Science. Author manuscript; available in PMC 2010 May 12.

Published in final edited form as:
Science. 2007 June 22; 316(5832): 1718–1723. doi:10.1126/science.1138878.

Genome sequence of *Aedes aegypti*, a major arbovirus vector

Sp Sinkins

Abstract

We present a draft sequence of the genome of *Aedes aegypti*, the primary vector for yellow fever and dengue fever, which at ~1.38 Gbp is ~5-fold larger in size than the genome of the malaria vector, *Anopheles gambiae*. Nearly 50% of the *Aedes aegypti* genome consists of transposable elements. These contribute to a ~4-6 fold increase in average gene length and the size of intergenic regions relative to *Anopheles gambiae* and *Drosophila melanogaster*. Nevertheless, chromosomal synteny is generally maintained between all three insects although conservation of orthologous gene order is higher (~2-fold) between the mosquito species than between either of them and fruit fly. Three methods have provided transcriptional evidence for 80% of the 15,419 predicted protein coding genes in *Aedes aegypti*. An increase in genes encoding odorant binding, cytochrome P450 and cuticle domains relative to *Anopheles gambiae* suggests that members of these protein families underpin some of the biological differences between them.

Mosquitoes are vectors of many important human diseases, with transmission of arboviruses largely associated with the subfamily Culicinae, lymphatic filarial worms with both the Culicinae and the subfamily Anophelinae, and transmission of malaria causing parasites with the Anophelinae (1). *Aedes aegypti* is the best characterized species within the Culicinae (2), primarily due to its easy transition from the field to laboratory culture, and has provided much of the existing information on mosquito biology, physiology, genetics and vector competence (3, 4). It maintains close association with human populations and it is the principal vector of the etiological agents of yellow fever and dengue fever (5, 6), as well as for the recent Chikungunya fever epidemics in countries in the Indian Ocean area (7). Despite an effective vaccine, yellow fever remains a disease burden in Africa and parts of South America with ~200,000 cases per year resulting in ~30,000 deaths (5). About 2.5 billion people are at risk for dengue, with ~50 million cases per year and ~500,000 cases of dengue hemorrhagic fever, the more serious manifestation of disease. The incidence of dengue, for which mosquito management is currently the only prevention option, is on the increase (8). Thus, there is an urgent need to improve the control of these diseases and their vector.

Availability of a draft sequence of the ~ 278 Mbp genome of *Anopheles gambiae* (9) has accelerated research to develop new mosquito and malaria control strategies. Recent studies have led to the identification of mosquito genes that regulate malaria parasite infection in the mosquito (10) and those involved in the ability to find and feed on blood of human hosts (11, 12). Genome features such as chromosomal inversions or specific ‘speciation islands’ that are involved in population differentiation (13, 14) can now be studied and the entire repertoire of genes encoding metabolic detoxification mechanisms that may underlie insecticide resistance can be rapidly screened (15). To provide a basis for similar platforms for research in *Ae. aegypti* and to harness the power of comparative genome analyses we have undertaken a project to sequence the genome of this mosquito species.

(steven.sinkins@zoo.ox.ac.uk)

Europe PMC Funders Author Manuscripts

Comparisons between *An. gambiae* and *Drosophila melanogaster* (16) revealed significant genomic differences between the two insects that reflect their divergence ~250 million years ago (MYA) (17), yet both genomes retain significant remnants of homology among chromosome arms. *Anopheles* mosquitoes radiated from the *Aedes* and *Culex* lineages ~150 million years ago (18), and *Ae. aegypti* and *An. gambiae* share similar characteristics such as anthropophily, but they exhibit variation in morphology and physiology, mating behavior, oviposition preferences, dispersal and biting cycle (1). Both mosquito species have 3 pairs of chromosomes, but *Ae. aegypti* lacks heteromorphic sex chromosomes (19) and by C_0t reassociation kinetics its genome size was estimated to be ~813 Mbp (20). Here we report on a draft sequence of the genome of *Ae. aegypti*, but it is ~70% bigger than expected.

Assembly of a draft genome sequence of *Aedes aegypti*

Whole-genome shotgun sequencing was performed on DNA purified from newly-hatched larvae of an inbred sub-strain (LVP^{ib12}) of the Liverpool strain of *Ae. aegypti* which is tolerant to inbreeding while maintaining relevant phenotypes (21). Approximately 98% of the sequence, assembled using Arachne (22), is contained within 1,257 scaffolds with an N50 scaffold size of ~1.5 Mbp. Assembly statistics of the ~1.38 gigabase-pair (Gbp) genome are provided as supplementary information (Table S1). The discrepancy between the genome size based on C_0t data and that determined by sequencing is not clear. Assessment of assembly parameters, sequence coverage and other metrics do not indicate a gross inflation in assembled genome size due to potential haplotype effects (21). Assembled sequences that are potentially "under-collapsed" are estimated to be <5% of the total genome size (Figure S1). Data related to the genome project have been deposited in GenBank™ (project accession AAGE00000000).

Genetic and physical mapping data allowed assignment, but without order or orientation, of 63, 48, 39, 43 and 45 scaffolds to *Ae. aegypti* chromosome 1 and chromosome arms 2p, 2q, 3p and 3q, respectively (21). These scaffolds total ~430 Mbp in length and represent ~31% of the genome (Table S2). Thus, development of high-resolution physical mapping techniques and generation of additional random or targeted sequence data represent priorities for improving the quality of the current genome assembly and size estimate, and to permit unambiguous differentiation between regions of segmental duplications and residual haplotype polymorphism.

The genome of *Aedes aegypti* is riddled with transposable elements

Transposable elements (TEs) have contributed significantly to the ~5-fold size difference of the *Ae. aegypti* and *An. gambiae* genomes. Approximately 47% of the *Ae. aegypti* genome consists of TEs (Figure 1 and Table S3, see Table S3 legend for definition of TE family, element and copy). *Aedes aegypti* harbors all known types of TEs that have been reported in *An. gambiae* with the exception of two DNA transposons, *merlin* (23) and *gambol* (24). Simple and tandem repeats occupy ~6% of the genome and an additional ~15% consists of repetitive sequences that remain to be classified.

Most eukaryotic TE families characterized to date (25) are present in *Ae. aegypti* and more than 1,000 TEs have been annotated, representing a diverse collection of TEs in a single genome (Table S3). Although the majority of protein coding TEs appear to be degenerate, more than 200 elements have at least one copy with an intact ORF and other features suggesting recent transposition. Approximately 3% of the genome is composed of ~13,000 copies of the element *Juan-A* in the Jockey family of non-long terminal repeat (LTR) retrotransposons. A tRNA-related SINE element, *Feilai-B,* has the highest copy number, with approximately 50,000 copies per haploid genome. Only one highly degenerate *mariner* element is found in *Ae. aegypti* while at least 20 *mariner* elements, many with intact ORFs,

were found in *An. gambiae*. TEs present in *Ae. aegypti* but missing from *An. gambiae* include the *LOA* family of non-LTR retrotransposons, the *Osvaldo* element of the *Ty3/gypsy* LTR retrotransposons (26) and a unique family, *Penelope* (27). Comparison of *Ae. aegypti* and *An. gambiae* TE sequences is consistent with the interpretation of an overall lack of apparent horizontal transfer events as a single candidate for such events was identified (21); one full-length copy of the *ITmD37E* DNA transposon in *Ae. aegypti* is 93% identical at the nucleotide level to a similarly classified TE in *An. gambiae*.

MITEs (miniature inverted repeat transposable elements) and MITE-like elements of non-protein coding TEs in *Ae. aegypti* have terminal inverted repeat sequences and target site duplications, features characteristic of transposition of DNA transposons. Such TEs can be mobilized to transpose in *trans*, by transposases encoded by DNA transposons (28). The latter TEs occupy only 3% of the *Ae. aegypti* genome and they are less numerous than non-protein coding DNA elements which occupy 16% of the genome (Table S3). Thus, DNA transposons may have made a significant contribution to the expansion in size and organization of the *Ae. aegypti* genome through cross-mobilization of MITEs and MITE-like TEs.

Annotation of the draft genome sequence

The fragmented nature of the assembled genome sequence, an asymmetric distribution of intron lengths within genes (Figure S2, S3), and the frequent occurrence of TE-associated ORFs close to genes and within introns complicated the process of automated gene modeling and often led to prediction of split or chimeric gene models. Thus we developed an extensive multi-stage genome masking strategy prior to gene finding (resulting in masking ~70% of the genome sequence) and optimized gene finding programs via iterative manual inspection of predicted gene models relative to a training set (21).

Two independent automated pipelines for structural annotation resulted in the prediction of 17,776 and 27,284 gene models (21). Extensive use was made of a large collection of ~265,000 *Ae. aegypti* ESTs and dipteran protein and cDNA/EST sequences in producing and then merging the two datasets into a single high confidence gene set which consists of 15,419 gene models (AaegL1.1). Alternative splice forms derived from these genes are predicted to generate at least 16,789 transcripts. Table 1 lists some of the genome and protein coding characteristics of *Ae. aegypti* and those of *D. melanogaster* and *An. gambiae*.

Gene descriptions and molecular function Gene Ontology (GO) codes were assigned computationally to predicted protein sequences using BLASTP comparison searches with protein databases (21). The functional annotation pipeline included analyses of protein domains, and secretion signal sequence and trans-membrane motifs. A total of 8,332 proteins were assigned a description, 9,335 proteins were assigned GO terms, 2,796 proteins were assigned as “hypothetical proteins” and 5,027 as “conserved hypothetical proteins”.

Genes encoding proteins < 50 amino acid residues in length were not included in this annotation release unless they encoded known small proteins. However, these and other genes are captured in a set of lower-confidence gene models that is available for analyses as a supplementary release (21). On the basis of transcriptional mapping data and limited manual examination we anticipate that ~5-10% of the second-tier models or modified versions of them represent “real” genes.

TEs contribute to complex protein coding gene structures in *Aedes aegypti*

A striking feature of protein coding genes in *Ae. aegypti* is the 4 to 6 fold increase in the average length of a gene relative to *An. gambiae* and *D. melanogaster*, which is due to

longer intron lengths rather than longer exons or an increased number of introns (Table 1). The increased length of introns is primarily due to infiltration by TEs; a plot of intron size before and after masking repeat sequences reveals a shift to shorter intron lengths (Figure S2). A more global perspective of the genome expansion was revealed by the difference in genomic span (~4.6-fold) of conserved gene arrangements between *Ae. aegypti* and *An. gambiae* that occupy ~33% of each genome (Table S4, Figure S4), providing evidence that TE-mediated expansion in both genic and intergenic regions have contributed to the increased size of the *Ae. aegypti* genome. Long introns, in particular those in 5′ and 3′ untranslated regions are likely to complicate *in silico* driven studies to define *cis*-acting transcription and translational regulatory elements as they may be distant from coding sequences (Figure S3).

Transcriptional analyses

Data derived from three different transcript profiling platforms, namely ESTs, massively-parallel signature sequencing (MPSS) and 60-mer oligonucleotide-based microarrays were used to experimentally confirm predicted protein coding gene models and to gain insight into differential transcription profiles (21). In total, the platforms identified transcripts from 12,350 (80%) of 15,419 genes. Mapping of ~265,000 ESTs and cDNA sequences and MPSS signature sequence tags to the genome sequence as well as gene models provided evidence for transcription of 9,270 and 3,984 genes, respectively, while microarray data identified transcripts from 9,143 genes (Table S5). The lower number of genes identified by MPSS may in part be explained by the observation that only ~2/3 of the genes can be assayed by MPSS as this approach required the presence of a *Dpn*II restriction enzyme site within the transcribed region (21). The platforms identified a common set of 2,558 genes and each platform identified a unique set of genes (Figure S5), highlighting the importance of utilizing a multi-platform approach. The data provides empirical support for ~76% of genes annotated as hypothetical (Table S6), underscoring the validity of *ab initio* gene finding programs in identifying novel genes.

The MPSS platform also provided quantitative data on transcript abundance for 7,421 unique sequence tags which mapped to the genome and gene models. The most abundant MPSS tag (129,296 tags per million [tpm]) present in a pool of total RNA made from all mosquito developmental stages, adults of both sexes and 2 day post-blood fed females (21) was derived from the gene encoding the mitochondrial large subunit rRNA (Table S7). About 40% of all the MPSS sequences corresponded to just 14 tags and only 146 tags were detected at a level greater than 1,000 tpm indicating a huge spread in transcript abundance. In addition to house-keeping genes, MPSS tags present at >10,000 tpm mapped to one gene encoding trypsin and two genes encoding rhodopsin (AAEL005625 and AAEL006259). The trypsin gene (AAEL007818) is a member of large multi-gene family and was identified previously as encoding an endoprotease expressed after adult emergence in the midgut of the female mosquito, post-transcriptionally regulated by blood feeding (29). Efficient digestion of a blood meal and oogenesis are closely linked and both processes are suppressed by trypsin modulation oostatic factor (TMOF), a decapeptide hormone, leading to its potential use in mosquito control programs (30). The rhodopsin genes identified by MPSS code for 2 members (AaOp2 and AaOp5) of a family of 6 putative visual receptors sensitive to long-wavelength light. This gene family is expanded 2-fold in size in *Ae. aegypti* and *An. gambiae* relative to *D. melanogaster* (Figure S6, Tables S8), and it is tempting to speculate that the highly transcribed genes play a prominent role in photoreception, perhaps linked to mosquito activity at dusk and dawn when long-wavelength light predominates.

Differences in transcript abundance between a pool of RNA from non-adult developmental stages and 4 day-old, non-fed adult females were revealed by the microarray analyses which

identified 398 and 208 pre-adult stage and adult female enriched transcripts, respectively (Table S9). Functional categorization of these transcripts differed mainly with regard to cytoskeletal, structural and chemosensory functions (Figure 2). Differential transcription of genes with putative implication in chemosensory processes between these stages was striking with 17 transcripts highly-enriched in mosquito developmental stages and only 3 enriched in adult females. A larger number of immune gene transcripts were also enriched in pre-adult stages, and may reflect a broader microbial exposure of larvae and pupae in their aqueous environments. A large number of highly expressed genes encoding cuticle proteins in adults may be indicative of their function in a variety of processes, including immunity, and suggests continuous growth of the cuticle.

Aedes aegypti gene families and domain composition

Consistent with evolutionary distance estimates (18), there is a higher degree of similarity between the *Ae. aegypti* and *An. gambiae* proteomes than between the mosquito and *D. melanogaster* proteomes. Orthologous proteins were computed among the three genomes, with 67% of the *Ae. aegypti* proteins having an ortholog in *An. gambiae* and 58% having an ortholog in *D. melanogaster* (Figure 3A). Analysis of three-way, single-copy orthologs revealed average amino acid identity of 74% between the mosquito proteins, in contrast with ~58% identity between mosquito and fruit fly proteins (Figure S7). Approximately 2,000 orthologs are shared only between the mosquitoes and may represent functions central to mosquito biology. Although most of these proteins are of unknown function, ~250 can be assigned a predicted function, of which 28% are involved in gustatory/olfactory systems, 12% are members of the cuticular gene family and 8% are members of the cytochrome P450 family (21).

Mapping of protein domains using Interpro revealed an expansion of Zn-fingers, insect cuticle, cytochrome P450, odorant binding protein (OBP) A10/OS-D, insect allergen related and HMG-I and MHG-Y domains in *Ae. aegypti* relative to *An. gambiae*, *D. melanogaster* and the honey bee, *Apis mellifera* (Table S10). These constitute large *Ae. aegypti* gene families as revealed by two independent clustering methods (Table S11). Genes containing Zinc-finger like domains could be of transposon or retroviral origin, and remain to be assessed (21).

Species-specific differences in the number of members within a multi-gene family often provide clues on biological adaptation to environmental challenges. In this context, cuticle-domain containing proteins have been described to play diverse roles in exoskeleton formation and wound healing and are expressed in hemocytes, a major cell type that mediates innate immunity (31). Cuticular proteins also are implicated in arbovirus transmission (32). Expansion of olfactory receptors and OBPs in *Ae. aegypti* may contribute to an elaborate olfactory system, which in turn may be linked to the expansion in detoxification capacity. The latter and “insect allergen-related” genes, suggested to have a digestive function, may contribute to the relative robustness of *Ae. aegypti*, and also could manifest in a higher insecticide resistance. In this context, the genome and EST data have lead to the development a specific microarray to identify candidate genes among members of multi-gene families (cytochrome P450, glutathione-S-transferase and carboxylesterase) associated with metabolic resistance to insecticides (33). This platform will provide a means to rapidly survey mechanisms of insecticide resistance in mosquito populations and represents an important tool in managing insecticide deployment and development programs.

G protein-coupled receptors (GPCRs) that are expected to function in signal transduction cascades in *Ae. aegypti* have been manually identified (21). This superfamily of proteins

includes 111 non-sensory class A, B and C GPCRs, 14 atypical class D GPCRs, 10 opsin photoreceptors (Table S8 and S12). *Aedes aegypti* possesses orthologs for >85% of the *An. gambiae* and *D. melanogaster* non-sensory GPCRs suggesting significant conservation of GPCR-mediated neurological processes across Diptera. Many *Ae. aegypti* GPCRs have sequence similarity to known drug targets (34), and may reveal new opportunities for the development of novel insecticides.

Metabolic potential and membrane transporters

Aedes aegypti and *An. gambiae* are predicted to contain similar metabolic profiles as judged by assigning an Enzyme Commission (EC) number to both mosquito proteomes (Table S13), although some biochemical pathways are incomplete (21). Given the early stages of annotation, it is premature to draw conclusions from missing enzymes in predicted metabolic pathways as preliminary computes on the supplemental *Ae. aegypti* gene set resulted in the assignment of 12 EC numbers not present in Release 1.0 (Table S14).

An automated pipeline (35) was used to predict potential membrane transporters for *Ae. aegypti* and *An. gambiae,* and their transport capacity resembles that of *D. melanogaster* (Table S15). Similar to other multi-cellular eukaryotes, ~ 32% of all three insect transporters code for ion channels and probably function to maintain haemolymph homeostasis under different environmental conditions by modulating the concentrations of Na^+, K^+ and Cl^- ions. *Aedes aegypti* encodes more paralogs of voltage-gated potassium ion channels, epithelial sodium channels and ligand-gated ion channels (LIC) such as the glutamate-gated ion channel than *An. gambiae* and *D. melanogaster*. These channels play important roles in the signal transduction pathway and cell communication in the central nervous system and at neuromuscular junctions. The binding of neurotransmitters (glutamate, γ-aminobutyric acid, glycine, histamine and acetylcholine) to the superfamily of LIC is responsible for the movement of cations and anions across the plasma membrane to mediate excitatory or inhibitory synaptic transmission. A collection of 64 putative ATP-binding cassette transporters was identified, including subgroups that encode multi-drug efflux proteins. *Aedes aegypti* encodes more members of 4 different types of amino acid transporters than *An. gambiae* and *D. melanogaster.* Mosquito larvae cannot synthesize *de novo* all the basic, neutral or aromatic L-amino acids (3), and must rely on uptake of these essential amino acids. Aromatic amino acids phenylalanine and tryptophan are particularly important because they are precursors for the synthesis of neurotransmitters. The richer repertoire of membrane transport systems in *Ae. aegypti* is likely to intersect with the apparent increase in odorant reception and detoxification capacity.

Autosomal sex determination and sex specific gene expression

Heteromorphic sex chromosomes are absent in *Ae. aegypti* and other culicine mosquitoes (19). Instead sex is controlled by an autosomal locus wherein the male-determining allele, *M*, is dominant. The primary switch mechanism at the top of the mosquito sex determination cascade is different than that of *D. melanogaster*, where the X-chromosome to autosome ratio controls sex differentiation. However, we expect conservation of function in mosquito orthologs of *Drosophila* genes that are further downstream of the cascade (36). We verified the presence of a number of these in *Ae. aegypti*, including orthologs for *doublesex*, *transformer-2*, *fruitless*, *dissatisfaction* and *intersex* (Table S16).

To define gene expression differences between the sexes, we have analyzed microarray transcription profiles of 4 day, non-fed adult female and male mosquitoes (Figure 2); 669 and 635 transcripts were enriched in females and males, respectively, and 6,713 transcripts were expressed at similar levels in both sexes (Table S17). An additional 373 and 534 transcripts generated exclusive hybridization signals (with signal intensity below cutoff

threshold level in one channel) in females and males, respectively, and may therefore represent sex specific transcripts. Functional categorization of female and male enriched transcripts were remarkably similar, with male mosquitoes expressing a slightly larger number of immune and redox/stress related transcripts while females expressed a larger number of putative blood digestive enzyme transcripts. This particular pattern of immune gene expression is surprising considering a predicted lower need for immune defense in the males, due to the lack of pathogen exposure that results from blood feeding. By comparing these data with previously described *An. gambiae* sex specific microarray analyses (37) we identified 144 orthologous genes displaying the same sex specific transcription pattern in *An. gambiae*, while 74 orthologs showed an opposite profile, suggesting differences in certain sex specific functions between the two mosquito species.

Conserved synteny with *Anopheles gambiae* and *Drosophila melanogaster*

The assignment of 238 *Ae. aegypti* scaffolds containing ~ 5,000 genes, about 1/3 of the predicted gene set, to a chromosomal location based on genetic and physical mapping data (21), allowed us to compare ortholog position and identify conserved evolutionary associations between *Ae. aegypti* and *An. gambiae* or *D. melanogaster* chromosomes (Table S2, S19). Most of the *Ae. aegypti* chromosome arms, with the exception of 2p and 3q, exhibited a distinct one-to-one correlation with *An. gambiae* and *D. melanogaster* chromosome arms based on the proportion of orthologous genes conserved between chromosome arm pairs (Figure 3B). These findings confirm and extend previous results that compared a small number (~75) of *Ae. aegypti* genes with orthologs in *An. gambiae* and *D. melanogaster* (38).

Maps of conserved local gene arrangements (microsynteny) were computed by identifying blocks of at least 2 neighboring single-copy orthologs in each pair of genomes and allowing not more than two intervening genes (21). In line with the species divergence times, twice as many orthologs are similarly arranged between these mosquito species than between either of them and fruit fly (Table S18) (39). 1,345 microsyntenic blocks were identified between *Ae. aegypti* and *An. .gambiae*, containing 5,265 out of total 6,790 single-copy orthologs (Table S4, S18). When *D. melanogaster* is used as an out group to count synteny breaks that have occurred in each mosquito lineage since their radiation, the data indicates a ~2.5-fold higher rate of genome shuffling in the *Ae. aegypti* lineage than in the *An. gambiae* lineage (21). However, this estimate may be inflated due to the fragmented nature of the current *Ae. aegypti* genome assembly in which half of all single-copy orthologs are found in scaffolds with less than 5 such orthologs. Thus, while the highly repetitive nature of the *Ae. aegypti* genome appears to have facilitated local gene rearrangements it does not appear to have had a gross influence on chromosomal synteny.

Concluding remarks

The draft genome sequence of *Ae. aegypti* represents a significant technical achievement which will stimulate efforts to elucidate interactions at the molecular level between mosquitoes and the pathogens they transmit. This already is evidenced, for example, in dissecting components of the Toll immune signaling pathway (40) and identification of genes encoding insulin-like hormone peptides (41). In addition, a phylogenomic study described in an accompanying paper provides new insight in the complex pattern of evolution of genes coding for different components of the mosquito innate immune system (42).

We expect the sequence data will facilitate the identification of *Ae. aegypti* genes encoding recently described mid-gut receptors for dengue virus (43). Dengue vector competence is a quantitative trait and multiple loci determine virus mid-gut infection and escape barriers

(44). Unfortunately, the fragmented nature of the genome sequence and low gene density has precluded its use in the identification of a comprehensive list of candidate genes for vector competence phenotypes or sex determination. The sequence may be used to improve the resolution of the current genetic map (45) and for integrating transcriptional profiling data with genetic studies (46), but filling gaps in the assembled sequence remains a high priority, especially when exploring genetic variations between the sequenced strain and field populations of *Ae. aegypti*.

Our data highlight potentially important differences in gene expression profiles between different life-cycle stages of *Ae. aegypti* and gene content between two mosquito species with the most impact on human health. The on-going genome project on *Culux pipiens quinquefasciatus*, a vector for West Nile virus, will provide additional resources to underpin studies to systematically study common and mosquito species specific gene function. Such analyses should improve our understanding of mosquito biology and the complex role of mosquitoes in the transmission of pathogens, and result in the development of new approaches for vector-targeted control of disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The *Aedes aegypti* genome sequencing project at the microbial sequencing centers and VectorBase was funded by NIH/NIAID contracts (HHSN266200309D266030071, HHSN266200400001C and HHSN266200400039C), and supported in part by NIH/NIAID grants UO1 AI50936 (D.W.S.), RO1 AI059492 (A.S.R., G.D.), R37 AI024716 (A.S.R.) and Swiss National Science Foundation grant SNF 3100A0-112588/1 (E.M.Z.). We would like to acknowledge the excellent work of the Broad Genome Sequencing Platform and the Venter Institute Joint Technology Center. We thank Chris Town, Neil Hall, Ewen Kirkness for critical comments and the *Aedes aegypti* research community for their enthusiastic support and willing assistance in this project.

References

1. Beaty, BJ.; Marquardt, WC. Biology of Disease Vectors, 1st Edition. University Press of Colorado; Niwot, Colorado: 1996. p. 612
2. Christophers, SR. Aedes aegypti (L.): The Yellow Fever Mosquito, Its Life History, Bionomics and Structure. Cambridge University Press; 1960. p. 739
3. Clements, AN. The Biology of Mosquitoes. Chapman & Hall; London: 1992. p. 595
4. Severson DW, Brown SE, Knudson DL. Annu. Rev. Entomol. 2001; 46:183. [PubMed: 11112168]
5. Tomori O. Crit Rev Clin Lab Sci. 2004; 41:391. [PubMed: 15487593]
6. WHO. Dengue and dengue haemorrhagic fever. World Health Organization; Geneva: 2002. Fact sheet No. 117
7. Ligon BL. Semin Pediatr Infect Dis. Apr.2006 17:99. [PubMed: 16822471]
8. Mackenzie JS, Gubler DJ, Petersen LR. Nat Med. Dec.2004 10:S98. [PubMed: 15577938]
9. Holt RA, et al. Science. Oct 4.2002 298:129. [PubMed: 12364791]
10. Riehle MM, et al. Science. Apr 28.2006 312:577. [PubMed: 16645095]
11. Kwon HW, Lu T, Rutzler M, Zwiebel LJ. Proc Natl Acad Sci U S A. Sep 5.2006 103:13526. [PubMed: 16938890]
12. Hallem EA, Nicole Fox A, Zwiebel LJ, Carlson JR. Nature. Jan 15.2004 427:212. [PubMed: 14724626]
13. Sharakhov IV, et al. Proc Natl Acad Sci U S A. Apr 18.2006 103:6258. [PubMed: 16606844]
14. Turner TL, Hahn MW, Nuzhdin SV. PLoS Biol. Sep.2005 3:e285. [PubMed: 16076241]
15. David JP, et al. Proc Natl Acad Sci U S A. Mar 15.2005 102:4080. [PubMed: 15753317]
16. Zdobnov EM, et al. Science. Oct 4.2002 298:149. [PubMed: 12364792]

17. Gaunt MW, Miles MA. Mol Biol Evol. May.2002 19:748. [PubMed: 11961108]
18. Krzywinski J, Grushko OG, Besansky NJ. Mol Phylogenet Evol. May.2006 39:417. [PubMed: 16473530]
19. Craig, GBJ.; Hickey, WA. Genetics of Aedes aegypti. In: Wright, JW.; Pal, R., editors. Genetics of Insect Vectors of Disease. Elsevier; New York: p. 67-131.
20. Warren AM, Crampton JM. Genet Res. Dec.1991 58:225. [PubMed: 1802804]
21. Supplementary on line material.
22. Jaffe DB, et al. Genome Res. Jan.2003 13:91. [PubMed: 12529310]
23. Feschotte C. Mol Biol Evol. Sep.2004 21:1769. [PubMed: 15190130]
24. Coy MR, Tu Z. Insect Mol Biol. Oct.2005 14:537. [PubMed: 16164609]
25. Craig, N.; Craige, R.; Gellert, M.; Lambowitz, A. Mobile DNA II. American Society for Microbiology Press; Washington, DC:
26. Tubio JM, Naveira H, Costas J. Mol Biol Evol. Jan.2005 22:29. [PubMed: 15356275]
27. Arkhipova IR, Pyatkov KI, Meselson M, Evgen'ev MB. Nat Genet. Feb.2003 33:123. [PubMed: 12524543]
28. Zhang X, Jiang N, Feschotte C, Wessler SR. Genetics. Feb.2004 166:971. [PubMed: 15020481]
29. Noriega FG, Wells MA. J Insect Physiol. Jul.1999 45:613. [PubMed: 12770346]
30. Borovsky D. J Exp Biol. 2003; 206:3869. [PubMed: 14506222]
31. Bartholomay LC, et al. Infect Immun. Jul.2004 72:4114. [PubMed: 15213157]
32. Sanders HR, et al. Insect Biochem Mol Biol. Nov.2005 35:1293. [PubMed: 16203210]
33. Ranson H. personal communication.
34. Wise A, Gearing K, Rees S. Drug Discov Today. Feb 15.2002 7:235. [PubMed: 11839521]
35. Ren Q, Kang KH, Paulsen IT. Nucleic Acids Res. Jan 1.2004 32:D284. [PubMed: 14681414]
36. Schutt C, Nothiger R. Development. Feb.2000 127:667. [PubMed: 10648226]
37. Marinotti O, et al. Insect Mol Biol. Feb.2006 15:1. [PubMed: 16469063]
38. Severson DW, et al. J Hered. Mar-Apr;2004 95:103. [PubMed: 15073225]
39. Zdobnov EM, Bork P. Trends Genet. Nov 8.2006
40. Shin SW, Bian G, Raikhel AS. J Biol Chem. Dec 22.2006 281:39388. [PubMed: 17068331]
41. Riehle MA, Fan Y, Cao C, Brown MR. Peptides. Aug 23.2006
42. Waterhouse RM, et al. Science. Submitted.
43. Mercado-Curiel RF, et al. BMC Microbiol. 2006; 6:85. [PubMed: 17014723]
44. Bosio CF, Fulton RE, Salasek ML, Beaty BJ, Black W. C. t. Genetics. Oct.2000 156:687. [PubMed: 11014816]
45. Severson DW, Meece JK, Lovin DD, Saha G, Morlais I. Insect Mol Biol. Aug.2002 11:371. [PubMed: 12144703]
46. Jansen RC, Nap JP. Trends Genet. Jul.2001 17:388. [PubMed: 11418218]

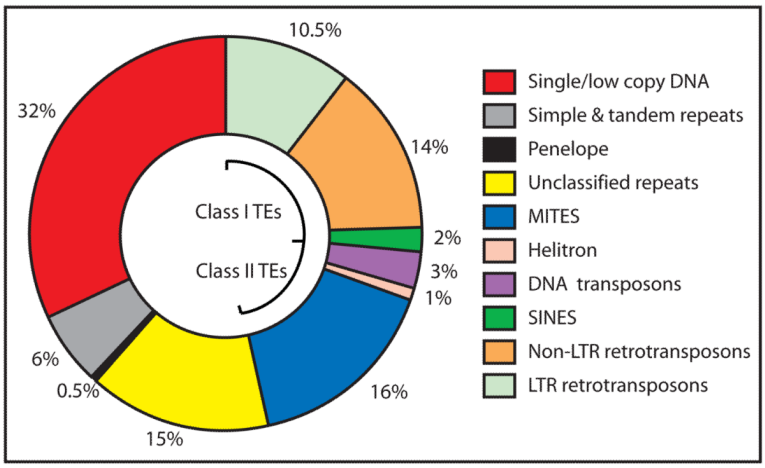

Figure 1.
Relative genomic content of annotated TEs and other sequences in *Aedes aegypti*. TEs have been deposited in TEfam, a relational database for submission, retrieval, and analysis of TEs (http://tefam.biochem.vt.edu).

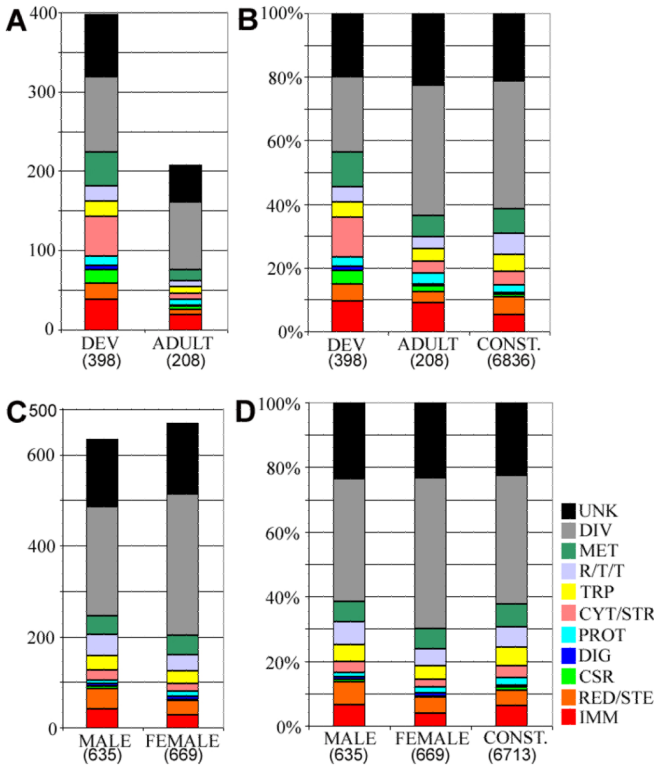

Figure 2.

Transcriptome analyses of *Aedes aegypti.* **A**. Functional class distributions of genes that are enriched in pre-adult stages (DEV) and the adult female stage (ADULT) (Table S9). **B.** Proportions of functional gene classes expressed as the percentage of the total umber of genes that are enriched in pre-adult stages (DEV), adult female stage (ADULT) and constitutively expressed genes (CONST.). C and D. same as A and B for genes enriched in the male, female and genes common (CONST.) for both sexes (Table S17). Functional classes are: Immunity (IMM), redox, oxidoreductive stress (RED/STE), chemosensory reception (CSR), blood and sugar food digestive (DIG), proteolysis (PROT), cytoskeletal and structural (CYTISTR), transport (TRP), replication, transcription and translation (R/T/T), metabolism (MET), diverse functions (DIV), unknown functions (UNK). The total number of genes in each category is indicated in parenthesis.

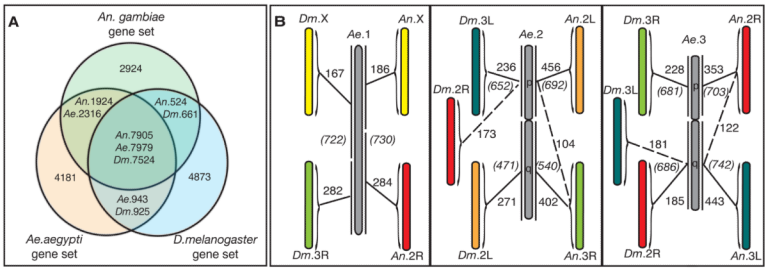

Figure 3.

Orthology and chromosomal synteny between *Ae. aegypti, An. gambiae* and *D. melanogaster*. **A.** Each circle represents a gene set for *Ae. aegypti* (*Ae*), *An .gambiae* (*An*) and *D. melanogaster* (*Dm*). Because a gene can be involved in several homologies, gene sets do not always have the same number of genes within intersections; e.g.: in the *Ae/Dm* comparison, 943 Ae genes are similar to *Dm* while 925 *Dm* genes are similar to *Ae*. **B.** *Aedes aegypti* chromosomes are represented in grey (not to scale). Chromosome arms are designated as 'p' and 'q' - with no arm distinctions for chromosome 1. Colored chromosomes represent the syntenic chromosome from *An. gambiae* (*An*) or *D. melanogaster* (*Dm*) (not to scale). Solid lines link the *Ae. aegypti* chromosome to their primary syntenic chromosome and dashed lines to their secondary syntenic chromosome. The number of *Ae* orthologs to *An* and *Dm* chromosome arms is indicated and the total number of orthologs on the *Ae* chromosome arm to *Ae* or *Dm* is shown in italic in parenthesis.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Comparative statistics of *Ae. aegypti* nuclear genome coding characteristics

	Species		
Features	***Ae. aegypti***	***An. gambiae*** [c]	***D. melanogaster*** [c]
Size (Mbp)	1,376	272.9	118
Number of chromosomes	3	3	4
Total G+C composition (%)	38.2	40.9	42.5
Number of protein coding genes	15,419	13,111	13,718
Average gene length[a] (bp)	14,587	5,124	3,460
Average protein coding gene length[b] (bp)	1,397	1,154	1,693
Percent genes with introns	90.1	93.6	86.2
Average number of exons/gene	4.0	3.9	4.9
Average intron length (bp)	4,685	808	1,175
Longest intron (bp)	329,294	87,786	132,737
Average length of intergenic region (bp)	56,417	17,265	6,043

[a] includes introns but not UTRs.

[b] not including introns.

[c] statistics were derived from genome updates for these species *An. gambiae* R-AgamP3 and *D. melanogaster* R-4.2.