Insect Molecular Biology (2008) **17**(1), 1–8

The *Aedes aegypti* genome: a comparative perspective

R. M. Waterhouse*, S. Wyder†‡ and E. M. Zdobnov*†‡

**Imperial College London, South Kensington Campus, London, UK;* †*Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; and* ‡*Swiss Institute of Bioinformatics, Geneva, Switzerland*

Abstract

The sequencing of the second mosquito genome, *Aedes aegypti*, in addition to *Anopheles gambiae*, is a major milestone that will drive molecular-level and genome-wide high-throughput studies of not only these but also other mosquito vectors of human pathogens. Here we overview the ancestry of the mosquito genes, list the major expansions of gene families that may relate to species adaptation processes, as exemplified by CYP9 cytochrome P450 genes, and discuss the conservation of chromosomal gene arrangements among the two mosquitoes and fruit fly. Many more invertebrate genomes are expected to be sequenced in the near future, including additional vectors of human pathogens (see www.vectorbase.org), and further comparative analyses will become increasingly refined and informative, hopefully improving our understanding of the genetic basis of phenotypical differences among these species, their vectorial capacity, and ultimately leading to the development of novel disease control strategies.

Keywords: comparative genomics, gene family, insect, mosquito, synteny, vector.

Introduction

Malaria research has greatly benefited from the sequencing of the genome of the primary mosquito vector, *Anopheles gambiae* (Holt *et al*., 2002). This has facilitated molecular-level studies and prompted the development of subsequent large-scale functional genomics tools for transcriptome analysis and reverse genetics, particularly in the field of insect innate immunity and the vector's response to the malaria parasite (eg Riehle *et al*., 2006). Our understanding of the mosquito genome was fuelled by comparative analysis (Zdobnov *et al*., 2002) with the fruit fly, *Drosophila melanogaster*, the best studied insect laboratory model organism, and the only other insect genome available at the time (Adams *et al*., 2000). The sequencing of the second mosquito genome, *Aedes aegypti* (Nene *et al*., 2007), now provides further opportunities to elucidate important features of vector biology in more detail. *Aedes aegypti* is primarily of concern as a viral vector; nevertheless, most *Aedes* strains are susceptible to the avian malaria parasite *Plasmodium gallinaceum* and some are susceptible to filariatic nematodes, facilitating the investigation of multiple mosquito–pathogen interactions. This will be complemented by references to the more distantly related genomes of the recently sequenced red flour beetle (*Tribolium castaneum*) and the honey bee (*Apis mellifera*). The field will continue to grow with the ever-accelerating high-throughput sequencing, and the projects underway for additional malaria-transmitting *Anopheles* species, as well as the *Culex pipiens* mosquito, the body louse *Pediculus humanus*, the blood-sucking bug *Rhodnius prolixus*, tsetse flies, sand flies and other arthropods. The Vectorbase resource (www.vectorbase.org; Lawson *et al*., 2007) will act as the central gateway for genomic resources of such invertebrate vectors of human pathogens.

The initial analysis of the *Aedes* genome (Nene *et al*., 2007) revealed 15 419 genes, 80% of which could be confirmed by large-scale transcriptional analyses. Approximately one third of the genes can be mapped to chromosomal locations based on genetic and physical mapping data, suggesting that, as between *Anopheles* and *Drosophila* (Zdobnov *et al*., 2002), it will be possible to map most *Aedes* chromosomal arms to their orthologous chromosomal elements in other mosquitoes and flies (Nene *et al*., 2007). The second mosquito genome provides a comparative perspective of 145–200 Mya divergence (Krzywinski *et al*., 2006) between the mosquitoes (representing the *Culicinae* and *Anophelinae* subfamilies), *ca.* 250 Mya of divergence between the mosquitoes and the drosophilids, as well as

Received 1 May 2007; accepted after revision 31 August 2007. Correspondence: Evgeny M. Zdobnov, Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland. Tel.: +41 22 379 59 73; fax: +41 22 379 57 06; e-mail: zdobnov@medecine.unige.ch

across the *ca.* 300 Mya that separate them from the honey bee (Diptera and Hymenoptera orders) (Grimaldi & Engel, 2005). However, dating these species radiations can be deceiving, as among insects, and particularly within Diptera, evolution proceeds at a faster rate than among vertebrates (Zdobnov *et al.*, 2002; HGSC, 2006; Zdobnov & Bork, 2007). In terms of objective measures of divergence such as protein sequence identity and retained chromosomal gene arrangements, these two mosquitoes are more divergent than human and chicken (300 Mya), but less than human and fish (450 Mya). Comparative genome analyses will provide important insights into insect evolution and facilitate the identification of genes and gene functions which may be common across many species. Of major interest with regard to the prevention of mosquito-borne diseases are genes specific to the activities of blood-feeding and host-seeking, as well as the molecular processes of insecticide resistance and innate immunity.

Gene repertoire evolution

The accurate characterisation of protein coding genes remains a priority, and total gene numbers tend to be adjusted as our knowledge increases, with the removal of orphan predictions (eg derived from transposons and pseudogenes), and the identification of previously overlooked genes or the correction of gene models (eg the splitting of artificially fused gene predictions). Over the last few years remarkable progress has been achieved with the development of gene prediction tools, from *ab initio* and comparative prediction methods to consensus gene model finding approaches (Eyras *et al.*, 2005; Guigo *et al.*, 2006; HGSC, 2006; Elsik *et al.*, 2007) that when applied concertedly can identify almost complete gene repertoires with a reasonable quality for the vast majority of the genes. However, it is still hard to underestimate the importance of primary sequence data from cDNA and expressed sequence tag (EST) libraries for the refinement of gene models and the identification of fast-evolving and species-specific genes (Kriventseva *et al.*, 2005), and the *Aedes aegypti* genome project greatly benefited from the sequencing of about 265 000 ESTs.

Equipped with relatively robust sets of gene models, the focus shifts to understanding their chromosomal organisation and inter-relationships; that is, how do the gene repertoires compare across different species? This can be approached in two ways: (1) by drawing parallels between orthologous genes that are ancestrally related and therefore represent 'corresponding' genes in different species; and (2) by delineating broader families of genes that exemplify the functional diversity of the gene repertoire. Such comparisons can provide insights into the evolutionary trends that shape the gene sets of each species and help to elucidate the genetic basis of both ubiquitous and species-specific traits.

Remarkably, the much larger genome of *Aedes* does not seem to encode considerably more genes than *Anopheles*, or in fact any other sequenced insect genome, as shown in our recent comparison across five insects (*Drosophila melanogaster*, *Anopheles gambiae*, *Aedes aegypti*, *Tribolium castaneum* and *Apis mellifera*) and five vertebrates [human (*Homo sapiens*), mouse (*Mus musculus*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*) and pufferfish (*Tetraodon nigroviridis*)], performed as part of the initial *Tribolium* genome analysis (TGSC, in press). Summarized from the mosquito perspective in Fig. 1, this shows that almost 90% of *Aedes* genes exhibit orthologous or homologous relationships to genes in the species listed above (Fig. 1A) (for category definitions see Fig. 1 legend). Of the set of *Aedes* genes that exhibit orthologous relationships (Fig. 1B), over 70% stem from ancient Urbilateria genes, with almost a quarter being single-copy (1:1:1) orthologues and 40% are multiple-copy (N:N:N) orthologues present in all metazoans. Over half of the 3190 *Aedes* genes which are members of insect-specific orthologous groups (ie without clearly identifiable vertebrate counterparts) seem to be common to all holometabolous insects, and about two thirds of the remaining Diptera-specific genes appear to be mosquito-specific. Orthologous relationships are unclear for the remaining 4217 genes, of which about half show weaker homology to proteins from other species and the rest have no detectable sequence similarity (Fig. 1A). Most of these have probably diverged beyond recognition as genes in all categories show variable rates of evolution, and even the most conserved fraction of single-copy orthologues displays a very broad distribution of protein identities (Zdobnov *et al.*, 2002; Nene *et al.*, 2007). Interestingly, our recent analysis of the phyletic distribution of genes forming orthologous groups suggests that hundreds of ancient genes have been lost during the evolution of Bilateria in each of the lineages, and at extremes even losses of universal single-copy orthologues can be tolerated (Wyder, in press).

The most specialized gene families

Expanded protein families may reflect species adaptations and can therefore reveal underlying biological differences between individual species or clades. The two major approaches to define protein families are either based on whole-length protein sequence similarity, eg applying clustering techniques to all-against-all sequence comparisons, or on using profiles of known protein domains, eg identification of InterPro domain families (Zdobnov & Apweiler, 2001). Here we survey gene family expansions in both mosquitoes compared to *Drosophila*, complementing the InterPro-based analysis reported for *Aedes* (Nene *et al.*, 2007) by applying single-linkage (nearest neighbour) clustering to all-against-all BLAST sequence comparisons using BLASTCLUST (Dondoshansky & Wolf, unpubl. data,

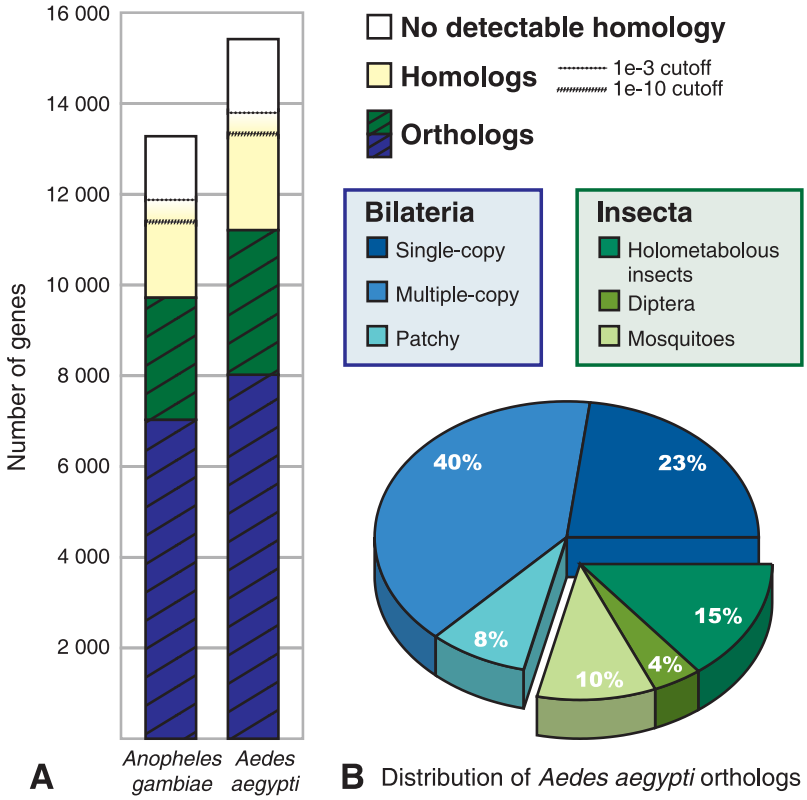

Figure 1. Distribution of orthologous and homologous genes in *Anopheles gambiae* and *Aedes aegypti*. (A) Coverage of the mosquito gene repertoires from metazoan orthology (bottom) to genes with undetectable similarity (top) (with Smith-Waterman *e*-value cutoffs of $1e^{-10}$ and $1e^{-3}$). (B) Partitioning of the Metazoan-wide orthologues (blue) and the insect-specific orthologues (green) on the basis of a five insect, five vertebrate comparison (TGSC, in press). Single-copy orthologues (1:1:1) are defined as having exactly one gene in at least nine out of 10 species (eg allowing absence or duplication in one genome to account for incomplete genome annotations or recent duplications). Similarly, multiple-copy (N : N : N) orthologues have representatives in all species (allowing an absence in one species) without a restriction on the gene copy-number. Patchy orthologous groups are less constrained and can accommodate losses in multiple lineages, but are required to have at least one insect and two vertebrate or two insect and one vertebrate gene members. The holometabolous insect orthologous groups were required to span at least four out of five insects considered, but without any vertebrate counterparts. The about twofold higher number of genes found to be orthologous only between the two mosquitoes compared to those found only in the mosquitoes and the fruit fly may be slightly biased by the derived state of the *Drosophila* genome.

http://biowulf.nih.gov/apps/blast/doc/blastclust.html). Requiring at least 60% coverage across both genes, 30% sequence identity between them, and a BLAST *e*-value cutoff of $1e^{-10}$ resulted in 158 clusters from *Aedes*, *Anopheles* and *Drosophila* proteomes with at least four proteins in each species. Examining the difference in the numbers of genes in each family between the mosquitoes and the fruit fly highlighted several prominent mosquito expansions of these multigene families (Fig. 2). The principal InterPro domain detected for proteins in each cluster provides clues as to the function of a given family, and in most cases the domain is identified for all, or the majority, of family members. This analysis reveals the most striking expansions in *Aedes* of proteins containing Zinc-finger, insect pheromone-binding, cytochrome P450 and insect cuticle domains, as well as highlighting additional expansions occurring in families implicated in insect immunity and insecticide resistance. The results of a large collaborative effort focussing on immune-related genes and pathways describe in detail further variations among these three Diptera (http://cegg.unige.ch/Insecta/immunodb), relating distinct functional categories of gene families to different modes of evolution (Waterhouse *et al*., 2007).

The expansion of fibrinogen-related proteins (FREPs) in *Anopheles* has been previously noted in comparison to *Drosophila* (Zdobnov *et al*., 2002), and while not as large, *Aedes* also exhibits an enrichment of FREPs. Fibrinogen is involved in platelet aggregation and blood clotting processes, the large expansions in mosquitoes have been speculated to be partly linked to blood-feeding (Kairies *et al*., 2001), and several *Anopheles* FREPs have been shown to be up-regulated by bacterial challenge and malarial infection (Christophides *et al*., 2002; Dimopoulos *et al*., 2002; Dong *et al*., 2006). Several additional domains define members of immune-related gene families with prominent expansions in *Aedes*. MD-2-related proteins interact with antigens through their lipid-recognition domain and may specifically recognize malaria parasites in *Anopheles* (Dong *et al*., 2006). The serine protease domain (IPR001254) and the disulphide knot CLIP domain (not shown in Fig. 2) combine to form the expanded family of CLIP-domain serine proteases and their homologues which are involved in the modulation of immune signalling. Infections can activate CLIP-protease zymogens present in the haemolymph, resulting in proteolytic activation of prophenoloxidases (PPOs) (Kanost *et al*., 2004; Tang *et al*., 2006). These PPO effectors are included in the set of proteins identified through their haemocyanin domains (which also include larval storage proteins), and are expanded in the mosquitoes. The insect family of thioester-containing proteins (TEPs), identified through their alpha-2-macroglobulin domains, are related to vertebrate complement factors C3/C4/C5. *Anopheles* TEP1, which has no orthologue in *Aedes*, binds to pathogen surfaces

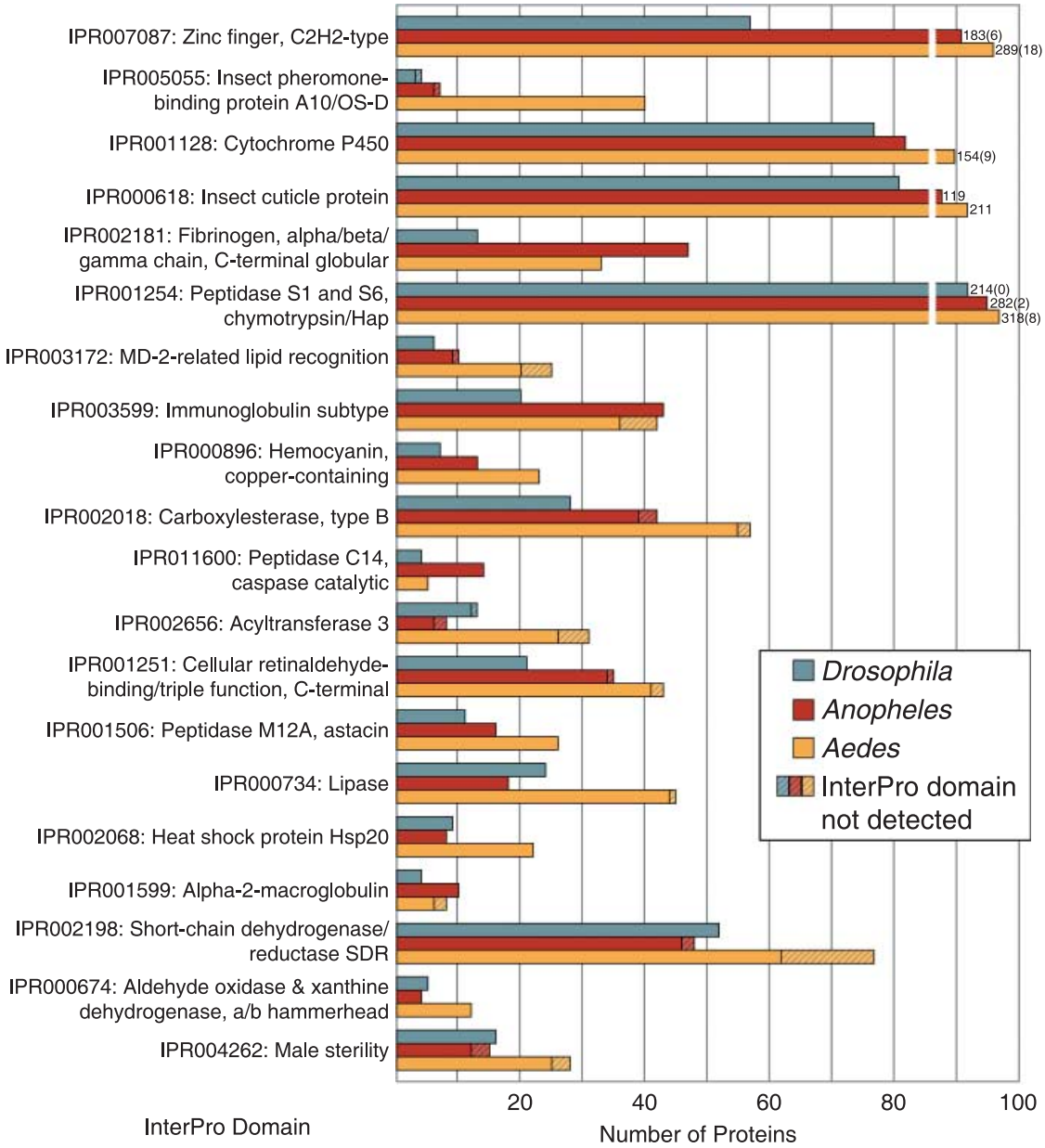

Figure 2. Gene family expansions in *Drosophila melanogaster* (blue), *Anopheles gambiae* (red), and *Aedes aegypti* (yellow). Multigene families are defined by applying BLASTCLUST (Dondoshansky & Wolf, unpubl. data) with 60% coverage, 30% identity and BLAST *e*-value cutoff of $1e^{-10}$, which identified 158 non-overlapping clusters with at least four proteins in each species. The numbers of proteins in each species are compared for the most significant expansions of these multigene families, where the size of the cluster is consistent with the number of domains identified by InterProScan. The principal InterPro domain detected for proteins in each cluster provides clues as to the function of a given family. The numbers of cluster members for which the respective InterPro domain was not detected are patterned accordingly. The largest clusters are truncated to fit on the chart, with the number of 'domainless' members indicated in parentheses. The cluster which captures the insect cuticle proteins is greatly expanded through the inclusion of proteins which exhibit similar low complexity sequences as those present near the N-terminal of cuticle proteins, these are therefore not shown on the chart. Among these Diptera, *Aedes* noticeably exhibits the majority of the largest expansions.

and promotes bacterial phagocytosis and killing of *Plasmodium* parasites (Levashina *et al*., 2001; Blandin *et al*., 2004; Moita *et al*., 2005). These and other multigene families of recognition receptors, signal modulators, and effectors of immune responses show significant diversity among these three insects (Waterhouse *et al*., 2007). In contrast, the honey bee appears to have a reduced immune repertoire compared to dipterans, which may reflect the environment created by the complex social existence of bees (Evans *et al*., 2006).

Common mechanisms of insecticide resistance involve altered amounts or activities of detoxification enzymes

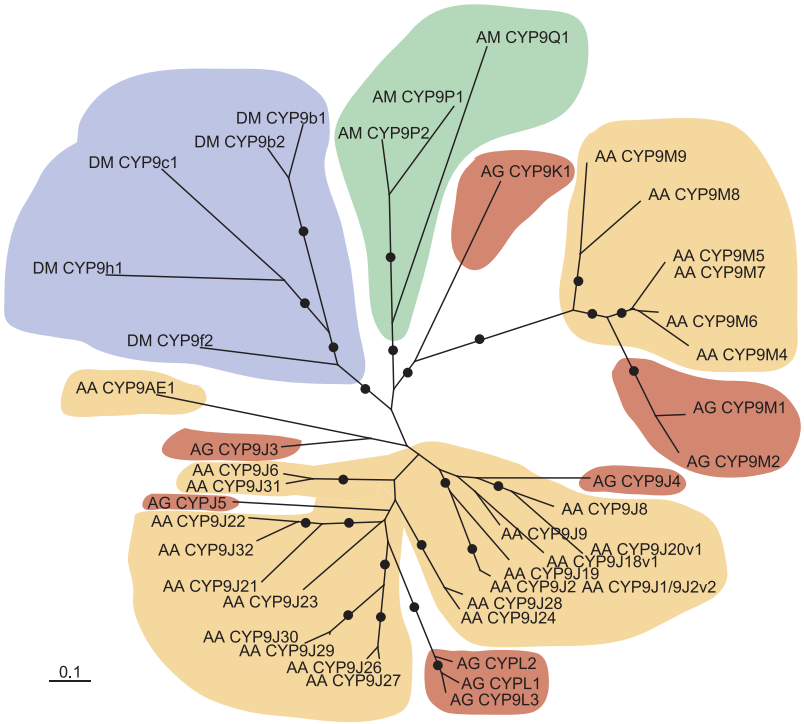

Figure 3. Phylogenetic relationships of the CYP9 cytochrome P450 genes from *Drosophila melanogaster* (blue), *Anopheles gambiae* (red), *Aedes aegypti* (yellow), and the honey bee, *Apis mellifera* (green). The unrooted maximum likelihood tree was calculated from the conserved protein sequence core using Phyml (JTT+G(4)+F+I, 100 bootstraps) (Guindon & Gascuel, 2003). Separations with bootstrap support > 70% are indicated with a dot.

such as glutathione-S-transferases (GSTs), cytochrome P450 monooxygenases (CYPs), and carboxylesterases (COEs) (Hemingway *et al*., 2004). The CYPs exhibit a dramatic expansion in *Aedes* (see discussion below), and the COEs are expanded in both mosquitoes, but to a much greater extent in *Aedes*. Resistance can develop through the amplifications of esterases, as in the case of the most common insecticide resistance-associated esterases, *est*α2[1] and *est*β2[1], which are co-amplified up to 80 times in resistant *Culex pipiens quinquefasciatus* (Vaughan *et al*., 1997; Paton *et al*., 2000). The likely orthologues of this pair of *Culex* esterases are also found in a neighbouring head-to-tail orientation in *Anopheles*, and although they are not linked in the current *Aedes* assembly, their separation is a result of an assembly error and the arrangement is in fact conserved (Ranson, pers. comm.).

In all three mosquitoes, an aldehyde oxidase (AO) gene is located adjacent to the α-esterase, and in *Culex* this AO is coamplified with the esterases (Hemingway *et al*., 2000). The family of aldehyde oxidase (AO) and xanthine dehydrogenase (XDH) enzymes is expanded in *Aedes*. As well as being a key enzyme in purine degradation, XDH is also involved in reactive oxygen species (ROS) production, and although the roles of AOs are less clear, they have a much wider substrate specificity and have been implicated in the metabolism of hormones, xenobiotics and ROS in plants and animals (Mendel & Bittner, 2006). Three *Aedes* genes cluster with *Drosophila* XDH (rosy), while a much more striking expansion is observed with respect to the AOs. The selective advantage of the *Culex* amplicon is unlikely to be a result of esterase activity alone and different alleles of the neighbouring AO gene are also likely to influence insecticide susceptibility (Coleman *et al*., 2002).

CYPs comprise a large superfamily of heme-thiolate proteins that metabolize a wide range of both endogenous and exogenous hydrophobic compounds (Werck-Reichhart & Feyereisen, 2000). The massive expansion of *Aedes* CYPs identified by BLASTCLUST and InterPro analysis can be partitioned through the classification of proteins into orthologous groups. A notable example of an expansion within such a CYP family is that of the CYP9 genes, with 26 in *Aedes* compared with nine in *Anopheles*, five in *Drosophila* and three in the honey bee (Fig. 3). The phylogenetic analysis reveals multiple gene duplications in *Aedes*, mostly occurring after the split between the two mosquito lineages. Although the majority of CYPs linked to insecticide resistance belong to the CYP6 family, the CYP9s and CYP6s are closely related and together make up the CYP3 clan, and CYP9 genes from Lepidoptera were found to be constitutively overexpressed in insecticide resistant strains (Yang *et al*., 2006) and to be induced after feeding of xenobiotics (Stevens *et al*., 2000).

Aedes genome size and the evolution of gene order

Comparison of the arrangements of genes along the chromosomes can reveal conserved gene orders and highlight putative clusters of co-regulated genes. Nevertheless, functional

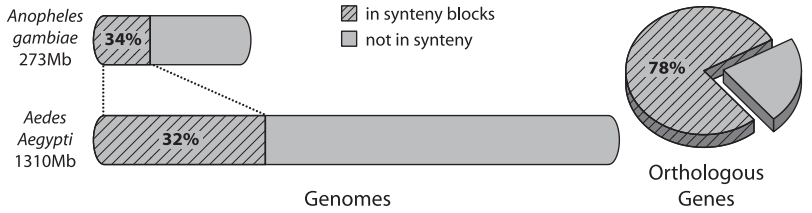

Figure 4. Regions with conserved gene arrangements (synteny blocks) between the genomes of *Anopheles gambiae* and *Aedes aegypti*. Left, the scaled schematic illustration of the mosquito genomes shows similar (32 and 34%) summed spans of syntenic regions. Right, 78% of mosquito single-copy (1 : 1) orthologues are confined to the syntenic regions.

interpretations of conserved gene arrangements should be made with caution as the recent comparison of 12 species from Diptera, Lepidoptera, Coleoptera and Hymenoptera suggested almost no constraints on gene arrangements over large evolutionary distances (Zdobnov & Bork, 2007).

Chromosomal rearrangements, such as inversions and translocations, lead to the decay of the ancestral gene order into mosaics of orthologous chromosomal regions, commonly termed 'syntenic'. Such local synteny can be identified as genomic blocks of proximal arrangements of orthologues across two or more species. Comparison of gene arrangements between *Aedes* and *Anopheles* mosquitoes revealed that almost 80% of their 1 : 1 orthologues are retained in the same genomic neighbourhood; however, these are confined to regions which make up only about a third of each genome (Fig. 4). The role and biological relevance of the gene-rich regions and gene deserts remain to be elucidated. The span of the syntenic regions is ~4.6-fold larger in *Aedes* than in *Anopheles* and correlates well with the radical fivefold difference in their total genome sizes. Taken together with the consistently longer intronic and intergenic sequences in *Aedes*, this suggests that the expansion was facilitated by the infiltration of transposons, which, with their recognisable remnants, constitute almost half of the *Aedes* genome (Nene *et al*., 2007).

Extending the synteny analysis to the three-way comparison of *Aedes–Anopheles–Drosophila* using 5358 1 : 1 : 1 orthologues shows that more than double the number of gene arrangements are preserved between the two mosquitoes than between either mosquito and the fruit fly (Fig. 5). It also shows that there are a substantial number of lineage-specific breaks of synteny; eg when genes remain similarly arranged between *Anopheles* and *Drosophila* but not between *Aedes* and *Drosophila*. In total, there are about two times more genes that are in exclusive *Anopheles–Drosophila* synteny than there are in exclusive *Aedes–Drosophila* synteny. More strictly, counting the number of blocks kept in synteny between *Anopheles–Drosophila* that have been broken into different blocks in *Aedes–Anopheles* reveals 81 traceable *Aedes*-specific breaks, and correspondingly there are only 31 traceable *Anopheles*-specific breaks. This is indicative of the about twofold higher rate of genome shuffling in *Aedes* compared to *Anopheles* that would be consistent with the higher transposon content of *Aedes*, as they are known to facilitate chromosomal

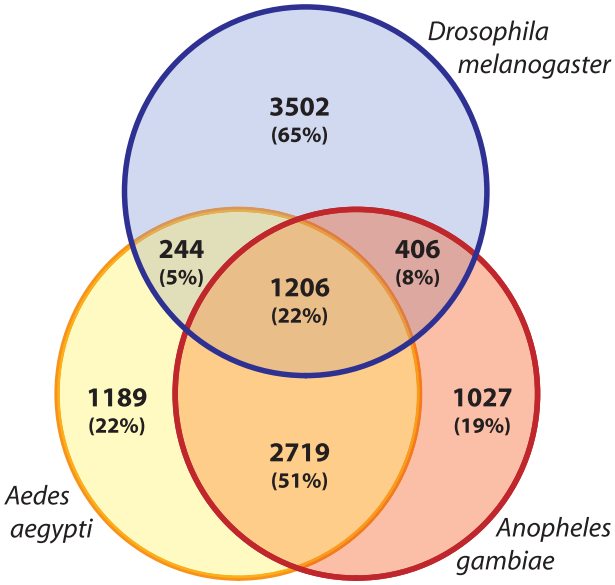

Figure 5. Three-way *Aedes–Anopheles–Drosophila* comparison of orthologues found in synteny. Each circle represents the *Aedes aegypti* (yellow), *Anopheles gambiae* (red), and *Drosophila melanogaster* (blue) members of the 5358 identifiable single-copy orthologous trios. Pairwise and three-way synteny blocks require at least two single-copy orthologues to be next to each other in the compared genomes with no more than one intervening gene as described in Zdobnov & Bork (2007). The Venn diagram indicates the number of orthologues falling in pairwise or three-way synteny for each species, with the percentages in parentheses. Of all trios, ~73% of mosquito genes are found in synteny, while ~27% of *Aedes* genes and ~30% of *Anopheles* genes retain their gene neighbourhoods with their orthologues in *Drosophila*.

rearrangements; however, this estimate might be biased by the currently fragmented *Aedes* assembly (1612 supercontigs, only 710 of which contain more than two 1 : 1 : 1 orthologues, 589 more than three, etc). Interestingly, the higher rate of *Aedes* genome shuffling contrasts with a slightly slower rate of accumulation of protein substitutions in well conserved cores of orthologues as measured using quantitative phylogeny reconstruction with maximum-likelihood methods (Zdobnov & Bork, 2007).

Conclusions

The recent sequencing of the *Aedes* genome, and the prior availability of the *Anopheles* genome, has provided the first

opportunity to carry out detailed comparative genomic analyses between two mosquito species with distinct vectorial capacities. Although the *Aedes* genome is about fivefold larger than that of *Anopheles*, their encoded gene repertoires are remarkably similar, with over 50% of their genes exhibiting traceable orthology across metazoa and a further ~20% across holometabolous insects. Several prominent protein family expansions are evident between the two mosquitoes and in comparison to *Drosophila*, some of these have been tentatively linked to mosquito or species-specific characteristics while others remain to be investigated further. The two mosquitoes display about twice the number of conserved gene arrangements than between either mosquito and the fruit fly, reflecting the difference in evolutionary time span between the radiation of the mosquito lineages and their divergence from flies. The availability of the genome sequences and gene annotations will drive studies to dissect the niceties of mosquito biology and interrogate gene function, particularly with respect to pathogen–vector interactions. With the imminent analysis of the *Culex pipiens* genome and the advancing efforts on sequencing additional mosquito species, comparative analyses will become increasingly complex and refined; eg genomes of more closely-related species will enable the use of currently inapplicable techniques to measure selection pressures (K_A/K_S). Hopefully, such studies will accelerate progress towards our understanding of the vector biology, producing testable hypotheses of the genetic basis of phenotypical differences, and will ultimately lead to the development of novel control strategies.

Acknowledgements

The authors would like to thank Evgenia Kriventseva for contributing the orthology classification and InterPro analysis, George Christophides and Hilary Ranson for discussions, and David Nelson for sharing his P450 annotations. We gratefully acknowledge the Wellcome Trust PhD Fellowship to RMW and Swiss National Science Foundation funding (SNF 3100A0-112588 to EMZ).

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., *et al*. (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.

Blandin, S., Shiao, S.H., Moita, L.F., Janse, C.J., Waters, A.P., Kafatos, F.C., *et al*. (2004) Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae. Cell* **116**: 661–670.

Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., *et al*. (2002) Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298**: 159–165.

Coleman, M., Vontas, J.G. and Hemingway, J. (2002) Molecular characterization of the amplified aldehyde oxidase from insecticide resistant *Culex quinquefasciatus*. *Eur J Biochem* **269**: 768–779.

Dimopoulos, G., Christophides, G.K., Meister, S., Schultz, J., White, K.P., Barillas-Mury, C., *et al*. (2002) Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge, and malaria infection. *Proc Natl Acad Sci USA* **99**: 8814–8819.

Dong, Y., Aguilar, R., Xi, Z., Warr, E., Mongin, E. and Dimopoulos, G. (2006) *Anopheles gambiae* immune responses to human and rodent *Plasmodium* parasite species. *PLoS Pathog* **2**: e52.

Elsik, C.G., Mackey, A.J., Reese, J.T., Milshina, N.V., Roos, D.S. and Weinstock, G.M. (2007) Creating a honey bee consensus gene set. *Genome Biol* **8**: R13.

Evans, J.D., Aronstein, K., Chen, Y.P., Hetru, C., Imler, J.L., Jiang, H., *et al*. (2006) Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Mol Biol* **15**: 645–656.

Eyras, E., Reymond, A., Castelo, R., Bye, J.M., Camara, F., Flicek, P., *et al*. (2005) Gene finding in the chicken genome. *BMC Bioinformatics* **6**: 131.

Grimaldi, D. and Engel, M.S. (2005) *Evolution of the Insects*. Cambridge University Press, Cambridge, UK.

Guigo, R., Flicek, P., Abril, J.F., Reymond, A., Lagarde, J., Denoeud, F., *et al*. (2006) EGASP: the human ENCODE genome annotation assessment project. *Genome Biol* **7** (Suppl. 1): S21–31.

Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.

Hemingway, J., Coleman, M., Paton, M., McCarroll, L., Vaughan, A. and Desilva, D. (2000) Aldehyde oxidase is coamplified with the World's most common *Culex* mosquito insecticide resistance-associated esterases. *Insect Mol Biol* **9**: 93–99.

Hemingway, J., Hawkes, N.J., McCarroll, L. and Ranson, H. (2004) The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem Mol Biol* **34**: 653–665.

HGSC (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931–949.

Holt, R.A., Subramanian, G.M., Halpem, A., Sutton, G.G., Charlab, R., Nusskem, D.R., *et al*. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129–149.

Kairies, N., Beisel, H.G., Fuentes-Prior, P., Tsuda, R., Muta, T., Iwanaga, S., *et al*. (2001) The 2.0-A crystal structure of tachylectin 5A provides evidence for the common origin of the innate immunity and the blood coagulation systems. *Proc Natl Acad Sci USA* **98**: 13519–13524.

Kanost, M.R., Jiang, H. and Yu, X.Q. (2004) Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol Rev* **198**: 97–105.

Kriventseva, E.V., Koutsos, A.C., Blass, C., Kafatos, F.C., Christophides, G.K. and Zdobnov, E.M. (2005) AnoEST: toward *A. gambiae* functional genomics. *Genome Res* **15**: 893–899.

Krzywinski, J., Grushko, O.G. and Besansky, N.J. (2006) Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol Phylogenet Evol* **39**: 417–423.

Lawson, D., Arensburger, P., Atkinson, P., Besansky, N.J., Bruggner, R.V., Butler, R., *et al*. (2007) VectorBase: a home for invertebrate vectors of human pathogens. *Nucleic Acids Res* **35** (Database issue): D503–505.

Levashina, E.A., Moita, L.F., Blandin, S., Vriend, G., Lagueux, M. and Kafatos, F.C. (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* **104**: 709–718.

Mendel, R.R. and Bittner, F. (2006) Cell biology of molybdenum. *Biochim Biophys Acta* **1763**: 621–635.

Moita, L.F., Wang-Sattler, R., Michel, K., Zimmermann, T., Blandin, S., Levashina, E.A., *et al*. (2005) In vivo identification of novel regulators and conserved pathways of phagocytosis in *A. gambiae*. *Immunity* **23**: 65–73.

Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z.J., *et al*. (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*.

Paton, M.G.. Karunaratne, S.H., Giakoumaki, E., Roberts, N. and Hemingway, J. (2000) Quantitative analysis of gene amplification in insecticide-resistant *Culex* mosquitoes. *Biochem J* **346** (Pt 1): 17–24.

Riehle, M.M., Markianos, K., Niare, O., Xu, J., Li, J., Toure, A.M., *et al*. (2006) Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region. *Science* **312**: 577–579.

Stevens, J.L., Snyder, M.J., Koener, J.F. and Feyereisen, R. (2000) Inducible P450s of the CYP9 family from larval *Manduca sexta* midgut. *Insect Biochem Mol Biol* **30**: 559–568.

Tang, H., Kambris, Z., Lemaitre, B. and Hashimoto, C. (2006) Two proteases defining a melanization cascade in the immune system of *Drosophila*. *J Biol Chem* **281**: 28097–28104.

TGSC (submitted) The genome sequence of the red flour beetle, *Tribolium castaneum*.

Vaughan, A., Hawkes, N. and Hemingway, J. (1997) Co-amplification explains linkage disequilibrium of two mosquito esterase genes in insecticide-resistant *Culex quinquefasciatus*. *Biochem J* **325** (Pt 2): 359–365.

Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z., Alvarez, K.S., Bartholomay, L.C., *et al*. (2007) Evolutionary dynamics of immune-related genes and pathways in disease vector mosquitoes. *Science* **316**(5832): 1738–1743.

Werck-Reichhart, D. and Feyereisen, R. (2000) Cytochromes P450: a success story. *Genome Biol* **1**: REVIEWS3003.

Wyder, S., Kriventseva, E.V., Schröder, R., Kadowaki, T. and Zdobnov, E.M. (2007) Quantification of ortholog losses in insects and vertebrates. *Genome Biol* (in press).

Yang, Y., Chen, S., Wu, S., Yue, L. and Wu, Y. (2006) Constitutive overexpression of multiple cytochrome P450 genes associated with pyrethroid resistance in *Helicoverpa armigera*. *J Econ Entomol* **99**: 1784–1789.

Zdobnov, E.M. and Bork, P. (2007) Quantification of insect genome divergence. *Trends Genet* **23**: 16–20.

Zdobnov, E.M. and Apweiler, R. (2001) InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**: 847–848.

Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R.R., *et al*. (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**: 149–159.