Malaria parasite and vector genomes: partners in crime

Alister Craig[1], Susan Kyes[2], Hilary Ranson[1] and Janet Hemingway[1]

[1]Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK
[2]Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK

The publication of the genome sequences of the malaria parasite *Plasmodium falciparum* and the insect vector *Anopheles gambiae* paves the way for scientists to study these organisms by using technologies developed to observe global changes in transcription and translation, as well as computational tools. Researchers are now able to investigate complex changes involved in development, growth and reaction to external factors. Given the medical importance of these organisms, much of this work is targeted on drug or insecticide discovery (including mechanisms of resistance to existing treatments), but the genome information also provides the opportunity to develop novel therapies.

Despite the technological advances, sequencing a much smaller genome than that of a mammal [1] is still not a trivial matter. Several centers have developed complex programs that can take on these tasks efficiently. One of the major differences in the type of information that has been released is between 'draft' and 'finished' sequences. Both are based on sequencing many random segments of the genome and using computers to assemble these pieces into a whole sequence. This phase can be very fast and can produce a large amount of information about the genes present in a particular organism in a draft format. 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 can be 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, and the (pre)release of a draft sequence is a major resource for the research community. The *Plasmodium falciparum* genome sequence is finished [2,3], but draft sequences have been available on the web from an early phase of the project. The *Anopheles gambiae* sequence is in draft form [4], but had the scaffold of the related insect *Drosophila melanogaster* to assist in creating a useable framework. Completing the *An. gambiae* sequence will depend on the longer-term efforts of the research community because there is, at present, no budget for this.

Corresponding author: Alister Craig (agcraig@liv.ac.uk).

For malaria researchers, the combination of draft genomes for the insect and human hosts, and the complete sequence of the parasite provides a major impetus in the holistic investigation of the biology of this system and the development of new treatments to overcome this terrible disease (Box 1).

The parasite genome

The nuclear genome of *P. falciparum* parasite clone 3D7 is 22.8 Mb (the mitochondrial and apicoplast genomes are 6 kb and 35 kb, respectively). 3D7 was chosen for the sequencing project because it can make gametocytes *in vitro*, whereas some other *P. falciparum* lines cannot, particularly those established in laboratory culture for some time. The gene identification process was adjusted specifically for this organism and, of the 5268 genes predicted by various algorithms, half have introns (Table 1). Expressed sequence tag (EST) and proteomic analyses detected 70% of the predicted genes [2,5,6].

Apart from the genes that encode proteins, the nuclear genome has 43 transfer RNAs (tRNAs). This relatively small number gives low redundancy, with only a limited repertoire of codon usage for each amino acid. The mitochondrial genome has no tRNAs, whereas the apicoplast genome encodes enough tRNAs for local protein synthesis, although many of the proteins within the apicoplast are encoded by nuclear genes. The ribosomal RNA (rRNA) genes are 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, might relate to this organizational difference. No identifiable retrotransposons or transposable elements were found.

The 14 chromosomes vary from 0.643 Mb to 3.29 Mb in length, and potential centromeric regions have been identified for most chromosomes. The subtelomeric regions of all the chromosomes were very similar, divided into five blocks of conserved tandem repeats or non-repeat regions. The highly variable multigene families, *var*, *rif* (repetitive interspersed family) and *stevor* (subtelomeric variant open reading frame), are contained within the fourth block from the telomere.

Web-based resources

The data from the Malaria Genome Sequencing Project is available to anyone with internet access (Box 2). Much of

 1471-4922/03/$ - see front matter doi:10.1016/S1471-4922(03)00146-6

Box 1. Malaria: the disease

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

Mortality from *P. falciparum* malaria has been estimated at between one and two 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 [47], the economic burden on the poorest countries is also substantial. Despite an overall reduction in childhood mortality in Africa over the past few decades [48], the proportion of deaths attributable to malaria might have actually risen due to increasing drug resistance and lack of effective vector control programs. Fig I was modified from Ref. [49].

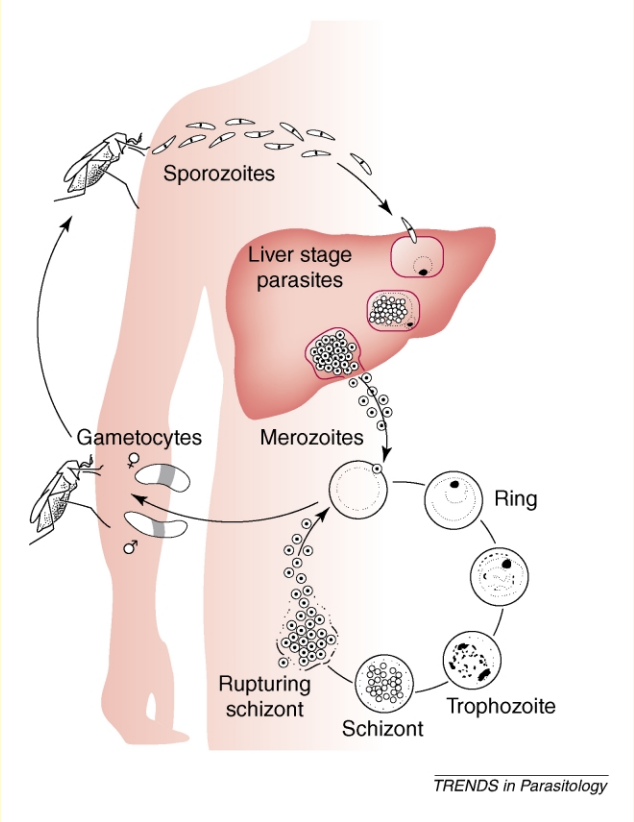

Fig. I.

Box 2. Websites of interest

(a) PlasmoDB: the official database of the malaria parasite genome project. http://www.plasmodb.org/
(b) National Institutes of Health (GenBank) 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
(c) Ensembl mosquito genome assembly and annotation (also includes rat, mouse, human, fugu and zebrafish) http://www.ensembl.org/Anopheles_gambiae/
(d) The Institute of Genome Research gene index site http://www.tigr.org/tdb/tgi/
(e) AnoBase: a database containing genomic/biological information on anopheline mosquitoes with emphasis on *Anopheles gambiae* http://www.anobase.org/AnoDB
(f) European Bioinformatics Institute (EBI) sites (for comparisons with other parasites) http://www.ebi.ac.uk/blast2/parasites.html http://www.ebi.ac.uk/parasites/paratable.html
(g) MR4 site (malaria repository containing many physical resources for malaria research) http://www.malaria.mr4.org/index.html
(h) South African National Bioinformatics Institute BLAST searcher comparing *Plasmodium falciparum*, *Plasmodium vivax* and *Plasmodium berghei* genome survey sequences (GSS) and expressed sequence tags (EST). http://www.sanbi.ac.za/malaria-genesearch/
(i) For general molecular biology of malaria parasites http://www.wehi.edu.au/MalDB-www/who.html
(j) For general information on *P berghei* http://www.lumc.nl/1040/research/malaria/genomics_proteomics.html
(k) BLAST against many different parasite genomes http://www.sanger.ac.uk/Projects/Protozoa/
(l) GO: Gene Ontology database: controlled vocabulary for genes and functions in all sequenced organisms. http://www.geneontology.org/

the data are also available on the GenePlot CD (available from PlasmoDB, Box 2a [7]). Specific questions, such as: 'is my favorite protein expressed at a particular stage of development?' can be asked using a simple BLAST or text search.

There are a few considerations for using genome information. Much of the gene data are for predicted proteins, some of which have homology to proteins with known functions in other organisms. However, it should be noted that many of the 'genes' in the database are actually '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 might have 'missed' your gene of interest; hence, 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 cannot be comprehensive because of the very nature of the parasite. The sequenced 3D7 genome represents one of countless naturally occurring *P. falciparum* genotypes, which differ from 3D7 by several deletions, duplications and polymorphisms.

What did we learn?

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, ~50% of proteins in *Anopheles* share some homology with *Drosophila* proteins). This could reflect the evolutionary distance of *Plasmodium* from other available sequenced organisms, but it might also reflect a true difference in the biology of this organism.

Of the 40% of proteins with an assigned function (gene ontology; GO [8]), a comparison with *Saccharomyces cerevisiae* shows that 1.3% of proteins in *P. falciparum* are involved in cell adhesion and invasion, which is to be expected from the parasite's life cycle (Fig I in Box 1). The categories cell organization, cell cycle and transcription factors were reduced in *P. falciparum* compared with *S. cerevisiae*. The lack of predicted proteins identified in these classes does not mean that these functions do not

Table 1. Comparison of genome features between malaria parasites of humans and rodents[a]

	Plasmodium falciparum	*Plasmodium yoelii yoelii*	*Plasmodium vivax*	*Anopheles gambiae*
Mammalian host	Human	Rodent	Human	NA
No. of nuclear chromosomes	14	14	14	3
Nuclear genome size (Mbp)	22.3	23.1	ND	278
Overall GC content (%)	19.4	22.6	46	35.2
No. of predicted genes	5268	5878	ND	13 683
Genome coding regions (%)	52.6	ND	ND	7
Mean length of genes (kb)	2.3	1.3	ND	4.54
GC content in coding sequence (%)	23.7	24.8	47	ND
Variant parasite multigene families, mainly subtelomeric[b] (no of copies)	*var* (59); *rif* (149); *stevor* (28)	*yir* (700)	*Vir* (600–1000)	NA
Refs	[2,3]	[3]	[44]	[4]

[a]Abbreviations: NA, not applicable; ND, not determined; *rif*, repetitive interspersed family; *stevor*, subtelomeric variant open reading frame.
[b]The large multigene families form non-overlapping subsets of *Plasmodium*-specific coding sequence. In *P. falciparum*, there are a total of ~200 *var* and *rif* gene copies; no similar proteins have been found in any other *Plasmodium* spp. *yir* and *vir* are homologues, also similar in gene structure and size to *P. chabaudi cir* [45] and to *P. berghei bir* genes [44], and are thought to be the *var* equivalents for these species. *Sicavar* in *Plasmodium knowlesi* is another large gene family that encodes variant surface antigens [46], but has no homology to the *var* genes in *P. falciparum*.

exist in *P. falciparum*, but is probably due to little being known about the pathways involved.

In addition to the sequence data, information about the proteins (often termed proteomics) involved in various stages of the parasite life cycle has also been published [5,6]. The relationship between the genome sequence and the technologies that are able to measure peptide masses to a high degree of accuracy (mass spectrometry) is an important one. Databases based on predicted peptide masses (based on translation of the gene models) can be interrogated using spectra derived from mass spectrometry, allowing the researcher to identify specific proteins from complex mixtures [9]. Using this information, several metabolic pathways in *P. falciparum* appear to be stage-specific. Asexual stages generate energy through glycolysis and conversion of pyruvate to lactate, whereas sporozoite and gametocyte stages express enzymes for the mitochondrial tricarboxylic acid (TCA) cycle and oxidative phosphorylation. Some genes associated with standard mitochondrial function (two subunits of ATP synthase and some components involved in creating an electrochemical gradient) have not been identified, suggesting specific use of components of the TCA cycle within erythrocytic stages.

The apicoplast is an organelle unique to the phylum Apicomplexa, with a role in anabolic synthesis of fatty acids and isoprenoids, in addition to heme and/or iron regulation. The function of this organelle is not clear, but it clearly plays an important role evidenced in part by the discovery that ~10% of the nuclear-encoded proteins have a targeting signal for the apicoplast [10]. Metabolic pathways unique to the apicoplast are obvious targets for drug therapy.

Subtelomeric regions of the chromosomes contain several multigene families. *Var* genes (59 copies in clone 3D7) encode *P. falciparum* erythrocyte membrane protein (PfEMP)1, a family of proteins involved in both host immune evasion and malaria pathogenesis during the disease-causing asexual intra-erythrocytic stages [11]. Rifins (encoded by 149 *rif* gene copies) have also been demonstrated on the surface of asexual-stage infected red blood cells, but 3D7 was poor at *rif* expression possibly due to some defects in transport pathways [12]. *stevor* is thought to be expressed in gametocytes [13] and asexual forms, where a subset of the members of this gene family are expressed for a short time and are found within the Maurer's clefts [14]. Interestingly, the proteomics studies confirmed the poor expression of rifin in the trophozoite stages, but showed that, unexpectedly, *var*, *rif* and *stevor* protein products are all highly expressed in 3D7 sporozoites [5]. Why the parasite should express these variant proteins at this stage of the life cycle raises many questions, such as what selective advantage can expressing nearly the entire variant gene family repertoire give to sporozoites?

The proteomics studies also revealed groups of genes with coordinated regulation, some of which are also clustered together within the genome. Investigation of the non-coding regions associated with coordinately expressed genes might reveal regulatory sequences. Also, using stage-specificity and examples of genes with known function within a cluster of coordinately expressed genes, functions for hypothetical proteins could be proposed based on the hypothesis that genes involved in specific pathways can undergo coordinated regulation.

Comparative genomics among *Plasmodium* spp

The finished sequence for the *P. falciparum* genome provides a framework for related *Plasmodium* spp. genomes to be modeled upon. However, this process is not as one-sided as it might seem and information from draft sequencing of other *Plasmodium* spp. has facilitated gene discovery in *P. falciparum*. The status of genome sequencing in all *Plasmodium* spp. was reported recently [15]. Synteny, the conserved organization of similar genes within a chromosomal context in different species, occurs within species of rodent malaria parasites [16] and within the four species of human malaria parasites [17]. Extensive synteny facilitates assembly of sequences and identification of genes, and demonstrates evolutionary relationships [17,18]. Although exons tend to be more conserved than intergenic sequences, in comparisons of *Plasmodium* spp., conserved regulatory sequences might be evident within syntenic regions containing coordinately expressed genes.

Predicted proteins from the genomes of rodent parasites *Plasmodium yoelii yoelii* and *Plasmodium chabaudi*

(a partial set from a genome survey sequencing project) were compared with those of *P. falciparum*. Out of 5878 *P. y. yoelii* genes [3] and 766 *P. chabaudi* genes [19], ~50% 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. In addition, discovering parallel pathways in other Apicomplexan parasites more amenable to experimentation, such as *Toxoplasma gondii* [20], can lead to a better understanding of *Plasmodium* spp.

Therapeutics

An integrated map of metabolic pathways and transporters allows researchers to identify properties unique to the parasite. Examples of recently identified drug targets with orthologues in *P. y. yoelii* [3] include: (1) the apicoplast-specific enzymes, 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DOXPR), inhibited by fosmidomycin (in clinical trials [21]); (2) enoyl-acyl carrier protein (ACP) reductase (FabI), inhibited by triclosan [22]; and (3) the cytosolic enzyme protein farnesyl transferase (FTase), inhibited by parasite-specific peptidomimetics [23]. Genome information can also highlight potential difficulties, such as multiple proteins involved in specific pathways, and can also contribute to the resolution of these problems (e.g. plasmepsins [24,25]). The challenge now will be to use the genome information to ask the right questions to optimize our understanding of this parasite and its interactions with the mosquito vector and human host, and to develop novel therapeutics, novel vaccine candidates or strategies.

The vector genome

Anopheles gambiae s.s. was selected for full genome sequencing from the 60 or so anopheline mosquito species that transmit malaria, largely because of the vast number of malaria deaths attributed to bites from this mosquito.

Anopheles gambiae s.s. is a member of the closely related *An. gambiae* complex of species, some chromosomal variants of which are apparently in the process of incipient speciation within sub-Saharan Africa [26]. The sequence of the nuclear genome of the PEST strain of *An. gambiae* s.s. comprises 278 Mbp [4] (the mitochondrial genome is 15.3 kb and contains 27 genes [27]). The gaps and regions of potential mis-assembly are highlighted within the draft sequence, most notably on the largely non-coding, highly repetitive Y chromosome. Despite these deficiencies, the immediate opportunities to develop urgently required novel methods of mosquito control afforded by this genome dataset fully justifies the publication of the genome in its draft format.

The PEST strain of *An. gambiae* is a hybrid strain derived from crossing a laboratory colony originating from Nigeria with the offspring of wild-caught mosquitoes from Kenya [28]. This strain was chosen partly because it lacked the large-scale chromosomal rearrangements that are typical of many *An. gambiae* populations. However, during genome assembly, 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 valuable for population studies [4].

At 278 Mb, the *An. gambiae* genome is considerably larger than the 122 Mbp assembled sequence (or its estimated size of 180 Mb) of *Drosophila melanogaster* [29], but an order of magnitude smaller than the predicted size of many other mosquito disease vectors such as *Aedes aegypti* [30]. The difference in size between *An. 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 *An. gambiae* versus 13 379 in *D. melanogaster* [4,31]). In addition to the protein-encoding genes, 1506 putative transposable elements were identified within the *An. gambiae* genome.

Web-based resources

The entire genome assembly is available through GenBank or on CD-ROM format (freely available from *Science* [4]). Both the GenBank (Box 2b) and Ensembl (Box 2c) 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 *An. gambiae* EST projects have been completed that will assist in this process. A useful resource that amalgamates complementary DNA (cDNA) sequence data from various sources with information from automated annotation pipelines can be found within The Institute for Genomic Research (TIGR) website (Box 2d). A database, AnoBase (Box 2e), incorporating genomic and biological information, and a comprehensive list of related published literature is under development (other websites of interest are included in Box 2f–l).

What did we learn?

The sibling species within the *An. gambiae* complex differ in their ecology and ability to transmit disease. The efficiency of *An. gambiae* s.s. as a vector of malaria is due to its remarkable adaptation to living in close association with humans and its preference for feeding on them. What is it about humans that this mosquito species finds so attractive? The identification of 79 *An. gambiae* genes encoding putative odorant receptors [32] might 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 might prove highly effective in reducing human malaria transmission.

Present mosquito control programs rely on the use of a limited number of chemical insecticides. Within Africa, there is widespread support for the use of pyrethroid-impregnated bednets in malaria control. However, insecticide resistance has jeopardized malaria control efforts in the past and there are fears that the emergence of pyrethroid resistance in *Anopheles* vectors will reduce the efficacy of current control measures. There is clearly a

need to prolong the effectiveness of existing control measures while, at the same time, searching for novel insecticidal targets. Mining of the *An. gambiae* genome has produced leads on both these fronts.

Effective management of insecticide resistance requires detailed knowledge of the mechanisms involved. The development of PCR-based assays for one of the major pyrethroid resistance mechanisms, knockdown resistance (*kdr*), has enabled alleles responsible for resistance to be detected at very low frequencies and has facilitated population-based studies to assess their impact on bednet efficacy [33]. 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-*S*-transferases, cytochrome P450s or carboxylesterases [34]. The challenge now is to determine the specific members of these supergene families that are involved in insecticide detoxification. This task will be helped by the integration of the genome sequence data with genetic mapping data that has already broadly defined the boundaries within which many of the major loci associated with metabolic resistance lie [34].

Thirty-five genes encoding regulatory peptides governing key physiological pathways in the mosquito have already been described [35]. Interfering with one or more of these pathways could also give rise to effective mosquito control strategies.

The application of genetic engineering to mosquito control has received much attention recently and the availability of the *An. gambiae* genome sequence will undoubtedly accelerate the engineering of laboratory strains that are unable to transmit pathogens. A description of the 242 *An. gambiae* genes implicated in innate immune responses accompanied the genome release; this will help to delineate the mosquito's response to pathogens [36]. In addition, an analysis of global changes in gene expression following a blood meal or parasite infection will assist in the search for additional effector molecules [4]. Further studies of the 40 different classes of transposable elements present in the mosquito genome [4] could identify improved means of driving parasite-resistant genes through natural populations. While this is all good news for the development of transgenic mosquito populations as powerful laboratory research tools, the biggest challenges to the success of this strategy for practical disease control will manifest themselves during field trials and operational implementation. The challenge to proponents of this scheme is to use the genome sequence to complement field studies of disease transmission ecology. The communities into which these genetically modified mosquitoes might ultimately be released also need to be fully engaged with, and accepting of, such novel technologically driven disease control interventions.

Comparative genomics among vectors

Comparative analysis of the mosquito proteome with that of *Drosophila* reveals a large amount of evolutionary relatedness, which is not surprising given that both insects belong to the same order *Diptera*. Almost 50% of the genes in both *Anopheles* and *Drosophila* have one clear orthologue in other species. A further 40% have good matches to proteins in *Drosophila* (32%) or another species (8%), but are not necessarily orthologues. Only ~10% of the putative *Anopheles* proteins have no detectable similarity to proteins in any other sequenced genome [37]. This type of comparative analysis enables gene families that are specific to, or absent from, insects to be identified, which will be valuable for the development of novel classes of insecticides. In addition, the identification of protein classes that are over-represented in the mosquito when compared with *Drosophila* could provide clues to the biochemical pathways involved in the seeking and processing of blood meals.

If successful global mosquito control strategies are to be developed, they must tackle other malaria vectors, in addition to *An. gambiae*. The ability to use the *An. 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 *An. gambiae* and *Anopheles funestus*, which occur sympatrically across much of sub-Saharan Africa and are thought to have diverged ~5 million years ago, found extensive synteny at the level of chromosome arms, but widespread variation in local gene order [38]. Although problematic for the transfer of functional genomics studies between species, further comparative studies of these two genomes will probably provide valuable insights into the mechanism and effects of chromosomal inversions and rearrangements within *Anopheles* spp. [38].

Chromosomal inversions on chromosome 2R are widespread within the *An. gambiae* complex and are believed to indicate adaptations to different environmental niches [39]. Identification of genes encoded within these inversions could provide clues to factors determining mosquito behavior and vectorial capacity [40].

Interactions

While often treated as individual subjects, the biology of malaria disease is based on interactions between the parasite and its human and insect hosts (Fig. 1), and all three have been subject to co-evolution. Some examples of this include the progression to severe disease in *P. falciparum* malaria due to cytoadherence and the processes contributing to oxidative stress in the mosquito. For cytoadherence, much is known about the primary processes involved in adhesion, but the actual mechanisms causing disease are not known [41]. In oxidative stress in the mosquito, specific sets of genes have been identified as being activated during malaria infection, but the significance of this has yet to be elucidated [42]. The genome data will allow us to study both sides of the equation in these relationships by looking at changes at the transcriptional and protein level associated with infection. Even the relationship between humans and the anopheline mosquitoes that feed on them is open to investigation, promising new insights into the way in which insects locate, feed and co-exist with human populations.

These interactions are an integral part of the parasite's life cycle, but also render it vulnerable at several stages.

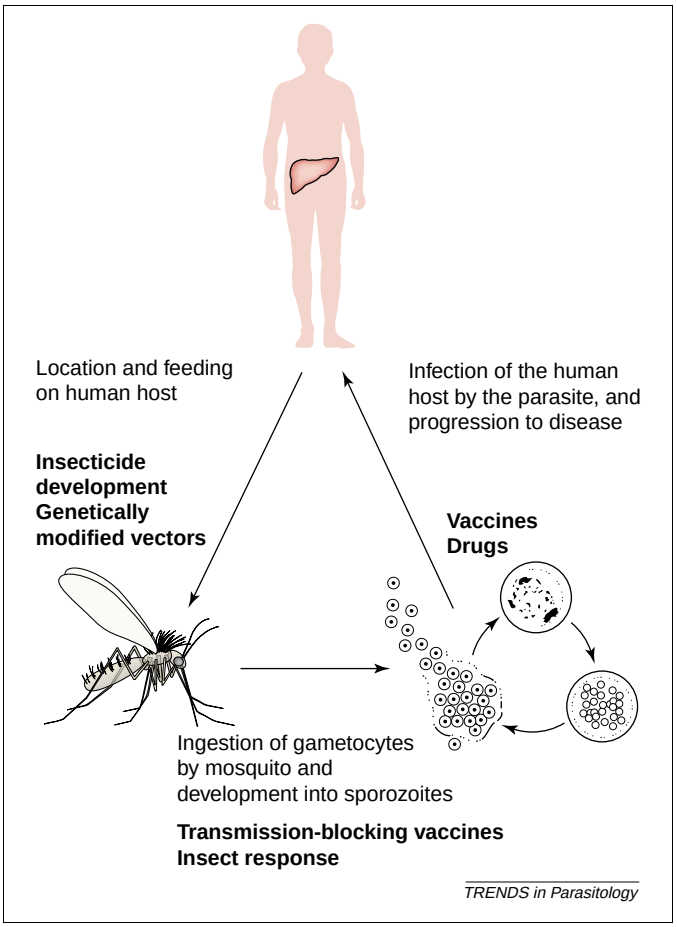

Fig. 1. Interactions and therapeutics based on genome resources. The dependence of the malaria parasite on two hosts, the human and the mosquito, to complete its transmission cycle presents a range of opportunities to develop countermeasures based on parasite-specific pathways that have been identified from genome sequencing information.

A realistic attack on malaria will require interventions at different stages running in parallel to reduce transmission in hyper-endemic regions to levels that lead to significant reductions in morbidity and mortality.

The future

Perhaps the greatest testament to the impact of the genome sequencing in malaria is that, despite the short time in which the information has been available, new interventions are already being developed. However, given the long lag between laboratory research and the appearance of therapeutics in the field, this should be no reason for complacency.

Genome sequence information has occasionally been criticized as lacking biological significance. By itself, data concerning the gene and/or DNA content and genome composition can have limited appeal. However, the definition of functional pathways using bioinformatics tools (e.g. BLAST searches), microarrays, proteomics and transfection could have a potentially dramatic impact on exploring biology of the parasite and vector when carried out in association with a strong research community. For pathogens, the further challenge will be the translation of this knowledge into alleviation of disease. For malaria, in addition to new drugs based on parasite-specific pathways, more information on the basis of resistance to existing drugs will facilitate the development of co-treatments to reduce or even reverse drug resistance, rescuing previously highly successful treatment regimens for future use (e.g. chloroquine and the *P. falciparum* chloroquine-resistance transporter gene [43]). Similarly, compounds aimed at the control of the insect vector or reversal of resistance to current insecticides will reduce the malaria burden. Finally, although more controversial, genetic manipulation of both the parasite and the insect vector to create either attenuated strains or insects refractory to parasite invasion could also produce therapeutic benefits.

The malaria parasite has evolved over millennia to evade or subvert host defences to infection. The genome sequence information provides an opportunity to create a fundamental leap in antimalarial treatment, tipping the balance in favor of the human host.

Acknowledgements

We acknowledge all the members of the sequencing teams who contributed to producing the genome sequence information and to the funding agencies who provided the resources and occasionally acted as referees.

References

1 Fleischmann, R.D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512
2 Gardner, M.J. *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511
3 Carlton, J.M. *et al.* (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 419, 512–519
4 Holt, R.A. *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129–149
5 Florens, L. *et al.* (2002) A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 419, 520–526
6 Lasonder, E. *et al.* (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 419, 537–542
7 Bahl, A. *et al.* (2003) PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Res.* 31, 212–215
8 Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29
9 Yates, J.R. 3rd (2000) Mass spectrometry. From genomics to proteomics. *Trends Genet.* 16, 5–8
10 Roos, D.S. *et al.* (1999) Origin, targeting, and function of the apicomplexan plastid. *Curr. Opin. Microbiol.* 2, 426–432
11 Craig, A. and Scherf, A. (2001) Molecules on the surface of the *Plasmodium falciparum* infected erythrocyte and their role in malaria pathogenesis and immune evasion. *Mol. Biochem. Parasitol.* 115, 129–143
12 Kyes, S.A. *et al.* (1999) Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9333–9338
13 Sutherland, C.J. (2001) Stevor transcripts from *Plasmodium falciparum* gametocytes encode truncated polypeptides. *Mol. Biochem. Parasitol.* 113, 331–335
14 Kaviratne, M. *et al.* (2002) Small variant STEVOR antigen is uniquely located within Maurer's Clefts in *Plasmodium falciparum*-infected red blood cells. *Eukaryot Cell* 1, 926–935
15 Carlton, J.M. and Carucci, D.J. (2002) Rodent models of malaria in the genomics era. *Trends Parasitol* 18, 100–102
16 Thompson, J. *et al.* (2001) Comparative genomics in *Plasmodium*: a tool for the identification of genes and functional analysis. *Mol. Biochem. Parasitol.* 118, 147–154
17 Carlton, J.M. *et al.* (1999) Karyotype and synteny among the chromosomes of all four species of human malaria parasite. *Mol. Biochem. Parasitol.* 101, 23–32
18 Tchavtchitch, M. *et al.* (2001) The sequence of a 200 kb portion of a *Plasmodium vivax* chromosome reveals a high degree of conservation

with *Plasmodium falciparum* chromosome 3. *Mol. Biochem. Parasitol.* 118, 211–222
19 Janssen, C.S. *et al.* (2001) Gene discovery in *Plasmodium chabaudi* by genome survey sequencing. *Mol. Biochem. Parasitol.* 113, 251–260
20 Roos, D.S. *et al.* (1999) Transport and trafficking: *Toxoplasma* as a model for *Plasmodium*. *Novartis Found. Symp.* 226, 176–195
21 Missinou, M.A. *et al.* (2002) Fosmidomycin for malaria. *Lancet* 360, 1941–1942
22 Surolia, N. and Surolia, A. (2001) Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*. *Nat. Med.* 7, 167–173
23 Ohkanda, J. *et al.* (2001) Peptidomimetic inhibitors of protein farnesyltransferase show potent antimalarial activity. *Bioorg. Med. Chem. Lett.* 11, 761–764
24 Banerjee, R. *et al.* (2002) Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine. *Proc. Natl. Acad. Sci. U. S. A.* 99, 990–995
25 Nezami, A. and Freire, E. (2002) The integration of genomic and structural information in the development of high affinity plasmepsin inhibitors. *Int. J. Parasitol.* 32, 1669–1676
26 Powell, J.R. *et al.* (1999) Population structure, speciation, and introgression in the *Anopheles gambiae* complex. *Parassitologia* 41, 101–113
27 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
28 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
29 Adams, M.D. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195
30 Severson, D.W. *et al.* (2001) Genetic and physical mapping in mosquitoes: molecular approaches. *Annu. Rev. Entomol.* 46, 183–219
31 Misra, S. *et al.* (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol* 3, 83
32 Hill, C.A. *et al.* (2002) G protein-coupled receptors in *Anopheles gambiae*. *Science* 298, 176–178
33 Martinez-Torres, D. *et al.* (1998) Molecular characterization of pyrethroid knockdown resistance (*kdr*) in the major malaria vector *Anopheles gambiae* s.s. *Insect Mol. Biol.* 7, 179–184
34 Ranson, H. *et al.* (2002) Evolution of supergene families associated with insecticide resistance. *Science* 298, 179–181
35 Riehle, M.A. *et al.* (2002) Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science* 298, 172–175
36 Christophides, G.K. *et al.* (2002) Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 298, 159–165
37 Zdobnov, E.M. *et al.* (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* 298, 149–159
38 Sharakhov, I.V. *et al.* (2002) Inversions and gene order shuffling in *Anopheles gambiae* and *An. funestus*. *Science* 298, 182–185
39 Coluzzi, M. *et al.* (2002) A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* 298, 1415–1418
40 della Torre, A. *et al.* (2002) Speciation within *Anopheles gambiae* – the glass is half full. *Science* 298, 115–117
41 Ho, M. and White, N.J. (1999) Molecular mechanisms of cytoadherence in malaria. *Am. J. Physiol.* 276, C1231–C1242
42 Dimopoulos, G. *et al.* (2002) Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge, and malaria infection. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8814–8819
43 Wellems, T.E. (2002) *Plasmodium* chloroquine resistance and the search for a replacement antimalarial drug. *Science* 298, 124–126
44 Carlton, J.M. *et al.* (2001) Profiling the malaria genome: a gene survey of three species of malaria parasite with comparison to other apicomplexan species. *Mol. Biochem. Parasitol.* 118, 201–210
45 Janssen, C.S. *et al.* (2002) A large gene family for putative variant antigens shared by human and rodent malaria parasites. *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* 269, 431–436
46 al-Khedery, B. *et al.* (1999) Antigenic variation in malaria: a 3′ genomic alteration associated with the expression of a *P. knowlesi* variant antigen. *Mol. Cell* 3, 131–141
47 Chima, R.I. *et al.* (2003) The economic impact of malaria in Africa: a critical review of the evidence. *Health Policy* 63, 17–36
48 Snow, R.W. *et al.* (2001) The past, present and future of childhood malaria mortality in Africa. *Trends Parasitol.* 17, 593–597
49 Kumar, S. *et al.* (2002) A multilateral effort to develop DNA vaccines against falciparum malaria. *Trends Parasitol.* 18, 129–135