Plasmodium, human and *Anopheles* genomics and malaria

Stephen L. Hoffman*, **G. Mani Subramanian***, **Frank H. Collins**† & **J. Craig Venter***

**Celera Genomics, 45 West Gude Drive, Rockville, Maryland 20850, USA (e- mail: stephen.hoffman@celera.com)*
†*University of Notre Dame, Notre Dame, Indiana 46556, USA*

The *Plasmodium* spp. parasites that cause malaria are transmitted to humans by *Anopheles* spp. mosquitoes. Scientists have now amassed a great body of knowledge about the parasite, its mosquito vector and human host. Yet this year there will be 300–500 million new malaria infections and 1–3 million deaths caused by the disease. We believe that integrated analyses of genome sequence, DNA polymorphisms, and messenger RNA and protein expression profiles will lead to greater understanding of the molecular basis of vector–human and host–parasite interactions and provide strategies to build upon these insights to develop interventions to mitigate human morbidity and mortality from malaria.

By the end of 2002, the genome sequences of many of the key species in the life cycle of the malaria parasite will be available[1–3]. This includes: (1) the 25–30 million base pairs (megabase or Mb) of the genomes of *Plasmodium falciparum* — the parasite responsible for more than 95% of all malaria deaths — and *P. yoelii*, a rodent parasite; (2) the 280 Mb of the genome of *Anopheles gambiae,* the most important mosquito vector of *P. falciparum* in sub-Saharan Africa, where malaria causes the most deaths; and (3) the 2,900 and 2,600 Mb of the human and mouse genomes, respectively. Here, we summarize the current status of the *Plasmodium*, human and *Anopheles* genome sequencing projects. We show how a number of fields of biology, based on genomics, enabled and integrated by bioinformatics, and often referred to collectively as 'functional genomics', are being brought to bear on malaria research, and how we think these approaches will improve the chances of success in our battle against malaria.

Status of *Plasmodium* spp. sequencing projects

To sequence the genome of *P. falciparum*[4], a consortium of genome centres and funders was established in 1996, which included The Institute for Genomic Research and Naval Medical Research Center (TIGR/NMRC), Sanger Institute and Stanford University (Table 1). The ~25-Mb *P. falciparum* (clone 3D7) genome, comprising 14 chromosomes and an estimated 5,000 genes, has been sequenced using a whole-chromosome shotgun strategy. Additional sequence information includes the 6-kilobase (kb) mitochondrial genome[5] and a 35-kb circular DNA that localizes to a novel organelle called the apicoplast[6]. The sequences of chromosomes 2 and 3 were published in 1998 and 1999 respectively[7,8], and a draft sequence and annotation of the entire genome will be published in 2002.

Experimental and computational hurdles had to be overcome to complete sequencing and assembly of this genome[7–9]. Large genomic fragments of *P. falciparum* DNA are not stable in *Escherichia coli*, a problem thought in large part to be due to the skewing of nucleotide content from the 59% adenine and thymine (A+T) of the human genome and 49% of the *E. coli* genome to the 80% found in the *P. falciparum* genome. The presence of A+T

repeats/homopolymers has also made gap closure and accurate assembly difficult. New techniques[9] and annotation software[10] had to be developed to overcome these problems. Additional resources used to facilitate physical gap closure included use of sequence-tagged site markers derived from end sequences of yeast artificial chromosomes (YACs) previously mapped to the genome, microsatellite markers and a high-resolution linkage map[11]. Additionally, groups of contigs were ordered on the chromosomes by reference to optical restriction maps[12]. The genome-wide, high-resolution linkage map comprises 901 markers that fall into 14 inferred linkage groups with a high rate of uniform meiotic crossover activity[11].

Progress in other *Plasmodium* spp. sequencing initiatives is summarized in Table 1. Based on analysis of the *P. falciparum* genome sequencing effort, it was estimated that a three- to fivefold shotgun sequence coverage of the entire genomes of other *Plasmodium* spp. would give 90–95% of the information provided by the significantly more expensive and time-consuming 10–15-fold coverage on a chromosome-by-chromosome basis of the *P. falciparum* genome. A fivefold shotgun sequence of the genome of the rodent parasite *P. yoelii* has been completed, and will be published in 2002. It is anticipated that a threefold shotgun sequence of the genomes of the rodent parasites *P. chabaudi* and *P. berghei* will be completed soon, as will a fivefold coverage of the human parasite *P. vivax*, and of the simian parasite *P. knowlesi*. Additional ongoing efforts are directed at sequencing expressed sequence tags (ESTs; Table 1) of *P. falciparum*, *P. yoelii*[13], *P. berghei*[14] and *P. vivax*[14]. As genomic sequence from other related apicomplexans, such as *Cryptosporidium* and *Toxoplasma*, become available, comparative genomic analyses will enhance understanding of these protozoan pathogens.

To facilitate interpretation and dissemination of genomic sequence data to researchers worldwide, informatics capabilities that integrate sequence data, automated analyses and annotation data emerging from the *P. falciparum* genome project have been developed (ref. 15; and Table 1).

Insights from *P. falciparum* genome data

The biological insights gleaned from publicly available genomic sequence data have already led to the identification of important leads for antimalarial drugs and potential
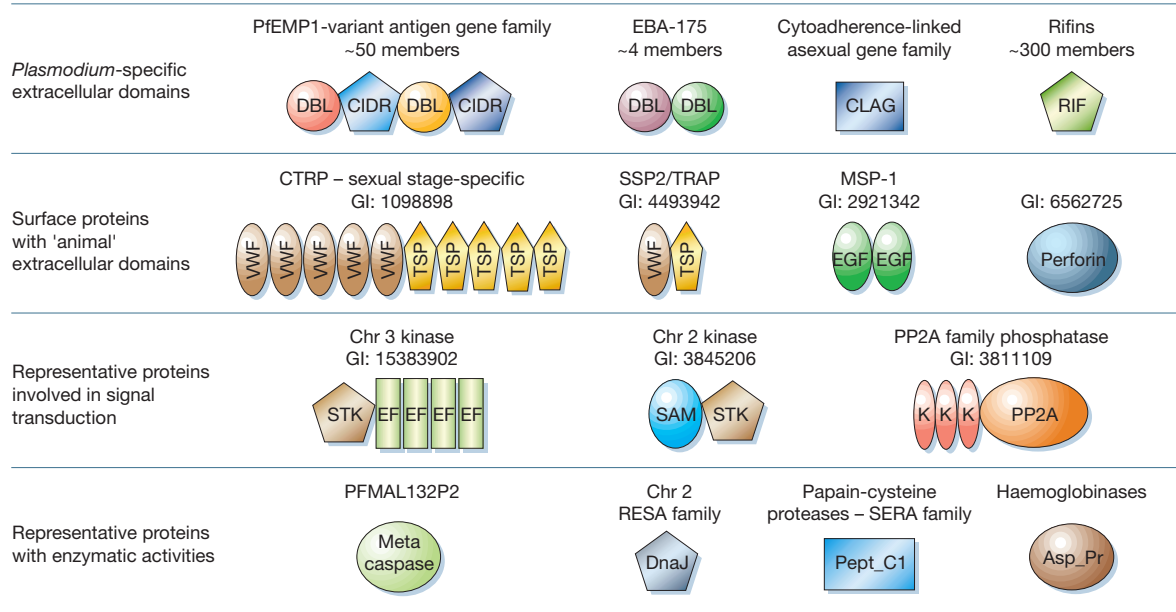
Figure 1 Protein domain architecture in representative *Plasmodium* proteins. Domain name abbreviations: DBL, Duffy-binding-like; CIDR, cysteine-rich interdomain region; CLAG, cytoadherence-linked asexual gene; RIF, Rifins; VWF, von-Willebrand factor A; TSP, thrombospondin; EGF, epidermal growth factor-like; Perforin, membrane-activated complex-perforin; STK, serine/threonine kinase; EF, calcium-binding EF-hand; SAM, sterile alpha-motif; K, kelch repeats; PP2A, PP2A phosphatase; DnaJ, DnaJ family of chaperones; Pept_C1, papain cysteine protease; Asp_Pr, aspartyl protease.

targets for vaccines (refs 16–18; and see Table 2). Furthermore, computational analysis of published genomic sequence data has improved our understanding of the evolutionary and functional adaptations of *P. falciparum*.

Target and lead identification for drugs based on genome sequence data

A rapid demonstration of how genomic sequence facilitates target and lead identification occurred almost as soon as the preliminary, uncompleted sequence of chromosome 2 was posted on the TIGR website. Using homology to known genes in plants and algae, researchers identified several lipid biosynthesis genes in *P. falciparum* and demonstrated their targeting to the apicoplast[19]. These genes are attractive drug targets given their critical role in parasite viability and the absence of the type II fatty acid biosynthesis pathway in humans;

at least one major pharmaceutical company has assessed the effect on malaria parasites of compounds known to be active in this pathway. In follow-up, using sequence data on the Sanger website, researchers identified a key enzyme in this pathway — enoyl–acyl-carrier protein (enoyl–ACP) reductase (FabI) — and demonstrated *in vitro* activity against *P. falciparum* using an inhibitor to this enzyme[17].

Another effort was based on unpublished data from chromosome 14 posted on the TIGR website. Researchers used genomic sequence data to identify the enzymes 1-deoxy-D-xylulose-5-phosphate (DOXP) synthase and DOXP reductoisomerase, and provided evidence that isoprenoid biosynthesis in *P. falciparum* depends on the DOXP pathway and is critical for parasite viability[18]. They identified two drugs known to be active in this pathway, fosmidomycin and its derivative FR-900089, and demonstrated

Table 1 Current status of *Plasmodium* sequencing projects

Plasmodium species	Sequencing approach	*n*-Fold coverage	Status	Institution publications*	Anticipated full publication
P. falciparum Chromosomes 2,10,11,14	Whole-chromosome shotgun sequencing	10–15×	Completed	TIGR/NMRC[7]	Mid-2002
P. falciparum Chromosomes 1,3,4,5–9,13	Whole-chromosome shotgun sequencing; includes YAC shotgun sequence	10-15×	Near completion	Sanger[8]	Mid-2002
P. falciparum Chromosome 12	Whole-chromosome shotgun sequencing; includes YAC shotgun sequence	10-15×	Completed	Stanford	Mid-2002
P. yoelii	Whole-genome shotgun sequencing	5×	Completed	TIGR/NMRC	Mid-2002
P. chabaudi	Whole-chromosome shotgun sequencing	3×	In progress	Sanger	?
P. berghei	Whole-chromosome shotgun sequencing	3×	In progress	Sanger	Mid-2002
P. knowlesi	Whole-chromosome shotgun sequencing	5×	In progress	Sanger	Mid-2002
P. vivax	Whole-genome shotgun sequencing	5×	In progress	TIGR/NMRC	End 2002
P. yoelii	Sporozoite cDNA sequencing	>1,900 ESTs	Completed	TIGR/NMRC/NYU[13]	August 2001
P. vivax	Genome sequence tags (genome survey sequences; GSSs)	>10,000 GSSs	In progress	Univ. Florida[14]	December 2001
P. berghei ANKA clone	EST sequencing; genome sequence tags	>5,000 ESTs >5,000 GSSs	In progress	Univ. Florida[14]	December 2001

*Institutional websites and additional Internet database resources are as follows. *Plasmodium* genome database at TIGR/NMRC: http://www.tigr.org; http://www.nmri.nnmc.navy.mil. *Plasmodium* genome database at Sanger: http://www.sanger.ac.uk/Projects/P_falciparum/. *Plasmodium* genome database at Stanford: http://www-sequence.stanford.edu/group/malaria. Malaria genome tag sequencing projects at University of Florida: http://parasite.vetmed.ufl.edu/. The University of Pennsylvania provides access to all of the genome data produced by the *P. falciparum* genome project (PlasmoDB): http://plasmodb.org/. WHO/TDR database: http://www.wehi.edu.au/MalDB-www/who.html. National Center for Biotechnology Information Malaria Genetics and Genomics: http://www.ncbi.nlm.nih.gov/Malaria/. Metabolic pathways in *Plasmodium*: http://sites.huji.ac.il/malaria/.

Table 2 Highlights of data generated from the *Plasmodium* genome project

	Implications on malaria research (diagnostics and drug or vaccine development)
High A + T content[7]	Technical difficulties in cloning, sequencing and assembly. May confound DNA-based diagnostics.
High percentage of low-complexity proteins[7]	Variable specificity of antibody reagents. May confound serum-based diagnostics.
	Need for altered strategies in immunogen production and vaccine design.
Surface protein families (also refer to Fig. 1)	
Plasmodium-specific expansions (vars, rifins, stevors and clags)[7,8,26–28]	Central role in cytoadherence and pathology associated with malaria. The antigenic variations in these families pose a huge challenge for vaccine development.
Several protein families with animal-specific extracellular domains[7,8,35–39]	Mediate interactions with host receptors and constitute some of the most promising vaccine candidates.
Metabolic enzymes (plastid derived)	The cyanobacterial inheritance (absent in humans) makes them attractive as drug targets.
Type II fatty acid biosynthesis[17,19]	Inhibitors of enoyl–ACP reductase have shown *in vitro* and *in vivo* activity.
Isoprenoid biosynthesis[18]	Inhibitors of the DOXP pathway have shown *in vitro* and *in vivo* activity.
Shikimate pathway[16]	Preliminary evidence supports the importance of this pathway for parasite viability.
Unusual features of regulatory proteins	
Proteases	
Papain family cysteine proteases[7]	The serine-rich antigen (SERA) family of proteases are related structurally to the cysteine proteases. Candidates for therapeutic (drug or vaccine) development.
Aspartyl proteases[7,8,21]	Central role in haemoglobin digestion in the red cell stages. Attractive drug targets.
Metacaspases[23]	Recently described family of caspases (cysteine proteases) that are absent in animals.
DnaJ family chaperones[7]	Significant expansion of this family suggests diverse roles for this protein domain in the *Plasmodium* life cycle.
Serine/threonine kinases[7,8]	Several Ser/Thr kinases with many containing calcium-binding EF-hands; indicative of a prominent role for calcium mediated signal transduction. Other features of signal transduction are consistent with the early branching of *Plasmodium* from the eukaryotic radiation.
Likely absence of tyrosine kinases/phosphatases[7,8]	
Paucity of transcription factors[7,8]	
Genetic map[11]	The high-resolution map has greatly helped validate assembly of the genome and markers to identify genetic correlates of drug resistance.
Genetics of drug resistance[77,78]	Genetic and biochemical approaches have been used to identify mutations in plasmodial genes to dissect the molecular basis for drug resistance. The comprehensive catalogue of these mutations will enhance efforts in monitoring the spread of drug resistance.
Chloroquine	
Dihydrofolate reductase antagonists	
Stage-specific RNA expression	
Asexual versus sexual stages[43]	Identified several stage-specific proteins, several of which are vaccine candidates.
Sporozoite stages (*P. yoelii*)[13]	Use of syntenic relationships to identify the orthologous gene in *P. falciparum*. Provides a set of genes to further investigate as vaccine candidates.

activity *in vitro* against *P. falciparum*, including against multidrug-resistant *P. falciparum*, and *in vivo* against *P. vinckei* in mice[18].

The value of genomic sequence data and comparative analyses to identify new drug targets was shown recently with the identification in *P. falciparum* of the metal-dependent RNA triphosphatase protein family, members of which are crucial in mRNA cap formation and eukaryotic gene expression[20]. The structure of the active site and catalytic mechanism of this protein family in *P. falciparum* and fungi are completely different from the RNA triphosphatase domain of the metazoan (human) capping enzymes, and metazoans encode no identifiable homologues of the fungal or *Plasmodium* RNA triphosphatases. The structural similarity between the plasmodial and the fungal RNA triphosphatases raises the exciting possibility of achieving antifungal and antimalarial activity with a single class of mechanism-based inhibitors[20]. Protein families that are expanded in *P. falciparum* and are considered attractive drug targets include the cysteine and aspartyl class of proteases, with both classes constituting a parasite adaptation to its need for haemoglobin digestion[7,21,22]. Of special interest is the representation of several members of the recently described metacaspase family of cysteine proteases in *Plasmodium*[23]. The absence of this class of proteases in humans makes the *P. falciparum* metacaspases particularly attractive as drug targets.

Target identification for vaccines based on genome sequence data
P. falciparum proteins that traffic to the erythrocyte surface are important in the pathogenesis of malaria, and are potentially important targets for vaccine development[24]. Genomic sequence data have demonstrated the extent of paralogous expansion of genes encoding such parasite proteins. Several members of these multi-gene families are being considered as vaccine candidates given their prominent role in malarial pathogenesis (refs 24,25; and Table 2, Fig. 1). Prominent among these are the variant gene families clustered in the subtelomeric regions of the *P. falciparum* genome (*var* genes encoding *P. falciparum* erythrocyte membrane protein (PfEMP), and *rifin*, *stevor* and *clag* genes)[7,26–28] and *P. vivax* genome (*vir* genes)[29]. Furthermore, the availability of genomic sequence data has provided invaluable clues into the regulation of the large variant-antigen PfEMP gene family[7,8,30]. Other attractive targets for vaccines are parasite proteins important for parasite invasion of erythrocytes

(see reviews in this issue by Richie and Saul, pages 694–701, and Miller *et al.*, pages 673–679). Here again, genomic sequence data have enabled the identification and characterization of paralogues of the Duffy-binding-like domain-containing, EBA-175-like gene family members[31,32], and the rhoptry[33] and reticulocyte binding protein[34] families involved in red cell invasion.

Although most of these merozoite and erythrocyte surface proteins have plasmodial-specific domains, it is of interest to note that there are several examples of plasmodial proteins with extracellular binding domains that are found predominantly in animal genomes (Table 2, Fig. 1). These include the epidermal growth factor domain-containing merozoite surface proteins and Pfs25/28 (*P. falciparum* gametocyte surface protein) that have homologues in other plasmodium species[35,36]. Likewise, thrombospondin and von-Willebrand factor A domain-containing proteins such as circumsporozoite protein (CSP), sporozoite surface protein 2/thrombospondin-related adhesive protein (SSP2/TRAP) and CSP-TRAP-related protein (CTRP) are involved in the gametocyte and sporozoite stages of the life cycles[37,38]. Many of these proteins are central to initiating infection of the host, elicit a protective immune response, and are being developed as vaccine candidates (see reviews by Richie and Saul and by Miller *et al.* in this issue) The complete genomic sequence allows for rapid identification of all members of these families and their investigation as potential targets for vaccine development. Other interesting examples of animal extracellular adhesion domains detected in predicted plasmodium proteins include the LCCL (domain identified in *Limulus* factor C, Coch-5b2 and LGL1) module[39] and perforin domain-containing proteins[40] in the *P. falciparum* genome.

Finally, over 40% of predicted proteins in the genome are predominantly low-complexity or non-globular proteins with several instances of low-complexity inserts in the midst of globular protein domains[7]. This may reflect an adaptive mechanism of the parasite to evade an antibody-driven immunity and may impact vaccine design to improve specificity of antibody response.

mRNA and protein expression to identify drug and vaccine targets
Genomic sequence data and new technologies have provided the foundation for studies to determine mRNA and protein expression at different stages of the parasite life cycle[41]. Over 3,000 random inserts

from a *P. falciparum* mung bean nuclease genomic library were used in a DNA microarray to identify differences in gene expression between the asexual blood-stage trophozoite and the sexual blood-stage gametocyte form of the parasite[42]. These experiments increased the list of stage-specific transcripts in *P. falciparum* by an order of magnitude, provided potential targets for developing drugs and vaccines, and yielded clues as to how to explore differences in the metabolic machinery of the parasite as it transits from humans to mosquitoes[43]. A DNA microarray based on genes from chromosomes 2, 3, 12 and 14 was used to assess differential parasite gene responses to antimalarial drugs in drug-sensitive and drug-resistant strains at different times during the asexual erythrocytic stage of the life cycle (ref. 41; and A. Whitney, unpublished data). A different approach, based on sequencing of a *P. yoelii* sporozoite complementary DNA library, helped identify genes expressed in the sporozoite stage of the life cycle[13]. In addition to identifying over 1,300 new sporozoite-expressed genes, this study documented the sporozoite-stage expression of potentially important surface molecules.

Work has also begun to identify proteins expressed at specific stages of the parasite life cycle[41]; data that will be especially important for vaccine development. There are multiple methods available for conducting such studies, but the most progress to date has used recent technological advances in liquid chromatography–mass spectrometry (LC-MS), wherein mass spectral fingerprints that are experimentally derived may be matched to computationally generated profiles derived from the genomic sequence (refs 41,44; and D. Carucci, personal communication). The availability of an accurately assembled and annotated complete genome is critical for the success of these ongoing efforts.

Status of the human genome project
The 2,900-Mb human genome, with approximately 30,000 predicted protein-coding genes[1,2] and a genome-wide catalogue of several million single nucleotide polymorphisms (SNPs), is now available for researchers to better define the genetic determinants of human disease[1,45]. Furthermore, ongoing efforts at haplotype analysis will help discern disease-related SNPs from the background of irrelevant genetic variation[46]. Likewise, efforts are underway to obtain portraits of gene and protein expression profiles from human tissue from a wide range of anatomic compartments in the body during different states of health and disease and over time. These projects are in large part dependent on the availability of the complete genomic sequence. The relevance of the human genome sequence to human disease and medicine has been discussed in a recent publication[47].

Relevant insights from the human genome project
Human genetic variations are one of the principal determinants of susceptibility to many common infectious diseases[48]. Malaria was one of the first infectious diseases to be studied extensively and many susceptibility and resistance loci have been identified over the past few decades[49]. The most prominent human polymorphism that protects against severe malaria is the sickle cell trait, although other haemoglobinopathies or blood group antigen variants have been shown to influence clinical outcomes by altering the efficiency of invasion of erythrocytes by the parasites or the development of the parasites within erythrocytes.

Other determinants of clinical severity include receptor polymorphisms in genes encoding proteins such as intercellular adhesion molecule 1 (CD54) and CD36, which are involved in the sequestration of infected red cells in vascular endothelium, or polymorphisms in the human tumour-necrosis factor-α gene promoter that are associated with cerebral malaria[50–52]. Although a dozen or more specific susceptibility determinants have been defined so far, including both structural and regulatory polymorphisms of red cell proteins and genes of the immune system, it is most likely that many more are yet to be discovered[49]. It must also be emphasized that efforts made so far do not represent a methodical and comprehensive

survey of the genome. As the human genome sequence and the extensive catalogue of SNPs have been publicly available only since early 2001, we have yet to reap the full benefits of this tremendous resource for malaria research.

Status of the *Anopheles* genome project
The *A. gambiae* genome project is funded largely by the National Institute of Allergy and Infectious Diseases and the French Government, with the World Health Organization/Tropical Disease Research programme providing a coordinating role and assisting in database development. Celera Genomics and Génoscope have recently completed a shotgun sequencing effort (10× coverage), which has been assembled by Celera. Two independent annotations, one by Celera and the other by the European Bioinformatics Institute using Ensembl should be completed by early 2002. The annotated genome may be presented publicly as early as March 2002 either through Ensembl or through an adapted version of the *Drosophila* GadFly. End sequences of two different *A. gambiae* bacterial artificial chromosome (BAC) libraries (>10× coverage), the physical map location of ~2,000 BAC clones, perhaps as many as 80,000 ESTs, and several sequenced and annotated genomic DNA contigs (100–500+ kb) will facilitate assembly and annotation. Celera Genomics, the European Molecular Biology Laboratory, Génoscope, the Institute of Molecular Biology and Biotechnology in Crete, the Institut Pasteur in Paris, TIGR, and the Universities of Iowa, Rome and Notre Dame are contributing.

All genomic libraries used in this project have been made from a strain of *A. gambiae* (PEST) that has a sex-linked, pink-eye mutation[53] and a standard chromosome arrangement[54]. The shotgun sequence has been obtained from plasmid libraries with 2-kb, 10-kb and 50-kb inserts produced at Celera. All BAC libraries, sequenced cDNAs and the end-sequenced 10-kb plasmid clones will be archived at the American Type Culture Collection as part of the National Institutes of Health-supported Malaria Research and Reference Reagents Resource Center (http://www.malaria.mr4.org/ mr4pages/index.html).

Relevant insights from the *Anopheles* genome project
The 120-Mb genome of fruitfly, *Drosophila melanogaster*, provided the first complete invertebrate insect genome sequenced using the shotgun sequence and assembly approach[55]. The completion of the 280-Mb *Anopheles* genome will undoubtedly enhance our understanding of the evolutionary adaptations of mosquitoes since their divergence from the fly approximately 200 million years ago, including traits such as the physiology of blood feeding and blood meal digestion and behaviours associated with the selection of humans as the blood meal source and oviposition in aquatic sites produced by human agricultural activity. Most of the specific EST efforts in this project are focused on mosquito tissues with which malaria parasites interact[56], including the midgut, haemocytes, fat body and salivary glands, or on genes expressed in tissues such as the head and antennae that may mediate behaviours such as host selection and mating. A number of current research efforts are motivated by an interest in the mosquito's response to infection with the malaria parasite, as well as the basis whereby several selected strains of *A. gambiae* are able to mount a defensive response to *Plasmodium*[57,58].

Comparative analysis of the *Drosophila*, *Anopheles* and human genomes has already provided several practical insights. The shared aspects of innate immunity between the two insects and humans have been confirmed by computational[1,59] and experimental evidence[56,60]. Of special interest are the specific adaptations of haematophagous insects (including *Anopheles*) that promote transmission by enhancing blood feeding. These include the convergent evolution of proteins that interact extensively with mammalian proteins which regulate haemostasis and the complement cascade[61]. On a larger comparative scale, preliminary examinations of conserved gene arrangement between *Drosophila* and *A. gambiae* suggest that substantial gene order arrangement may be preserved at a scale of up to
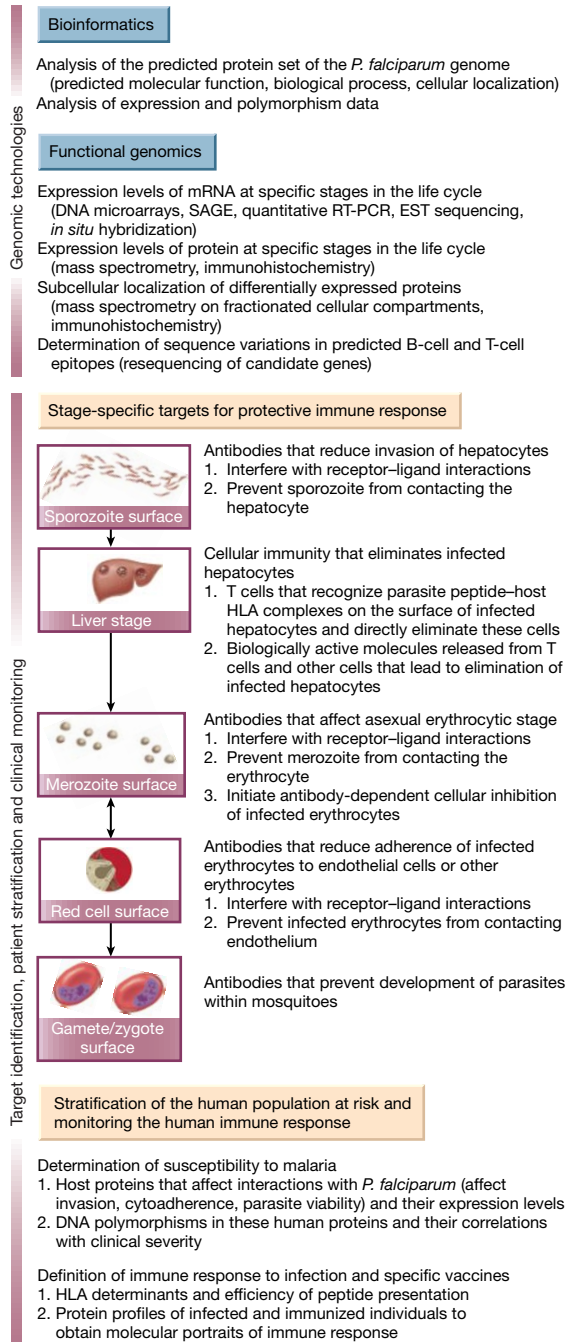
Figure 2 Malaria 'vaccinomics'. The potential impact of genomics and functional genomics on identification of new functionally or anatomically important targets for malaria vaccine development, and optimizing selection of human populations for immunization with specific vaccines.

several hundred kilobases, and very large blocks of most chromosome arms may be shared[62]. The ability to compare mosquito and *Drosophila* genes in terms of sequence similarity and location will be important in guiding researchers towards an understanding of mosquito gene function.

What does genomics offer for the control of malaria?

There is currently no vaccine for malaria, no optimal way to sustainably reduce or eliminate contact between humans and infective mosquitoes, a poor system for early diagnosis and recognition of those at greatest risk for developing severe disease, and a disturbing,

ever increasing resistance to the drugs used to treat malaria. We believe that genomics and related disciplines provide the scientific foundation for improving our chances to address and solve each of these problems[63].

Vaccine development

Effective vaccines elicit immune responses that destroy the infectious agent or the host cells in which they reside, or inhibit a function critical for its survival. Immune responses against *P. falciparum* can protect through both mechanisms. Immunization with radiation-attenuated sporozoites protects more than 90% of recipients against experimental challenge for at least 10 months[64]. Protection is thought to be mediated primarily by T cells against peptides from parasite proteins expressed in infected hepatocytes (function independent), although antibodies that reduce sporozoite invasion of hepatocytes (function inhibiting) also have a role (refs 65–67; and Fig. 2). In areas of Africa with the most intense malaria transmission, children who survive to the age of 7–10 rarely develop life-threatening *P. falciparum* infections. They become infected frequently, but their immune systems limit the infections, thereby preventing severe disease. Antibodies against parasite proteins expressed on erythrocytes, which prevent sequestration in the microcirculation (function inhibiting)[25], and on the surface of merozoites, which prevent invasion of erythrocytes (function inhibiting)[68] or initiate antibody-dependent cellular inhibition[69] activity (function independent), are thought to be primary in this 'naturally acquired' protective immunity.

During the past 20 years there has been considerable work to develop subunit vaccines that provide protection after exposure to irradiated sporozoites and naturally acquired immunity (see review in this issue by Richie and Saul, pages 694–701). Such a vaccine has yet to be developed. There are numerous potential explanations for the lack of current success. One is that exposure to the whole parasite elicits a more potent, protective immune response than do the subunit vaccines tested thus far. However, in the case of people protected by the irradiated sporozoite vaccine, or by naturally acquired immunity, T-cell and antibody responses are generally modest, and in many instances lower than after subunit vaccination. It is more likely that immunization with only a few parasite proteins cannot duplicate the immunity elicited by exposure to a parasite that has thousands of proteins; modest immune response against tens, hundreds or thousands of parasite proteins may be additive or synergistic. In the case of naturally acquired immunity, this 'breadth' would be expanded by exposure to so many polymorphic strains of *P. falciparum*. If this is the case, then the malaria genome and SNP projects may provide the essential foundation for duplicating this whole-organism immunity.

Analysis of the sequence of the *P. falciparum* genome has dramatically expanded our knowledge regarding the paralogous expansion in the genome of parasite proteins expressed on the erythrocyte surface (Table 2, Fig. 1). But genomics on its own will not be enough[70]. Immune responses are directed against proteins, not genes, and the parasite expresses different proteins at different stages of its life cycle[70].

We believe that the first step in duplicating whole parasite-induced protective immunity is to catalogue expression of proteins at each stage of the parasite life cycle. The first step to accomplishing this could involve generating antibodies to every parasite protein and determining their subcellular localization[70]. This cannot be done rapidly and without great expense, so other approaches have been initiated. Assessment of mRNA and/or protein expression at different stages of the parasite life cycle would provide comprehensive data. Unfortunately, neither is 'the best' way. Immune responses recognize proteins, not mRNA, so gene expression profiles must be confirmed at the protein level. Furthermore, there are many instances (for example, alternative splice sites or post-translational modifications) in which mRNA expression data will not be adequate. The inability to culture infective *P. falciparum* sporozoites and the poor performance of hepatocyte cultures make it difficult (sporozoites) or impossible (liver-stage parasites) to acquire enough parasite material for

conventional, non-amplified gene chip or DNA microarray analysis of these critical stages of the life cycle. For these reasons there has been increasing emphasis on LC-MS methods for establishing differential protein expression profiles at each stage of the life cycle[41].

We anticipate the generation of comprehensive databases that catalogue stage-specific expression of all genes and proteins in the *P. falciparum* genome, and the polymorphisms in the DNA sequences of these genes in field isolates. Concurrent efforts to establish the functional importance of each gene, and the significance of their protein–protein interactions will take longer. Nonetheless, characterization of genes, their transcripts, and the proteins they encode will provide the foundation for the development of effective vaccines.

In subsequent sections we address how genomics-based data may be helpful in focusing vaccination strategies. This will include the need to identify patients at high risk, as well as a 'systems biology/immunology' approach to monitor immune response, wherein molecular kinetic 'portraits' of individuals should provide a more comprehensive indication of the response to infection and immunization than do currently available cross-sectional single immune response measurements (Fig. 2, Table 3). Significant difficulties lie ahead in developing vaccines that duplicate the protection provided by immunization with irradiated sporozoites or naturally acquired immunity (or that work through other mechanisms), delivering those vaccines optimally to the individuals who need them most, and effectively monitoring responses to those vaccines. In addition to increased understanding of the target of immune responses (the parasite proteins and epitopes within those proteins), work in vaccinology that takes advantage of the human genome project to develop improved methods for maximizing magnitude, quality and longevity of protective immune responses will be fundamental to fielding effective malaria vaccines.

Reducing contact between humans and infective mosquitoes

The sequence of the *A. gambiae* genome will accelerate development of new malaria control tools to supplement existing insecticides and bednet strategies, not only by advancing current research efforts, but also, and perhaps most important, by opening lines of research not currently envisaged. The dire need for new vector-related tools and strategies for malaria control is obvious given that most of the world's successful malaria control programmes have utilized primarily transmission-reduction tools such as insecticides and bednets[71].

Preserving the efficacy of insecticides in the face of resistance

The most important vector-targeted control strategies today involve insecticide-impregnated bednets. Unfortunately, resistance to the pyrethroid insecticides used in bednets has now been recorded in *A. gambiae* populations in both east and west Africa as well as in *A. funestus* (another major vector in Africa) from southern Africa[71]. One of the main causes of resistance is upregulation of groups of insecticide-detoxifying enzymes such as oxidases, esterases and glutathione *S*-transferases[72]. The *A. gambiae* genome will enable rapid identification of the genes involved and the molecular changes that lead to resistance, information that may guide in the selection and/or development of alternative insecticides. It will also provide tools that can be used in population studies to monitor the emergence and spread of important traits like insecticide resistance.

Increasing the efficiency of vector-targeted control efforts

A. gambiae is a member of a complex of what are now recognized as seven related but genetically and ecologically distinct species[71]. Decades of field work in Africa, where the vector was identified as *A. gambiae sensu lato*, have produced a picture of *A. gambiae* ecology that is probably the composite of data from at least two or even more members of this complex. Recent studies in west Africa suggest *A. gambiae sensu stricto* may comprise at least two and possibly as many as five distinct, reproductively isolated cytogenetic forms[73]. An immediate outcome of the *A. gambiae* genome project should be development of simple, DNA-based diagnostics for these different forms that can be used to more fully understand cytotype-specific differences in vector behaviour, ecology and pathogen transmission.

Development of a genetic control strategy for *A. gambiae*

The *A. gambiae* genome project builds on a decade of progress in developing tools for the genetic control of malaria transmission[74]. Robust germline-transformation tools now exist for *Anopheles* mosquitoes[75]. Significant progress has been made in the study of parasite development in the *Anopheles* vector[58,71]. For example, it has now been shown that normal parasite development can be almost totally blocked in *Anopheles* mosquitoes transformed with a construct that binds to both midgut and salivary glands (M. Jacobs-Lorena, personal communication). Considerable progress has also been made in studies of *A. gambiae* population structure, using tools that emerged from laboratory cloning and mapping efforts[71]. The *A. gambiae* genome project will help solve the problems that must be overcome before even a limited genetic-control field trial can be implemented.

Early diagnosis and predicting risk of developing severe malaria

To achieve the goal of reducing morbidity and mortality caused by malaria, it will be imperative to target expensive resources to those at

Table 3 Integrative analysis of malaria using genome-based computational and experimental approaches

Genomic technologies	Vaccine development	Drug development	Early diagnosis and predicting risk of developing severe malaria	Reducing and eliminating contact between humans and infective mosquitoes

Genomic sequence analysis
  Predict gene structure and function
  Comparative genomics
	Target identification: predict surface proteins; identify homologues across *Plasmodium* spp.; mouse models of malaria to identify and validate vaccine candidates.	Target identification: identify new or divergent members in a biological pathway in *Plasmodium*; structure-based rational drug design; mouse models of malaria to identify drug targets and to model efficacy of new therapies.	Comparative analysis across *Plasmodium* spp. to define determinants of drug resistance; mouse models of malaria to identify genetic determinants of resistance and clinical severity.	Target identification: identify new or divergent members in a biological pathway in *Anopheles* (insecticide)

Resequencing Catalogue genetic variations in *Plasmodium*, *Anopheles* and human populations	Evaluate polymorphisms in vaccine epitopes to optimize vaccine design.	Evaluate polymorphisms in genes that are drug targets (in *Plasmodium*) or in drug modifiers (in humans)	Identify polymorphisms in *Plasmodium* genes that impart drug resistance, and in human genes that impact susceptibility to severe malaria.	Evaluate polymorphisms in genes in *Anopheles* that will allow detection of resistance to insecticides and efficiency of transmission.

Differential expression
  RNA (microarray)
  Protein (LC-MS)
  Metabolite (LC-MS)
	Target identification of proteins that are relevant for T-cell or B-cell vaccines based on stage-specific differential expression.	Target identification and validation based on stage-specific expression	In *Plasmodium*, identify surrogate markers for diagnosis and predictors of clinical response to therapy; in humans, monitor immune response profiles and severity of illness.	Identify genes in *Anopheles* that modify the innate immune response to *Plasmodium* and resistance to insecticides.

Protein-interaction maps
  Yeast two-hybrid genetic screens
  Mass spectrometry
  Antibody or protein chips
	Identification of receptor–ligand interactions to define critical binding motifs for vaccine design.	Increased numbers of potential drug targets based on the understanding of biological pathways.	Not applicable.	Increased numbers of potential insecticide targets and targets to inhibit sporogonic development based on the understanding of biological pathways.

highest risk of suffering severe disease and death. We believe that genomics-based efforts may facilitate this approach.

Early diagnosis of infection

Conventional approaches for diagnosing *P. falciparum* infection include microscopic examination of blood smears, the use of dipsticks that assay *P. falciparum* histidine-rich protein 2, and the less practical amplification of *P. falciparum* genes using the polymerase chain reaction (PCR)[76]. Gene expression and proteomics profiling studies will probably provide additional markers. We speculate that several of these protein- or metabolite-based markers could be adapted in antibody-based assays to enhance detection as well as provide surrogate markers of drug resistance and clinical outcomes (Table 3).

Rapid identification and monitoring of drug resistance

It is imperative that drug-resistant strains of *P. falciparum* be detected when they first emerge. Genetic and genomic sequence has been the basis of PCR-based screens against individual *P. falciparum* genes used to document and monitor resistance to chloroquine and sulpha-based treatment regimens[77,78]. The genome-wide, high-resolution linkage map[11] is complementary to sequence data and will lead to determination of most drug-resistant genes and their polymorphisms. The capacity to detect *P. falciparum* metabolite and protein levels during the erythrocytic stages of its life cycle using mass spectrometry should lead to the development of clinically relevant, rapid methods for determining drug resistance, and early recognition of failing therapy.

Identification of individuals at risk of developing severe disease

Case–control and association studies that have been used successfully to dissect the genetics of human disease will benefit significantly from the genomic sequence and availability of genetic markers[48]. Additionally, the recent completion of the mouse genome and the comparative genomic map of the two species will complement ongoing efforts to dissect determinants of host susceptibility and immune response in murine models of malaria[3,79]. Future efforts aimed at identifying genetic correlates in populations at risk with biological correlates of parasite invasion, parasite development, cytoadherence efficiency and clinical outcomes will be critical in identifying at-risk populations to better direct vaccination and chemoprophylaxis strategies (Table 3, Fig. 2). Similar efforts are needed to develop pharmacogenetics in malaria to improve predictors of drug efficacy and toxicity[80]. We foresee a time when such data will be used to direct allocation of resources as well as establish appropriate diagnostic, therapeutic and prevention guidelines in malaria control programmes.

Drug development and new therapeutics

Advances in proteomics will be crucial in identifying differential expressed proteins, which along with comparative genomic analysis, utilization of protein interaction maps and an understanding of metabolic pathways, will help identify and prioritize targets for therapeutic intervention (Table 3, Fig. 2). The ability to target regulatory regions as a therapeutic strategy may now be feasible given developments in computational and experimental approaches to identify potential regulatory regions in *P. falciparum* genes[30], and highly selective methodologies that are able to target these highly AT-rich segments[81].

Recent advances in gene-targeting technologies to manipulate *P. falciparum* genes as well as those in other rodent *Plasmodium* species will significantly aid target validation[82,83]. Of immediate interest will be *Plasmodium* proteins that are amenable to interrogation using currently available small-molecule design platforms and target-family chemical libraries, in addition to being sufficiently divergent or distinct from their human homologue. Structure-based modelling, the identification of critical catalytic and substrate-recognition motifs, polymorphism detection to target invariable regions, and the ability to predict determinants of drug resistance will provide the robust framework to rationalize and streamline the development

of the next generation of therapeutics. Knowledge can be gleaned from the recent success in developing small-molecule inhibitors of the *Trypanosome cruzi* cysteine proteases[84], and the *P. falciparum* homologues might also be attractive targets[22]. Equally tantalizing are the metacaspase family of cysteine proteases[23] that may provide a new class of drug targets (absent in humans) that are amenable to high-throughput screening and the rational drug design process.

An integrated, systems-biology approach in malaria

There are numerous views on how to solve the problems of malaria[63,85]. We have outlined our perspective on how genomics and genomics-dependent science will be used to develop antimalarial vaccines; reduce or eliminate contact between humans and infective mosquitoes; rapidly diagnose and recognize those at greatest risk for developing severe disease; and identify and treat multidrug-resistant malaria. However, we have not discussed the integration of these efforts, or their integration with complementary fields of biomedicine.

The availability of genomic sequence data, use of sequence-based, high-throughput technologies, and advances in bioinformatics to analyse and interpret genomic data will ultimately provide an integrated picture of malarial biology, pathogenesis and epidemiology. We anticipate a time when we will routinely measure real-time biological profiles, including molecular phenotypic expression and genetic polymorphisms of the vector, host and parasite, and clinical outcomes (Table 3, Fig. 2), to create molecular kinetic portraits. Such complex biological systems have properties that cannot be predicted *a priori*. The development of mathematical modelling methods to describe biological systems at the molecular and/or population level, and predict behaviours of these systems in response to various virtual treatments, will be crucial in developing a new generation of interventions to attack malaria. The promise of a genome-sequence-based platform thereby lies in providing an integrated reconstruction of the spectrum of molecular and cellular interactions among parasite, vector and human host, with the ultimate goal of eradicating malaria. □

1. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
2. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
3. Marshall, E. Genome sequencing. Celera assembles mouse genome: public labs plan new strategy. *Science* **292,** 822 (2001).
4. Hoffman, S. L. *et al.* Funding for malaria genome sequencing. *Nature* **387,** 647 (1997).
5. Feagin, J. E., Gardner, M. J., Williamson, D. H. & Wilson, R. J. The putative mitochondrial genome of *Plasmodium falciparum. J. Protozool.* **38,** 243–245 (1991).
6. Wilson, R. J. *et al.* Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum. J. Mol. Biol.* **261,** 155–172 (1996).
7. Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum. Science* **282,** 1126–1132 (1998).
8. Bowman, S. *et al.* The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum. Nature* **400,** 532–538 (1999).
9. Gardner, M. J. The genome of the malaria parasite. *Curr. Opin. Genet. Dev.* **9,** 704–708 (1999).
10. Salzberg, S. L., Pertea, M., Delcher, A. L., Gardner, M. J. & Tettelin, H. Interpolated Markov models for eukaryotic gene finding. *Genomics* **59,** 24–31 (1999).
11. Su, X. *et al.* A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum. Science* **286,** 1351–1353 (1999).
12. Lai, Z. *et al.* A shotgun optical map of the entire *Plasmodium falciparum* genome. *Nature Genet.* **23,** 309–313 (1999).
13. Kappe, S. H. *et al.* Exploring the transcriptome of the malaria sporozoite stage. *Proc. Natl Acad. Sci. USA* **98,** 9895–9900 (2001).
14. Carlton, J. M. *et al.* 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 (2001).
15. The Plasmodium Genome Database Collaborative. PlasmoDB: an integrative database of the *Plasmodium falciparum* genome. Tools for accessing and analyzing finished and unfinished sequence data. *Nucleic Acids Res.* **29,** 66–69 (2001).
16. Roberts, F. *et al.* Evidence for the shikimate pathway in apicomplexan parasites. *Nature* **393,** 801–805 (1998).
17. Surolia, N. & Surolia, A. Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum. Nature Med.* **7,** 167–173 (2001).
18. Jomaa, H. *et al.* Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285,** 1573–1576 (1999).
19. Waller, R. F. *et al.* Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum. Proc. Natl Acad. Sci. USA* **95,** 12352–12357 (1998).
20. Ho, C. K. & Shuman, S. A yeast-like mRNA capping apparatus in *Plasmodium falciparum. Proc. Natl Acad. Sci. USA* **98,** 3050–3055 (2001).
21. Francis, S. E., Sullivan, D. J. Jr & Goldberg, D. E. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum. Annu. Rev. Microbiol.* **51,** 97–123 (1997).

22. Singh, A. & Rosenthal, P. J. Comparison of efficacies of cysteine protease inhibitors against five strains of *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **45,** 949–951 (2001).

23. Uren, A. G. *et al.* Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol. Cell* **6,** 961–967 (2000).

24. Brown, G. V. & Rogerson, S. J. in *Malaria Vaccine Development: a Multi-Immune Response Approach* (ed. Hoffman, S. L.) 145–166 (ASM Press, Washington DC, 1996).

25. Duffy, P. E., Craig, A. G. & Baruch, D. I. Variant proteins on the surface of malaria-infected erythrocytes—developing vaccines. *Trends Parasit.* **17,** 354–356 (2001).

26. Su, X. Z. *et al.* The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82,** 89–100 (1995).

27. Kyes, S. A., Rowe, J. A., Kriek, N. & Newbold, C. I. Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **96,** 9333–9338 (1999).

28. Trenholme, K. R. *et al. clag9*: A cytoadherence gene in *Plasmodium falciparum* essential for binding of parasitized erythrocytes to CD36. *Proc. Natl Acad. Sci. USA* **97,** 4029–4033 (2000).

29. del Portillo, H. A. *et al.* A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* **410,** 839–842 (2001).

30. Deitsch, K. W., Calderwood, M. S. & Wellems, T. E. Cooperative silencing elements in *var* genes. *Nature* **412,** 875–876 (2001).

31. Mayer, D. C., Kaneko, O., Hudson-Taylor, D. E., Reid, M. E. & Miller, L. H. Characterization of a *Plasmodium falciparum* erythrocyte-binding protein paralogous to EBA-175. *Proc. Natl Acad. Sci. USA* **98,** 5222–5227 (2001).

32. Narum, D. L., Fuhrmann, S. R., Luu, T. & Sim, B. K. A novel *Plasmodium falciparum* erythrocyte binding protein-2 (EBP2/BAEBL) involved in erythrocyte receptor binding. *Mol. Biochem. Parasitol.* (in the press).

33. Preiser, P. R., Jarra, W., Capiod, T. & Snounou, G. A rhoptry-protein-associated mechanism of clonal phenotypic variation in rodent malaria. *Nature* **398,** 618–622 (1999).

34. Triglia, T., *et al.* Identification of proteins from *Plasmodium falciparum* that are homologous to reticulocyte binding proteins in *Plasmodium vivax*. *Infect. Immun.* **69,** 1084–1092 (2001).

35. Black, C. G., Wu, T., Wang, L., Hibbs, A. R. & Coppel, R. L. Merozoite surface protein 8 of *Plasmodium falciparum* contains two epidermal growth factor-like domains. *Mol. Biochem. Parasitol.* **114,** 217–226 (2001).

36. Kaslow, D. C. *et al.* A vaccine candidate from the sexual stage of human malaria that contains EGF-like domains. *Nature* **333,** 74–76 (1988).

37. Sultan, A. A. *et al.* TRAP is necessary for gliding motility and infectivity of plasmodium sporozoites. *Cell* **90,** 511–522 (1997).

38. Templeton, T. J., Kaslow, D. C. & Fidock, D. A. Developmental arrest of the human malaria parasite *Plasmodium falciparum* within the mosquito midgut via CTRP gene disruption. *Mol. Microbiol.* **36,** 1–9 (2000).

39. Trexler, M., Banyai, L. & Patthy, L. The LCCL module. *Eur. J. Biochem.* **267,** 5751–5757 (2000).

40. Ponting, C. P. Chlamydial homologues of the MACPF (MAC/perforin) domain. *Curr. Biol.* **9,** R911–R913 (1999).

41. Carucci, D. J. & Hoffman, S. L. The Malaria Genome Project: from sequence to functional analysis. *Nature Med.* **6** <http://www.nature.com/nm/special_focus/malaria/commentaries/carucci.html> (2000).

42. Hayward, R. E. *et al.* Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Mol. Microbiol.* **35,** 6–14 (2000).

43. Hayward, R. E. *Plasmodium falciparum* phosphoenolpyruvate carboxykinase is developmentally regulated in gametocytes. *Mol. Biochem. Parasitol.* **107,** 227–240 (2000).

44. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* **405,** 837–846 (2000).

45. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409,** 928–933 (2001).

46. Stephens, J. C. *et al.* Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293,** 489–493 (2001).

47. Subramanian, G., Adams, M. D., Venter, J. C. & Broder, S. Implications of the human genome for understanding human biology and medicine. *J. Am. Med. Assoc.* **286,** 2296–2307 (2001).

48. Hill, A. V. The genomics and genetics of human infectious disease susceptibility. *Annu. Rev. Genomics Hum. Genet.* **2,** 373–400 (2001).

49. Kwiatkowski, D. The molecular genetic approach to malarial pathogenesis and immunity. *Parassitologia* **41,** 233–240 (1999).

50. Knight, J. C. *et al.* A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nature Genet.* **22,** 145–150 (1999).

51. Fernandez-Reyes, D. *et al.* A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum. Mol. Genet.* **6,** 1357–1360 (1997).

52. Pain, A. *et al.* A non-sense mutation in Cd36 gene is associated with protection from severe malaria. *Lancet* **357,** 1502–1503 (2001).

53. Benedict, M. Q. & Chang, H. Rapid isolation of anopheline mosquito eye-colour mutants, based on larval colour change. *Med. Vet. Entomol.* **10,** 93–96 (1996).

54. Mukabayire, O. & Besansky, N. J. 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 (1996).

55. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287,** 2185–2195 (2000).

56. Dimopoulos, G. *et al. Anopheles gambiae* pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines. *Proc. Natl Acad. Sci. USA* **97,** 6619–6624 (2000).

57. Collins, F. H. *et al.* Genetics in the study of mosquito susceptibility to *Plasmodium*. *Parassitologia* **41,** 163–168 (1999).

58. Zheng, L. *et al.* Quantitative trait loci for refractoriness of *Anopheles gambiae* to *Plasmodium cynomolgi B*. *Science* **276,** 425–428 (1997).

59. Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287,** 2204–2215 (2000).

60. Schneider, D. & Shahabuddin, M. Malaria parasite development in a *Drosophila* model. *Science* **288,** 2376–2379 (2000).

61. Ribeiro, J. M. Blood-feeding arthropods: live syringes or invertebrate pharmacologists? *Infect. Agents Dis.* **4,** 143–152 (1995).

62. Bolshakov, V. N. *et al.* A comparative genomic analysis of two distant Diptera, the fruit fly, *Drosophila melanogaster*, and the malaria mosquito, *Anopheles gambiae*. *Genome Res.* **12,** 57–66 (2002).

63. Hoffman, S. L. Research (genomics) is crucial to attacking malaria. *Science* **290,** 1509 (2000).

64. Hoffman, S. L. *et al.* Protection of humans against malaria by immunization with radiation attenuated *Plasmodium* sp. sporozoites. *J. Infect. Dis.* (in the press).

65. Schofield, L. *et al.* Gamma interferon, CD8$^+$ T cells and antibodies required for immunity to malaria sporozoites. *Nature* **330,** 664–666 (1987).

66. Weiss, W. R., Sedegah, M., Beaudoin, R. L., Miller, L. H. & Good, M. F. CD8$^+$ T cells (cytotoxic/suppressors) are required for protection in mice immunized with malaria sporozoites. *Proc. Natl Acad. Sci. USA* **85,** 573–576 (1988).

67. Doolan, D. L. & Hoffman, S. L. The complexity of protective immunity against liver-stage malaria. *J. Immunol.* **165,** 1453–1462 (2000).

68. Sim, B. K. *et al.* Induction of biologically active antibodies in mice, rabbits, and monkeys by *Plasmodium falciparum* EBA-175 region II DNA vaccine. *Mol. Med.* **7,** 247–254 (2001).

69. Bouharoun-Tayoun, H., Oeuvray, C., Lunel, F. & Druilhe, P. Mechanisms underlying the monocyte-mediated antibody-dependent killing of *Plasmodium falciparum* asexual blood stages. *J. Exp. Med.* **182,** 409–418 (1995).

70. Hoffman, S. L., Rogers, W. O., Carucci, D. J. & Venter, J. C. From genomics to vaccines: malaria as a model system. *Nature Med.* **4,** 1351–1353 (1998).

71. Collins, F. H., Kamau, L., Ranson, H. A. & Vulule, J. M. Molecular entomology and prospects for malaria control. *Bull. World Health Organ.* **78,** 1412–1423 (2000).

72. Hemingway, J. & Ranson, H. Insecticide resistance in insect vectors of human disease. *Annu. Rev. Entomol.* **45,** 371–391 (2000).

73. Gentile, G., Slotman, M., Ketmaier, V., Powell, J. R. & Caccone, A. Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae* s.s. *Insect Mol. Biol.* **10,** 25–32 (2001).

74. Enserink, M. Malaria research: two new steps toward a 'better mosquito'. *Science* **293,** 2370–2371 (2001).

75. Catteruccia, F. *et al.* Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* **405,** 959–962 (2000).

76. Mills, C. D., Burgess, D. C., Taylor, H. J. & Kain, K. C. Evaluation of a rapid and inexpensive dipstick assay for the diagnosis of *Plasmodium falciparum* malaria. *Bull. World Health Organ.* **77,** 553–559 (1999).

77. Djimde, A. *et al.* A molecular marker for chloroquine-resistant falciparum malaria. *N. Engl. J. Med.* **344,** 257–263 (2001).

78. Kublin, J. G. *et al.* Molecular assays for surveillance of antifolate-resistant malaria. *Lancet* **351,** 1629–1631 (1998).

79. Fortin, A. *et al.* Identification of a new malaria susceptibility locus (Char4) in recombinant congenic strains of mice. *Proc. Natl Acad. Sci. USA* **98,** 10793–10798 (2001).

80. Roses, A. D. Pharmacogenetics and the practice of medicine. *Nature* **405,** 857–865 (2000).

81. Dervan, P. B. & Burli, R. W. Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.* **3,** 688–693 (1999).

82. Menard, R. & Janse, C. Gene targeting in malaria parasites. *Methods* **13,** 148–157 (1997).

83. Mota, M. M., Thathy, V., Nussenzweig, R. S. & Nussenzweig, V. Gene targeting in the rodent malaria parasite *Plasmodium yoelii*. *Mol. Biochem. Parasitol.* **113,** 271–278 (2001).

84. Caffrey, C. R., Scory, S. & Steverding, D. Cysteine proteinases of trypanosome parasites: novel targets for chemotherapy. *Curr. Drug Targets.* **1,** 155–162 (2000).

85. Curtis, C. F. Infectious disease. The case for deemphasizing genomics in malaria control. *Science* **290,** 1508 (2000).