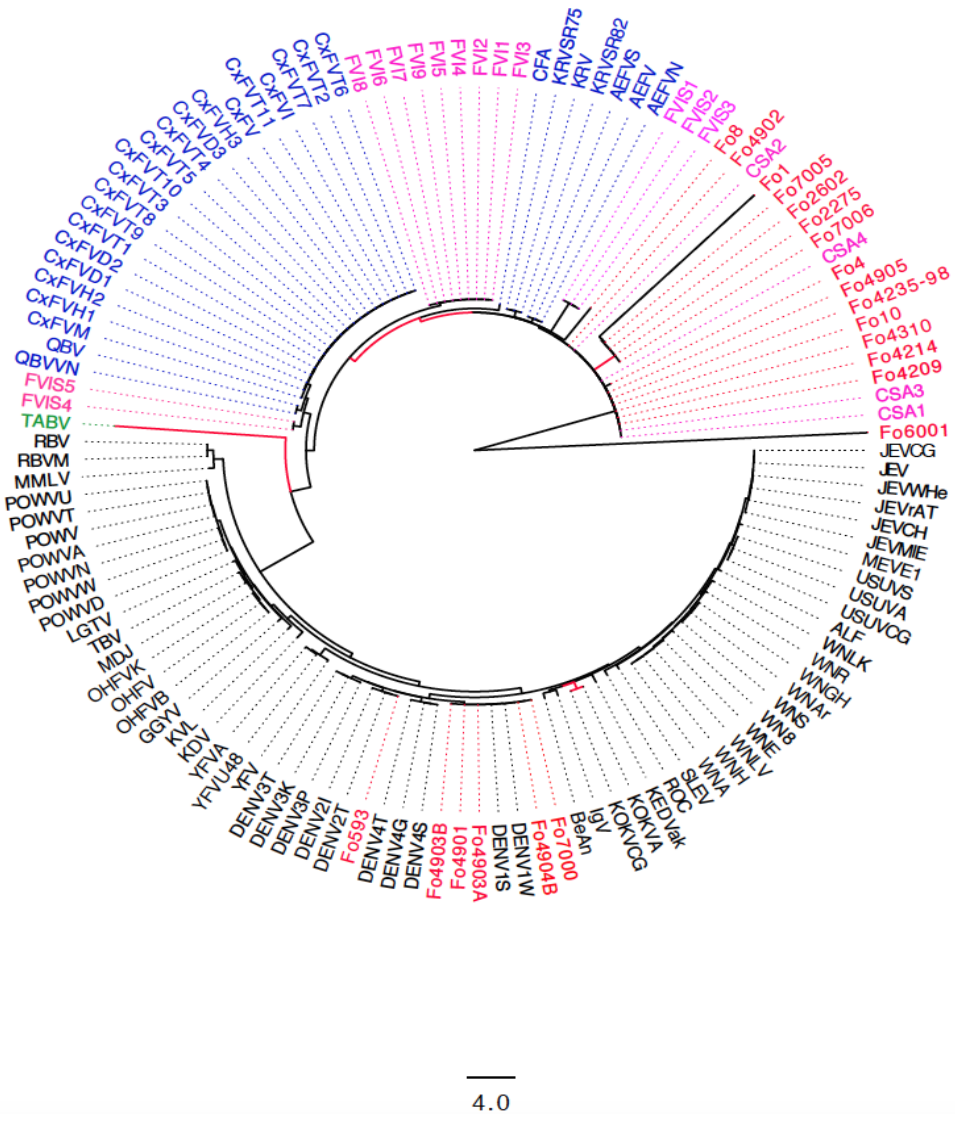

Figure S3.1

Phylogenetic relationships between NIRVs and representative members of the flavivirus lineage. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (136). The tree with the highest log likelihood (-429325.6628) is shown. Nodes where less than 50% of trees in which the associated taxa clustered together are shown in red. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter = 2.2133)]. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 129 nucleotide sequences. Codon positions included were Noncoding. All ambiguous positions were removed for each sequence pair. There were a total of 11851 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [39]. MBVs, TBVs and NBVs are in black, ISFs are in blue, previously identified NIRVs in pink and NIRVs identified from the *Ae. albopictus* genome assembly of the Foshan strain are red. The recently discovered Tamana Bat Virus (TABV) is shown in green. Abbreviations are as reported in Table S3.1.

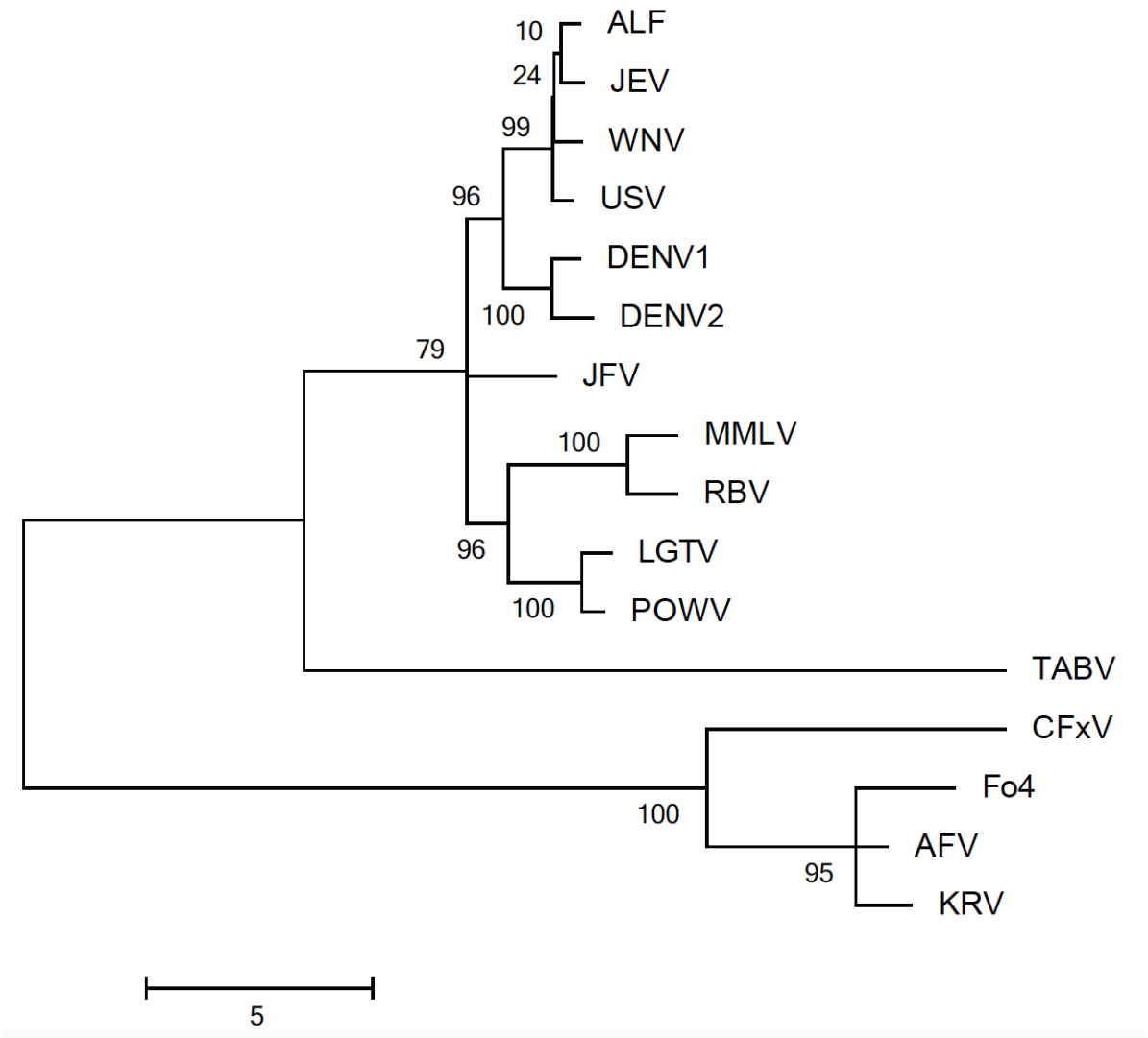

Figure S3.2

Phylogenetic relationship between NIRVs encompassing the envelope gene and the E-encoding gene of representative members of the flavivirus lineage. Functional annotation of the Foshan viral integrations identified an ORF for the flaviviral E protein in Fo4 (Table S3.4). This NIRV was aligned to the E encoding sequence from 14 flaviviruses and TABV. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones et al. w/freq. model (137). The tree with the highest log likelihood (-24048.1276) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.4863)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 amino acid sequences. There were a total of 1625 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (41).

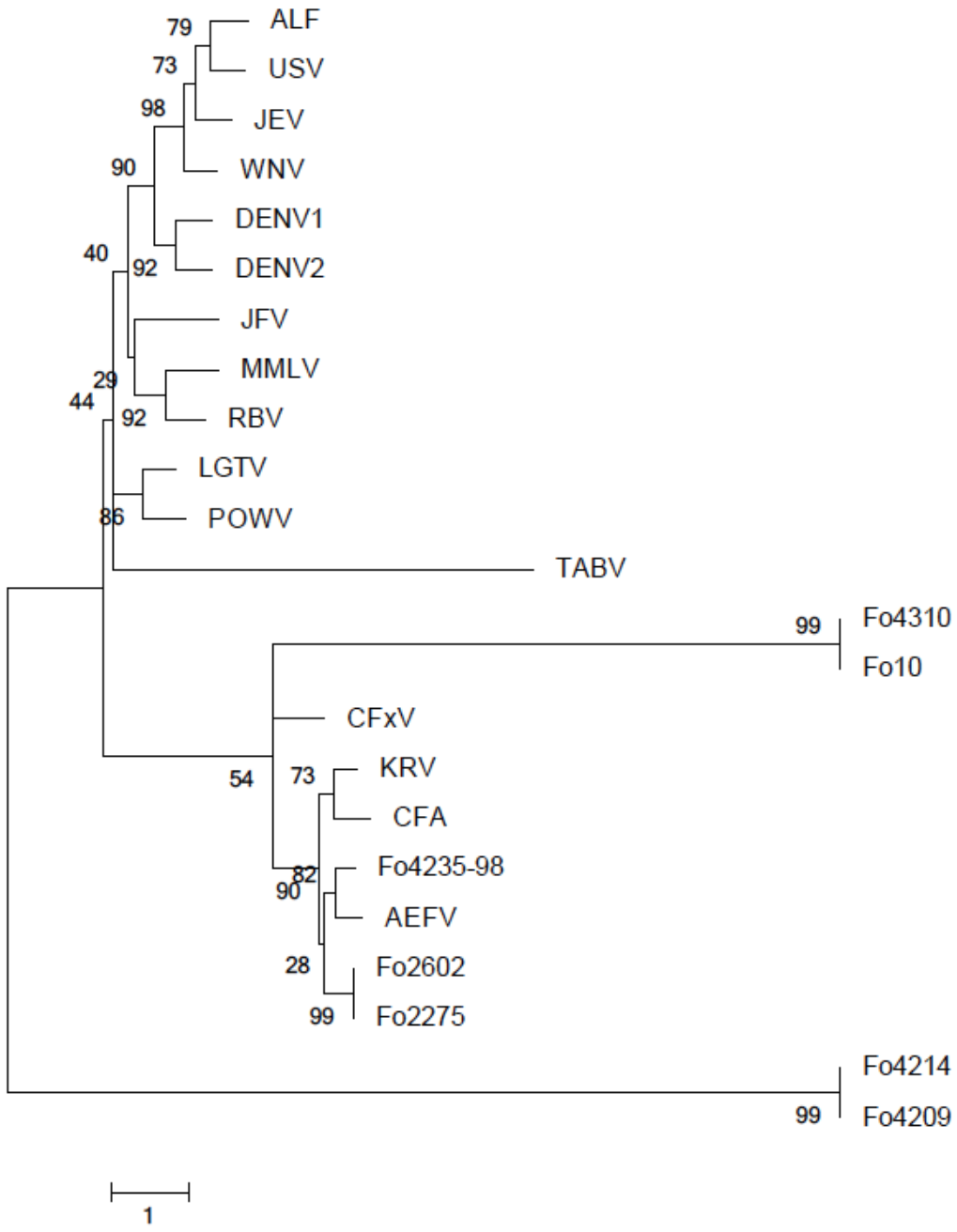

Figure S3.3

Phylogenetic relationship between NIRVs encompassing NS1 and the NS1-encoding gene of representative members of the flavivirus lineage. Functional annotation of the Foshan viral integrations identified ORFs for the flaviviral NS1 protein in 7 NIRVs (i.e. Fo4235-98, Fo4310, Fo10, Fo2602, Fo2275, Fo4214, Fo4209, Table S3.4). These NIRVs were aligned to the NS1 encoding sequence from 15 flaviviruses and TABV. The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan and Goldman model (138). The tree with the highest log likelihood (-9705.2976) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT

model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 4.2172)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 23 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 488 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (41).Abbreviations are as reported in Table S3.1.

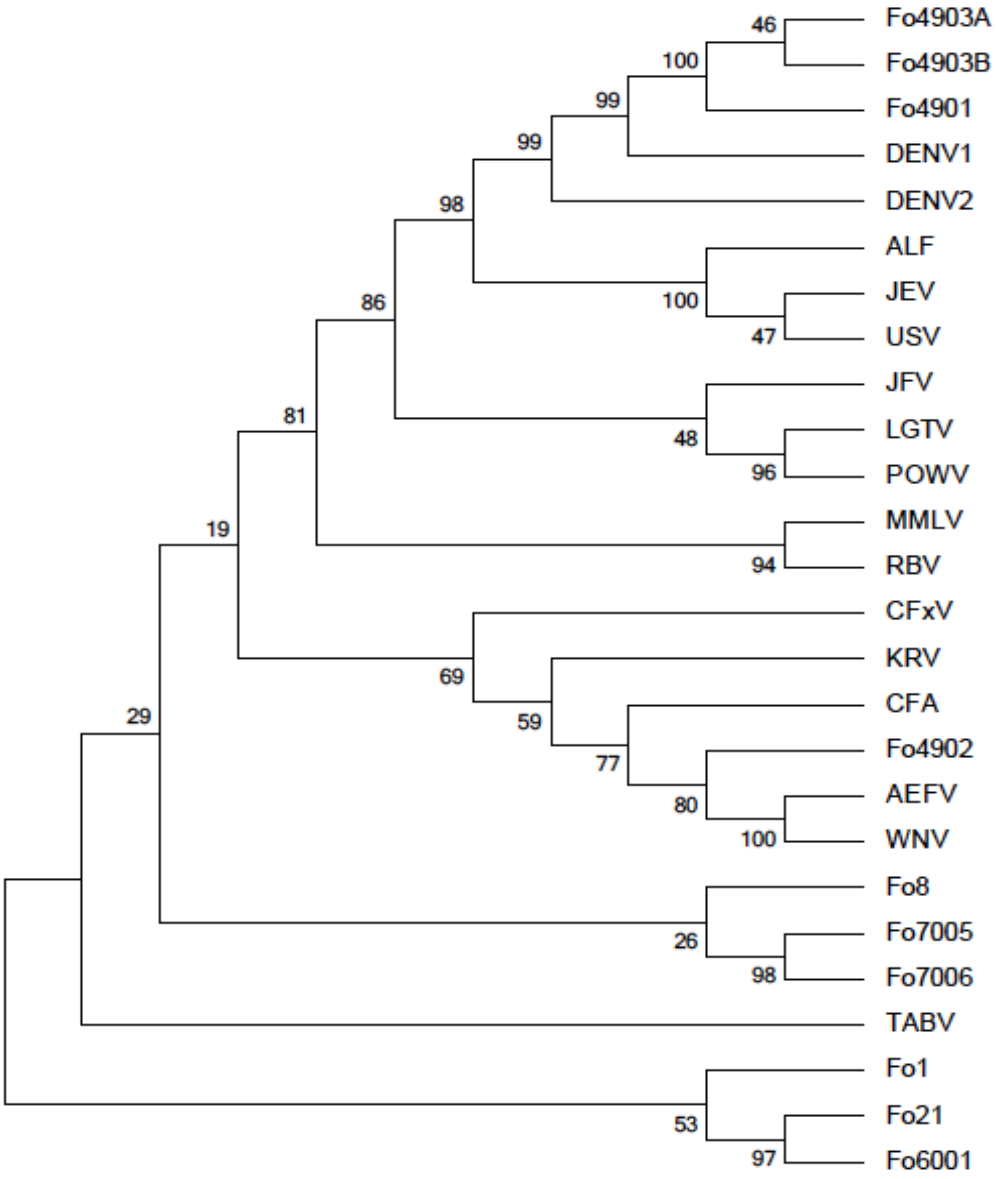

Figure S3.4

Phylogenetic relationship between NIRVs encompassing NS5 and the NS5-encoding gene of representative members of the flavivirus lineage. Functional annotation of the Foshan viral integrations identified ORFs for the flaviviral NS5 protein in 10 NIRVs (i.e. Fo4901, Fo4903A, Fo4903B, Fo8, Fo21, Fo1, Fo7005, Fo7006, Fo6001, Table S3.4). These NIRVs were aligned to the NS5 encoding sequence from 15 flaviviruses and TABV. The tree with the highest log likelihood (-21387.5132) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 3.1297)). The analysis involved 26 amino acid sequences. The coding data was translated assuming a N/A genetic code table. All ambiguous positions were removed for each sequence pair. There were a total of 656 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [39]. Abbreviations are as reported in Table S3.1.

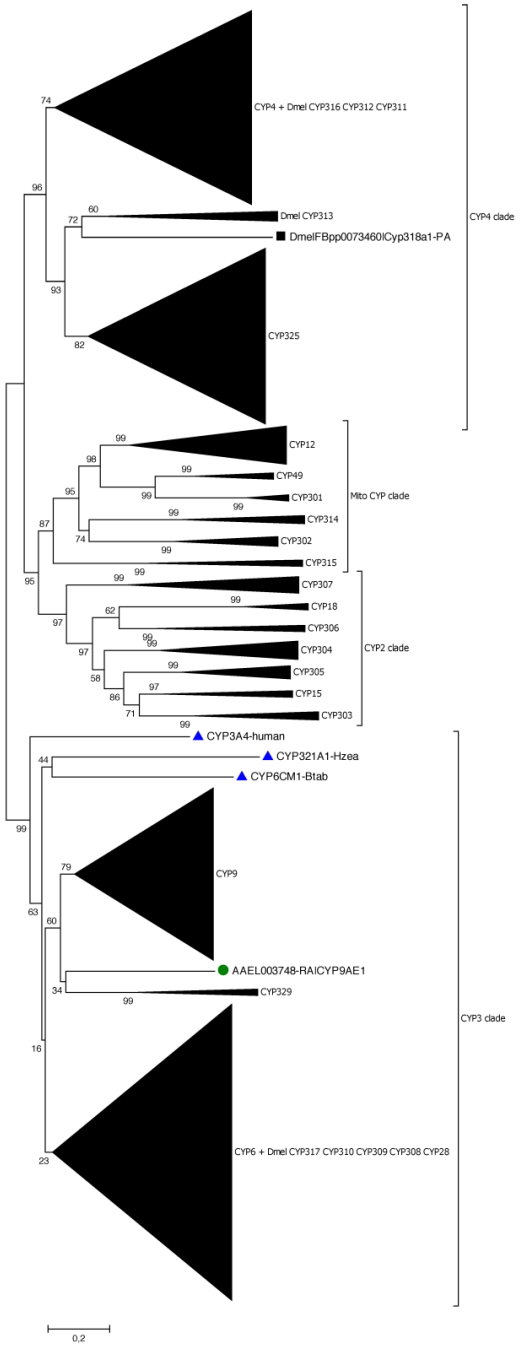

Figure S5.1.

Phylogenetic relationships between P450s of *Ae. albopictus, Ae. aegypti, An. gambiae* and *D. melanogaster*. The evolutionary history was inferred using the Neighbor-Joining method, bootstrapped with 1000 pseudoreplicates (bootstrap values are shown next to the branches) and the resulting tree was midpoint rooted. The scale bar represents the number of substitutions per site. Color and shape codes are as follows: black square, *D. melanogaster*, red dot, *Ae. albopictus*, green dot, *Ae. aegypti*, yellow dot, *An. gambiae* and blue triangle, CYP marker sequence. For accession numbers see Table S5.1 and Table S5.2.

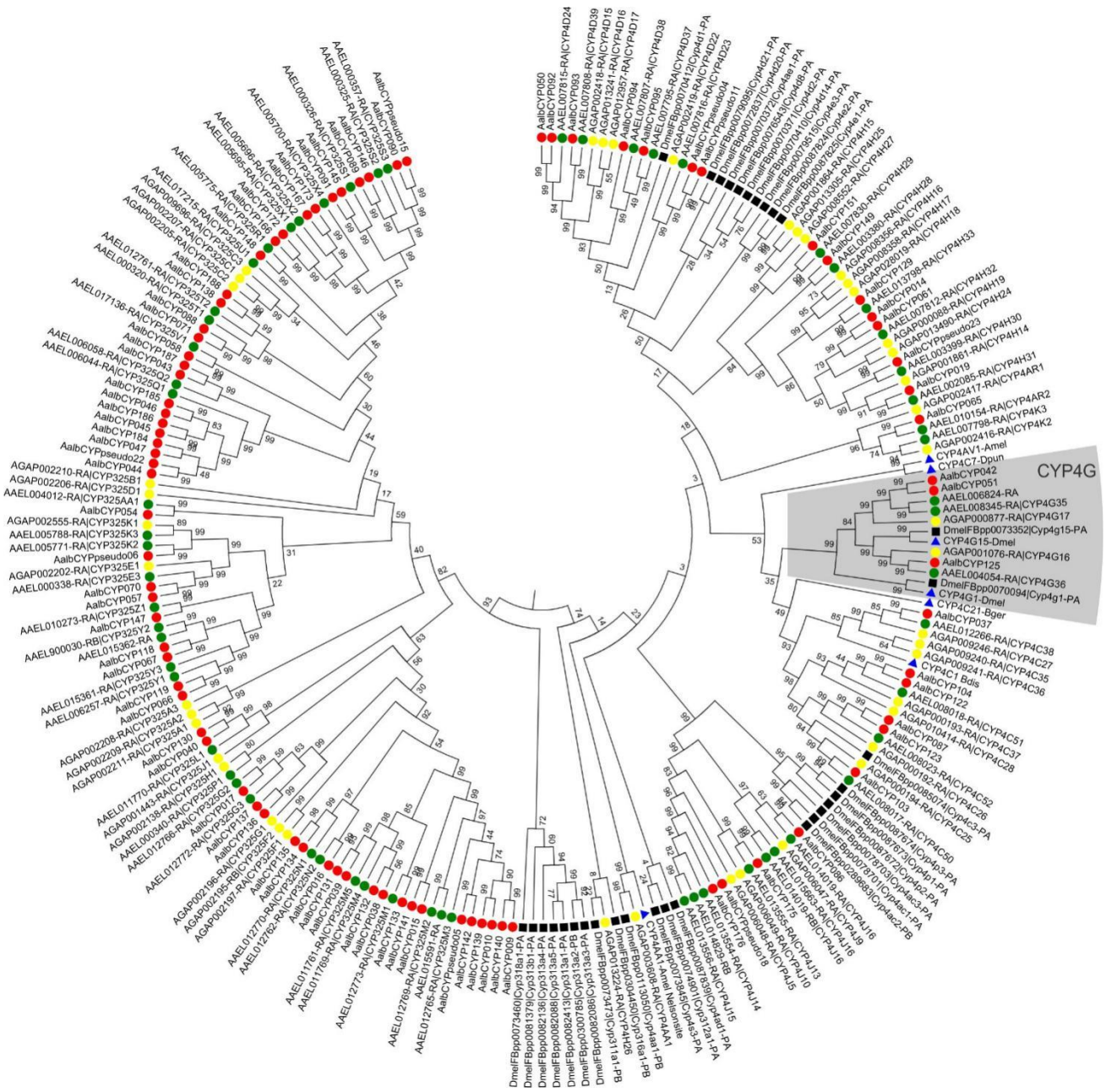

Figure S5.2.
Detailed (expanded) circular view of the CYP4 clan as shown in Figure S5.1. Only the topology is shown. The CYP4G family is shaded in grey.

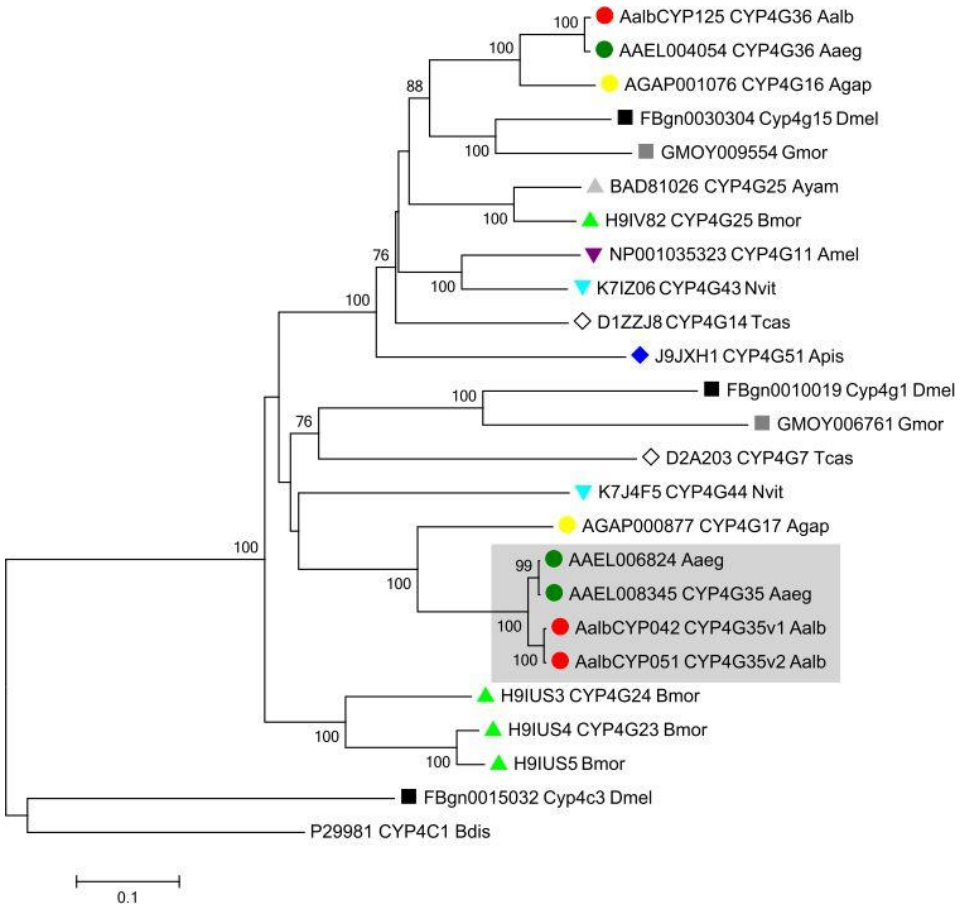

Figure S5.3.

Phylogenetic analysis of the CYP4G family in insects. Insect CYP4G protein sequences were aligned using MUSCLE and used in a neighbor joining phylogenetic analysis. The tree was rooted with CYP4C sequences of *D. melanogaster* and the cockroach, *Blaberus discoidalis*, and bootstrapped with 1000 replicates. Only bootstrap values > 75% are shown. The scale bar represents 0.1 substitutions per site. Color and shape codes: red circle, *Ae. albopictus*; green circle, *Ae. aegypti;* yellow circle, *A. gambiae*; black square, *D. melanogaster*; grey square, *Glossina morsitans*; upward grey triangle, *Antheraea yamamai*; upward green triangle, *Bombyx mori*; downward blue triangle, *Nasonia vitripennis*; downward purple triangle, *Apis mellifera*; white rhombus, *Tribolium castaneum*; blue rhombus, *Acyrthosiphon pisum*. For each insect CYP4 sequence the accession number and, in case a name has been assigned, the CYP name is shown. The duplications of CYP4G genes in *Aedes* species are shaded in grey.

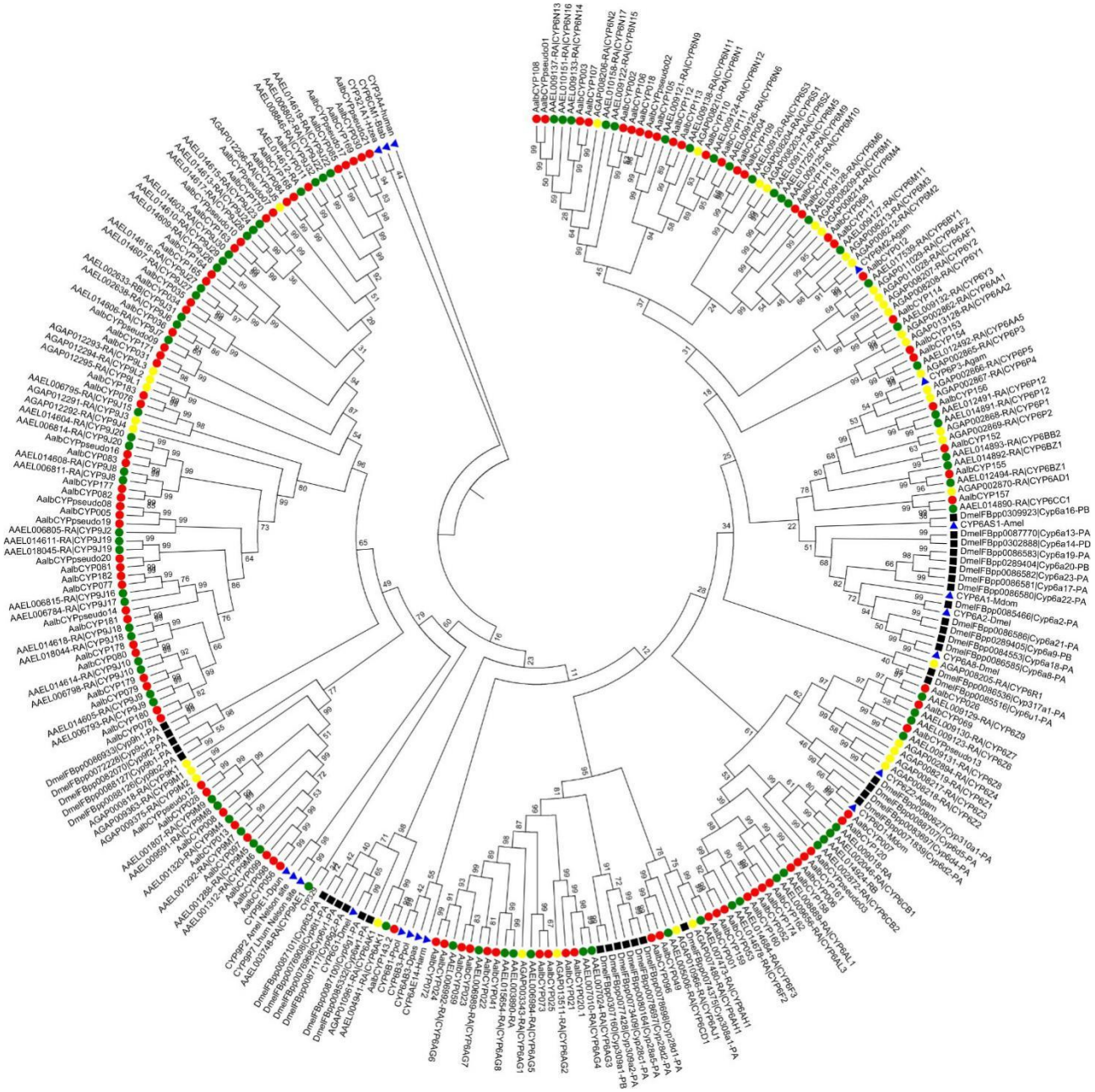

Figure S5.4.
Detailed (expanded) circular view of the CYP3 clan as shown in Figure S5.1. Only the topology is shown.

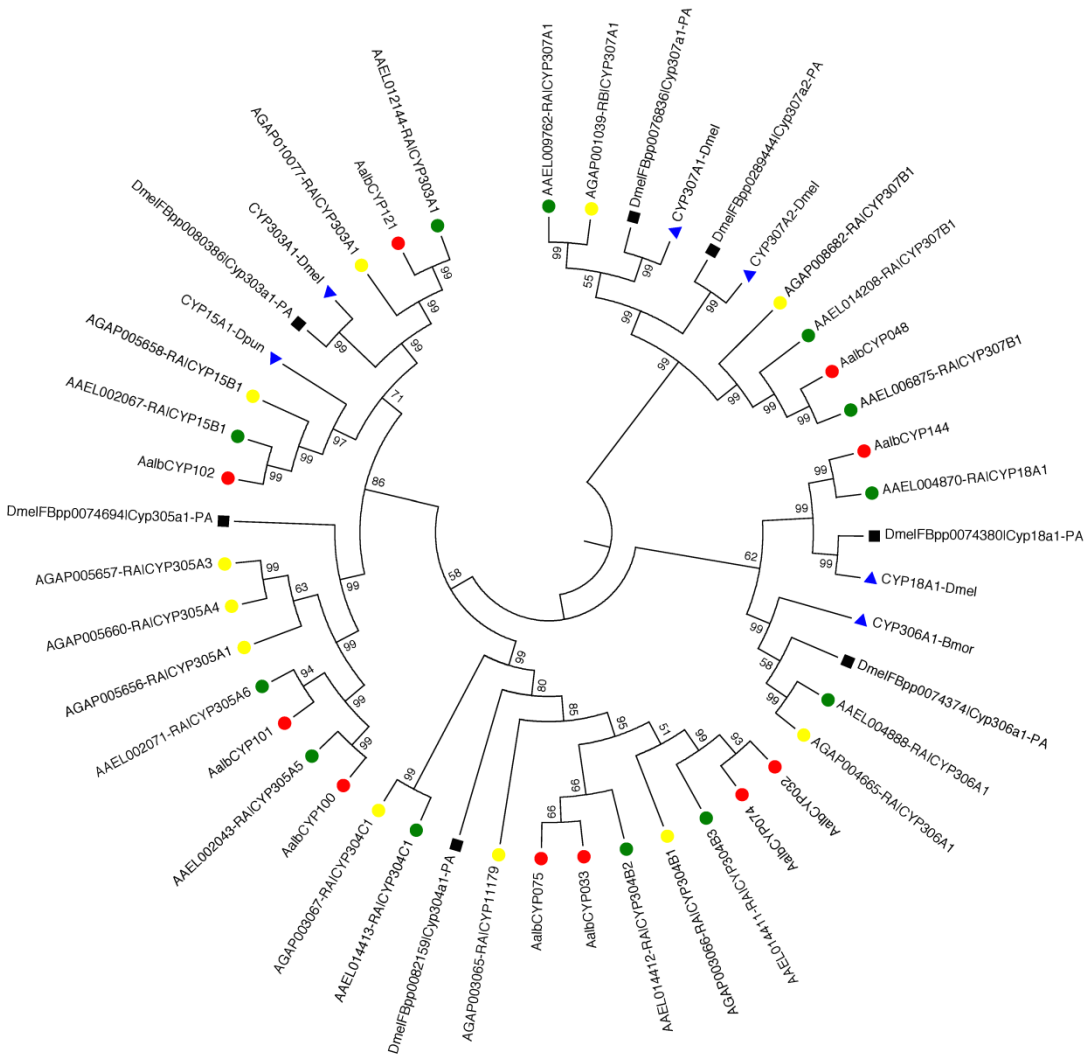

Figure S5.5.
Detailed (expanded) circular view of the CYP2 clan as shown in Figure S5.1. Only the topology is shown.

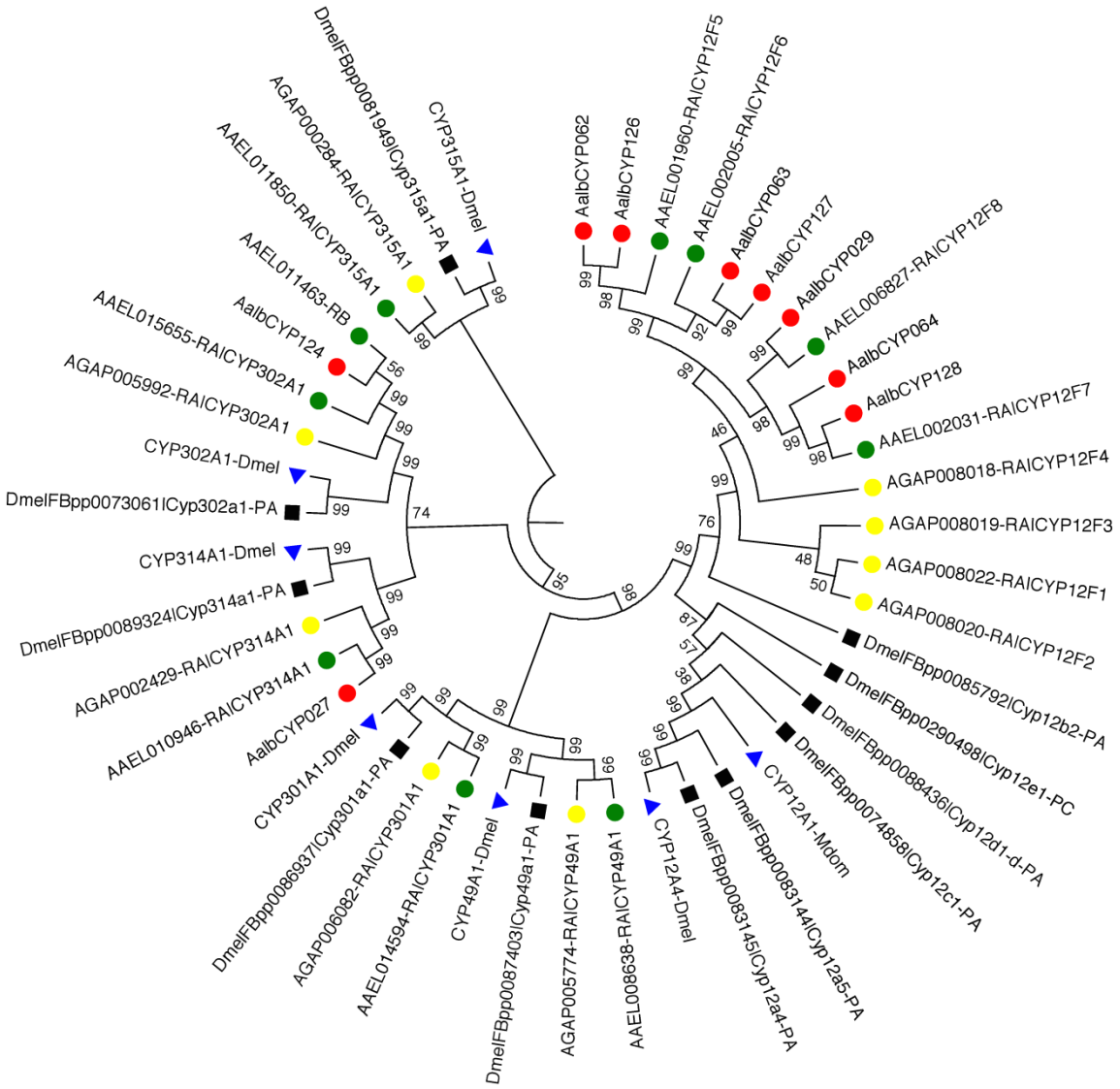

Figure S5.6.
Detailed (expanded) circular view of the mitochondrial CYP clan as shown in Figure S5.1. Only the topology is shown.

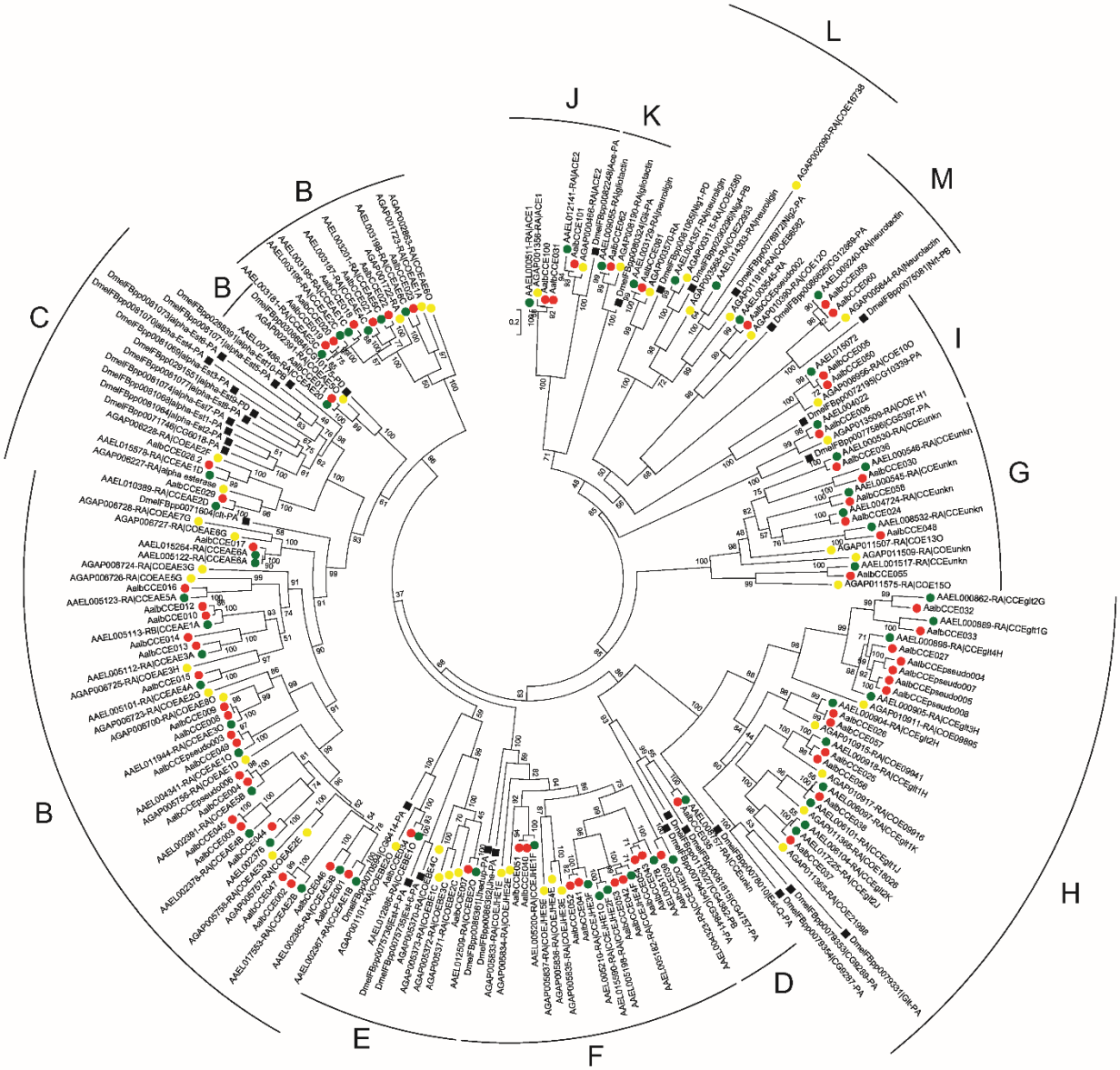

Figure S5.7.

Phylogenetic relationships between CCEs of *Ae. albopictus*, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. The evolutionary history was inferred by using the Maximum Likelihood method, bootstrapped with 1000 pseudoreplicates (bootstrap values are shown next to the branches). The scale bar represents the number of substitutions per site. The resulting tree was midpoint rooted. Color and shape codes are as follows: black square, *D. melanogaster*, red dot, *Ae. albopictus*, green dot, *Ae. aegypti*, yellow dot, *An. gambiae*. For accession numbers see Table S5.4 and Table S5.5.

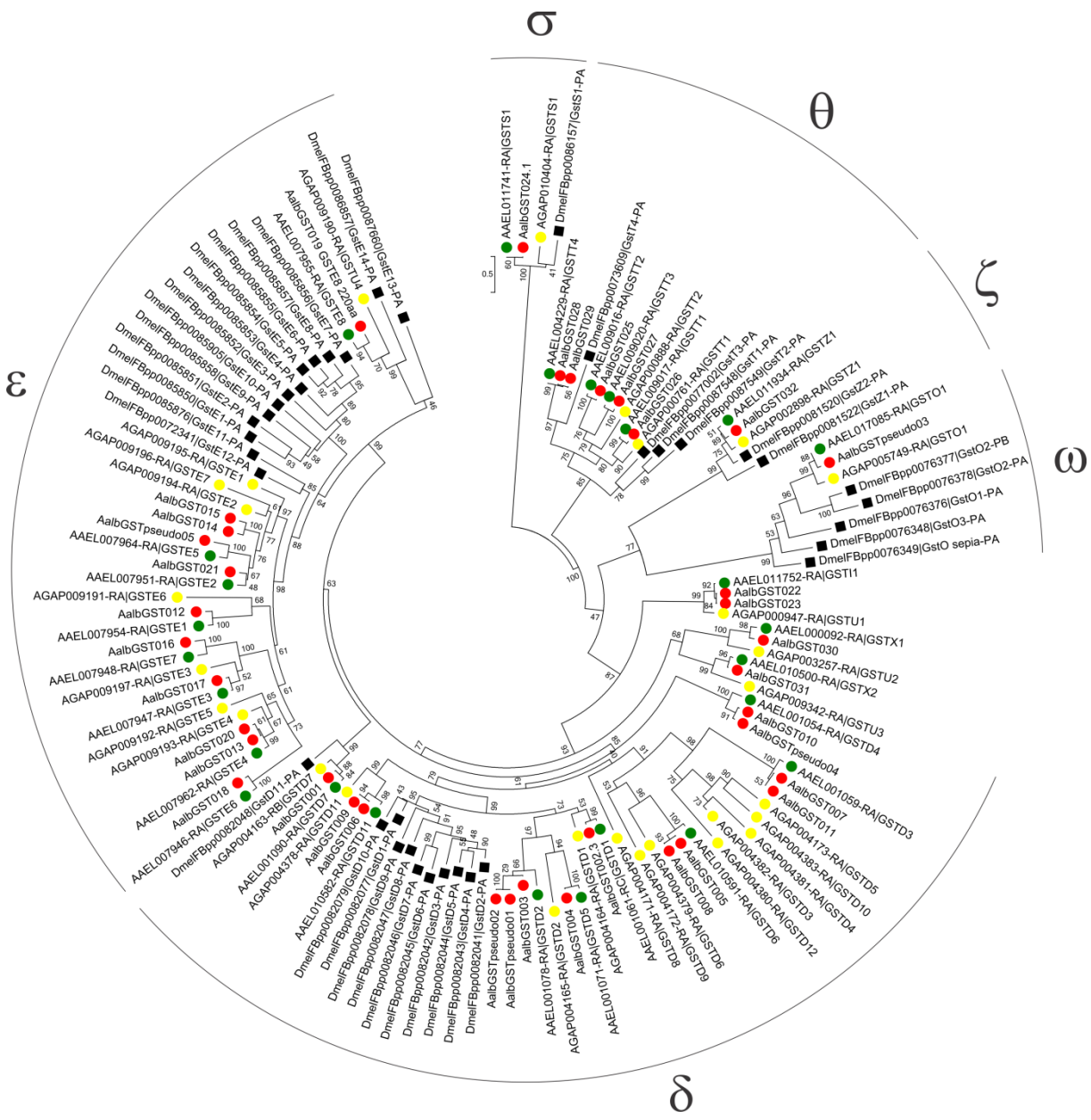

Figure S5.8.

Phylogenetic relationships between cytosolic GSTs of *Ae. albopictus*, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. The evolutionary history was inferred by using the Maximum Likelihood method, bootstrapped with 1000 pseudoreplicates (bootstrap values are shown next to the branches). The scale bar represents the number of substitutions per site. The resulting tree was midpoint rooted. Color and shape codes are as follows: black square, *D. melanogaster*, red dot, *Ae. albopictus*, green dot, *Ae. aegypti*, yellow dot, *An. gambiae*. For accession numbers see Table S5.7 and Table S5.8.

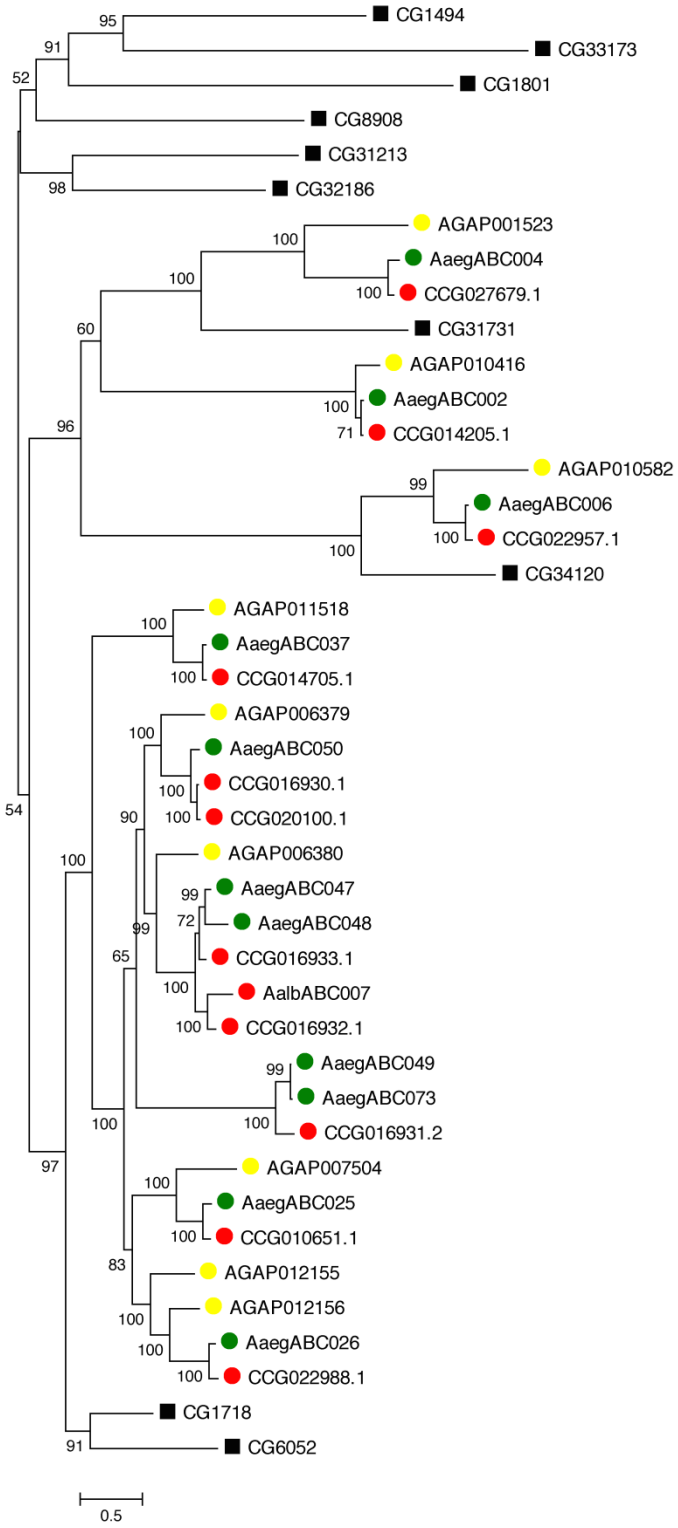

Figure S5.9.

Phylogenetic relationships between ABCAs of *Ae. albopictus*, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. The evolutionary history was inferred using the Maximum Likelihood method, bootstrapped with 1000 pseudoreplicates (bootstrap values are shown next to the branches). The scale bar represents the number of substitutions per site. The resulting tree was midpoint rooted. Color and shape codes are as follows: black square, *D. melanogaster*, red dot, *Ae. albopictus*, green

dot, *Ae. aegypti*, yellow dot, *An. gambiae*. For accession numbers see Table S5.10/Table S5.11 and Dermauw and Van Leeuwen 2014. (97)

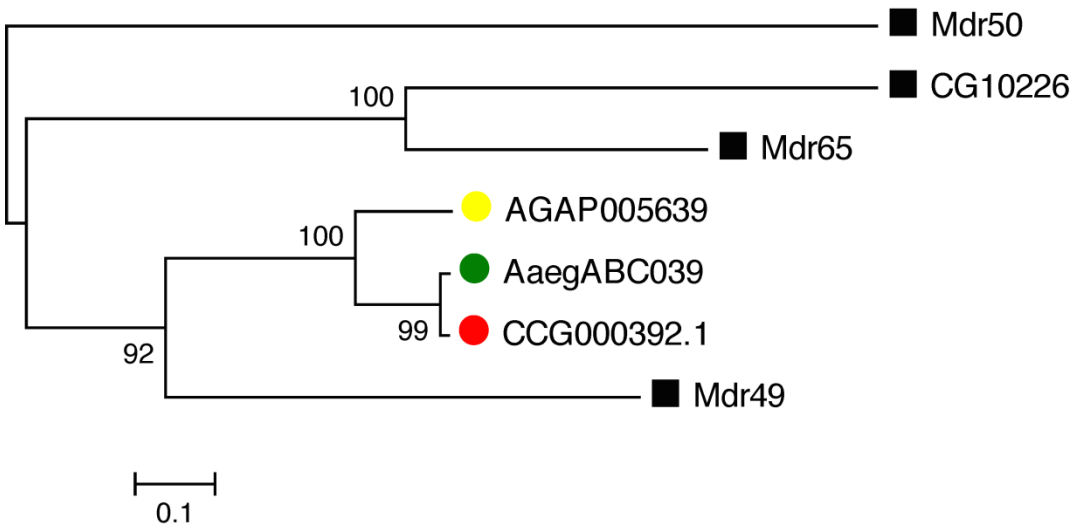

Figure S5.10.
Phylogenetic relationships between ABCB full transporters of *Ae. albopictus*, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. For procedure of phylogenetic analysis and tree details see Figure S5.9.

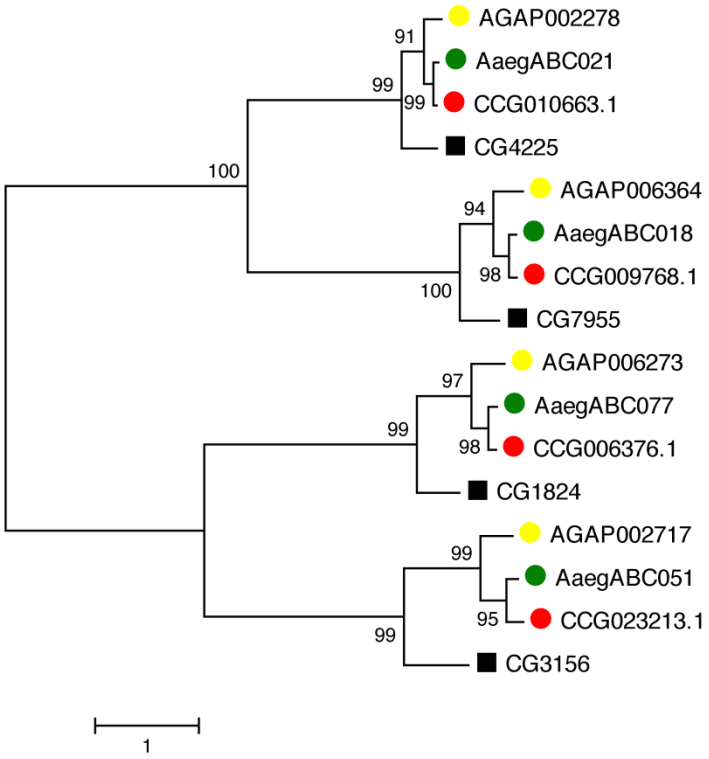

Figure S5.11.
Phylogenetic relationships between ABCB half transporters of *Ae. albopictus*, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. For procedure of phylogenetic analysis and tree details see Figure S5.9.

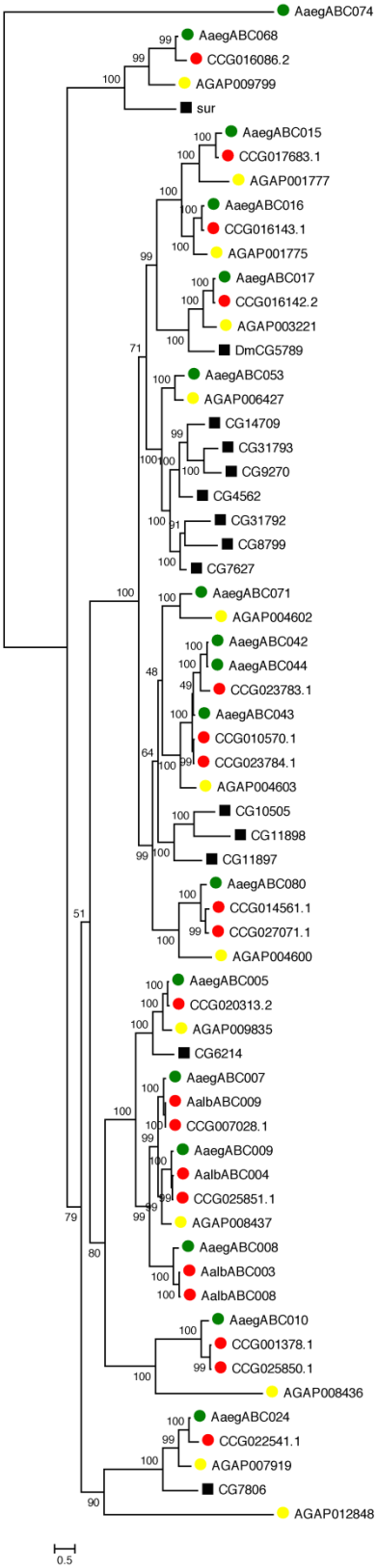

Figure S5.12.
Phylogenetic relationships between ABCCs of *Ae. albopictus, Ae. aegypti, An. gambiae* and *D. melanogaster*. For procedure of phylogenetic analysis and tree details see Figure S5.9.

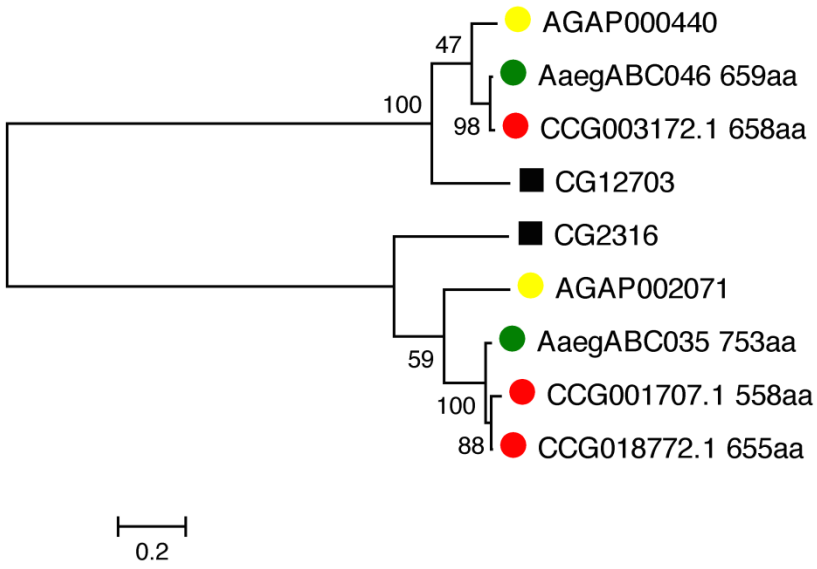

Figure S5.13.
Phylogenetic relationships between ABCDs of *Ae. albopictus, Ae. aegypti, An. gambiae* and *D. melanogaster*. For procedure of phylogenetic analysis and tree details see Figure S5.9

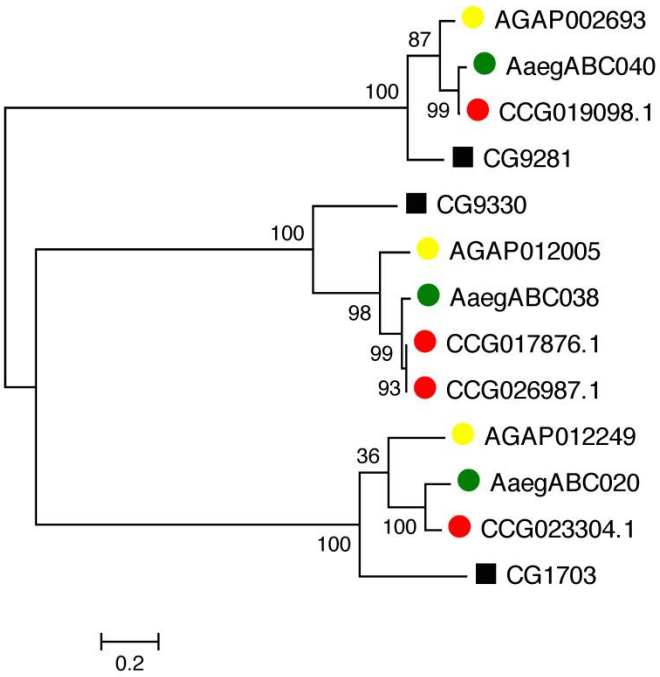

Figure S5.14.
Phylogenetic relationships between ABCFs of *Ae. albopictus*, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. For procedure of phylogenetic analysis and tree details see Figure S5.9.

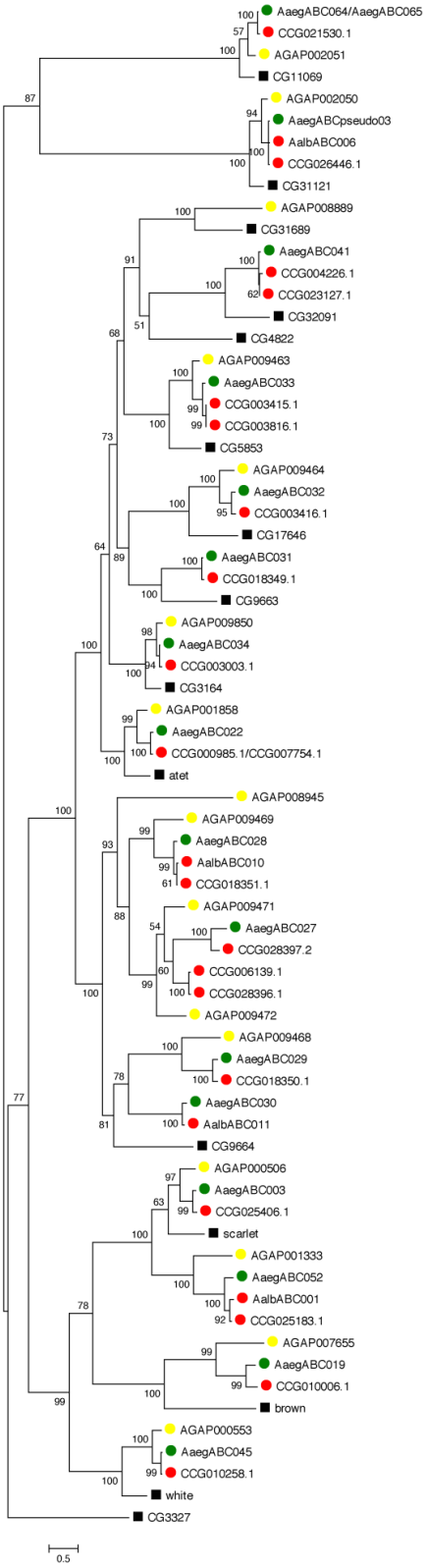

Figure S5.15.
Phylogenetic relationships between ABCGs of *Ae. albopictus*, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. For procedure of phylogenetic analysis and tree details see Figure S5.9.

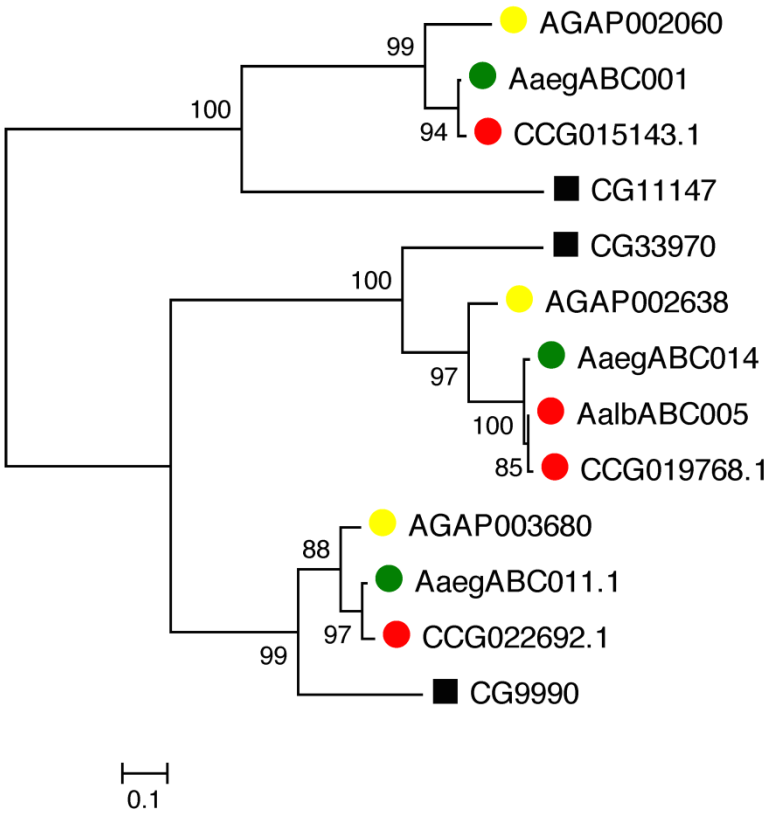

Figure S5.16.
Phylogenetic relationships between ABCHs of *Ae. albopictus*, *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. For procedure of phylogenetic analysis and tree details see Figure S5.9.

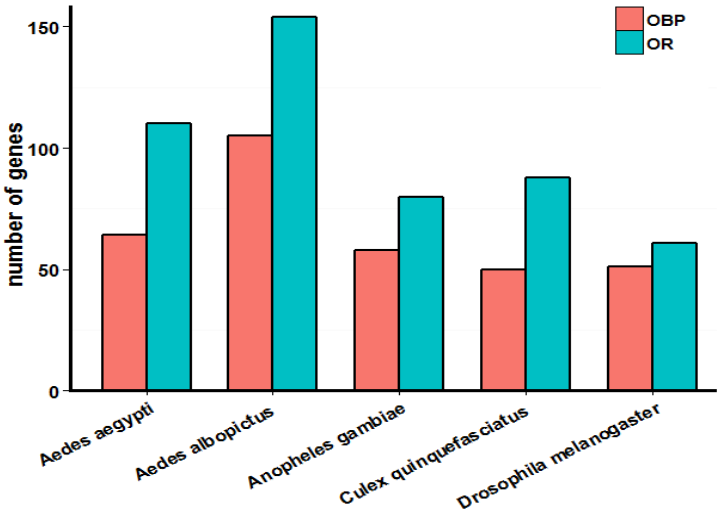

Figure S6.1.
Comparison of numbers of OBP and OR genes among mosquito species and *Drosophila melanogaster*. OBP and OR were colored in red and blue, respectively.

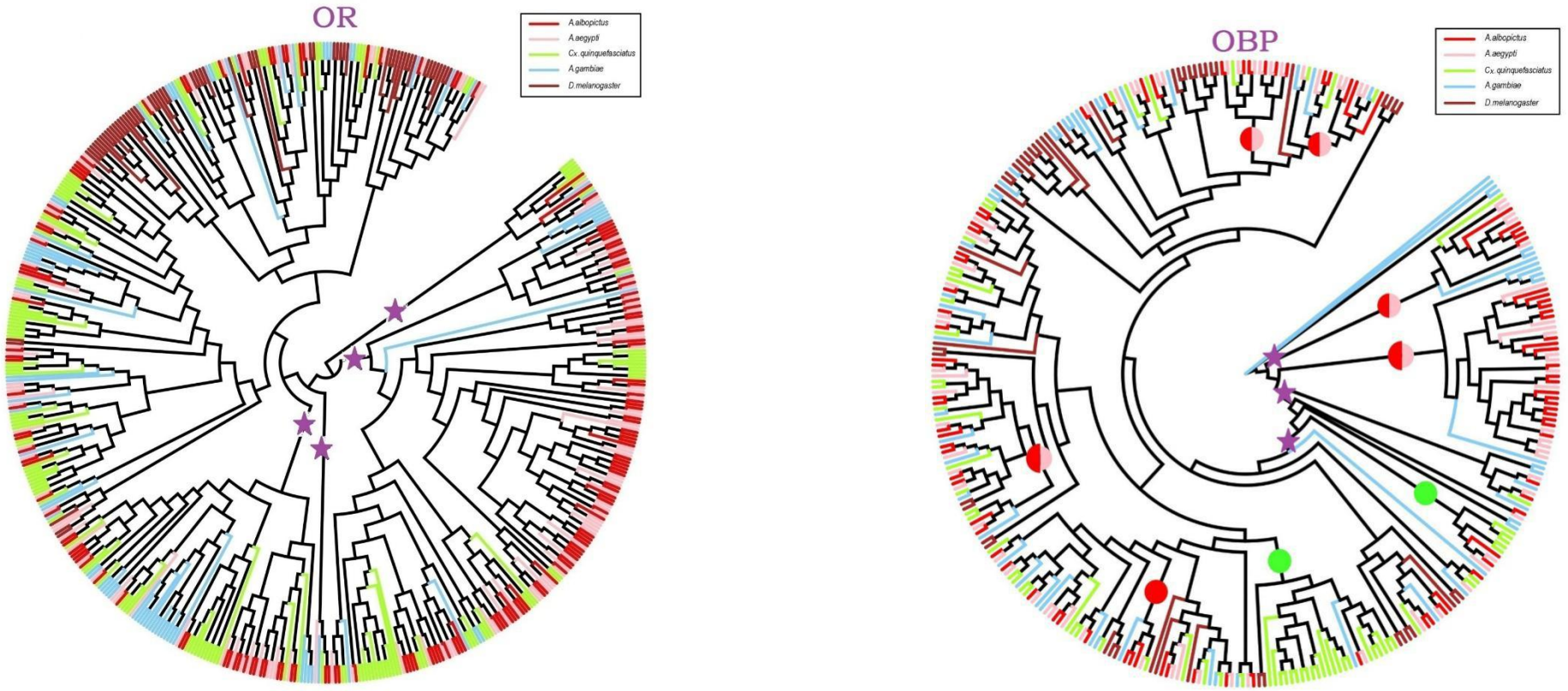

Figure S6.2.

Phylogenetic relationships among genes encoding Dipteran odorant-receptor (OR) and odorant binding proteins (OBP). Left panel (OR): All predicted OR genes from each species were used in the phylogenetic analysis. Purple stars indicate hematophagous-specific clades. Different species are depicted in different colors. Right panel (OBP): All predicted OBP genes from each species were used in the phylogenetic analysis. Here again, purple stars indicate hematophagous-specific clades. , indicates the copy numbers of genes in the phylogenetic clade are primarily from *Aedes* species; , indicates the copy numbers of genes of *Culex quinquefasciatus* are higher in this clade; , indicates that the copy number of genes of *Aedes* predominate, but those of *Ae. albopictus* are higher than *Ae. aegypti*; , indicated the genes of *Aedes* predominate, but those of *Ae. aegypti* are higher than *Ae. albopictus*.

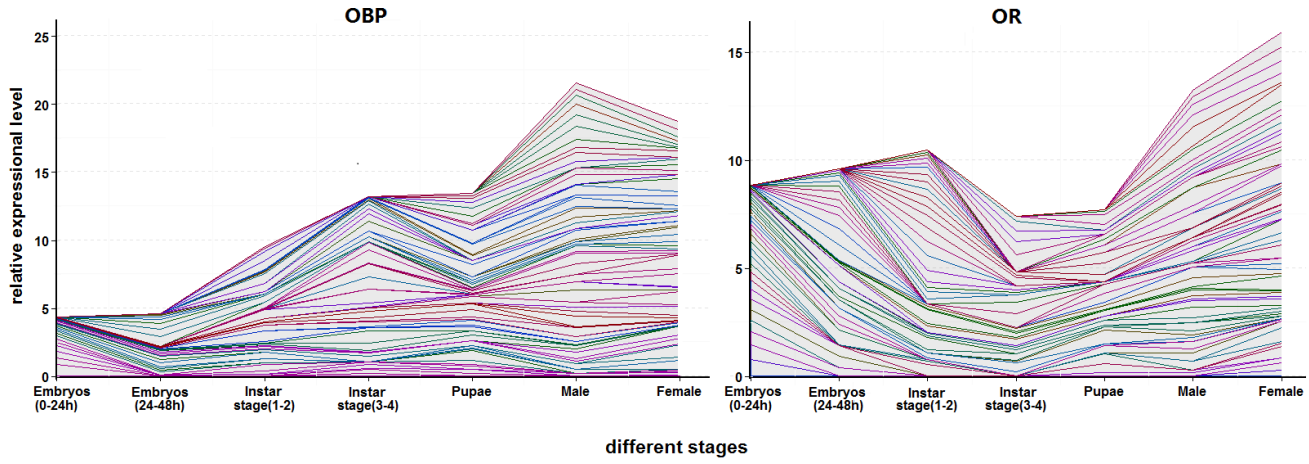

Figure S6.3.

Stage-specific abundance of OBP and OR transcripts. Relative transcriptional abundance of each gene in each stage was normalized and calculated after discarding those that were specific to a single stage or were not transcribed at each mosquito stage. Relative abundance is defined as the RPKM of individual gene in each specific stage comparing to the total RPKM of the same gene in all stages.

Figure S6.4.

Conservative structure of novel OBPs. Putative conservative cysteines are marked in a red-line box.

Figure S6.4.

Conservative structure of novel OBP. Putative conservative cysteines are marked in a red-line box (continuous).

Figure S7.1.
GO enrichment of sex-biased genes for (A) Biological processes, (B) Cellular compartments and (C) Molecular functions. See additional data files S21, S22 & S23 (separate files).

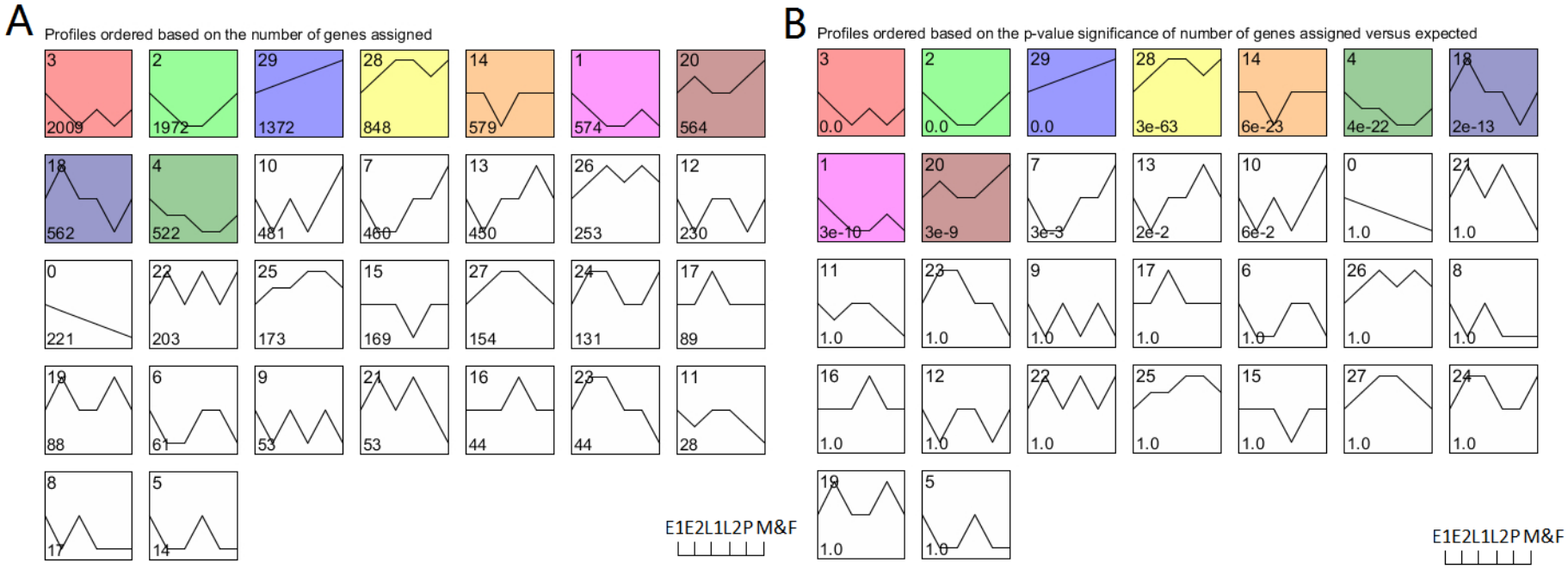

Figure S7.2.

Relative differential transcript abundance of genes exhibiting sex-biased expression. (A) The expression profiles are ordered based on the number of genes assigned. **(B)** The expression profiles are ordered based on the p-value significance of the number of genes assigned versus expected. The X axis is a developmental stage axis and the y- axis displayed the relative transcript abundance compared with the E1 stage. Abbreviations: E1, embryos 0-24 hours post-deposition; E2, embryos 24-48 hours post-deposition; L1, mixed 1st and 2nd instar larvae; L2, mixed 3rd and 4th instar larvae; P, pupae; M+F, adult males and females.

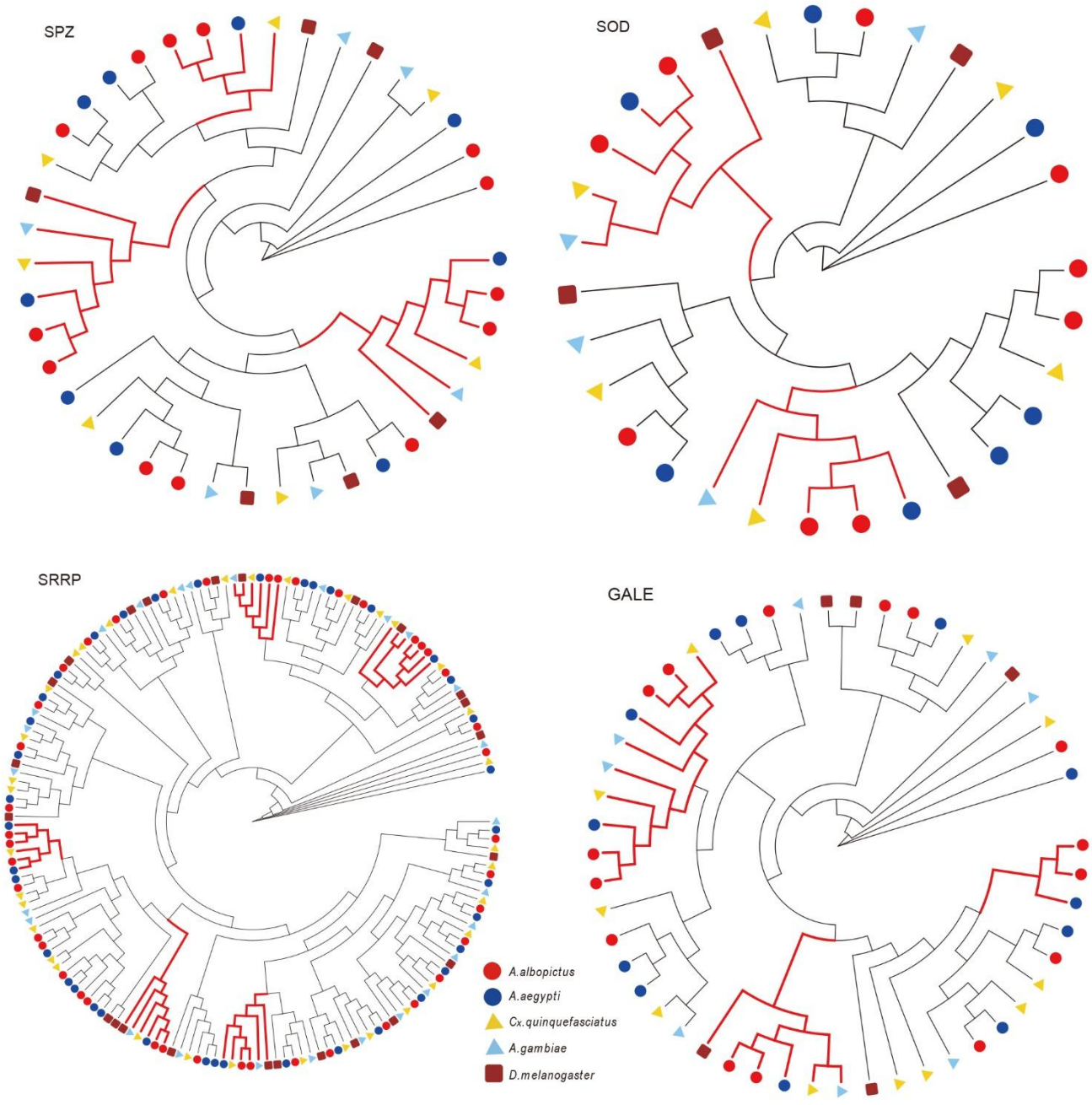

Figure S8.1.
Part 1 of the expanded immune-related families. The red branches are that have most members in *Aedes albopictus.* The *Aedes albopictus* gene ID in the red branches in the Table S8.2.

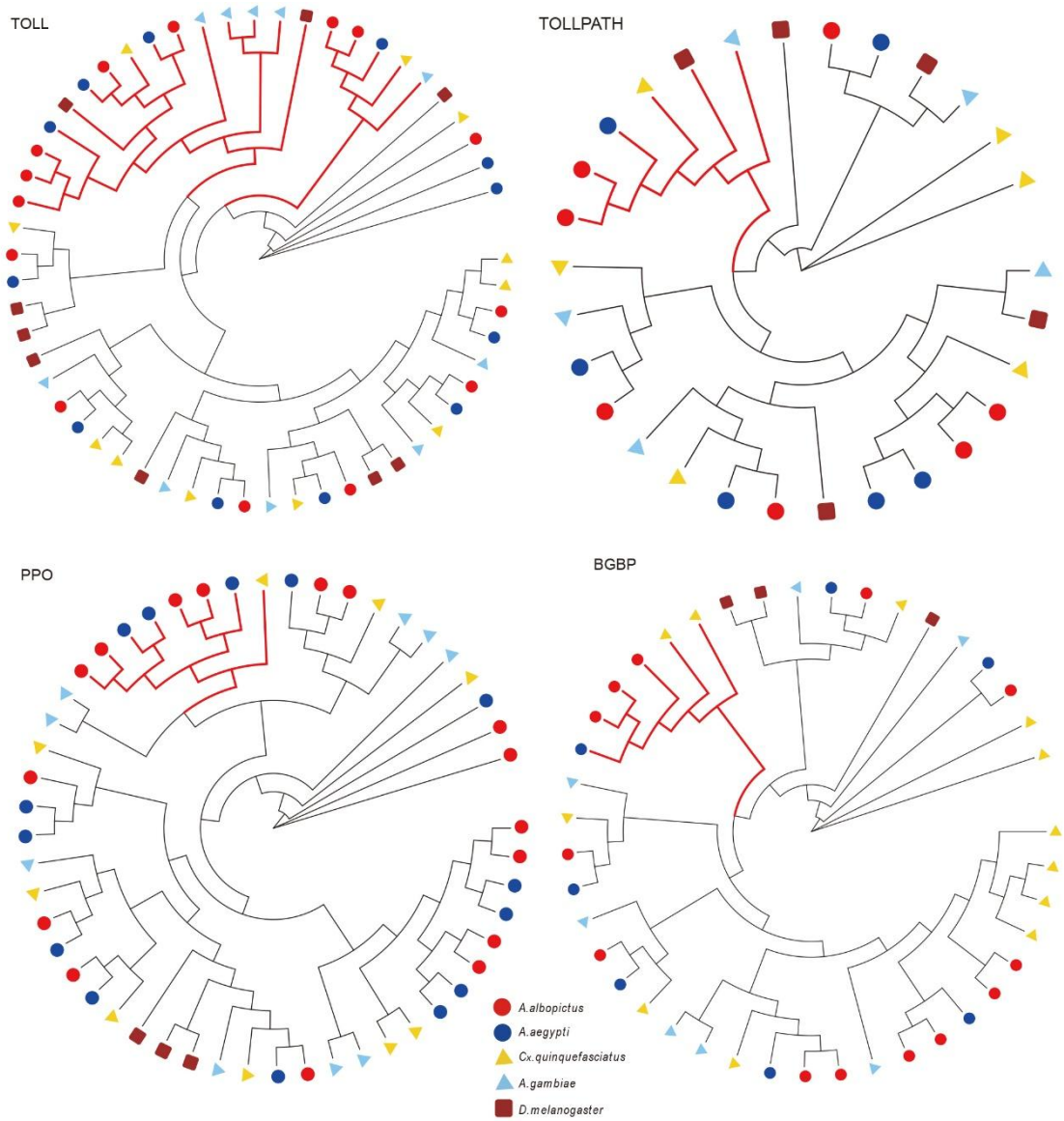

Figure 8.2.
Part 2 of the expanded immune-related families. The red branches are that have most members in *Aedes albopictus.* The *Aedes albopictus* gene ID in the red branches in the Table S8.2.

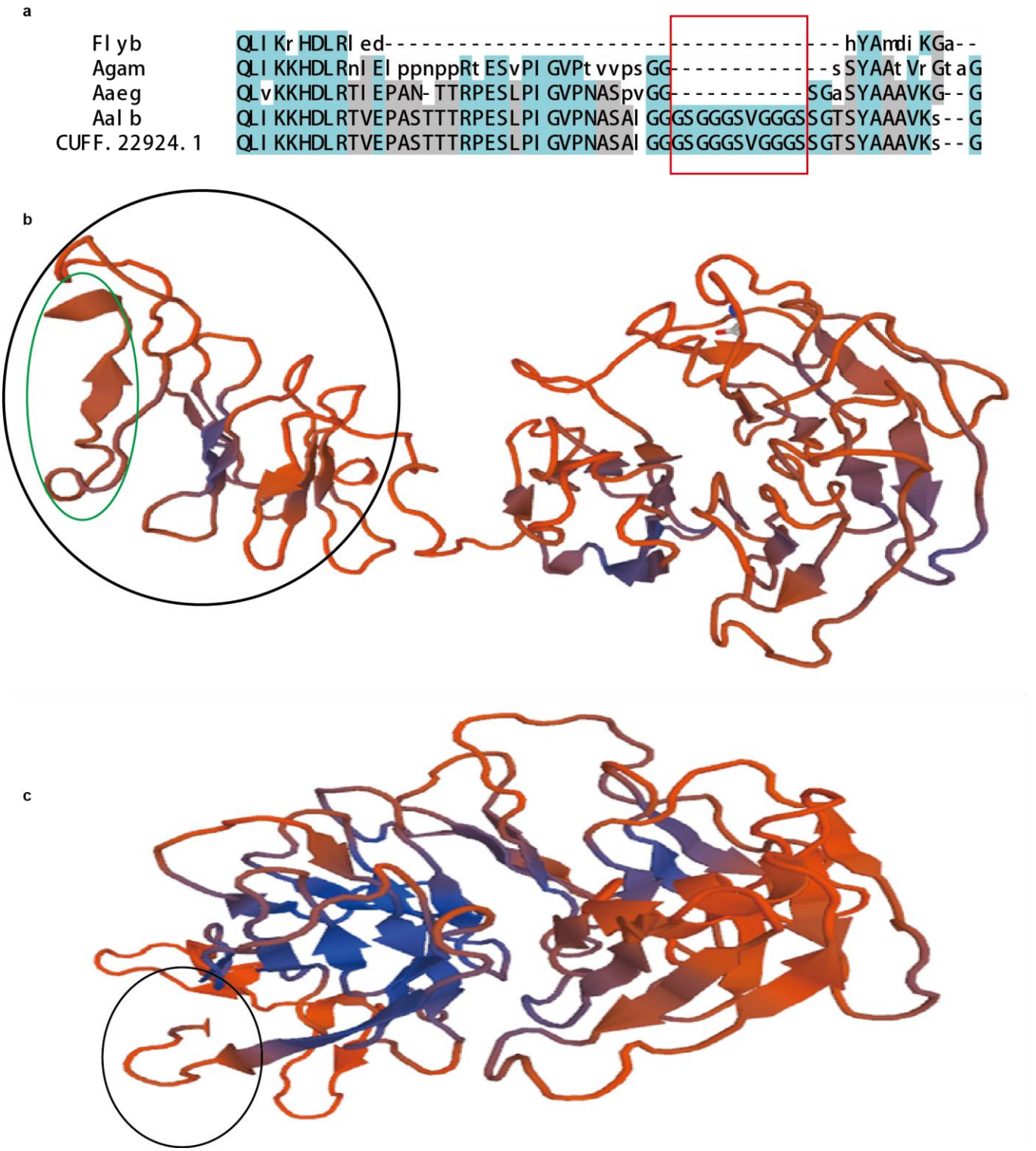

Figure S8.3.

The variation of the APG18B gene in *Aedes albopictus.* (a), the alignment of the APG18B gene among three mosquito and *Drosophila melanogaster.* The CUFF.22924.1 is the *Aedes albopictus* transcripts from "Tophat+Cufflinks", which can show that there is no assemble error for the gene. The red rect shows that there are 11 AAs insert in *Aedes albopictus.* (b), the 3D structure of the *Aedes albopictus* APG18B gene. The green circle region is the 11 AAs insert region. (c), the 3D structure the hypothetical *Aedes albopictus* APG18B gene which have deleted the 11 insert AAs. The

11 insert AAs extend 45 AAs in the C terminal to participate in the domain.

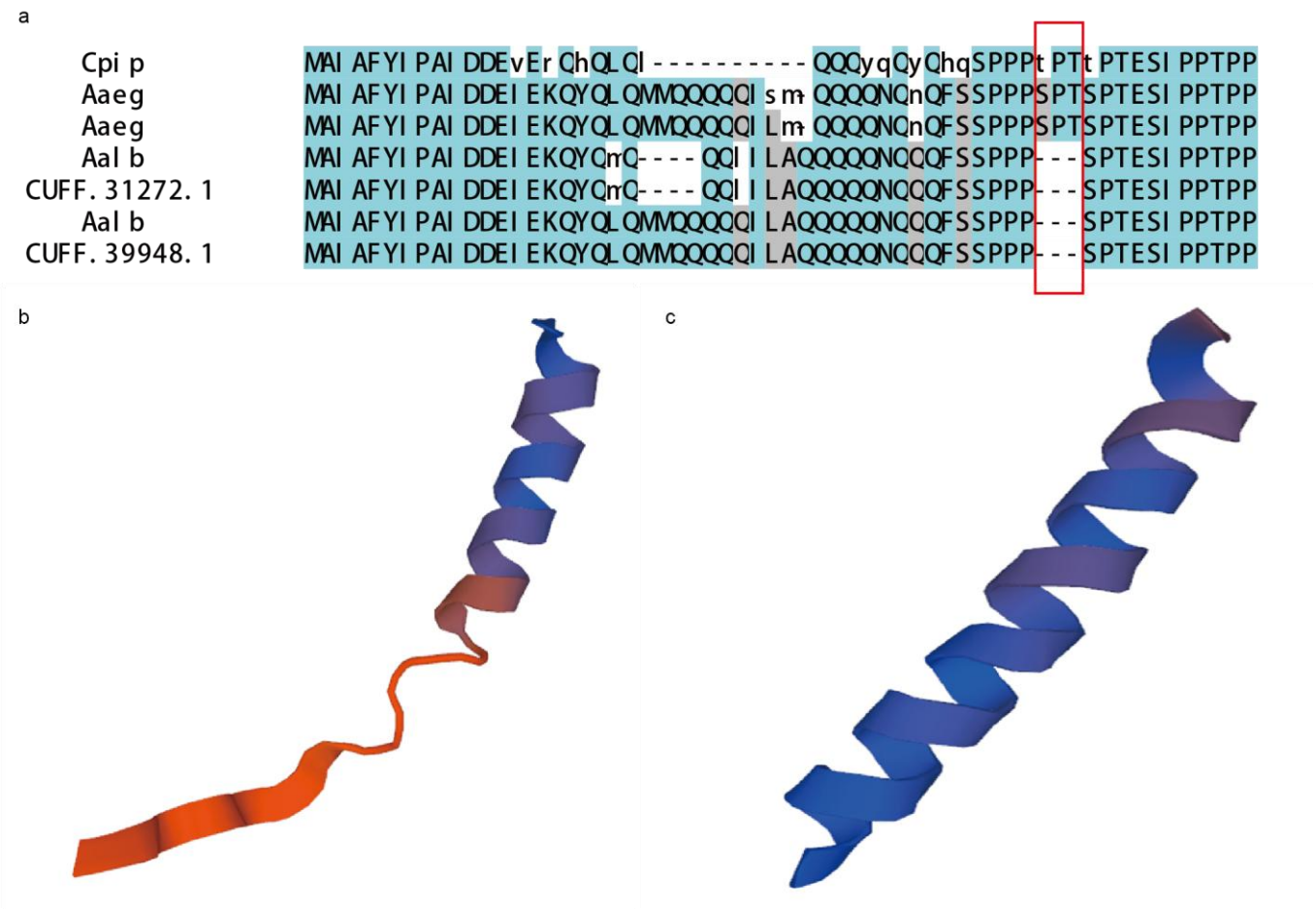

Figure S8.4.

The variation of the Michelob_x gene in *Aedes albopictus.* (a), the alignment of the Michelob_x gene among three mosquito. The CUFF* is the *Aedes albopictus* transcripts from "Tophat+Cufflinks", which can show that there is no assemble error for the gene. The red rect shows that there are 3 AAs deletion in *Aedes albopictus.* (b), the 3D structure of the *Aedes albopictus* Michelob_x gene. (c), the 3D structure the hypothetical *Aedes albopictus* Michelob_x gene which have insert the 3 deletion AAs.

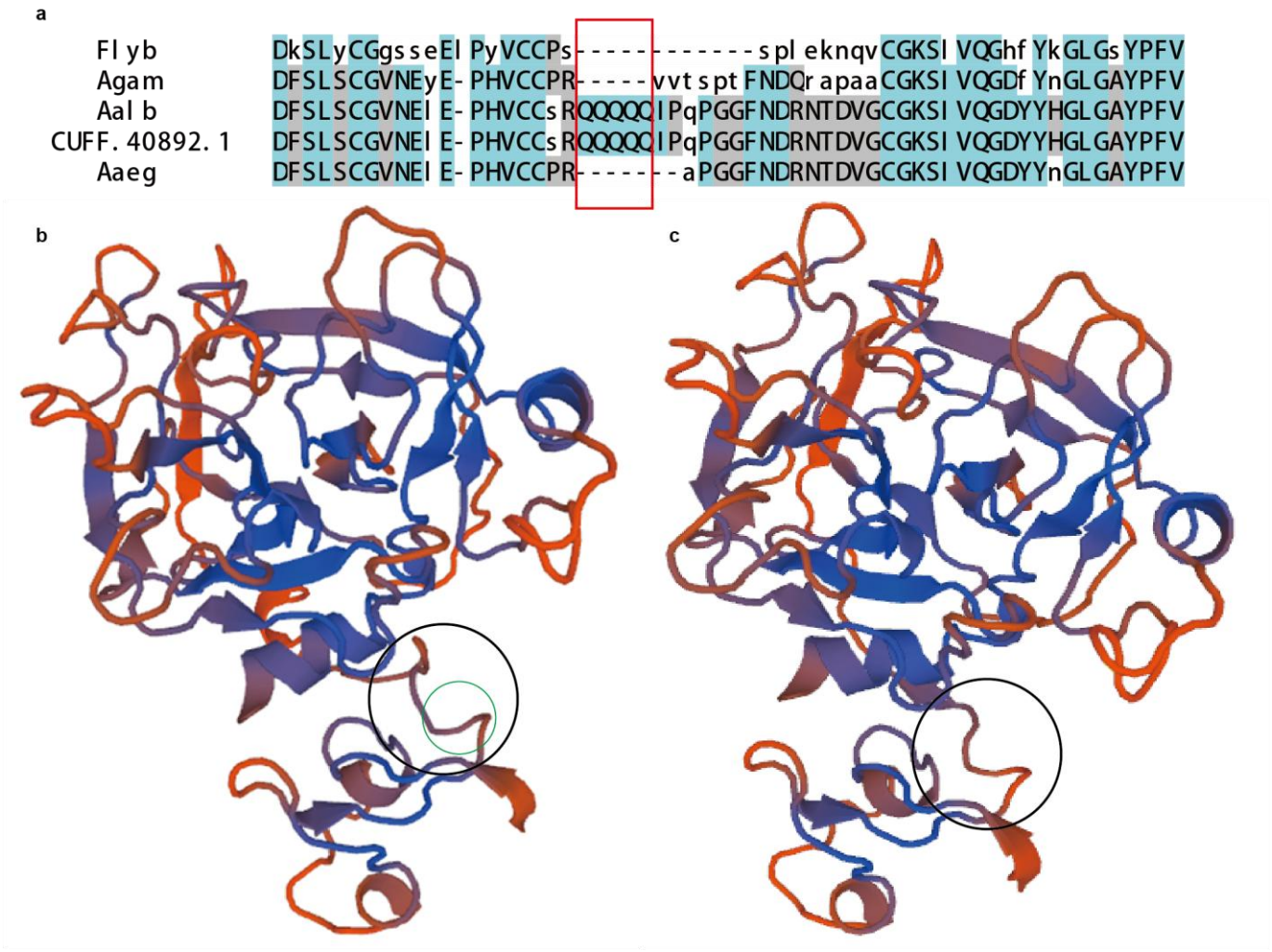

Figure S8.5.
The variation of the CLIPB16 gene in *Aedes albopictus.* (a), the alignment of the CLIPB16 gene. The CUFF.40892.1 is the *Aedes albopictus* transcripts from "Tophat+Cufflinks", which can show that there is no assemble error for the gene. The red rect shows that there are 5 "QQQQQ" AAs insertion in *Aedes albopictus.* (b), the 3D structure of the *Aedes albopictus* CLIPB16 gene. The green circle region is the 5 AAs insert region. (c), the 3D structure the hypothetical *Aedes albopictus* CLIPB16 gene which have deleted the 5 insertion AAs. The 5 insert AAs extend truss arm which connects the two parts.

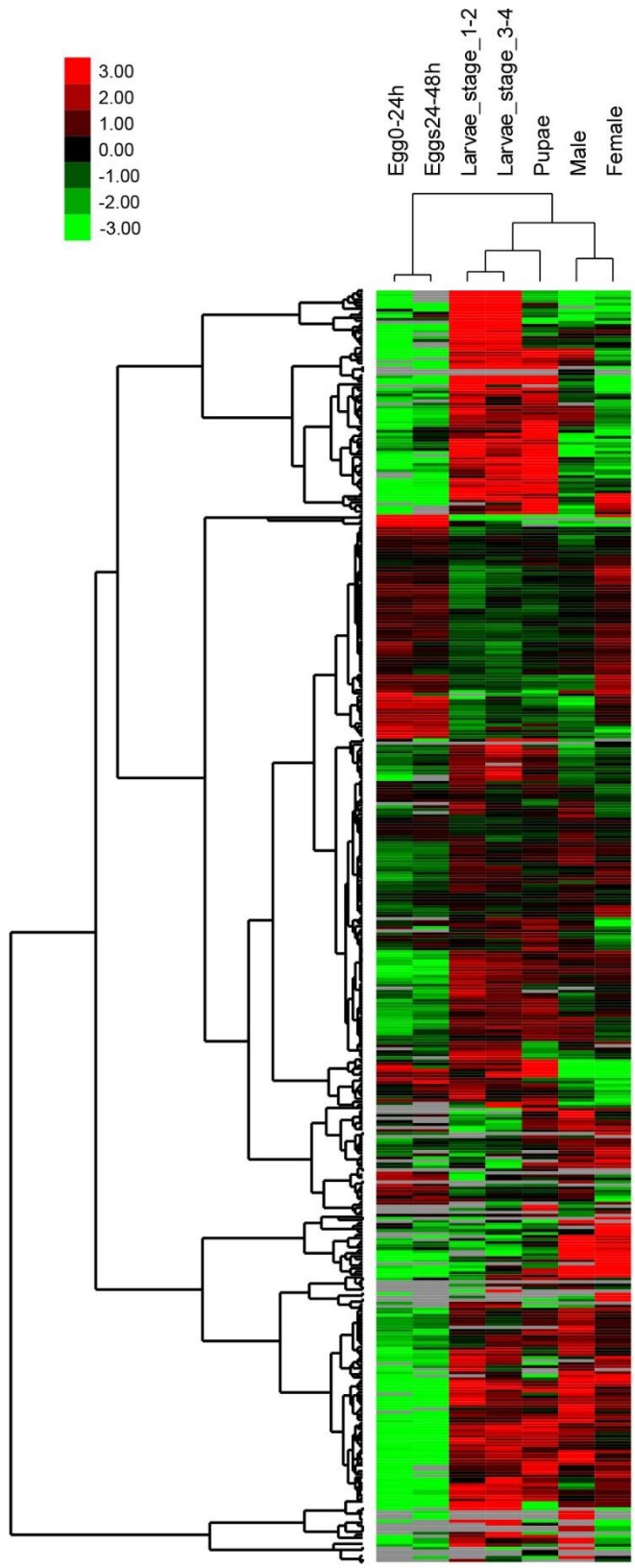

Figure S8.6

Clustering of transcriptome expression profiles. Genes whose products vary in abundance among developmental stages (early embryos [Eggs0-24h]; late embryos [Eggs24-48h], first- and second-instar larvae [larvae_ stage 1-2]; third- and fourth-instar larvae [larvae_stage 3-4]; pupae,

adult males and adult females) and show relative increases (red) or decreases (green) are clustered. The scale bar indicates the log2 (Subtract the mean of each row).

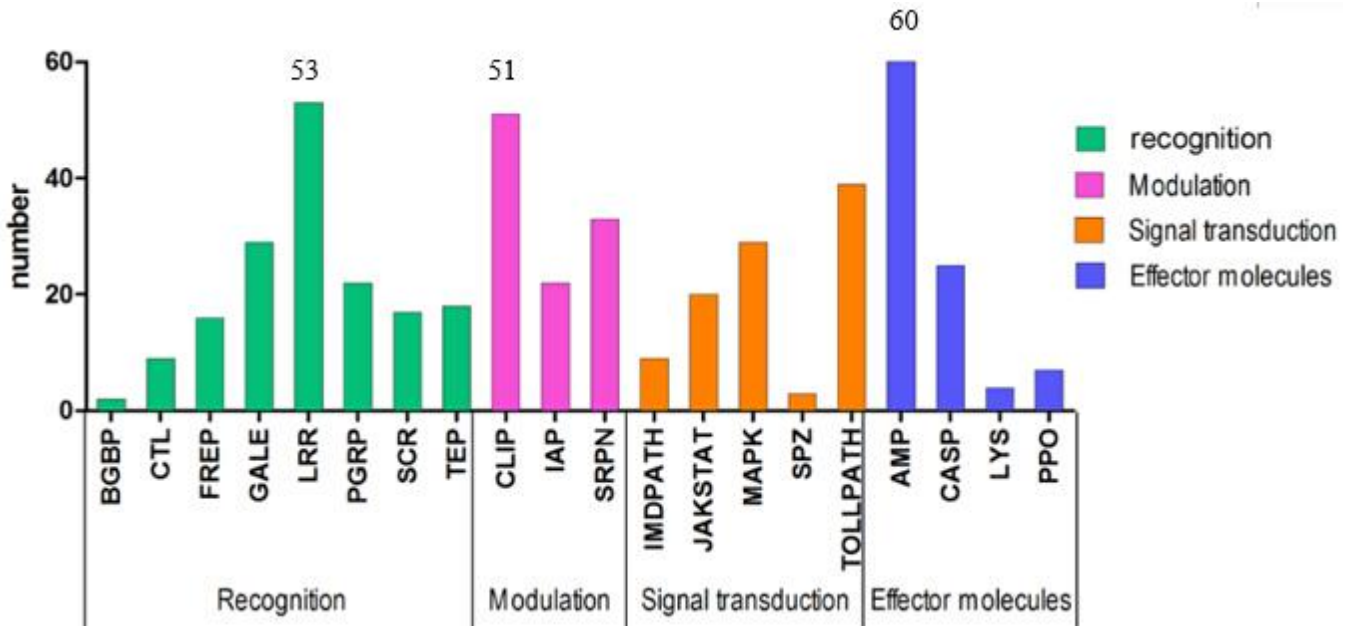

Figure S8.7

Classification of immune-related transcripts. A total of 468 immune-related genes were identified in a transcriptome analysis of the *Aedes albopictus*, and these include 166 immune recognition transcripts (including BGBP, CTL, FREP, LRR, PGRP, SCR, TEP), 106 modulation transcripts (including CLIP, IAP, SRPN), 100 signal transduction transcripts (including IMD PATH, JAK STAT, MAPK, SPZ, TOLL PATH) and 96 effector molecules (including AMP, CASP, LYS, PPO). The top three gene families are effector AMP (60 transcripts)，recognition LRR (53 transcripts) and modulation CLIP (51 transcripts).

Table S1.1.
Global statistics of *Aedes albopictus* genome sequencing.

Pair-end Libraries	Insert Size (bp)	Average Read Length (bp)	Total Data (Gb)	Fold-coverage
Solexa Reads	170	90	247.19	82.39
	500	90	205.72	68.57
	800	90	194.35	64.78
	2,000	49	116.24	38.75
	5,000	49	94.12	31.37
	10,000	49	53.84	17.95
	20, 000	49	32.13	10.71
Total	----	---	943.59	314.52

Table S1.2.
Aedes albopictus 17-k-mer statistics.

Kmer length (bp)	Total Kmers	Peak depth	Genome size (bp)	Fold-coverage
17	69,801,535,654	24	2,908,397,318	30

Table S1.3.
Statistics of high quality data

Pair-end Libraries	Insert Size (bp)	Average Read Length (bp)	Total Data (Gb)	Fold-coverage
Solexa Reads	170	85	216.47	74.39
	500	80	164.03	56.37
	800	80	156.48	53.77
	2,000	49	76.66	26.34
	5,000	49	55.32	19.01
	10,000	49	18.05	6.20
	20,000	49	2.58	0.89
Total	----	----	689.59	236.97

Table S1.4.
Statistics of the assembled genome.

	Contig		Scaffold	
	Length	Number	Length	Number
N90	2,055	130,270	13,051	13,737
N80	6,099	81,737	63,408	7,940
N70	9,759	58,422	109,290	5,605
N60	13,381	42,560	153,387	4,095
N50	17,284	30,600	195,549	2,960
Total Size	150,187	----	1,305,381	----
Longest	1,819,488,353	----	1,966,787,862	----
Total Number(>100 bp)	---	607,139	----	401,027
Total Number(>2 kb)	---	131,405	----	27,206

Table S1.5.
Assembly quality validation by Fosmid coverage estimation.

fosmid id	length (bp)	coverage ratio (%)	alignment block (N)	assemble block (N)	scaffold (N)	scaffold length (bp)	Gap (N)	gap length (bp)	gap ratio (%)
fosmid0	37,729	97.74	8	3	1	404,777	2	586	10.73
fosmid1	31,450	91.13	6	3	1	140,232	2	2559	4.46
fosmid3	29,703	98.46	9	5	1	115,931	1	205	3.90
fosmid4	31,789	96.45	9	3	1	171,540	4	1124	5.40
fosmid5	34,120	94.44	12	3	1	265,638	7	1084	7.79
fosmid6	37,539	96.63	3	1	1	220,223	1	1226	5.87
fosmid7	38,041	98.02	7	2	1	404,777	2	586	10.64
fosmid8	39,259	90.99	9	3	1	198,638	4	3159	5.06
fosmid9	38,037	98.02	7	2	1	404,777	2	586	10.64
Average	34,954	95	8	3	1	240,220	3	1,316	6.73
Total	317,667	-	70	25	9	2,326,533	25	11,115	-

Table S1.6.
Assembly quality validation by EST coverage estimation.

Dataset (> X bp)	Number	Total length	% bases covered by assembly	% sequences covered by assembly	Number with >90% sequence in one scaffold (%)	Number with >50% sequence in one scaffold (%)
0	68,133	62,243,828	94.49	98.16	52,749 (77.42)	64,974 (95.36)
200	68,133	62,243,828	94.49	98.16	52,749 (77.42)	64,974 (95.36)
500	56,807	57,353,398	94.75	98.53	44,603 (78.52)	54,459 (95.87)
1,000	18,883	31,496,387	94.74	99.39	14,101 (74.68)	18,104 (95.87)

Table S1.7.

Statistics of TEs in the genomes of four species.

Type	***Aedes albopictus***		***Aedes aegypti***		***Anopheles gambiae***		***Culex quinquefasciatus***	
	length[1]	%[2]	length[1]	%[2]	length[1]	%[2]	length[1]	%[2]
DNA	268,995,073	13.68	331,621,708	23.96	24,754,765	9.06	199,940,240	34.53
LINE	810,551,543	41.21	556,339,638	40.2	20,599,143	7.54	80,583,704	13.92
SINE	2,035,962	0.1	462,440	0.03	4,368,426	1.6	2,987,569	0.52
LTR	422,621,193	21.49	344,730,123	24.91	29,128,459	10.67	113,405,059	19.58
Total	**1,400,580,949**	**71.21**	**990,328,449**	**71.56**	**62,040,492**	**22.72**	**325,988,315**	**56.3**

[1]length in base-pairs.

[2]percent of genome represented.

Abbreviations: DNA - DNA Transposon, LINE - long interspersed nuclear elements, SINE - short interspersed nuclear elements, LTR - long terminal repeats.

Table S1.8.

Transcriptome sequencing data statistics.

Sample	Total Reads (M)	Total Base (G)	Total Map Reads (M)	Total Map Rate (%)	Unip Map Reads (M)	Uniq Map Rate (%)
Egg 0-24h	79.50	7.15	66.48	83.63	50.79	63.89
Eggs 24-48h	76.80	6.91	64.71	84.26	49.18	64.04
Larval stage 1-2	77.40	6.97	65.74	84.93	49.28	63.66
Larval stage 3-4	77.40	6.97	65.74	84.93	49.28	63.66
Pupae	73.65	6.63	61.17	83.05	45.18	61.35
Female	74.82	6.73	63.30	84.60	47.27	63.17
Male	82.58	7.43	69.13	83.72	48.22	58.39

Abbreviations: Egg 0-24h - mixed sex samples of embryos 0-24 hours post-deposition (hpd) - Eggs 24-48h - mixed sex samples of embryos 24-48 (hpd), Larval stage 1-2 - a combined pool of 1^{st}-and 2^{nd}-instar larvae, Larval stage 3-4 - a combined pool of 3^{rd}- and 4^{th}-instar larvae, Pupae - mixed sex pupae of all stages, Male and Female - adult males and sugar-fed adult females, respectively.

Table S1.9.
Statistics of predicted protein-coding genes.

Species	GN	SE	SE%	ACL	AEG	AEL	AIL
Ae. albopictus	17,539	4,367	24.90	1,380.99	3.32	416.23	3,086.92
Ae. aegypti	15,986	2,205	13.79	1,384.49	3.84	360.92	4,599.13
An. gambiae	12,665	1,601	12.64	1,572.02	4.04	389.19	1,209.47
Cx. quinquefasciatus	18,882	1,839	9.74	1,312.27	3.75	349.89	1,564.43
D. melanogaster	13,689	2,761	20.17	1,621.49	3.97	408.18	888.01

Abbreviations: GN - gene number, SE - single exon gene number, SE% - percent of SE present in all genes, ACL - average CDS length in base-pairs, AEG - average exons per gene, AEL - average transcript length in base-pairs, AIL - average intron length in base-pairs.

Table S1.10.
Functional classification of ***Aedes albopictus*** genes with various methods.

	Databases	Gene Number	Percent (%)
Total		22,543	100%
Annotated	SwissProt	20,835	92.4%
	TrEMBL	21,060	93.4%
	KEGG	16,616	73.7%
	InterPro	18,129	80.4%
	GO	14,988	66.5%
Unannotated		1,436	6.4%

Table S1.11.

GO enrichment of *Aedes albopictus* specific families.

GO ID	GO Term	GO Class	Gene Number	P value
GO:0005488	binding	MF	146	2.12E-09
GO:0004221	ubiquitin thiolesterase activity	MF	8	5.18E-05
GO:0016790	thiolester hydrolase activity	MF	8	0.00068
GO:0006511	ubiquitin-dependent protein catabolic process	BP	8	0.00178
GO:0019941	modification-dependent protein catabolic process	BP	8	0.00215
GO:0043632	modification-dependent macromolecule catabolic process	BP	8	0.00215
GO:0017154	semaphorin receptor activity	MF	3	0.0024
GO:0044257	cellular protein catabolic process	BP	8	0.00367
GO:0051603	proteolysis involved in cellular protein catabolic process	BP	8	0.00367
GO:0030163	protein catabolic process	BP	8	0.00946
GO:0044265	cellular macromolecule catabolic process	BP	8	0.03956

Abbreviations: BP - biological process, MF - molecular function.

Table S1.12.

KEGG pathway enrichment of *Aedes albopictus* specific families.

Pathway	**Gene Number**	**P value**	**Pathway ID**
ABC transporters	89	1.51E-49	ko02010
Bile secretion	89	2.12E-43	ko04976
Progesterone-mediated oocyte maturation	61	1.07E-32	ko04914
Purine metabolism	69	3.60E-29	ko00230
Cell cycle	61	2.65E-26	ko04110
Vasopressin-regulated water reabsorption	44	1.15E-25	ko04962
Ubiquitin mediated proteolysis	62	1.83E-25	ko04120
Drug metabolism - other enzymes	44	4.20E-25	ko00983
Cytokine-cytokine receptor interaction	29	1.25E-18	ko04060
Glycosaminoglycan biosynthesis - heparan sulfate	19	2.72E-18	ko00534
Primary immunodeficiency	18	3.13E-18	ko05340
Metabolic pathways	128	2.90E-17	ko01100
RNA polymerase	25	6.38E-12	ko03020
Chemokine signaling pathway	27	8.56E-12	ko04062
Toxoplasmosis	23	3.73E-10	ko05145
T cell receptor signaling pathway	20	8.99E-10	ko04660
Endocytosis	33	2.09E-09	ko04144
Cell adhesion molecules (CAMs)	19	2.23E-09	ko04514
Fc gamma R-mediated phagocytosis	22	6.16E-09	ko04666
Neuroactive ligand-receptor interaction	51	5.32E-08	ko04080
Spliceosome	30	1.60E-07	ko03040
Pyrimidine metabolism	25	1.62E-07	ko00240
Huntington's disease	29	2.15E-07	ko05016
Epstein-Barr virus infection	26	5.27E-06	ko05169
Notch signaling pathway	14	6.99E-05	ko04330
Terpenoid backbone biosynthesis	6	5.24E-04	ko00900
Pentose and glucuronate interconversions	10	7.23E-04	ko00040
Fanconi anemia pathway	9	2.25E-03	ko03460
Glycerophospholipid metabolism	14	4.52E-03	ko00564
Carbohydrate digestion and absorption	16	1.08E-02	ko04973
Galactose metabolism	16	1.67E-02	ko00052
Starch and sucrose metabolism	16	2.76E-02	ko00500

Table S1.13.
Summary of representation of RNA types in *Aedes albopictus*.

Type		**Copy number**	**Average length**[1]	**Total length**[1]	**% of genome**
miRNA		193	77.57	14,971	0.000761
tRNA		603	77.49	46,726	0.002376
rRNA	**18S**	485	93.52	45,356	0.002306
	28S	15	108.60	1,629	0.000083
	5.8S	18	48.00	864	0.000044
	5S	170	76.91	13,074	0.000665
snRNA	**snRNA**	64	147.11	9,415	0.000479
	CD-box	10	131.40	1,314	0.000067
	HACA-box	0	0.00	0	0
	splicing	54	150.02	8,101	0.000412

[1]lengths in base-pairs.

Table S1.14.

GO enrich of *Aedes albopictus* significant expansion families.

GO ID	GO Term	GO Class	Gene Number	P value
GO:0050660	flavin adenine dinucleotide binding	MF	35	1.64E-16
GO:0008152	metabolic process	BP	284	8.47E-15
GO:0016310	phosphorylation	BP	71	1.07E-14
GO:0050662	coenzyme binding	MF	39	1.61E-12
GO:0016042	lipid catabolic process	BP	20	3.92E-10
GO:0008609	alkylglycerone-phosphate synthase activity	MF	9	4.86E-10
GO:0016301	kinase activity	MF	67	5.41E-09
GO:0003995	acyl-CoA dehydrogenase activity	MF	14	1.06E-08
GO:0048037	cofactor binding	MF	39	2.20E-08
GO:0008611	ether lipid biosynthetic process	BP	8	7.45E-08
GO:0046485	ether lipid metabolic process	BP	8	7.45E-08
GO:0097384	cellular lipid biosynthetic process	BP	8	7.45E-08
GO:1901503	ether biosynthetic process	BP	8	7.45E-08
GO:0008762	UDP-N-acetylmuramate dehydrogenase activity	MF	9	2.88E-07
GO:0003869	4-nitrophenylphosphatase activity	MF	8	3.93E-07
GO:0006629	lipid metabolic process	BP	42	8.99E-07
GO:0031409	pigment binding	MF	7	1.44E-06
GO:0055114	oxidation-reduction process	BP	72	1.45E-06
GO:0008812	choline dehydrogenase activity	MF	12	2.30E-06
GO:0008043	intracellular ferritin complex	CC	6	2.95E-06
GO:0070288	ferritin complex	CC	6	2.95E-06
GO:0022627	cytosolic small ribosomal subunit	CC	10	3.24E-06
GO:0003824	catalytic activity	MF	309	3.98E-06
GO:0015935	small ribosomal subunit	CC	12	2.03E-05
GO:0016772	transferase activity, transferring phosphorus -ontaining groups	MF	67	3.50E-05
GO:0006796	phosphate-containing compound metabolic process	BP	83	5.16E-05
GO:0008198	ferrous iron binding	MF	6	9.50E-05
GO:0008234	cysteine-type peptidase activity	MF	17	9.70E-05
GO:0006793	phosphorus metabolic process	BP	83	0.00011
GO:0031062	positive regulation of histone methylation	BP	5	0.00011
GO:0035103	sterol regulatory element binding protein cleavage	BP	5	0.00011
GO:0051569	regulation of histone H3-K4 methylation	BP	5	0.00011
GO:0051571	positive regulation of histone H3-K4 methylation	BP	5	0.00011
GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	MF	12	0.00012
GO:0044710	single-organism metabolic process	BP	118	0.00022
GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	MF	14	0.00037

GO:0050878	regulation of body fluid levels	BP	8	0.00042
GO:0042381	hemolymph coagulation	BP	6	0.00087
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	MF	22	0.00129
GO:0004197	cysteine-type endopeptidase activity	MF	9	0.00192
GO:0018904	ether metabolic process	BP	8	0.00238
GO:0031012	extracellular matrix	CC	11	0.00298
GO:0035101	FACT complex	CC	3	0.00384
GO:0044391	ribosomal subunit	CC	15	0.00456
GO:0008233	peptidase activity	MF	60	0.00472
GO:0006991	response to sterol depletion	BP	5	0.00569
GO:0031060	regulation of histone methylation	BP	5	0.00569
GO:0032933	SREBP signaling pathway	BP	5	0.00569
GO:0071501	cellular response to sterol depletion	BP	5	0.00569
GO:0005840	ribosome	CC	18	0.00605
GO:0005777	peroxisome	CC	8	0.00685
GO:0042579	microbody	CC	8	0.00685
GO:0042302	structural constituent of cuticle	MF	20	0.00686
GO:0010165	response to X-ray	BP	5	0.01235
GO:0004035	alkaline phosphatase activity	MF	6	0.01244
GO:0016491	oxidoreductase activity	MF	72	0.01252
GO:0022626	cytosolic ribosome	CC	10	0.0129
GO:0006915	apoptotic process	BP	13	0.02207
GO:0031058	positive regulation of histone modification	BP	5	0.02384
GO:0046670	positive regulation of retinal cell programmed cell death	BP	5	0.02384
GO:0046672	positive regulation of compound eye retinal cell programmed cell death	BP	5	0.02384
GO:2001252	positive regulation of chromosome organization	BP	5	0.02384
GO:0005506	iron ion binding	MF	28	0.02786
GO:0006066	alcohol metabolic process	BP	12	0.0341
GO:0016740	transferase activity	MF	97	0.03458
GO:0007599	hemostasis	BP	6	0.03968
GO:0050817	coagulation	BP	6	0.03968
GO:0008745	N-acetylmuramoyl-L-alanine amidase activity	MF	5	0.04254
Abbreviations: BP, biological process; cellular component, CC; MF, molecular function.				

Table S1.15.

KEGG pathway enrich of *Aedes albopictus* significant expansion families

Pathway	**Gene Number**	**P value**	**Pathway ID**
MAPK signaling pathway	78	2.13E-33	ko04010
Axon guidance	68	3.60E-33	ko04360
alpha-Linolenic acid metabolism	43	1.01E-30	ko00592
Glycerophospholipid metabolism	59	1.20E-30	ko00564
Circadian rhythm - mammal	25	7.93E-17	ko04710
TGF-beta signaling pathway	29	6.44E-13	ko04350
Glycerolipid metabolism	36	4.97E-12	ko00561
Ether lipid metabolism	16	8.62E-09	ko00565
Rheumatoid arthritis	16	5.35E-07	ko05323
Oocyte meiosis	25	7.94E-07	ko04114
Wnt signaling pathway	28	4.09E-06	ko04310
Steroid biosynthesis	14	1.37E-05	ko00100
Fatty acid metabolism	14	2.90E-05	ko00071
Protein processing in endoplasmic reticulum	30	2.03E-04	ko04141
Pentose phosphate pathway	11	2.29E-04	ko00030
Herpes simplex infection	29	5.89E-04	ko05168
Ribosome	18	5.89E-04	ko03010
Tyrosine metabolism	16	8.02E-04	ko00350
Complement and coagulation cascades	20	8.02E-04	ko04610
Antigen processing and presentation	11	1.42E-03	ko04612
Bladder cancer	8	1.50E-03	ko05219
Mucin type O-Glycan biosynthesis	7	2.93E-03	ko00512
Nicotine addiction	9	2.93E-03	ko05033
Glycine, serine and threonine metabolism	12	3.17E-03	ko00260
Collecting duct acid secretion	7	3.17E-03	ko04966
Proximal tubule bicarbonate reclamation	7	3.17E-03	ko04964
Phototransduction - fly	11	4.09E-03	ko04745
Staphylococcus aureus infection	10	5.58E-03	ko05150
Folate biosynthesis	6	7.50E-03	ko00790
Fatty acid elongation	9	9.48E-03	ko00062
Arachidonic acid metabolism	7	1.36E-02	ko00590
Lysosome	21	3.25E-02	ko04142
Gastric acid secretion	14	3.42E-02	ko04971
Peroxisome	17	3.86E-02	ko04146
Amphetamine addiction	10	4.08E-02	ko05031

Table S1.16.

GO enrich of *Aedes albopictus* positive-selection genes.

GO ID	GO Term	GO Class	Gene Number	Pvalue
GO:0016569	covalent chromatin modification	BP	12	0.00014
GO:0016570	histone modification	BP	12	0.00014
GO:0009987	cellular process	BP	142	0.00027
GO:0016568	chromatin modification	BP	13	0.00036
GO:0044267	cellular protein metabolic process	BP	48	0.00091
GO:0005524	ATP binding	MF	33	0.0045
GO:0032559	adenyl ribonucleotide binding	MF	33	0.00482
GO:0030554	adenyl nucleotide binding	MF	33	0.00505
GO:0006325	chromatin organization	BP	15	0.00647
GO:0004402	histone acetyltransferase activity	MF	5	0.00675
GO:0006464	cellular protein modification process	BP	32	0.01296
GO:0036211	protein modification process	BP	32	0.01296
GO:0018205	peptidyl-lysine modification	BP	7	0.01549
GO:0043412	macromolecule modification	BP	33	0.01841
GO:0008135	translation factor activity, nucleic acid binding	MF	9	0.01994
GO:0044237	cellular metabolic process	BP	87	0.01998
GO:0018193	peptidyl-amino acid modification	BP	10	0.02231
GO:0006475	internal protein amino acid acetylation	BP	6	0.02981
GO:0016573	histone acetylation	BP	6	0.02981
GO:0018393	internal peptidyl-lysine acetylation	BP	6	0.02981
GO:0035639	purine ribonucleoside triphosphate binding	MF	35	0.03099
GO:0001883	purine nucleoside binding	MF	35	0.03216
GO:0032550	purine ribonucleoside binding	MF	35	0.03216
GO:0032555	purine ribonucleotide binding	MF	35	0.03276
GO:0044763	single-organism cellular process	BP	108	0.03458
GO:0018394	peptidyl-lysine acetylation	BP	6	0.03543
GO:0017076	purine nucleotide binding	MF	35	0.0366
GO:0032549	ribonucleoside binding	MF	35	0.0366
GO:0001882	nucleoside binding	MF	35	0.03938
GO:0006473	protein acetylation	BP	6	0.04185
GO:0032553	ribonucleotide binding	MF	35	0.04471
GO:0051276	chromosome organization	BP	17	0.0471

Abbreviations: BP - biological process, CC -cellular component, MF - molecular function.

Table S2.

Transposable element families in the genomes of *Aedes albopictus* and *Aedes aegypti*.

Element	*Aedes albopictus*			*Aedes aegypti*		
	Number	**Length(bp)**	**Percent of genome**	**Number**	**Length(bp)**	**Percent of genome**
LINE/CR1	181,331	46,023,318	2.3400	102,000	29,617,029	2.1400
LINE/Dong-R4	15,028	3,079,357	0.1566	5,956	1,593,906	0.1152
LINE/I	152,198	35,635,039	1.8118	107,741	33,362,406	2.4106
LINE/I-Nimb	3,070	638,804	0.0325	1,278	309,325	0.0224
LINE/L1-Tx1	27,081	6,507,384	0.3309	9,161	3,413,990	0.2467
LINE/L2	70,907	18,945,695	0.9633	36,983	10,899,592	0.7876
LINE/LOA	365,673	85,024,251	4.3230	151,138	39,658,013	2.8655
LINE/other	277,553	77,470,059	3.9389	163,074	29,731,180	2.1483
LINE/Penelope	46,704	8,835,062	0.4492	23,244	4,293,947	0.3103
LINE/R1	337,244	73,633,124	3.7438	173,202	45,334,469	3.2757
LINE/Rex-Babar	629	111,326	0.0057	350	55,176	0.0040
LINE/RTE	8,951	1,351,920	0.0687	1,591	159,530	0.0115
LINE/RTE-BovB	1,533,792	308,685,101	15.6949	586,995	125,170,745	9.0443
LINE/RTE-RTE	2,315	521,071	0.0265	855	168,834	0.0122
LINE/RTE-X	1,809	406,105	0.0206	873	195,468	0.0141
LTR/Copia	174,684	43,328,099	2.2030	105,656	29,221,752	2.1114
LTR/ERV1	1,145	129,221	0.0066	522	39,663	0.0029
LTR/ERVK	1,446	148,362	0.0075	451	39,239	0.0028
LTR/Gypsy	350,691	108,573,489	5.5203	286,066	101,400,135	7.3267
LTR/other	344,019	79,453,981	4.0398	209,503	52,553,366	3.7973
LTR/Pao	272,556	87,136,649	4.4304	208,263	78,693,970	5.6861
DNA/Academ	934	328,336	0.0167	800	352,003	0.0254
DNA/CMC-Chapaev-3	21,753	6,677,244	0.3395	2,777	612,353	0.0442
DNA/Crypton	14,325	2,512,865	0.1278	4,113	349,451	0.0252
DNA/En-Spm	2,579	995,087	0.0506	528	380,423	0.0275
DNA/Ginger	36,802	7,732,823	0.3932	449	15,514	0.0011
DNA/Harbinger	1,474	419,412	0.0213	187	85,366	0.0062

DNA/hAT	1,898	503,014	0.0256	535	89,179	0.0064
DNA/hAT-Blackjack	2,791	618,445	0.0314	422	89,744	0.0065
DNA/hAT-Charlie	6,611	1,060,383	0.0539	3,685	651,723	0.0471
DNA/hAT-hAT5	2,531	546,700	0.0278	2,302	626,714	0.0453
DNA/hAT-hATm	24,101	6,182,657	0.3144	7,584	1,683,238	0.1216
DNA/hAT-hATw	326	111,413	0.0057	18	1,697	0.0001
DNA/hAT-hATx	7,567	1,558,358	0.0792	1,392	291,391	0.0211
DNA/hAT-Pegasus	959	284,006	0.0144	354	117,308	0.0085
DNA/hAT-Tip100	1,010	235,206	0.0120	666	99,044	0.0072
DNA/Maverick	2,228	217,284	0.0110	956	53,404	0.0039
DNA/MuDR	2,648	605,175	0.0308	343	149,582	0.0108
DNA/MULE-MuDR	1,881	139,717	0.0071	821	47,541	0.0034
DNA/Novosib	37,073	7,511,349	0.3819	70	53,408	0.0039
DNA/PiggyBac	2,610	738,472	0.0375	1,291	545,751	0.0394
DNA/TcMar-Ant1	1,562	264,721	0.0135	44	10,350	0.0007
DNA/TcMar-Fot1	9,693	2,012,676	0.1023	7,764	1,265,168	0.0914
DNA/TcMar-Tigger	1,055	123,094	0.0063	532	86,218	0.0062
DNA/Zator	22,330	3,754,413	0.1909	22,001	4,321,582	0.3123
Satellite	304,219	87,419,484	4.4448	45,426	20,363,159	1.4714
Simple_repeat	52,064	12,258,690	0.6233	4,855	3,430,168	0.2478
SINE/other	5,699	1,341,942	0.0682	17	4,487	0.0003
TRF	124,698	21,617,926	1.0991	51,425	10,669,453	0.7709
Unknown	114,784	49,009,943	2.4919	30,205	17,390,501	1.2566

Table S3.1.
List of sequences used as BLAST queries and number of BLAST hits in *Aedes albopictus* and *Aedes aegypti*. See additional data file S2 (separate file).

Table S3.2.
Output of BLAST analyses of the *Aedes albopictus* genome annotation (Foshan strain). See additional data file S3 (separate file).

Table S3.3.
Output of BLAST analyses of the *Aedes aegypti* genome, AaegL3 assembly. See additional data file S4 (separate file).

Table S3.4.
Mapping cohordinates of sequences spanning partial or complete flaviviral ORFs in the *Aedes albopictus* genome assembly of the Foshan strain. See additional data file S5 (separate file).

Table S3.5.
Argot2-based annotation of putative viral integrations. See additional data file S6 (separate file).

Table S3.6.
BLAST hits to viral sequences other than flavivirus-like sequences, such as the Negev virus and the Wuhan mosquito virus 8. See additional data file S7 (separate file).

Table S3.7.
piRNAs identified within NIRVs. See additional data file S8 (separate file).

Table S4.1.
Annotation of diapause related genes. See additional data file S9 (separate file).

Table S4.2.
211 gene models from the *Ae. albopictus* gene expansion family that are present in the *Ae. albopictus* diapause transcriptome. See additional data file S10 (separate file).

Table S4.3.
Differential expression of genes from *Ae. albopictus* gene expansion families and from complete diapause transcriptome at seven life-cycle stages.

Type	**Developing embryos**		**Pharate larvae**			**Adults**[1]	
	3dpo[2]	**6dpo**	**11dpo**	**21 dpo**	**40 dpo**	**blood fed**	**non-blood fed**
Number of genes from expansion families in diapause transcriptome	159	159	162	162	162	211	211
Number of DE[3] genes from expansion families	46	81	13	5	1	35	43
% DE genes from expansion families	**0.289**	**0.509**	**0.080**	**0.031**	**0.006**	**0.166**	**0.204**
Number of genes in diapause transcriptome	11397	11397	11207	11207	11207	11783	11783
Number of DE genes in diapause transcriptome	2721	4518	383	116	35	877	1817
% of DE genes in diapause transcriptome	**0.239**	**0.396**	**0.034**	**0.010**	**0.003**	**0.074**	**0.154**
Fisher's exact test	p=0.160	p=0.004	p=0.004	p=0.029	p=0.404	p<0.001	p=0.055

[1]12 days post eclosion.
[2] Days post oviposition.
[3]Differential expression under diapause vs. non-diapause photoperiod conditions. Note that the number of genes adds up to more than 140 because some genes are expressed differentially at multiple life-cycle stages.

Table 4.4.

Contrasting expression between pre-adult and adult life-stages for genes from *Ae. albopictus* gene expansion families according to protein super-family annotation.

Type	Protein super family category				
	Stress response	**Lipid metabolism**	**Gene expression regulation**	**Serine protease related**	**Others**
Number of genes from expansion family in category	12	12	18	10	44
Number of genes with contrasting expression	10	10	15	10	38
Proportion of genes with contrasting expression	0.83	0.83	0.83	1.00	0.86
Fisher's exact test[1]	$p = 0.043$	$p = 0.043$	$p = 0.012$	$p = 0.003$	$p < 0.001$

[1]One-tailed comparison versus proportion of genes with contrasting expression patterns from the complete diapause transcriptome (0.55).

Table S4.5.

Putative unique *Ae. albopictus* miRNAs with support from short non-coding RNA reads of mature oocytes. See additional data file S11(separate file).

Table S5.1.
Genome coordinates of the *Ae. albopictus* CYP gene family. See additional data file S12 (separate file).

Table S5.2.
Accession numbers of CYP sequences used for phylogenetic analysis. See additional data file S13 (separate file).

Table S5.3.
The number of genes belonging to different clans in the CYP gene family of *D. melanogaster*, *Ae. albopictus*, *Ae. aegypti* and *An. gambiae*[*].

CYP - clan	***D. melanogaster***	***An. gambiae***	***Ae. aegypti***	***Ae. albopictus*[1]**
CYP4 clan	**32**	**44**	**61**	**76 (84)**
CYP4	22	29	27	25 (28)
CYP325	0	15	34	52 (56)
Others	10	0	0	0
CYP3 clan	**37**	**40**	**85**	**92 (107)**
CYP9	5	9	38	38 (49)
CYP6	23	30	46	52 (56)
Others	9	1	1	2
CYP2 clan	**7**	**11**	**12**	**10**
Mito CYP clan	**11**	**9**	**10**	**8 (9)**
Total	87	104	168	186 (210)

[*]Numbers were derived from this study, Strode et al. 2008 (89), and from Vectorbase (87) and Flybase (90) (Table S5.2).
[1]The total number of genes including pseudogenes for *Ae. albopictus* is shown between brackets.

Table S5.4.
Genome coordinates of the *Ae. albopictus* CCE gene family. See additional data file S14 (separate file).

Table S5.5.
Accession numbers of CCE sequences used for phylogenetic analysis. See additional data file S15 (separate file).

Table S5.6.
The number of genes belonging to different clades in the CCE gene family of *D. melanogaster*, *Ae. albopictus*, *Ae. aegypti* and *An. gambiae*[*].

CCE - clade	*D. melanogaster*	*An.gambiae*	*Ae. aegypti*	*Ae. albopictus*[1]
B (Dipteran mitochondrial, cytosolic and secreted esterases)	2	14	21	25 (27)
C (Dipteran microsomal esterases)	11	2	2	2
D (Integument esterases)	3	0	1	1
E (Beta esterases)	3	5	2	2
F (Juvenile Hormone Esterases)	2	5	7	9
G (Mosquito specific CCE clade)	0	3	6	6
H (Glutactin)	4	5	10	9 (13)
I (Uncharacterized group)	2	2	2	3
J (Acetylcholinesterases)	1	2	2	3
K (Gliotactin)	1	1	1	1
L (Neuroligin)	3	5	3	1
M (Neurotactin)	2	2	2	2 (3)
Total	34	46	59	64 (71)

[*]Numbers were derived from this study, Strode et al. 2008 (89), and from Vectorbase (87) and Flybase (90) (Table S5.5).
[1]The total number of genes including pseudogenes for *Ae. albopictus* is shown between brackets.

Table S5.7.
Genome coordinates of the *Ae. albopictus* GST gene family. See additional data file S16 (separate file).

Table S5.8.
Accession numbers of GST sequences used for phylogenetic analysis. See additional data file S17 (separate file).

Table S5.9.
The number of genes belonging to different classes in the GST gene family of *D. melanogaster*, *Ae. albopictus*, *Ae. aegypti* and *An. gambiae*[*].

GST - class	*D.melanogaster*	*An. gambiae*	*Ae. aegypti*	*Ae. albopictus*[1]
Delta	11	12	8	11 (14)
Epsilon	14	8	8	10 (11)
Omega	5	1	1	0 (1)
Sigma	1	1	1	1
Theta	4	2	4	5
Zeta	2	1	1	1
Others	0	3	3	4
Total	37	28	26	32 (37)

[*]Numbers were derived from this study, Strode et al. 2008 (89), and from Vectorbase (87) and Flybase (90) (Table S5.8).

[1]The total number of genes including pseudogenes for *Ae. albopictus* is shown between brackets.

Table S5.10.
Genome coordinates of the *Ae. aegypti* ABC transporter gene family. See additional data file S18 (separate file).

Table S5.11.
Genome coordinates of the *Ae. albopictus* ABC transporter gene family. See additional data file S19 (separate file).

Table S5.12.
The number of putative ABC genes belonging to different subfamilies in the ABC gene family of *D. melanogaster*, *Ae. albopictus*, *Ae. aegypti* and *An. gambiae*[*].

ABC - subfamily	***D. melanogaster***	***An. gambiae***	***Ae. aegypti***	***Ae. albopictus***
A	10	9	11	12
B - FT	4	1	1	1
B - HT	4	4	4	4
C	14	13	17	19
D	2	2	2	3
E	1	1	1	1
F	3	3	3	4
G	15	16	16	23
H	3	3	3	4
Total	56	52	58	71

[*]Numbers were derived from Dermauw and Van Leeuwen 2014 and this study

Table S5.13.
Nucleotide sequences of *Ae. aegypti* ABC transporter genes and of *Ae. albopictus* genes involved in detoxification (CYPs, CCEs, GSTs, ABCs). See additional data file S20 (separate file).

Table S6.1.

Comparisons of numbers of OBP and OR predicted genes and domains among hematophagous and non-hematophagous Diptera. Abbreviation: NA - not available.

Species	OR[1]	Aligned[2] (sensitivity)	OR_domain[3] (sensitivity)	OBP[1]	OBP_Align[2] (sensitivity)	OBP_domain[3] (sensitivity)
Ae.aegypti	109	102(93.5%)	74(67.9%)	64	63(98.4%)	58(90.6%)
Ae.albopictus	NA	158	112	NA	86	83
An.gambiae	73	73(100&)	73(100%)	54	49(90.7%)	45(83.3%)
Cx.quinquefasciatus	88	88(100%)	78(88.6%)	47	46(97.9%)	46(98.4%)
D.melanogaster	61	59(96.7%)	59(96.7%)	51	51(100%)	44(86.3%)

[1]The numbers of *Ae. albopictus* genes were derived from this study and Vectorbase, and those of other insects were derived from Vectorbase.

[2]The number of OR and OBP were predicted from alignments with *Ae. albopictus.*

[3]The number of OR and OBP domains were predicted from predicted olfactory genes using InterProScan.

Table S6.2.
The expression of representative olfactory genes in developmental stages.

	Number of OBP/OR genes						
Expressional level	**Embryos (0-24h)**	**Embryos (24-48h)**	**Larvae (1-2 instar)**	**Larvae (3-4 instar)**	**Pupae**	**Male**	**Female**
>1	11/14	13/14	23/16	26/13	30/16	38/30	32/46
0.1-1	23/7	19/6	18/11	18/12	24/7	26/11	25/12
<0.1	74/132	75/133	68/126	64/128	55/130	44/112	50/95
Stage-specific	0/8		3/5		3/4	3/8	8/13

Table S6.3.

Developmental expressional profiles of AalbOBPs and AalbORs. Zero values indicate that transcripts are not detected.

AalbOBP							
gene name	**Eggs 0-24h**	**Eggs 24-48h**	**Larval stages 1-2**	**Larval stages 3-4**	**Pupae**	**Male**	**Female**
CCG024477.1	0.244299402	0.208342369	0.274665759	0.167410657	0.047712528	1.195270845	152.7218537
CCG007517.1	0.095005323	0.777811511	312.6611894	32.42186388	7.45905862	309.3427351	106.201701
CCG024478.1	0.275601027	0.040292038	0	0.080940288	0	0.481578093	73.23818089
CCG014341.1	139.4076668	298.6544392	108.5305192	2764.051716	1551.043682	121.370152	64.7582822
CCG027078.1	2.13863046	1.237051286	94.62527607	73.11749487	15.58135522	142.8504073	63.83972501
CCG025127.1	0.366733834	1.219749869	0	7.539333079	172.9729302	729.0462576	56.23679016
CCG023073.1	0.091046768	0	4.91346525	172.2004618	45.983612	45.21416323	51.9592305
CCG013191.1	0.198647493	0.609874935	20.8662745	36.03958281	37.2446889	514.3845064	47.16354857
CCG000094.1	0.108353178	0.258734821	0	0	0.08464702	1.060267802	46.19588944
CCG001999.1	0	0	0	0.137515897	0	3.109127139	43.36114752
CCG018602.1	0	0	0.377152983	0.536380413	2.904529844	99.4904402	41.31637934
CCG027709.1	13.77098738	46.71905255	25.26924986	755.2689631	651.8959334	71.41456664	13.14888979
CCG021771.1	2.250349962	5.106564751	1.414323686	2.413711859	0.687914963	19.98585026	13.042791
CCG005090.1	0.213761977	0.218759487	2.197326075	2.636717846	7.598220161	38.34827295	12.91306731
CCG021480.1	0.086254833	0.088271372	0	0	0	5.169677466	12.17461761
CCG005538.1	0	0	0	0.256696341	0.438955262	2.901851996	11.85826364
CCG023075.1	0.05555396	0	1.070730926	0	0.716092271	15.08522971	11.47150096
CCG028521.2	0.275050375	0.93826913	0	0.376966654	0	37.56814279	9.505934701
CCG008468.1	0.191397585	0.195872242	0	0.393476142	0.112141855	47.29026573	5.586478585
CCG026944.1	0	0	0	0	0	4.166977691	5.396739214
CCG014342.1	9.979020638	6.953068081	9.309723632	29.95396872	40.74109062	21.6850521	5.179704357
CCG023074.1	1.394758996	0.285473374	84.14122559	4471.349863	88.09458531	4.947446433	5.103937405
CCG014636.1	0.31890976	0.217577004	0.034147635	0.655616329	7.723239881	7.671539026	4.692732152
CCG026945.1	0	0	0	0	1.453297827	2.925586917	3.781971636
CCG019174.1	3.548770257	10.59256465	151.6154992	120.6810559	27.49246115	0.241151135	3.779070016
CCG009948.1	0.177172089	0.36262834	0	0	0	1.191905781	2.932957595
CCG018890.1	0	0	0.864257731	0.526770992	0.450393836	8.984647189	2.827867909
CCG005057.1	0	0.160685612	0.151312873	0	0	4.129171687	1.915249964
CCG009949.1	0	0	0	0	0	0	1.439130457
CCG024996.1	0	0	0.4749859	0	0.346543628	2.049784651	1.116543414
CCG004000.1	0.03524391	0.036067872	0.033964046	0.144909224	3.964757206	3.190066498	0.951924834
CCG016092.1	0.096758189	0	0	0	0	0	0.927336369
CCG024997.1	0.086254833	0	0	0.797954085	1.617203597	5.802699196	0.901823526
CCG015708.1	0.159887007	0.163624982	0	0	0.093679477	0	0.835836439
CCG013211.1	1.213956904	4.845117537	6.492793366	13.66373229	15.50568819	12.25014275	0.740385982
CCG020567.1	0.33790553	0.069161075	79.38973087	43.55567933	2.138209911	0.247987894	0.706583382
CCG007787.1	0	0	0	0	0	8.334786118	0.676367645
CCG019176.1	0	0	0	0.641740851	54.4629677	2.460634734	0.664836392

CCG019175.1	0	0.100128721	0.565729474	1.307427257	121.187686	14.00206276	0.596729
CCG008469.1	6.748172201	17.56081062	14.02814974	28.9349625	3.050093549	0.589579063	0.503960206
CCG017320.1	0.106591338	0	0.513602639	224.7188635	11.74116107	0.521513838	0.464353577
CCG013539.1	0	0	0	0	0	0.370074242	0.439349923
CCG003160.1	0	0	0.360989284	0.256696341	0.292636841	2.596393891	0.435165638
CCG007857.1	0	0.081070988	0	0.040714676	0.185660836	0	0.414130441
CCG017969.1	0	0.100128721	0	0	0.114652494	0.11967575	0.340988
CCG019178.1	0.478493962	1.07729733	9.037906884	12.00102234	27.47475454	0.819385792	0.33352111
CCG002302.1	0	0	0	0	0	0.659036322	0.312961589
CCG009587.1	0	0	0	0	0.200829205	0.733698389	0.223982314
CCG022643.1	0.038560984	0.078924992	0.074321323	0	0	0.09433265	0.201584082
CCG001738.1	0	0	0	0	0	0.548258137	0.195266633
CCG019177.1	0.852730703	1.854416468	35.23314106	0.87652409	5.745674569	0.260756919	0.185741431
CCG013541.1	0	0	0	0	0	1.252855508	0.178485906
CCG013003.1	0	0	0	0	0	0	0.115384828
CCG013137.1	0	0.173686066	0.040888754	0	0.049719852	0.051898222	0.110903864
CCG013138.1	0	0	0.243754822	0	0	0	0.110190656
CCG010901.1	0	0	0	0	0	0	0.085247
CCG017322.1	0.194233105	0.298161079	14.60001103	351.9877265	5.690160805	0	0.084615541
CCG021668.1	0.478493962	0.04896806	0.09222354	0.049184518	0.336425566	0	0.083380277
CCG000521.1	0	0	0	0	0	0	0.083380277
CCG008607.1	0	0	0	0.093587207	0	0	0.079327069
CCG011068.1	0.043413028	0.088855951	0	0	0	0.159303482	0.075649656
CCG013135.1	0	0	0	0	0	0	0.074660771
CCG022738.1	0.169170769	0	0.326054837	0	0.09911893	0	0.073697406
CCG017938.1	0	0	0	0	0	0	0.073460437
CCG002301.1	0	0	0	0	0	0	0.073224987
CCG013136.1	0	0	0	0	0	0	0.072070019
CCG003679.1	0	0	0	0	0.09063973	0	0.067392909
CCG020570.1	0	0	50.40264354	33.47403086	0	0	0.061414505
CCG015686.1	0	0	0	0	0.16172036	0.168805795	0.060121568
CCG007451.1	0	0.04896806	0.09222354	0.196738071	0	0.117055113	0.041690139
CCG015687.1	0.09398376	0.096180993	0.045285394	0	0.110132145	0	0.040943004
CCG008889.1	0.044899776	0	0	1.569188246	0	0.109839387	0.039120199
CCG003184.1	0	0	0	0.046152595	0	0.054919694	0.039120199
CCG010900.1	0	0	0	0	0.052256579	0	0.038854075
CCG006398.1	0	0	0.042829237	0.045683247	0.156238314	0.217444753	0.038722366
CCG013004.1	0	0.044873741	0	0.0450721	0	0	0.038204341
CCG009315.1	0.127288685	0	0.122666261	0.087226912	0.099439703	0	0.036967955
CCG022642.1	0.061746003	0	0	0.042312584	0	0.025175119	0.01793265
CCG008470.1	0	0.114677338	0.971894225	189.4781001	80.49389015	1.096516274	0
CCG016091.1	0	0	0	0	0	0.863506566	0
CCG006396.1	0.044443168	0	0.214146185	0.182732988	0.572873817	0.326167129	0

CCG009946.1	0	0	0	0	0	0.267275842	0
CCG020118.1	0	0	0	0	0.104869858	0.218929017	0
CCG009317.1	0.085691076	0.087694435	0.165158496	0	0.050207301	0.157221083	0
CCG017970.1	0	0	0.04417701	0	0	0.11214371	0
CCG006397.1	0	0	0.042829237	0.045683247	0.72911213	0.108722376	0
CCG017937.1	0	0	0	0	0	0.104133445	0
CCG017321.1	0.639548027	0	2.465292669	238.3049869	3.122649222	0	0
CCG017319.1	0	0	34.0542625	11.89727375	2.28050976	0	0
CCG009316.1	0.090418859	0.046266374	0	0	0.370841515	0	0
CCG022739.1	0.514146453	0	0	0.352328311	0.200829205	0	0
CCG014080.1	0	0	0	0	0.110132145	0	0
CCG009318.1	0	0	0	0	0.101075225	0	0
CCG013134.1	0	0.262226356	0	0	0.100087519	0	0
CCG019526.1	0	0	0.096447519	0	0.058639062	0	0
CCG014691.1	0	0	0	0	0.056070928	0	0
CCG017939.1	0.043702449	0.089448324	0.168461666	0	0.051211447	0	0
CCG008888.1	0	0	0	0.275970468	0	0	0
CCG015707.1	0	0	0	0.088082078	0	0	0
CCG022644.1	0.430501732	0	0	0.080457062	0	0	0
CCG017194.1	0	0.108203617	0.101892137	0	0	0	0
CCG021149.1	0	0	0.087740451	0	0	0	0

Table S6.3. (continued)

Developmental expressional profiles of AalbOBPs and AalbORs. Zero values indicate that transcripts are not detected.

AalbOR							
AalbOBP	**Egg 0-24h**	**Eggs 24-48h**	**Larval stages 1-2**	**Larval stages 3-4**	**Pupae**	**Male**	**Female**
CCG030001.1	239463.6015	17959.77011	201653.5592	200726.8425	425362.9764	393279.6376	324837.5812
CCG030024.1	211111.1111	416666.6667	102339.1813	208333.3333	109649.1228	6082.725061	50724.63768
CCG030035.1	0	0	0	94632.44757	33868.45492	28182.50993	39169.60439
CCG030019.1	0	26123.30199	36664.28349	0	0	7627.241456	36345.46363
CCG030071.1	79365.07937	0	104427.736	0	0	0	25879.91718
CCG030034.1	17993.7022	185560.054	0	0	0	19701.1338	23470.04635
CCG030012.1	11223.34456	0	0	0	0	0	21958.71761
CCG030042.1	0	0	0	0	0	0	18821.75795
CCG030009.1	0	0	0	24757.3777	0	12288.33346	18298.93134
CCG030002.1	43254.93377	0	0	0	10671.44748	11839.85413	17631.08713
CCG030031.1	0	0	0	26568.89314	0	0	15710.30203
CCG030068.1	0	0	0	0	0	0	15096.61836
CCG030028.1	11138.95851	0	0	0	0	0	14529.07631
CCG030053.1	0	0	0	0	21712.69759	12045.00012	14349.26101
CCG030063.1	0	0	0	0	0	5862.867529	13968.91915
CCG030060.1	0	0	0	0	0	0	12825.44568

CCG030039.1	0	0	0	0	0	6540.564582	11687.70453
CCG030033.1	22446.68911	0	0	0	0	0	10979.35881
CCG030054.1	0	0	27913.85783	0	0	5806.897433	10376.67324
CCG030081.1	0	0	32131.61108	13466.92523	24098.70831	0	7963.051441
CCG030079.1	0	0	15553.06706	0	23329.6006	0	7708.911502
CCG030008.1	0	0	15268.80735	0	0	0	7568.017558
CCG030085.1	0	0	0	0	0	55578.19853	7356.727728
CCG030023.1	0	21150.59222	0	0	0	0	7356.727728
CCG030006.1	0	0	0	0	0	0	7356.727728
CCG030049.1	0	0	0	0	0	0	7356.727728
CCG030048.1	0	0	0	0	0	0	7356.727728
CCG030004.1	0	0	0	0	0	0	7301.135327
CCG030075.1	0	0	205706.8971	147797.8126	0	0	7282.790765
CCG030046.1	0	0	14656.52435	0	21984.78653	24391.87994	7264.538157
CCG030076.1	0	0	0	0	0	0	7246.376812
CCG030010.1	0	20475.02048	0	0	0	0	7121.746252
CCG030036.1	0	10137.8751	71142.98317	23853.82377	0	23679.70827	7052.434853
CCG030057.1	0	0	0	23567.11916	0	0	6967.670011
CCG030047.1	0	0	0	0	0	5779.311222	6884.918586
CCG030017.1	0	0	15594.54191	0	23391.81287	0	3864.7343
CCG030050.1	11574.07407	21701.38889	0	0	0	19008.51582	3774.154589
CCG030041.1	172711.5717	107944.7323	0	0	0	0	3754.599384
CCG030040.1	0	0	0	0	0	0	3754.599384
CCG030022.1	0	0	29989.50367	0	0	12477.38474	3716.090673
CCG030045.1	0	0	0	0	0	0	3632.269078
CCG030030.1	11111.11111	0	0	0	0	0	3623.188406
CCG030051.1	0	0	0	0	0	0	3552.145496
CCG017217.3	0.285015969	0.291679317	0.274665759	0.732421624	0.222658466	0.639137882	4.428519878
CCG006802.1	0.044025301	0.045054562	2.469224885	1.013683333	2.692986111	8.820641279	3.60567902
CCG000221.1	0.704878202	0.057708596	0.434739783	0.550655053	1.486783952	2.586540404	2.554843424
CCG011758.1	0	0.135527763	0.159528093	0.102095135	0	0	0.576924142
CCG018383.1	0	0	0	0	0.043895526	0.091637431	0.489561343
CCG004497.1	0.067349664	0.160823185	0.454327266	0.530754848	0.92075376	1.839809733	0.430322185
CCG013368.1	0.520267245	1.437562346	0.250687003	0.74869766	0.243864034	0.063637105	0.362638032
CCG000551.1	0	0	0	0.234783238	0.267655648	0.167629448	0.318413882
CCG018937.1	0	0	0.247737744	0.052849247	0	0	0.268778777
CCG018936.1	0	0	0.067564839	0	0	0	0.244344342
CCG000550.1	0.337036878	0.689832831	0	0.034644108	0.315957515	0.37102559	0.205557033
CCG003920.1	0.104329453	0.213537112	0.268108752	0.142987351	0.529773592	0.425372692	0.181799968
CCG016854.1	0	0	0	0	0.08086018	0.253208692	0.180364705
CCG003919.1	0	0	0.02993987	0	0.072812484	0	0.162413716
CCG013214.1	0	0.06675248	0	0	0	0	0.113662667
CCG019854.1	0	0	0	0	0	0	0.099331287

CCG000278.1	0.037892296	0	0.146065028	0.155798357	0	0.092696824	0.099044202
CCG010412.1	0	0.081812491	0	0.164348267	0	0	0.069653037
CCG019852.1	0	0	0	0	0	0	0.06902174
CCG001781.1	0.072235452	0.073924235	0.278449034	0	0	0	0.062937179
CCG013594.1	0.068107712	0.069699993	0.984516228	0.280032371	0	0.124960134	0.059340769
CCG027084.1	0.260910141	0.03337624	0.031429415	0	0.038217498	0.279243417	0.056831333
CCG005181.1	0	0	0	0	0	0.078227076	0.055722429
CCG016235.1	0	0	0.030592312	0	0.074399197	0.15531768	0.055317666
CCG007974.1	0.034684483	0	0	0	0	0.042424737	0.030219836
CCG014651.1	0.034142538	0.069881503	0	0	0	0.125285551	0.029747651
CCG010978.1	0	0	0.062547648	0	0	0.079388864	0.028274995
CCG030026.1	0	0	14767.55863	0	0	30720.83364	0
CCG030043.1	0	0	0	0	0	24030.51876	0
CCG030082.1	0	0	0	0	0	13405.45468	0
CCG030016.1	0	0	0	0	0	12805.73697	0
CCG030020.1	11484.35257	0	15110.99022	25333.13067	0	12574.10865	0
CCG030044.1	0	0	0	0	0	12015.25938	0
CCG030074.1	0	0	0	0	0	12015.25938	0
CCG030052.1	0	18726.59176	0	0	19712.20185	10935.23606	0
CCG030038.1	25839.79328	0	0	0	25499.796	7072.936117	0
CCG030083.1	0	0	0	0	0	6629.673091	0
CCG016233.1	0	0.070617098	0.199494078	0.070929252	0	0.33761159	0
CCG003917.1	0	0	0.030154236	0	0	0.153093561	0
CCG025037.1	0.032613768	0	0.031429415	0.067047552	0	0.079783833	0
CCG019769.1	0	0	0	0	0	0.070028605	0
CCG014266.1	0	0	0	0	0	0.045173382	0
CCG008906.1	0.069368966	0	0	0.03565227	0.243864034	0.042424737	0
CCG008907.1	0.065227535	0.06675248	0	0	0.076434996	0.039891917	0
CCG030029.1	0	31565.65657	0	0	22151.33794	0	0
CCG030007.1	0	0	0	0	11131.89064	0	0
CCG030077.1	0	0	0	0	10467.69669	0	0
CCG011520.1	0.060279239	0	0.029045115	0	0.070636479	0	0
CCG002781.1	0	0	0	0	0.0584161	0	0
CCG012996.1	0	0	0	0	0.040644006	0	0
CCG015303.1	0	0.06742336	0.063490578	0.101582095	0.038601593	0	0
CCG030064.1	0	19984.01279	0	35265.90492	0	0	0
CCG030078.1	0	0	0	13248.54266	0	0	0
CCG003918.1	0	0	0	0.104469441	0	0	0
CCG001075.1	0	0	0	0.032163623	0	0	0
CCG012994.1	0	0	0.06285883	0	0	0	0
CCG006796.1	0	0	0.060743389	0	0	0	0
CCG030027.1	0	11022.92769	0	0	0	0	0
CCG010302.1	0.122149701	0.125005421	0	0	0	0	0

CCG019062.1	0.038335481	0.078463442	0	0	0	0	0
CCG007975.1	0.034053856	0.034849996	0	0	0	0	0
CCG000712.1	0	0.032175656	0	0	0	0	0
CCG030072.1	11028.39813	0	0	0	0	0	0
CCG026720.1	0.07011088	0	0	0	0	0	0
CCG020411.1	0.034053856	0	0	0	0	0	0
CCG000962.1	0.031516189	0	0	0	0	0	0

Table S6.4.

Gene annotation of AalbOBPs and AalbORs.

ID	Protein	Molecular Function	Biological Process	Cellular Component
CCG011068.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG000094.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG000521.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG001738.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG001999.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG002301.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG002302.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG003160.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG003184.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG003679.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG004000.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG004118.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG005057.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG005090.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG005538.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG006396.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG006397.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG006398.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG006399.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG007451.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG007517.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG007787.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG007857.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG008134.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG008468.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG008469.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG008470.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG008607.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG008888.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG008889.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG009314.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG009315.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG009316.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG009317.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG009318.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG009587.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG009946.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG009948.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG009949.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG010900.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space

CCG010901.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013003.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013004.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013134.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013135.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013136.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013137.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013138.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013191.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013211.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013539.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013541.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG013706.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG014080.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG014081.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG014341.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG014342.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG014636.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG014691.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG015686.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG015687.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG015707.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG015708.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG016091.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG016092.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017194.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017319.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017320.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017321.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017322.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017937.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017938.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017939.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017969.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG017970.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG018393.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG018602.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG018890.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG019174.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG019175.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG019176.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG019177.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG019178.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space

CCG019526.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG020118.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG020119.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG020567.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG020570.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG021149.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG021480.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG021668.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG021669.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG021771.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG022257.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG022642.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG022643.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG022644.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG022738.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG022739.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG023073.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG023074.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG023075.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG024477.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG024478.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG024996.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG024997.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG025127.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG026944.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG026945.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG027078.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG027709.1	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG028521.2	OBP	odorant binding	Pheromone/general odorant binding	extracellular space
CCG019852.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000084.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000085.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000086.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000087.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000088.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000221.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000278.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000550.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000551.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000712.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG000962.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG001075.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG001781.1	OR	olfactory receptor activity	sensory perception of smell	membrane

CCG001943.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG002781.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG003917.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG003918.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG003919.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG003920.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG004018.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG004496.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG004497.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG004843.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG005181.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG006796.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG006802.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG007388.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG007898.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG007974.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG007975.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG008015.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG008906.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG008907.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG010302.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG010412.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG010815.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG010978.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG011372.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG011520.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG011757.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG011758.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG012784.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG012785.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG012994.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG012995.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG012996.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG013214.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG013368.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG013594.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG014266.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG014651.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG015303.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG016233.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG016235.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG016854.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG017037.1	OR	olfactory receptor activity	sensory perception of smell	membrane

CCG017217.3	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG018383.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG018935.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG018936.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG018937.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG019061.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG019062.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG019149.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG019150.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG019151.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG019254.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG019769.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG019854.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG019855.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG020411.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG020881.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG021484.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG025037.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG026720.1	OR	olfactory receptor activity	sensory perception of smell	membrane
CCG027084.1	OR	olfactory receptor activity	sensory perception of smell	membrane

Table S6.5.

Novel *Aedes albopictus* OBP and OBP genes.

Gene ID	Name	Amino-acid/ signal peptide length	MW(kDa)	Signal peptide	CDD prediction(E-value)
CCG025127.1	OBP	142/20	16.171	88.6%	PBP_GOBP(7.56e-15)
CCG008468.1	OBP	136/18	15.614	86.5%	PBP_GOBP(1.80e-12)
CCG014636.1	OBP	369/26	42.325	89.2%	PBP_GOBP(4.47e-11)
CCG004000.1	OBP	371/26	42.468	89.4%	PBP_GOBP(5.39e-11)
CCG016092.1	OBP	270/21	30.248	66.2%	PBP_GOBP(7.38e-08)
CCG024997.1	OBP	151/0	17.100	No	PBP_GOBP(1.74e-18)
CCG015708.1	OBP	163/0	18.793	No	PBP_GOBP(5.89e-17)
CCG020567.1	OBP	193/19	20.837	75%	No CDD
CCG007787.1	OBP	151/23	17.150	75%	PBP_GOBP(2.05e-21)
CCG008469.1	OBP	135/18	15.494	82.3%	PBP_GOBP(3.22e-11)
CCG017320.1	OBP	122/0	14.244	No	PBP_GOBP(2.62e-15)
CCG013539.1	OBP	129/0	14.495	No	PBP_GOBP(4.43e-05)
CCG007857.1	OBP	330/17	35.796	76.6%	PBP_GOBP(6.57e-09)
CCG013003.1	OBP	296/19	32.996	61.1%	PBP_GOBP(3.16e-05)
CCG013137.1	OBP	308/22	34.032	66.9%	PBP_GOBP(1.42e-13)
CCG017322.1	OBP	134/17	15.539	85.5%	PBP_GOBP(7.16e-16)
CCG000521.1	OBP	273/18	31.004	48.3%	PBP_GOBP(4.61e-08)
CCG008607.1	OBP	287/18	32.407	65.1%	PBP_GOBP(2.35e-14)
CCG013135.1	OBP	305/20	34.329	84.0%	PBP_GOBP(7.84e-14)
CCG022738.1	OBP	154/22	17.794	75.2%	PBP_GOBP(4.91e-17)
CCG013136.1	OBP	316/32	35.479	77.2%	PBP_GOBP(9.47e-13)
CCG007451.1	OBP	273/17	31.283	82.0%	PBP_GOBP(3.68e-14)
CCG003184.1	OBP	291/0	33.004	No	PBP_GOBP(1.09e-09)
CCG006398.1	OBP	294/20	33.535	84.3%	PBP_GOBP(8.48e-08)
CCG009315.1	OBP	308/0	34.165	No	PBP_GOBP(2.03e-13)
CCG008470.1	OBP	116/17	13.512	81.8%	PBP_GOBP(4.00e-15)
CCG016091.1	OBP	259/22	29.112	62.7%	PBP_GOBP(5.06e-08)
CCG006396.1	OBP	294/19	33.595	77.9%	PBP_GOBP(1.53e-12)
CCG020118.1	OBP	292/21	32.981	89.2%	PBP_GOBP(6.89e-15)
CCG009317.1	OBP	305/20	34.333	90.2%	PBP_GOBP(6.59e-14)
CCG006397.1	OBP	294/19	33.336	78.7%	PBP_GOBP(6.27e-11)
CCG017321.1	OBP	122/0	14.549	No	PBP_GOBP(1.27e-13)
CCG017319.1	OBP	127/0	14.381	No	PBP_GOBP(5.63e-15)
CCG009316.1	OBP	289/0	32.700	No	PBP_GOBP(1.09e-12)
CCG022739.1	OBP	152/20	17.677	64.6%	PBP_GOBP(1.82e-16)
CCG014080.1	OBP	278/21	32.103	88.0%	PBP_GOBP(1.34e-11)
CCG009318.1	OBP	303/17	34.281	89.1%	PBP_GOBP(1.36e-15)
CCG013134.1	OBP	306/17	34.395	67.7%	PBP_GOBP(1.69e-15)
CCG014691.1	OBP	273/0	30.871	No	PBP_GOBP(1.38e-08)
CCG015707.1	OBP	152/20	17.687	85.2%	PBP_GOBP(2.58e-16)

CCG017194.1	OBP	247/0	28.746	No	PBP_GOBP(1.74e-14)
CCG021149.1	OBP	287/18	32.422	70.0%	PBP_GOBP(2.32e-14)
CCG008134.1	OBP	201/0	23.471	No	PBP_GOBP(1.25e-05)
CCG001781.1	OR	362/0	41.346	No	7tm_6 super familyOdorant receptor(8.64e-30)
CCG001943.1	OR	390/0	44.616	No	7tm_6 super familyOdorant receptor(1.78e-20)

Table S7.1.
Transcript analysis of adult male vs. adult female gene expression. See additional data file S24 (separate file).

Table S7.2.
Candidate male-specific gene expression. See additional data file S25 (separate file).

Table S7.3.
Candidate female-specific gene expression. See additional data file S26 (separate file).

Table S7.4.
Putative homologs of *Drosophila melanogaster* genes involved in sex determination

***Drosophila melanogaster* gene**	***Aedes albopictus* genes**	**locus**
doublesex	CCG017777.1	scaffold4592:42904-4351
fruitless	CCG011203.1	scaffold253:32810-33949
	CCG011203.1	scaffold253:32810-33949
	CCG011204.1	scaffold253:38036-39655
	CCG015477.1	scaffold380:24520-26644
	CCG005952.1	scaffold1524:39941-50757
	CCG007440.1	scaffold1721:55518-61312
	CCG007442.1	scaffold1721:118676-158367
	CCG007444.1	scaffold1721:175734-177859
transformer-2	CCG015404.1	scaffold377:370900371478
	CCG008877.1	scaffold1991:199579221331
	CCG008878.1	scaffold1991:222889-233758
transformer	No[1]	

[1]No homologs were found.

Table S7.5.
Full list of Biological Process Gene Ontology categories enriched in sex-biased genes (terms from the Function Ontology with p-value ≤ 1). See additional data file S27 (separate file).

Table S7.6.
Full list of Cellular Compartment Gene Ontology categories enriched in sex-biased genes (terms from the Function Ontology with p-value ≤ 1). See additional data file S28 (separate file).

Table S7.7.
Full list of Molecular Function Gene Ontology categories enriched in sex-biased genes (terms from the Function Ontology with p-value ≤ 1). See additional data file S29 (separate file).

Table S7.8.
Sex-bias genes of stable expression during development. See additional data file S30 (separate file).

Table S8.1.
Numbers of immune-related genes in specific families.

Family	***Aedes albopictus***	***Aedes aegypti***	***Anopheles gambiae***	***Culex quinquefasciatus***	***Drosophila melanogaster***
CASP	12	10(11)	15	16	7
SPZ	13	9	6	7	6
BGBP	13	7	7	12	3
SRRP	41 (42)	39	26	35	24
LYS	9	7	8	5	11
IAP	5	5	7 (8)	6	4
AMP	11	16	11	6	21
GALE	15	12	8	11	5
CAT	2	2	1	1	2
TEP	3 (9)	6 (9)	10 (12)	10	6
TOLL	14	12	9 (10)	9 (11)	9
SCR	20 (23)	18 (20)	16 (19)	19 (21)	22
SRPN	30	26	17 (18)	32 (44)	29
FREP	49 (51)	38	53 (54)	79 (83)	17
CTL	48 (50)	44	27	62	35
TOLLPATH	7	6	5	6	5
REL	4(5)	3 (7)	2	4	3
SOD	9	6 (7)	4	6	4
JAKSTAT	4	3	3 (4)	5	3
PGRP	13	10	7	10	13
APHAG	21(24)	20(22)	21	22	20 (21)
CASPA	4	4	2	3	5
IMDPATH	11	10(11)	7	8	8
PPO	16	14	9	9	3
PRDX	20(23)	23	23 (5)	20	21

CLIP	107(113)	82(84)	64 (66)	88 (90)	48
ML	26	27 (28)	18	21 (23)	10
Total	527 (554)	459 (476)	386 (400)	516 (536)	344 (345)

* The numbers not in parentheses are the number of functional genes. The numbers in parentheses are total genes that contains functional genes and candidate pseudogenes which are with premature termination codon or contain tiny (<=5bp) and non-triple length intron.

Table S8.2.

Aedes albopictus immune-related expansion genes. Abbreviation: Aaeg - *Aedes aegypti*, Agam - *Anopheles gambiae*, Cqui - *Culex quinquefasciatus*.

CLIP	PPO	SOD	TOLLPATH	TOLL	GALE	SRRP	BGBP	SPZ
Aaeg:CLIP46-D2	Aaeg:PPO8-D2	Aaeg:SOD7-D1	Aaeg:TOLLPATH4-D2	Aaeg:TOLL7-D1	Aaeg:GALE12-D1	Aaeg:SRRP17-D2	Aaeg:BGBP4-D3	Aaeg:SPZ3-D1
Aaeg:CLIP46-D1	Aaeg:PPO8-D1	Aaeg:SOD7-D4	Aaeg:TOLLPATH4-D1	Aaeg:TOLL7-D2	Aaeg:GALE12-D2	Aaeg:SRRP17-D1	Aaeg:BGBP4-D1	Aaeg:SPZ3-D2
Aaeg:CLIP40-D1	Aaeg:PPO7-D1	Aaeg:SOD5-D2		Aaeg:TOLL7-D3	Aaeg:GALE7-D1	Aaeg:SRRP6-D2	Aaeg:BGBP4-D4	Aaeg:SPZ4-D1
Aaeg:CLIP40-D2	Aaeg:PPO7-D2	Aaeg:SOD5-D1		Aaeg:TOLL11-D1	Aaeg:GALE7-D2	Aaeg:SRRP6-D1		Aaeg:SPZ4-D2
Aaeg:CLIP15-P1	Aaeg:PPO5-D1			Aaeg:TOLL11-D2	Aaeg:GALE4-D1	Aaeg:SRRP32-D1		Aaeg:SPZ8-D1
Aaeg:CLIP15-P2	Aaeg:PPO5-D2				Agam:GALE5-D2	Aaeg:SRRP32-D2		Aaeg:SPZ8-D2
Agam:CLIP50-D1					Aaeg:GALE5-D1	Aaeg:SRRP8-D1		Aaeg:SPZ2-D2
Agam:CLIP50-D2					Aaeg:GALE5-D2	Aaeg:SRRP8-D2		Aaeg:SPZ2-D1
Aaeg:CLIP10-D2					Aaeg:GALE2-D1	Cqui:SRRP11-D2		
Aaeg:CLIP10-D1					Aaeg:GALE2-D2	Cqui:SRRP11-D1		
Aaeg:CLIP19-D2						Cqui:SRRP11-D3		
Cqui:CLIP69-D1								
Aaeg:CLIP48-D1								
Aaeg:CLIP48-D2								
Aaeg:CLIP49-D6								
Aaeg:CLIP49-D9								
Aaeg:CLIP49-D5								
Cqui:CLIP62-P3								
Aaeg:CLIP49-D8								
Aaeg:CLIP29-D2								
Aaeg:CLIP29-D1								
Aaeg:CLIP62-D1								
Aaeg:CLIP62-D2								
Aaeg:CLIP62-D3								
Aaeg:CLIP28-D2								

Aaeg:CLIP28-D1								
Aaeg:CLIP25-D2								
Aaeg:CLIP25-D1								
Cqui:CLIP38-D1								
Cqui:CLIP38-D2								
Aaeg:CLIP35-D2								
Aaeg:CLIP35-D1								
Aaeg:CLIP44-D1								
Aaeg:CLIP44-D2								
Aaeg:CLIP66-P2								
Aaeg:CLIP66-D1								
Aaeg:CLIP65-D1								
Aaeg:CLIP65-D2								
Aaeg:CLIP53-D1								
Aaeg:CLIP53-D2								
Aaeg:CLIP37-D1								
Aaeg:CLIP37-D2								
Aaeg:CLIP9-D1								
Aaeg:CLIP9-D2								
Aaeg:CLIP9-D3								
Aaeg:CLIP4-D1								
Aaeg:CLIP4-D2								
Cqui:CLIP34-D2								
Cqui:CLIP34-D1								
Aaeg:CLIP16-D1								
Aaeg:CLIP16-D2								
Cqui:CLIP73-D2								
Cqui:CLIP73-D1								

Additional Data Files Content

Additional data file S1 (separate file)

Figure S2.3. Non-LTR retrotransposon consensus sequences in *Aedes albopictus*.

Additional data file S2 (separate file)

Table S3.1. List of sequences used as BLAST queries and number of BLAST hits in *Aedes albopictus* and *Aedes aegypti*.

Additional data file S3 (separate file)

Table S3.2. Output of BLAST analyses of the *Aedes albopictus* genome annotation (Foshan strain).

Additional data file S4 (separate file)

Table S3.3. Output of BLAST analyses of the *Aedes aegypti* genome, AaegL3 assembly.

Additional data file S5 (separate file)

Table S3.4. Mapping coordinates of sequences spanning partial or complete flaviviral ORFs in the *Aedes albopictus* genome assembly of the Foshan strain.

Additional data file S6 (separate file)

Table S3.5. Argot2-based annotation of putative viral integrations.

Additional data file S7 (separate file)

Table S3.6. BLAST hits to viral sequences other than flavivirus-like sequences, such as the Negev virus and the Wuhan mosquito virus 8.

Additional data file S8 (separate file)

Table S3.7. piRNAs identified within NIRVs.

Additional data file S9 (separate file)

Table S4.1. Annotation of diapause related genes.

Additional data file S10 (separate file)

Table S4.2 211 gene models from the *Ae. albopictus* gene expansion family that are present in the *Ae. albopictus* diapause transcriptome..

Additional data file S11 (separate file)

Table S4.5. Putative unique *Ae. albopictus* miRNAs with support from short non-coding RNA reads of mature oocytes.

Additional data file S12 (separate file)

Table S5.1. Genome coordinates of the *Ae. albopictus* CYP gene family.

Additional data file S13 (separate file)

Table S5.2. Accession numbers of CYP sequences used for phylogenetic analysis.

Additional data file S14 (separate file)

Table S5.4. Genome coordinates of the *Ae. albopictus* CCE gene family.

Additional data file S15 (separate file)

Table S5.5. Accession numbers of CCE sequences used for phylogenetic analysis.

Additional data file S16 (separate file)

Table S5.7. Genome coordinates of the *Ae. albopictus* GST gene family.

Additional data file S17 (separate file)

Table S5.8. Accession numbers of GST sequences used for phylogenetic analysis.

Additional data file S18 (separate file)

Table S5.10. Genome coordinates of the *Ae. aegypti* ABC transporter gene family.

Additional data file S19 (separate file)

Table S5.11. Genome coordinates of the *Ae. albopictus* ABC transporter gene family.

Additional data file S20 (separate file)

Table S5.13. Nucleotide sequences of *Ae. aegypti* ABC transporter genes and of *Ae. albopictus* genes involved in detoxification (CYPs, CCEs, GSTs, ABCs).

Additional data file S21 (separate file)

Figure S7.1. GO enrichment of sex-biased genes for **(A)** Biological processes.

Additional data file S22 (separate file)

Figure S7.1. GO enrichment of sex-biased genes for **(B)** Cellular compartments.

Additional data file S23 (separate file)

Figure S7.1. GO enrichment of sex-biased genes for **(C)** Molecular functions.

Additional data file S24 (separate file)

Table S7.1. Transcript analysis of adult male vs adult female gene expression.

Additional data file S25 (separate file)

Table S7.2. Candidate male-specific gene expression.

Additional data file S26 (separate file)

Table S7.3. Candidate female-specific gene expression.

Additional data file S27 (separate file)

Table S7.5. Full list of Biological Process Gene Ontology categories enriched in sex-biased genes (terms from the Function Ontology with p-value $\leqslant$ 1).

Additional data file S28 (separate file)

Table S7.6. Full list of Cellular Compartment Gene Ontology categories enriched in sex-biased genes (terms from the Function Ontology with p-value $\leqslant$ 1).

Additional data file S29 (separate file)

Table S7.7. Full list of Molecular Function Gene Ontology categories enriched in sex-biased genes (terms from the Function Ontology with p-value $\leqslant$ 1).

Additional data file S30 (separate file)

Table S7.8. Sex-bias genes of stable expression during development.

Author contributions

XC, XJ, JG, MX, HP, YD, CZ, MB, WD, TVL, JV, PA, XH, GY, ZT, and XF designed research; XC, XJ, JG, MX, HP, YD, CZ, MB, WD, TVL, JV, PA, XH, GY , RM, CS, RW, ZT, and XF performed research; XC, XJ, JG, MX, HP, YD, CZ, MB, WD, TVL, JV, PA, XH, GY, ZT, XF and AAJ analysed data and wrote the manuscript.

References

1. Spits C *et al.* (2006) Whole-genome multiple displacement amplification from single cells. *Nat Protoc* 1(4), 1965-1970.
2. Li R *et al.* (2010) The sequence and de novo assembly of the giant panda genome. *Nature* 463(7279), 311-317.
3. Luo R *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1(1), 18.
4. Boetzer M, Henkel CV, Jansen HJ, Butler D & Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4), 578-579.
5. Li R *et al.* (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25(15), 1966-1967.
6. Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12(4), 656-664.
7. Tarailo-Graovac M & Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* Chapter 4, 4-10.
8. Jurka J *et al.* (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1-4), 462-467.
9. Xu Z & Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35(Web Server issue), W265-W268.
10. Zhulidov PA *et al.* (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res* 32(3), e37.
11. Mortazavi A, Williams BA, McCue K, Schaeffer L & Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7), 621-628.
12. Chen Y, Liu M, Yan G, Lu H & Yang P (2010) One-pipeline approach achieving glycoprotein identification and obtaining intact glycopeptide information by tandem mass spectrometry. *Mol Biosyst* 6(12), 2417-2422.
13. Gertz EM, Yu YK, Agarwala R, Schaffer AA & Altschul SF (2006) Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol* 4, 41.
14. Birney E & Durbin R (2000) Using GeneWise in the Drosophila annotation experiment. *Genome Res* 10(4), 547-548.
15. Stanke M & Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19 Suppl 2, i215-i225.
16. Salamov AA & Solovyev VV (2000) Ab initio gene finding in Drosophila genomic DNA. *Genome Res* 10(4), 516-522.
17. Trapnell C, Pachter L & Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9), 1105-1111.
18. Roberts A, Pimentel H, Trapnell C & Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27(17), 2325-2329.
19. Mulder N & Apweiler R (2007) InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol Biol* 396, 59-70.
20. Bairoch A & Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28(1), 45-48.
21. Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1), 27-30.
22. Ashburner M *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1), 25-29.
23. Li H *et al.* (2006) TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res*

34(Database issue), D572-D580.
24. De Bie T, Cristianini N, Demuth JP & Hahn MW (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22(10), 1269-1271.
25. Wheeler TJ & Kececioglu JD (2007) Multiple alignment by aligning alignments. *Bioinformatics* 23(13), i559-i568.
26. Zhang Z *et al.* (2006) KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* 4(4), 259-263.
27. Jiang X *et al.* (2014) Genome analysis of a major urban malaria vector mosquito, Anopheles stephensi. *Genome Biol* 15(9), 459.
28. Fu L, Niu B, Zhu Z, Wu S & Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23), 3150-3152.
29. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16(2), 111-120.
30. Gaunt MW & Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* 19(5), 748-761.
31. Petrov DA (2002) Mutational equilibrium model of genome size evolution. *Theor Popul Biol* 61(4), 531-544.
32. Sun C, Lopez AJ & Mueller RL (2012) Slow DNA loss in the gigantic genomes of salamanders. *Genome Biol Evol* 4(12), 1340-1348.
33. Price AL, Jones NC & Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21 Suppl 1, i351-i358.
34. Petrov D (1997) Slow but Steady: Reduction of Genome Size through Biased Mutation. *Plant Cell* 9(11), 1900-1901.
35. Crochu S *et al.* (2004) Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. *J Gen Virol* 85(Pt 7), 1971-1980.
36. Rizzo F *et al.* (2014) Molecular characterization of flaviviruses from field-collected mosquitoes in northwestern Italy, 2011-2012. *Parasit Vectors* 7, 395.
37. Vazquez A *et al.* (2012) Novel flaviviruses detected in different species of mosquitoes in Spain. *Vector Borne Zoonotic Dis* 12(3), 223-229.
38. Tromas N, Zwart MP, Forment J & Elena SF (2014) Shrinkage of genome size in a plant RNA virus upon transfer of an essential viral gene into the host genome. *Genome Biol Evol* 6(3), 538-550.
39. Ballinger MJ, Bruenn JA & Taylor DJ (2012) Phylogeny, integration and expression of sigma virus-like genes in Drosophila. *Mol Phylogenet Evol* 65(1), 251-258.
40. Falda M *et al.* (2012) Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC Bioinformatics* 13 Suppl 4, S14.
41. Tamura K *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10), 2731-2739.
42. Roiz D, Vazquez A, Seco MP, Tenorio A & Rizzoli A (2009) Detection of novel insect flavivirus sequences integrated in Aedes albopictus (Diptera: Culicidae) in Northern Italy. *Virol J* 6, 93.
43. Blitvich BJ & Firth AE (2015) Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection, exclusion potential and genome organization. *Viruses* 7.
44. Lizeth Alcaraz-Estrada S, Yocupicio-Monroy M & Maria Del Angel R (2010) Insights into dengue virus genome replication. *Future Virology* 5(5), 575-592.
45. Goic B *et al.* (2013) RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model Drosophila. *Nat Immunol* 14(4), 396-403.
46. Geuking MB *et al.* (2009) Recombinantion of retrotransposon and exogenous RNA virus results in nonretroviral

cDNA integration. *Science* 323, 393-396.
47. Barron MG, Fiston-Lavier AS, Petrov DA & Gonzalez J (2014) Population genomics of transposable elements in Drosophila. *Annu Rev Genet* 48, 561-581.
48. Belyi VA, Levine AJ & Skalka AM (2010) Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathog* 6(7), e1001030.
49. Forster M *et al.* (2015) Vy-PER: eliminating flase positive detection of virus integration events in next generation sequencing data. *Sci Rep* 5, 11534.
50. Samina I, Margalit J & Peleg J (1986) Isolation of viruses from mosquitoes of the Negev, Israel. *Trans R Soc Trop Med Hyg* 80(3), 471-472.
51. Li CX *et al.* (2015) Unprecedented genomic diversity of RNA viruses in artrhopods revelas the ancestry of negative-sense RNA viruses. *eLIFE* 4, e5378.
52. Katzourakis A & Gifford RJ (2010) Endogenous viral elements in animal genomes. *PLoS Genet* 6(11), e1001191.
53. Severson DW & Behura SK (2012) Mosquito genomics: progress and challenges. *Annu Rev Entomol* 57, 143-166.
54. Eisen L *et al.* (2014) Temporal correlations between mosquito-based dengue virus surveillance measures or indoor mosquito abundance and dengue case numbers in Merida City, Mexico. *J Med Entomol* 51(4), 885-890.
55. Tuksinvaracharn R *et al.* (2004) Prevalence of dengue virus in Aedes mosquitoes during dry season by semi-nested reverse transcriptase-polymerase chain reaction (semi-nested RT-PCR). *J Med Assoc Thai* 87 Suppl 2, S129-S133.
56. Garcia-Rejon J *et al.* (2008) Dengue virus-infected Aedes aegypti in the home environment. *Am J Trop Med Hyg* 79(6), 940-950.
57. Calzolari M *et al.* (2015) Insect-specific flaviviruses, a worldwide widespread group of viruses only detected in insects. *Infect Genet Evol*.
58. Zhdanov VM (1975) Integration of viral genomes. *Nature* 256(5517), 471-473.
59. Klenerman P, Hengartner H & Zinkernagel RM (1997) A non-retroviral RNA virus persists in DNA form. *Nature* 390(6657), 298-301.
60. Tanne E & Sela I (2005) Occurrence of a DNA sequence of a non-retro RNA virus in a host plant genome and its expression: evidence for recombination between viral and host RNAs. *Virology* 332(2), 614-622.
61. Taylor DJ & Bruenn J (2009) The evolution of novel fungal genes from non-retroviral RNA viruses. *BMC Biol* 7, 88.
62. Koonin EV (2010) Taming of the shrewd: novel eukaryotic genes from RNA viruses. *BMC Biol* 8, 2.
63. Liu H *et al.* (2010) Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *J Virol* 84(22), 11876-11887.
64. Taylor DJ, Leach RW & Bruenn J (2010) Filoviruses are ancient and integrated into mammalian genomes. *BMC Evol Biol* 10, 193.
65. Horie M & Tomonaga K (2011) Non-retroviral fossils in vertebrate genomes. *Viruses* 3(10), 1836-1848.
66. Chiba S *et al.* (2011) Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLoS Pathog* 7(7), e1002146.
67. Fort P *et al.* (2012) Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. *Mol Biol Evol* 29(1), 381-390.
68. Maori E, Tanne E & Sela I (2007) Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. *Virology* 362(2), 342-349.
69. Maori E *et al.* (2007) Isolation and characterization of Israeli acute paralysis virus, a dicistrovirus affecting honeybees in Israel: evidence for diversity due to intra- and inter-species recombination. *J Gen Virol* 88(Pt 12), 3428-3438.
70. Flegel TW (2007) Update on viral accommodation, a model for host-viral interaction in shrimp and other arthropods.

Dev Comp Immunol 31(3), 217-231.
71. Aswad A & Katzourakis A (2012) Paleovirology and virally derived immunity. *Trends Ecol Evol* 27(11), 627-636.
72. Quenouille J, Vassilakos N & Moury B (2013) Potato virus Y: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity. *Mol Plant Pathol* 14(5), 439-452.
73. Cook S & Holmes EC (2006) A multigene analysis of the phylogenetic relationships among the flaviviruses (Family: Flaviviridae) and the evolution of vector transmission. *Arch Virol* 151(2), 309-325.
74. Lee E *et al.* (2013) Web Apollo: a web-based genomic annotation editing platform. *Genome Biol* 14(8), R93.
75. Cantarel BL *et al.* (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18(1), 188-196.
76. Poelchau MF, Reynolds JA, Elsik CG, Denlinger DL & Armbruster PA (2013) RNA-Seq reveals early distinctions and late convergence of gene expression between diapause and quiescence in the Asian tiger mosquito, Aedes albopictus. *J Exp Biol* 216(Pt 21), 4082-4090.
77. Poelchau MF, Reynolds JA, Elsik CG, Denlinger DL & Armbruster PA (2013) Deep sequencing reveals complex mechanisms of diapause preparation in the invasive mosquito, Aedes albopictus. *Proc Biol Sci* 280(1759), 20130143.
78. Poelchau MF, Reynolds JA, Denlinger DL, Elsik CG & Armbruster PA (2013) Transcriptome sequencing as a platform to elucidate molecular components of the diapause response in the Asian tiger mosquito. *Physiol Entomol* 38(2), 173-181.
79. Huang X, Poelchau MF & Armbruster PA (2015) Global Transcriptional Dynamics of Diapause Induction in Non-Blood-Fed and Blood-Fed Aedes albopictus. *PLoS Negl Trop Dis* 9(4), e3724.
80. Slater GS & Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 31.
81. Reynolds JA, Poelchau MF, Rahman Z, Armbruster PA & Denlinger DL (2012) Transcript profiling reveals mechanisms for lipid conservation during diapause in the mosquito, Aedes albopictus. *J Insect Physiol* 58(7), 966-973.
82. Urbanski JM, Benoit JB, Michaud MR, Denlinger DL & Armbruster P (2010) The molecular physiology of increased egg desiccation resistance during diapause in the invasive mosquito, Aedes albopictus. *Proc Biol Sci* 277(1694), 2683-2692.
83. Colbourne JK *et al.* (2011) The ecoresponsive genome of Daphnia pulex. *Science* 331(6017), 555-561.
84. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4), 357-359.
85. Gu J *et al.* (2013) miRNA genes of an invasive vector mosquito, Aedes albopictus. *PLoS One* 8(7), e67638.
86. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3), 403-410.
87. Megy K *et al.* (2012) VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Res* 40(Database issue), D729-D734.
88. Ranson H *et al.* (2002) Evolution of supergene families associated with insecticide resistance. *Science* 298(5591), 179-181.
89. Strode C *et al.* (2008) Genomic analysis of detoxification genes in the mosquito Aedes aegypti. *Insect Biochem Mol Biol* 38(1), 113-123.
90. St PS, Ponting L, Stefancsik R & McQuilton P (2014) FlyBase 102--advanced approaches to interrogating FlyBase. *Nucleic Acids Res* 42(Database issue), D780-D788.
91. Feyereisen R (2012) Insect CYP Genes and P450 Enzymes. Academic Press, pp 236-316. Gilbert, L.I. (Ed.), Insect Molecular Biology and Biochemistry. San Diego.
92. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids*

Res 32(5), 1792-1797.

93. Tamura K, Stecher G, Peterson D, Filipski A & Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30(12), 2725-2729.
94. Claudianos C *et al.* (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Mol Biol* 15(5), 615-636.
95. Abascal F, Zardoya R & Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21(9), 2104-2105.
96. Jobb G, von Haeseler A & Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4, 18.
97. Dermauw W & Van Leeuwen T (2014) The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. *Insect Biochem Mol Biol* 45, 89-110.
98. Feyereisen R (2011) Arthropod CYPomes illustrate the tempo and mode in P450 evolution. Proteins and Proteomics 1814, pp 19-28. Biochimica et Biophysica Acta (BBA).
99. Chung H *et al.* (2009) Characterization of Drosophila melanogaster cytochrome P450 genes. *Proc Natl Acad Sci U S A* 106(14), 5731-5736.
100. Qiu Y *et al.* (2012) An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc Natl Acad Sci U S A* 109(37), 14858-14863.
101. Yang P, Tanaka H, Kuwano E & Suzuki K (2008) A novel cytochrome P450 gene (CYP4G25) of the silkmoth Antheraea yamamai: cloning and expression pattern in pharate first instar larvae in relation to diapause. *J Insect Physiol* 54(3), 636-643.
102. Stevenson BJ, Pignatelli P, Nikou D & Paine MJ (2012) Pinpointing P450s associated with pyrethroid metabolism in the dengue vector, Aedes aegypti: developing new tools to combat insecticide resistance. *PLoS Negl Trop Dis* 6(3), e1595.
103. Rewitz KF, Rybczynski R, Warren JT & Gilbert LI (2006) The Halloween genes code for cytochrome P450 enzymes mediating synthesis of the insect moulting hormone. *Biochem Soc Trans* 34(Pt 6), 1256-1260.
104. Daimon T & Shinoda T (2013) Function, diversity, and application of insect juvenile hormone epoxidases (CYP15). *Biotechnol Appl Biochem* 60(1), 82-91.
105. Oakeshott JG, Claudianos C, Campbell PM, Newcomb RD & Russell RJ (2005) Biochemical Genetics and Genomics of Insect Esterases. Elsevier, pp 309-381. Gilbert, L.I. (Ed.), Comprehensive Molecular Insect Science. Amsterdam.
106. Despres L, David JP & Gallet C (2007) The evolutionary ecology of insect resistance to plant chemicals. *Trends Ecol Evol* 22(6), 298-307.
107. Li X, Schuler MA & Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annu Rev Entomol* 52, 231-253.
108. Huchard E *et al.* (2006) Acetylcholinesterase genes within the Diptera: takeover and loss in true flies. *Proc Biol Sci* 273(1601), 2595-2604.
109. Poupardin R, Srisukontarat W, Yunta C & Ranson H (2014) Identification of carboxylesterase genes implicated in temephos resistance in the dengue vector Aedes aegypti. *PLoS Negl Trop Dis* 8(3), e2743.
110. Campbell PM *et al.* (2001) Identification of a juvenile hormone esterase gene by matching its peptide mass fingerprint with a sequence from the Drosophila genome project. *Insect Biochem Mol Biol* 31(6-7), 513-520.
111. Shirras AD & Bownes M (1989) cricklet: A locus regulating a number of adult functions of Drosophila melanogaster. *Proc Natl Acad Sci U S A* 86(12), 4559-4563.
112. Mensch J *et al.* (2010) Stage-specific effects of candidate heterochronic genes on variation in developmental time along an altitudinal cline of Drosophila melanogaster. *PLoS One* 5(6), e11229.

113. Olson PF *et al.* (1990) Glutactin, a novel Drosophila basement membrane-related glycoprotein with sequence similarity to serine esterases. *Embo J* 9(4), 1219-1227.
114. Fakhouri M *et al.* (2006) Minor proteins and enzymes of the Drosophila eggshell matrix. *Dev Biol* 293(1), 127-141.
115. Marinotti O *et al.* (2013) The genome of Anopheles darlingi, the main neotropical malaria vector. *Nucleic Acids Res* 41(15), 7387-7400.
116. Meyer F, Flotenmeyer M & Moussian B (2013) The sulfonylurea receptor Sur is dispensable for chitin synthesis in Drosophila melanogaster embryos. *Pest Manag Sci* 69(10), 1136-1140.
117. Schinkel AH & Jonker JW (2003) Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. *Adv Drug Deliv Rev* 55(1), 3-29.
118. Kerr ID, Haider AJ & Gelissen IC (2011) The ABCG family of membrane-associated transporters: you don't have to be big to be mighty. *Br J Pharmacol* 164(7), 1767-1779.
119. Deng Y *et al.* (2013) Molecular and functional characterization of odorant-binding protein genes in an invasive vector mosquito, Aedes albopictus. *PLoS One* 8(7), e68836.
120. Petersen TN, Brunak S, von Heijne G & Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10), 785-786.
121. Gasteiger E *et al.* (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31(13), 3784-3788.
122. Marchler-Bauer A *et al.* (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 37(Database issue), D205-D210.
123. Larkin MA *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21), 2947-2948.
124. Chen S *et al.* (2004) Genetic, biochemical, and structural characterization of a new densovirus isolated from a chronically infected Aedes albopictus C6/36 cell line. *Virology* 318(1), 123-133.
125. Pelletier J & Leal WS (2009) Genome analysis and expression patterns of odorant-binding proteins from the Southern House mosquito Culex pipiens quinquefasciatus. *PLoS One* 4(7), e6237.
126. Zhou JJ, He XL, Pickett JA & Field LM (2008) Identification of odorant-binding proteins of the yellow fever mosquito Aedes aegypti: genome annotation and comparative analyses. *Insect Mol Biol* 17(2), 147-163.
127. Surget-Groba Y & Montoya-Burgos JI (2010) Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res* 20(10), 1432-1440.
128. Hao DC, Ge G, Xiao P, Zhang Y & Yang L (2011) The first insight into the tissue specific taxus transcriptome via Illumina second generation sequencing. *PLoS One* 6(6), e21220.
129. Benjamini Y & Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29, 1165-1188.
130. Huang DW *et al.* (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 8(9), R183.
131. Hall AB *et al.* (2015) A male-determining factor in the mosquito Aedes aegypti. *Science*.
132. Erickson JW & Quintero JJ (2007) Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila. *PLoS Biol* 5(12), e332.
133. Hoshijima K, Inoue K, Higuchi I, Sakamoto H & Shimura Y (1991) Control of doublesex alternative splicing by transformer and transformer-2 in Drosophila. *Science* 252(5007), 833-836.
134. Waterhouse RM *et al.* (2007) Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* 316(5832), 1738-1743.
135. de Hoon MJ, Imoto S, Nolan J & Miyano S (2004) Open source clustering software. *Bioinformatics* 20(9), 1453-1454.
136. Nei M & Kumar S (2000) *Molecular Evolution and Phylogenetics* (Oxford University Press, New York).

137. Jones DT, Taylor WR & Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8(3), 275-282.
138. Whelan S & Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18(5), 691-699.