The influence of genomics on the control of malaria and other vector-borne diseases

Janet Hemingway*

The genomes of the malaria parasite *Plasmodium falciparum* and a major vector, *Anopheles gambiae,* have now been sequenced. The former is in a finished format while the latter still requires complete closure and annotation. The challenge now is to use this major resource, which is freely searchable by all scientists, to develop new effective control methods for diseases such as malaria. Once these methods have been developed, a further large challenge is their operational implementation, given the lack of success internationally of achieving the Abuja targets with impregnated bednets for malaria.

Introduction

Malaria is still one of the world's biggest killers. Why, 105 years after Ross discovered the way malaria was transmitted via the mosquito vector and a century after he received the Nobel prize for this seminal discovery, is malaria still such a big problem? This is undoubtedly in part due to the lack of resources that have been devoted to solving this problem at a scientific and practical implementation level.

Malaria eradication was suggested in the 1960s, when indoor residual spraying with DDT and chloroquine prophylaxis was a powerful combination for reducing malaria transmission (see Box 1). On the fringes of malaria transmission in Europe and in parts of South-East Asia, this campaign was a spectacular success, but in the African and Asian malaria heartlands it soon became clear that eradication with available tools, expertise, manpower and funding was impractical. Resistance in the parasites to the available drugs and in the mosquito vectors to DDT also reduced the efficacy of the available tools, and the euphoria of proposed eradication gave way to efforts to institute sustainable control of malaria.

The poverty of the countries where malaria transmission is most intense and where donors are unwilling to underpin open-ended disease control programmes, mean that it is crucial to allocate resources for malaria control to clearly defined priorities that are based on established evidence. Good leadership and political will are essential to implement evidence-based malaria control interventions on a national scale, but are often lacking.

In Abuja, Nigeria, in 2000, delegations from 44 African nations met in the largest ever heads-of-state summit focused on a single health issue. They pledged to take decisive steps towards halving the world's malaria burden by 2010, and to ensure that 60% of those affected have access to treatment, protection during pregnancy, and sleep under insecticide-treated nets (ITNs). These promises were made as the African leaders signed up to 'Roll Back Malaria' (RBM), a global partnership created in 1998 by the World Health Organization, the United Nations Development Programme, the United Nations Children's Fund and the World Bank. Despite such initiatives, there is little sign of delivery or adequate progress towards the Abuja goals.

*Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, U.K.
E-mail: hemingway@liverpool.ac.uk

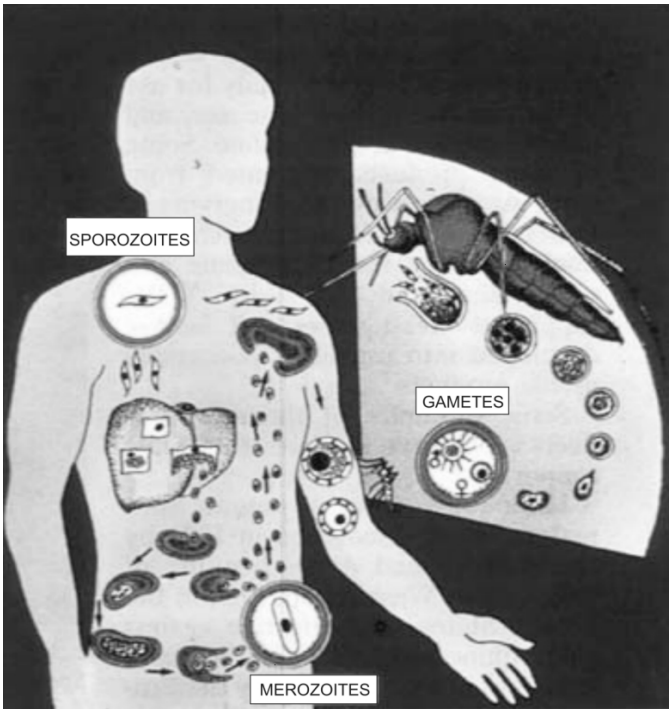

Fig. 1. The stages in the life cycle of the anopheline mosquito. Source: World Health Organization.

Major breakthroughs are now expected with the availability of the genome sequence of a number of human pathogens and their insect vectors. Since the publication of the whole genome sequence of *Haemophilus influenzae* in 1995, there has been a deluge of information based on sequencing pathogen genomes, including the recent release of the *Plasmodium falciparum* and *Anopheles gambiae* sequences. The rapid rise in the number of genomes sequenced has resulted from developments in both molecular and computational sciences, and from a realization of the impact of genome information on studying the biology of organisms. New opportunities to examine global changes in transcription and translation have also been derived from these

Box 1. Malaria — the disease

Malaria in humans is caused by the transmission of sporozoites from infected anopheline mosquitoes during feeding. The complex life cycle (see Fig. 1) results in the multiplication of parasites in the bloodstream following a 48-hour (for *P. falciparum*) cycle of invasion, replication and release. It is after the rupture of the infected red blood cells that the characteristic spike in fever is seen.

Mortality from *P. falciparum* malaria has been variably estimated at between 1 and 2 million deaths per year (mainly in children under five in sub-Saharan Africa), making it one of the big three infectious diseases in developing countries, the other two being tuberculosis and HIV/AIDS. In addition, although difficult to estimate,[26] the economic burden on the poorest countries is also substantial. Despite an overall reduction in childhood mortality in Africa over the last few decades,[27] the proportion of deaths attributable to malaria may have risen, due to increasing drug resistance and lack of effective vector control programmes.

studies, enabling a different type of approach to be added to the scientific armoury.

Despite the technological advances, sequencing a relatively small parasite or mosquito genome is not a trivial matter. Information is released either as a 'draft' or 'finished' sequence. Both are based on sequencing many random segments of the genome and using computers to assemble these pieces into the whole sequence. This phase is fast and produces a large amount of information, in a 'draft' format, about the genes present in a particular organism.

However, random sequencing leaves gaps in the sequence due to under-representation of segments in sequence libraries and difficulties in sequencing or assembling certain regions. Filling gaps is a major task and, unlike the highly automated random sequencing phase, requires input from skilled staff to identify and sequence these gaps. Having the 'finished' sequence of a genome allows a full picture of the genetic make-up of an organism to be developed, but at a cost.

Nevertheless, the release of a 'draft' sequence can make a substantial contribution to the research community. The *P. falciparum* genome sequence is 'finished' but 'draft' sequences have been available on the Web since the start of the project, providing investigators with valuable information. The *A. gambiae* sequence is in 'draft' form and will still take several years to finish, but is available in a robust searchable format.

Plasmodium falciparum genome information

The nuclear genome of *Plasmodium falciparum* parasite 3D7 is 22.8 mega-base pairs (Mb) (mitochondrial and apicoplast genomes are 6 and 35 kb, respectively). Of the 5268 genes predicted by various algorithms, half have introns. As well as genes coding for proteins, the nuclear genome contains 43 transfer RNAs (tRNAs), showing low redundancy. The mitochondrial genome has no tRNAs, whereas the apicoplast genome encodes enough tRNAs for local protein synthesis. The ribosomal RNA (rRNA) genes are relatively few in number and are spread throughout the genome, unlike the clustered organization found in other organisms. The developmental regulation of rRNA expression, divided into asexual (human host) and sexual (mosquito) stages, may relate to this organizational difference. No identifiable retrotransposons or transposable elements were found.

The fourteen chromosomes vary from 0.643 to 3.29 Mb in length and potential centromere regions have been identified for most chromosomes. The subtelomeric regions of all the chromosomes were highly similar, divided into five blocks of conserved tandem repeats or non-repeat regions. The highly variable multigene families *var, rif* and *stevor* are contained within the fourth block from the telomere. Strikingly, 60% of the *P. falciparum* predicted proteins are listed as 'hypothetical', with no homology to any sequenced functional protein to date. These proteins are apparently unique to *Plasmodium* (by comparison, approximately 50% of proteins in *Anopheles* share some homology with *Drosophila* proteins). This may reflect the evolutionary distance of *Plasmodium* from other available sequenced organisms, but it may also reflect a true difference in the biology of this organism.

Of the 40% of proteins with an assigned function,[1] *P. falciparum* has a much higher percentage of proteins involved in cell adhesion and invasion than *Saccharomyces cerevisiae,* as expected from the former's parasitic life cycle. The categories 'cell organization', 'cell cycle' and 'transcription factors' were reduced relative to *S. cerevisiae* but the lack of predicted proteins identified in these classes does not mean that these functions do not exist in *P. falciparum.*

Subtelomeric regions of the chromosomes contain several multigene families. *Var* genes (59 copies in 3D7) encode *P. falciparum* erythrocyte membrane protein 1, PfEMP1, a family of proteins involved in both host immune evasion and pathogenesis of malaria during the disease-causing asexual intra-erythrocytic stages. Rifins (repetitive interspersed family; encoded by 149 *rif* gene copies) have also been demonstrated on the surface of asexual-stage infected red blood cells, but 3D7 has relatively low *rif* expression.[2] 3D7 was chosen for the sequencing project because it can make gametocytes *in vitro,* whereas other *P. falciparum* lines cannot, particularly those established in laboratory culture for some time. Proteomic studies revealed groups of genes with coordinated regulation, some of which form clusters within the genome. Investigation of the non-coding regions associated with coordinately expressed genes may reveal regulatory sequences.

Comparative genomics

The provision of a 'finished' sequence for the *P. falciparum* genome provides a framework for related *Plasmodium* species genomes to be modelled on. Information from 'draft' sequencing of other *Plasmodium* species has also facilitated gene discovery in *P. falciparum.* The status of genome sequencing in all malaria species was reported recently.[3] Synteny, the conserved organization of similar genes within a chromosomal context in different species, occurs especially within species of rodent malaria parasites[4] and within the four species of human malaria parasites.[5] Extensive synteny facilitates assembly of sequences and identification of genes, and demonstrates evolutionary relationships.[5,6]

Predicted proteins from the genomes of rodent parasites *P. yoelii yoelii* and *P. chabaudi* (a partial set from a genome survey sequencing project) were compared with those of *P. falciparum.* In 5878 *P. y. yoelii* genes[7] and in 766 *P. chabaudi* genes,[8] roughly half were similar to *P. falciparum* predicted proteins. Identification of orthologues in biochemical pathways supports the use of rodent models to investigate mechanisms of drug activity. Additionally, discovering parallel pathways in other Apicomplexan parasites more amenable to experimentation, such as *Toxoplasma gondii,* may lead to a better understanding of *Plasmodium* species.

Therapeutics

An integrated map of metabolic pathways and of transporters allows researchers to identify properties unique to the parasite. Examples of recently identified drug targets with orthologues in *P. y. yoelii*[7] include the apicoplast-specific enzymes, 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DOXPR), inhibited by fosmidomycin (in clinical trials[9]) and enoyl-acyl carrier protein (ACP) reductase (FabI), inhibited by triclosan.[10] Genome information can also highlight potential difficulties, such as multiple proteins involved in specific pathways, and also contribute to the resolution of these problems (e.g. plasmepsins). The challenge now will be to use the genome information to ask the 'right' questions, efficiently, to optimize our understanding of this parasite and its interactions with the mosquito vector and human host, and to develop novel therapeutics or even novel vaccine candidates/strategies.

Anopheles gambiae genome information

The sequence of the nuclear genome of the PEST strain of *A. gambiae* consists of 278 Mb DNA[11] (the mitochondrial genome is 15.3 kb and contains 27 genes).[12] The sequence is currently in draft form; gaps and regions of potential misassembly are highlighted within the sequence, most notably on the largely non-coding, highly repetitive Y-chromosome. Yet despite this,

the immediate opportunities to develop urgently required novel methods of mosquito control afforded by this genome data-set fully justifies the publication of the genome in its draft format.

The PEST strain of *A. gambiae* is a hybrid strain derived from crossing a laboratory colony of Nigerian origin with the offspring of wild mosquitoes from Kenya.[13] This strain was chosen, in part, because it lacked the large-scale chromosomal rearrangements that are typical of many *A. gambiae* populations. During genome assembly, however, it became apparent that this strain possessed an unexpected amount of heterogeneity, posing problems for the assemblers, but at the same time providing a rich depository of single nucleotide polymorphisms (SNPs) that will be extremely valuable for population studies.[11] At 278 Mb, the *A. gambiae* genome is considerably larger than the 122 Mb assembled sequence of *Drosophila melanogaster,*[14] but smaller than the predicted size of many other mosquito disease vectors.[15] The difference in size between *A. gambiae* and *D. melanogaster* is largely due to intergenic DNA. Automatic annotation programs predicted a similar number of genes in both species (13 683 genes in *A. gambiae* versus 13 379 in *D. melanogaster*[11,16]).

Current malaria programmes cover both malaria prevention and treatment. Prevention of disease transmission is accomplished through control of the insect vectors at the population level and use of insecticide-impregnated nets and other materials to prevent mosquito biting, at the individual and household level. Malaria drug prophylaxis can be used to provide additional protection against malaria for groups at particular risk such as pregnant women living in, and travellers to, endemic countries.

Successful operational implementation of each of these malaria prevention strategies is subject to constraints, with problems building faster in some areas than others. For example, a central plank of RBM strategy is operational, large-scale use of ITNs. The only insecticide class that can currently be used on nets are the pyrethroids. In 1998, when large-scale ITN use was proposed, it was assumed that the major African vectors (*A. gambiae* and *A. funestus*) were fully susceptible to these insecticides. Since then, resistance networks, established through the Multilateral Initiative on Malaria (MIM) scheme, have demonstrated that serious resistance problems are present in both vectors. *A. gambiae* throughout much of West Africa has a high frequency of target site resistance (knockdown resistance, or kdr) while *A. funestus* has a metabolic mechanism, which produces high-level resistance to pyrethroids. We may have been lucky with *A. gambiae,* as the kdr mechanism is functionally recessive and the behaviour change it produces in the homozygous kdr mosquitoes result in only minor changes in ITN efficacy for reduction of malaria transmission.

The situation in southern African is more worrying because resistance of *A. funestus* has compromised the use of pyrethroids for both residual spraying and ITNs. An obvious fear is that this metabolic pyrethroid resistance mechanism will eventually be selected in *A. gambiae* and there is now evidence of this in parts of West Africa and in Kenya.

Effective management of insecticide resistance requires detailed knowledge of the mechanisms involved. The development of PCR-based assays for one of the main insecticide resistance mechanisms, kdr, has enabled resistant alleles to be detected at very low frequencies and has facilitated population-based studies to assess their impact on net efficacy.[17] But kdr is only one of several potential resistance mechanisms. An inventory of the three major enzyme families associated with metabolic resistance to insecticides has identified nearly 200 genes encoding glutathione transferases, cytochrome P450s or carboxylesterases.[18] The challenge now is to determine the specific members of these supergene families that are involved in insecticide detoxification. This task will be facilitated by the integration of the genome sequence information with genetic mapping data, which has already broadly defined the boundaries within which many of the major loci associated with metabolic resistance lie.[18]

The search for novel insecticidal targets will clearly be facilitated by the genome data-set. Thirty-five genes encoding regulatory peptides governing key physiological pathways in the mosquito have already been described.[19] Interfering with one or more of these pathways may give rise to effective future mosquito control strategies. The genome sequence availability may also allow us to develop novel synergists or 'resistance-breakers' rapidly to restore the efficacy of the pyrethroids and other insecticides. Targeting such molecules against the insect-specific regulators of the metabolic resistance mechanism should avoid the problems of increased human toxicity common to currently available synergists. Development and registration of such molecules, if they have little or no mammalian toxicity, should also be relatively rapid.

A. gambiae was selected for full genome sequencing from the 60 or so anopheline mosquito species that transmit malaria largely because of the vast number of deaths attributed to bites from this mosquito. It belongs to a group of sibling species known as the *A. gambiae* complex that differ in their ecology and ability to transmit disease. The efficiency of *A. gambiae sensu stricto* as a vector of malaria is due to its remarkable adaptation to living in close association with humans and its extreme preference for feeding on human blood. What is it about humans that this mosquito species finds so attractive? The identification of 79 *A. gambiae* genes encoding putative odorant receptors[20] may help answer this question. If, as predicted, members of this supergene family recognize chemical signals exuded exclusively by humans, a new generation of attractants or repellents acting as decoys to these odorant receptors may prove effective in reducing human malaria transmission.

Genetic engineering

The application of genetic engineering to mosquito control has received much attention recently, and the availability of the *A. gambiae* genome sequence will undoubtedly accelerate the development of laboratory strains that are unable to transmit pathogens. A description of the 242 *A. gambiae* genes implicated in innate immunity responses accompanied the genome release, and this will facilitate delineation of the mosquitoes' response to pathogens,[21] whereas analysis of global changes in gene expression following blood meal or infection will assist the search for additional effector molecules.[11] Further studies of the 40 different classes of transposable elements present in the mosquito genome[11] may identify improved means of driving parasite-resistant genes through natural populations. While this is all good news for the development of transgenic mosquito populations in the laboratory, the biggest challenges to the success of this strategy will manifest themselves during field trials. The challenge to proponents of this scheme is to use the genome sequence to complement field studies of disease transmission ecology.

Other potential uses of modern genomic and transgenic technology may provide novel solutions to malaria control in the longer term. Not all mosquito species are able to transmit the malaria parasite. The genes that encode this refractory phenotype are now being characterized. These genes, or single-chain variant fragment antibodies against the parasite, could be stably inserted into the genomes of major vectors to block transmission. While such transgenic technology was close to science fiction a decade ago, we now have all the available technology to

produce refractory mosquitoes and proof of principle in both *A. gambiae* and *Aedes aegypti* (the latter with bird malaria) that the technology will work. The major hurdles in introducing this new technology are to develop methods of rapid mosquito field population replacement and to gain acceptance of the technology by the human population living in the regions where malaria is endemic.

Web-based resources

The entire genome assembly of *Anopheles* is available through Genbank or in CD-ROM format (freely available from *Science*). Both the Genbank and Ensembl sites can be used to search for particular sequences of interest, either by gene name or by BLAST searching a particular sequence, although the caveats outlined above for the parasite genome annotation should be noted. Manual inspection and ultimately experimental verification are needed to confirm or rectify many of the automated annotations.

Several large-scale *A. gambiae* EST sequencing projects have been completed that could assist in this process. A useful resource that amalgamates cDNA sequence data from various sources with information from automated annotation pipelines can be found within The Institute for Genomic Research (TIGR) web site (see Box 2). The data from the Malaria Genome Sequencing Project is readily available to anyone with a computer to ask questions about the parasite genome.

The Internet is the best way to access the data (see Box 2 for details), but for those who are 'out in the field', many of the data are available on the GenePlot CD (available from PlasmoDB). Specific questions, such as 'is my favourite protein expressed at a particular stage of development?' can be asked easily using a simple BLAST or text search. There are a few considerations for using the information. Much of the gene data is for predicted proteins, some of which have homology to proteins with known functions in other organisms. However, many of the 'genes' in the database are 'gene models' based on computer algorithms and so the functions of these proteins in the parasite, and even the correctly spliced coding sequence, need to be verified experimentally. Conversely, gene prediction algorithms may have 'missed' genes of interest, so it pays to follow up any 'hits', even if a full gene does not come out of a search. Despite the large size of the database, it is still not comprehensive, as the sequenced genome represents only one of countless *P. falciparum* genotypes, which differ from one another due to deletions and duplications. Thus, the sequenced laboratory line 3D7 itself has sequence and expression features that distinguish it from other *P. falciparum* genotypes.

Comparative genomics

Comparative analysis of the mosquito proteome with that of *Drosophila* reveals an unexpected amount of evolutionary relatedness. Almost half of the genes in both *Anopheles* and *Drosophila* have one clear orthologue in the other species. A further 40% have good matches to proteins in *Drosophila* (32%) or another species (8%) but are not necessarily orthologous. Finally, only just over 10% of the putative *Anopheles* proteins have no detectable similarity to proteins in any other sequenced genome.[22]

This type of comparative analysis enables gene families that are specific to, or absent from, insects to be identified, information that will be valuable for the development of novel classes of insecticides. In addition, the identification of classes of proteins that are over-represented in the mosquito compared to the fly may provide clues to the biochemical pathways involved in seeking and processing blood meals.

Box 2. Web-based resources: web sites

PlasmoDB: the official database of the malaria parasite genome project.
http://www.plasmodb.org/

National Institutes of Health malaria and mosquito genome project information
http://www.ncbi.nlm.nih.gov/projects/Malaria/
http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=agambiae.inf

Ensembl mosquito genome assembly and annotation
http://www.ensembl.org/Anopheles_gambiae/

The Institute of Genome Research gene index site
http://www.tigr.org/tdb/tgi/

MR4 site
http://www.malaria.mr4.org/index.html

South African National Bioinformatics Institute BLAST searcher comparing *P. falciparum*, *P. vivax* and *P. berghei* GSS and EST
http://www.sanbi.ac.za/malaria-genesearch/

For general molecular biology of malaria parasites
http://www.wehi.edu.au/MalDB-www/who.html

For general information about *Anopheles* mosquitoes
http://www.anobase.org/

***P. berghei* information**
http://www.lumc.nl/1040/research/malaria/genomics_proteomics.html

BLAST against many different parasite genomes
http://www.ebi.ac.uk/blast2/parasites.html
http://www.sanger.ac.uk/Projects/Protozoa/

GO: Gene Ontology database: controlled vocabulary for genes and functions in all sequenced organisms.
http://www.geneontology.org/

If successful global mosquito control strategies are to be developed, they must tackle other malaria vectors in addition to *A. gambiae.* The ability to use the *A. gambiae* genome as a framework to assemble partial genome sequence data obtained from other anopheline mosquito species depends on the degree of conservation of gene order (microsynteny) between different species. A comparative genomic study between *A. gambiae* and *A. funestus,* which occur sympatrically across much of sub-Saharan Africa and are thought to have diverged approximately 5 million years ago, found extensive synteny at the level of chromosome arms but widespread variation in local gene order.[23] Although problematic for the transfer of functional genomics studies between species, further comparative studies of these two genomes are likely to provide valuable insights into the mechanism and effects of chromosomal inversions and rearrangements within *Anopheles* species.[23]

Chromosomal inversions on chromosome 2R are widespread within the *A. gambiae* complex and are believed to indicate adaptations to different environmental niches.[24] Identification of genes encoded within these inversions may provide clues to factors determining mosquito behaviour and vectorial capacity.[25]

After a decade or more of neglect and under-funding, malaria is now firmly back on the political and health agendas. New alliances and funding initiatives, such as RBM, the Global Fund for HIV/AIDS, TB and Malaria, Medicines for Malaria, and the Multi-lateral Initiative on Malaria, should provide the impetus for striving to achieve the Abuja targets and support development of new tools and techniques for malaria control and treatment. These initiatives also provide a much-needed opportunity to train the next generation of African malaria researchers, technicians and operational control specialists, and should start to redress the loss of capacity in this area, which has been apparent in the last decade.

1. Ashburner M. *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29.
2. Smith J.D. *et al.* (1998). Identification of a *Plasmodium falciparum* intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. *Proc. Natl Acad. Sci. USA* **97**, 1766–1771.
3. Carlton J.M. and Carucci D.J. (2002). Rodent models of malaria in the genomics era. *Trends Parasitol.* **18**, 100–102.
4. Thompson J. *et al.* (2001). Comparative genomics in *Plasmodium*: a tool for the identification of genes and functional analysis. *Mol. Biochem. Parasitol. Sci.* **118**, 147–154
5. Carlton J.M. *et al.* (1999). Karyotype and synteny among chromosomes of all four species of human malaria parasite *Mol. Biochem. Parasitol. Sci.* **101**, 23–32.
6. Tschavtchitch M. *et al.* (2001). The sequence of a 200 kb portion of a *P vivax* chromosome reveals a high degree of conservation with *P. falciparum* chr 3. *Mol. Biochem. Parasitol. Sci.* **118**, 211–222.
7. Carlton J.M. *et al.* (2002). Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii. Nature* **419**, 512–519.
8. Janssen C.S. *et al.* (2001). Gene discovery in *P. chabaudi* by genome survey sequencing. *Mol. Biochem. Parasitol. Sci.* **113**, 251–260.
9. Missinou M. *et al.* (2002). Fosmidomycin for malaria. *Lancet* **360**, 1941–1942.
10. Surolia A. and Surolia N. (2001). Triclosan offers protection against blood stages of malaria. *Nature Med.* **7**,167–173.
11. Holt R.A. *et al.* (2002). The genome sequence of the malaria mosquito *Anopheles gambiae. Science* **298**, 129–149.
12. Beard C.B *et al.* (1993). The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects *Insect Mol. Biol.* **2**, 103–124.
13. Mukabayire O. and Besansky N.J. (1996). Distribution of T1, Q, Pegasus and mariner transposable elements on the polytene chromosomes of PEST, a standard strain of *Anopheles gambiae. Chromosoma* **104**, 585–595.
14. Adams M.D. *et al.* (2000). The genome sequence of *Drosophila melanogaster. Science* **287**, 2185–2195.
15. Knudson D.L. *et al.* (1996). Genome organization of vectors. In *The Biology of Disease Vectors*, eds B.J. Beatty and W.C. Marquardt, chap. 13. University Press of Colorado, Denver.
16. Misra S. *et al.* (2002). Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* **3**, 83.1–83.22.
17. Matinez-Torres A. *et al.* (1998). Molecular characterisation of pyrethroid knock-down resistance (kdr) in the major malaria vector *Anopheles gambiae. Insect Mol. Biol.* **7**, 179–184.
18. Ranson H. *et al.* (2002). Evolution of supergene families associated with insecticide resistance. *Science* **298**, 179–181.
19. Riehle M.A. *et al.* (2002). Neuropeptides and peptide hormones in *Anopheles gambiae. Science* **298**, 172–175.
20. Hill C.A. *et al.* (2002). G Protein-coupled receptors in *Anopheles gambiae. Science* **298**, 176–178.
21. Christophides G.K. *et al.* (2002). Immunity-related genes and gene families in *Anopheles gambiae. Science* **298**, 159–165.
22. Zdobnov E.M. *et al.* (2002). Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster. Science* **298**, 149–159.
23. Sharakhov I.V. *et al.* (2002). Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus. Science* **298**, 182–185.
24. Coluzzi A. *et al.* (2002). A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**, 1415–1418.
25. della Torre A. *et al.* (2002). Speciation within *Anopheles gambiae* — the glass is half full. *Science* **298**, 115–117.