PERSPECTIVES

INNOVATION

Applied systems biology and malaria

Elizabeth A. Winzeler

Abstract | One of the goals of systems-biology research is to discover networks and interactions by integrating diverse data sets. So far, systems-biology research has focused on model organisms, which are well characterized and therefore suited to testing new methods. Systems biology has great potential for use in the search for therapies for disease. Here, the potential of systems-biology approaches in the search for new drugs and vaccines to treat malaria is examined.

Systems biology, also known as integrative biology, combines information from many different sources, including gene-expression data, two-hybrid genetic-interaction data, protein-interaction data from the systematic analysis of protein complexes, and protein-expression data. This enables gene functions to be predicted and networks to be elucidated. Through the use of genome sequences, automation and parallel technologies such as microarrays, large-scale data sets containing consistent, high-quality data can be created, assembled and evaluated at significant cost savings over single-gene analyses. Novel networks can also be revealed. Some of the genome-scale data sets that are present in model organisms such as *Saccharomyces cerevisiae* (TABLE 1) have resulted from the phenotypic analysis of a set of deletion strains for every gene[1,2], two-hybrid screens[3,4], gene-expression analyses[5,6], protein-complex co-precipitation studies[7], subcellular-localization and protein-expression studies[8,9], comparative sequencing and analysis of genetic diversity[10,11], and bioinformatic analysis of pathways[12]. Recent research in yeast has concentrated on synthesis of the combined data to characterize networks and protein interactions[13,14]. Because the yeast genome is well annotated, it is an excellent model for the development of new systematic methods. The challenge will be to translate these methods to organisms that affect human health, such as the human malaria parasite *Plasmodium falciparum*. This article describes how these approaches are accelerating malaria research, and speculates on how they might lead to the

development of new therapies or vaccines for malaria.

As the causative agents of human malaria, *Plasmodium* parasites are important contributors to the global morbidity and mortality rates — 300–500 million cases of malaria occur each year. In Africa, it is estimated that malaria causes approximately 1.5–2.7 million deaths annually, primarily in children under five (reviewed in REF. 15), and the disease also poses an important health threat to travellers. The economic cost of malaria to the developing world is enormous. It has been estimated that the gross national product per capita is reduced by more than 50% in malarious countries compared with non-malarious ones[16].

A malaria vaccine would impact the lives of many people who live or travel in disease-endemic areas, and much effort has been devoted to this end. The feasibility of a human-malaria vaccine is discussed in REFS 17,18. Studies have shown that infective, sporozoite-stage parasites (FIG. 1), which had been irradiated so that they could invade, but not replicate, in the host, could elicit an immune response that subsequently protected from reinfection[19,20]. Also, humans who live in malaria-endemic regions gradually cease to show symptoms of the disease, although parasites can still be found in their bloodstream. The observation that administration of gamma globulin isolated from adults with so-called 'naturally acquired' immunity can reduce parasitaemia in recipients is consistent with the development of antigen-specific acquired immunity in humans[21]. However, although a recent subunit-vaccine

formulation based on the sporozoite-expressed circumsporozoite protein (CSP) has shown some promise[22], there is currently no licensed vaccine despite extensive testing of several candidates[23]. There are several reasons why the development of a malaria vaccine poses a considerable challenge: blood-stage parasites alter the complement of antigens that are on the surface of the red blood cell, or on the surface of the merozoite during its brief extracellular phase; the parasite population is genetically diverse; our understanding of the human immune response to malaria is incomplete; or the right antigen has yet to be discovered.

Although vector-control strategies such as bednets can reduce the spread of malaria, drugs remain crucial weapons for preventing infections as well as reducing transmission, symptoms and mortality from the disease. However, resistance to inexpensive drugs such as chloroquine has emerged and has spread quickly, and multidrug-resistant *Plasmodium* strains are now common (reviewed in REF. 24).

Systems biology for malaria

To rationally develop new therapies for malaria, molecular details of parasite stage-specific development need to be understood. This is a difficult task compared with many other microbial pathogens. *Plasmodium* has a complex life cycle, and although *P. falciparum*, the species responsible for most human deaths, can be cultured in human erythrocytes[25], many stages of its life cycle cannot be easily maintained in cell culture. Cell culture of the mosquito stages has been described for rodent parasites[26], but is not yet widely used. Other human *Plasmodium* species such as *Plasmodium vivax*, *Plasmodium malariae* and *Plasmodium ovale* also cause disease, and these species are even more refractory to experimental manipulation. *P. vivax* can only be cultured in human reticulocytes, which are difficult to obtain[27]. Transient transfection and gene disruption has been achieved in several *Plasmodium* species[28–31] (see Supplementary information S1 (table)), but the process is labour intensive and, until recently, inefficient. Because the malaria parasites are haploid for most of their life cycle and because stable transfection has

been reported only for erythrocytic stages, mutants that bear deleterious mutations in genes that are essential for parasite growth in erythrocytes are not easily recovered. The DNA of most *Plasmodium* species is AT rich (up to 90% in non-coding regions), and long, homopolymeric tracts of A and T recombine when cloned in *Escherichia coli* or are unclonable in some circumstances, making vector construction, subcloning and sequencing time-consuming (reviewed in REF. 32).

When the complete genome sequence of *P. falciparum* was determined in 2002 (REF. 33), it was shown to encode more completely uncharacterized genes ('hypothetical proteins') than other lower single-celled eukaryotes such as *S. cerevisiae*, in which 75% of putative protein-coding open reading frames are now characterized[33–36]. At least 65% of the genes identified in the *P. falciparum* project are described as 'hypothetical', indicating that they showed no significant homology to characterized genes from other species. Many others are described only as a 'kinase' or 'putative cysteine protease'. Although divergence from the ancestral eukaryotic tree might have reduced the number of statistically significant matches to characterized, homologous genes from other species when conventional search methods are used[37], the low number of homologous genes might also reflect historically low levels of malaria-research funding. Also, some genes might be involved in

parasite-specific processes that are not found in other model organisms. Because of difficulties in malaria research, the number of genes that remain hypothetical is unlikely to be reduced quickly if conventional gene-by-gene approaches are used.

Having as much information as possible about potential drug targets, such as proteases, kinases or enzymes, is fruitful in informing the optimal selection of truly novel targets from this list of hypothetical or uncharacterized genes. Although finding chemical inhibitors of known, validated targets, such as dihydrofolate reductase, can lead to new drugs, in many cases the lack of chemical diversity in libraries used in drug-screening efforts can limit the development of novel classes of antimalarials against known targets. Finding a novel chemical scaffold would be ideal, because parasite resistance to one class of compounds might result in cross-resistance to similar classes.

The availability of genome sequences has been a tremendous boon to malaria researchers and has already facilitated the search for new drugs. Through comparative genome analysis, enzymatic pathways have been discovered that are not found in humans. Researchers in public–private partnerships (the Medicines for Malaria Venture) are working to discover novel inhibitors to interrupt some of these pathways, such as the non-mevalonate pathway of isoprenoid biosynthesis[36].

Expression analysis and gene function

For sequenced microorganisms, DNA microarrays can be used to inexpensively determine the expression programme of almost every gene in the genome. The expression programme can be used to begin to predict the functions of uncharacterized genes. For example, in both prokaryotes and eukaryotes, genes that encode components of multiprotein complexes or pathways are often transcriptionally co-regulated. In prokaryotes, such genes are members of an operon, with a single promoter. In eukaryotes, co-regulated genes might have similar promoter elements in their upstream regions. In *S. cerevisiae*, reliable functional annotations are available for most genes, so that the hypothesis that genes encoding proteins involved in similar processes are co-transcribed can be mathematically validated. Indeed, in *S. cerevisiae*, genes that encode members of protein complexes such as the ribosome, the proteasome and the DNA polymerase complex or genes involved in metabolic pathways such as electron transport are under tight transcriptional control and show similar patterns of induction or repression under different conditions[5,6]. Gene function might be assigned in *P. falciparum* by examining gene-expression profiles in different conditions and sorting genes into groups (clusters) that have similar expression patterns. If some gene functions are known, others can be inferred. For example, the rhoptry is a specialized

Table 1 | Methods for generating data that can be used in systematic analyses

Method	Description	Feasibility in *Plasmodium falciparum*	Refs
Gene expression	Transcript levels for all genes are determined for different life-cycle stages or conditions.	Yes	38,39
Systematic two-hybrid analysis	Libraries of cDNAs or amplified gene products are fused to either the GAL4 transcription-factor activation domain or the GAL4 DNA-binding domain of a transcription factor. If the two domains interact, transcriptional activity at a reporter gene will be reconstituted. Crosses between all prey and bait combinations are performed in yeast and scored for promoter activity.	Yes	50
Systematic protein-complex analysis	Genes are engineered so that proteins bear tags that are recognized by high-affinity antibodies. Recombinant genes are introduced into the genome. Protein complexes are immunoprecipitated with an antibody to the tag. The protein complexes are then analysed by mass spectrometry.	Likely to work	7
Localization	Genes are fused to easily visualized reporter genes and introduced into the cell. Reporter is immunolocalized.	Yes	9,75
Comparative sequencing	Sequencing of different species of the same genus can reveal conserved regulatory regions.	Yes	34,35
Shotgun proteomics	Whole proteomes are digested and analysed using liquid chromatography and mass spectrometry. Used to determine the relative protein abundance for most genome-encoded proteins.	Yes	42,45
Protein arrays	Proteins are spotted on solid supports. Used to identify the targets of a kinase or to study protein–protein interactions.	*P. falciparum* proteins difficult to overexpress	76
Array-based ChIP	Used to study the genome-wide binding patterns of proteins.	Likely to work for abundant proteins	
Systematic mutagenesis	Gene disruptions are either created for every gene based on sequence and gene annotations or created randomly and mapped to the chromosomal locus. Phenotypic data are systematically collected for mutants.	Increases in transfection efficiency might allow this approach in the future	2,77

ChIP, chromatin immunoprecipitation.

invasion organelle in *P. falciparum*. Clusters of genes that contain known rhoptry genes could be identified to pinpoint unassigned potential rhoptry genes.

Gene-expression profiling is without doubt one of the most efficient approaches to extract large amounts of functional information. With amplification methods, ample quantities of parasite RNA (10–100 ng) can be extracted from almost every stage of the organism's life cycle[35,38,39]. In contrast to methods such as sequencing cDNA libraries or differential display, these data sets are comprehensive, which is the main power of the approach. It is just as important to know when a gene is not expressed as when it is expressed. In other single-celled parasites, such as trypanosomes, RNA editing has an important role in the control of protein expression. This has led to speculation that gene expression might be regulated differently in *P. falciparum* compared with other protozoan species[40], especially as the *P. falciparum* genome seems to encode fewer conventional transcription factors and more RNA-binding proteins[41]. Preliminary analysis of gene-expression patterns has indicated that, similar to *S. cerevisiae*, transcripts that encode ribosomal proteins and other multiprotein complexes are tightly regulated during the erythrocytic life cycle although the genes are distributed across all 14 chromosomes[39]. Although adjacent co-regulated genes are observed[39,42] there is little evidence for a higher frequency of adjacent co-regulated genes than in *S. cerevisiae*[6]. As discussed below, there is preliminary evidence that genes with common expression patterns might also share sequence motifs upstream of their promoters.

There are probably limits to the amount of information revealed by expression profiling, as some proteins are likely to be controlled at the level of transcription, others at the level of RNA stability, others at the translational level[35,43] and others at the activity level[44]. Comprehensive proteomic analyses of different life-cycle stages of the *P. falciparum* and *Plasmodium berghei* genomes have been published[35,42,45,46]. In these so-called 'shotgun proteomics' experiments, protein extracts derived from relatively pure life-cycle-stage populations are digested and fractionated using chromatography or gel electrophoresis. The proteins are then analysed using tandem mass spectrometry. The sequences of detected peptides are then compared with predicted proteins. Although it might be difficult to infer much from proteins that are detected once or twice across the malaria life cycle,
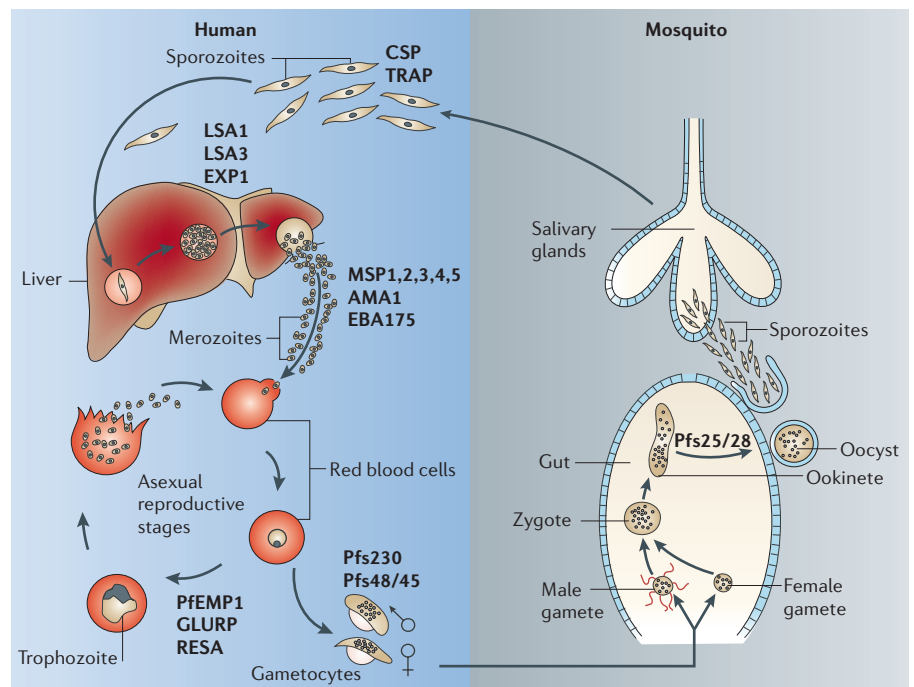

Figure 1 | **The malaria life cycle.** *Plasmodium* species are parasites of red blood cells and hepatocytes. Haploid sporozoites are injected into the vertebrate host when a female mosquito takes a blood meal. They rapidly invade hepatocytes, where they undergo asexual multiplication, generating several thousand merozoites. Merozoites released from hepatocytes then invade erythrocytes, where they again multiply and mature progressively from ring-stage parasites to trophozoites to schizonts. The erythrocyte eventually ruptures, releasing 8–32 new merozoites, which can infect new erythrocytes. The asexual cycle takes about 42–48 hours, during which time the human host has periodic cycles of fever and chills. In response to a cue that is not understood, some of the parasites exit the asexual cycle and form male and female gametocytes in a process known as gametocytogenesis. The mosquito takes up the mature sexual forms, and sexual reproduction of the parasite occurs in the insect's midgut. Sporozoites released from the oocyst move to the mosquito salivary glands, which facilitates parasite transmission. Pre-erythrocytic vaccines, which would prevent infection, are targeted against proteins that are present in the sporozoite or liver stages, for example, circumsporozoite protein (CSP), thrombospondin-related adhesion protein (TRAP), exported protein 1 (EXP1), liver-stage antigen-1 (LSA1) and LSA3 (REF. 15). Erythrocytic vaccines, which would reduce the severity of the disease, target proteins that are expressed in the merozoite phase of the life cycle: apical membrane antigen-1 (AMA1), merozoite surface protein-1 (MSP1), MSP2–5, *P. falciparum* erythrocyte membrane protein-1 (PfEMP1), ring-infected erythrocyte surface antigen precursor (RESA), erythrocyte-binding antigen (EBA) and glutamate rich protein (GLURP)[15]. Furthermore, some research has focused on developing vaccines against sexual-stage antigens, which would not reduce symptoms but could reduce transmission of the disease (Pfs230, Pfs48/45, Pfs25/28)[15]. All drugs target erythrocytic parasites, but some are active against liver and sexual stages as well[24]. Figure modified with permission from REF. 40 © (2002) Macmillan Magazines Ltd.

this analysis can provide information about potential protein function if multiple high-confidence spectra are detected in just one (stage-specific) or in all (housekeeping) stages. Identifying the proteins in purified subcellular organelles, such as the rhoptries, provides additional functional data[47].

A significant challenge to the identification of novel malarial drug targets lies in identifying genes that are likely to be essential to parasite viability in erythrocytes. Drugs are seldom 100% effective at inhibiting protein activity and, therefore, if the protein is not crucial for survival, it will not be a good target. This is more

important than determining whether or not a homologous enzyme is found in humans, because chemists can usually engineer small molecules that are selective for parasite protein and not human protein, providing a therapeutic window. Many effective and widely used antifungal drugs, for instance, target proteins that have close human homologues. Because *Plasmodium* is haploid for most of its life cycle, and stable integration of disruption constructs by stable transfection and homologous recombination has only been reported in erythrocytic stages, definitive genetic proof of a gene's essential function is difficult to

PERSPECTIVES

Box 1 | *Plasmodium*: a model genus for comparative genomics

One advantage that malaria parasites have over larger model organisms is that they present opportunities for sophisticated studies in comparative genomics. In addition to the complete sequence for the genome of *Plasmodium falciparum*[33], 5 × shotgun coverage of several rodent-parasite genomes are available[34,35]. Partial shotgun sequences of *Plasmodium vivax*, the non-human primate malaria parasites *Plasmodium knowlesi* and *Plasmodium reichenowi* and the avian parasite *Plasmodium gallinaceum* are likely to be published soon (see Online links box).

   Comparisons across species can be used to identify genes that are rapidly evolving and that might interact with the host immune system. Comparisons can also identify genes that are conserved and probably essential, and might therefore be promising drug targets. Comparative genome sequencing can also be used to identify phylogenetic footprints in regulatory regions. The same promoter motifs that are associated with common gene-expression patterns in *Saccharomyces cerevisiae* are often conserved in the upstream regions of homologous genes from related yeast species[10]. Similarly, the importance of the presence of a motif in *P. falciparum* can be determined by monitoring conservation in the related rodent parasites *Plasmodium yoelii* and *Plasmodium berghei*.

   Every individual gene-expression measurement that has been produced with an array cannot be verified independently, and false positives might exist, especially for weakly expressed genes. However, cross-species expression data can be used to gain confidence in the large-scale microarray results[72]. The expression programme of the rodent parasite *P. berghei* has been published[35]. Most of the genes in the human parasites have homologues in *P. berghei* and are expected to be active at similar times in the parasite life cycle. It is likely that when these data are eventually merged with the *P. falciparum* data, a more accurate list of biologically relevant genes will emerge. Furthermore, because transfections of rodent parasites are more efficient than human parasites[73], we can address questions that are more difficult or impossible to study in human parasites. Related apicomplexan parasites, such as *Toxoplasma gondii*, have many of the same specialized sub-cellular organelles as malaria parasites and are easier organisms to study in the laboratory[74].

obtain. The use of a tetracycline-regulatable promoter has been reported, and this might eventually allow gene-dosage experiments to be carried out[48]. On the other hand, essential proteins are more likely to be conserved in evolution[49] and to have more interaction partners in protein–protein interaction networks. A comprehensive two-hybrid study in which 32,000 searches of *P. falciparum* baits against *P. falciparum* activation domains were carried out identified 2,846 protein–protein interactions[50]. Whereas having many interacting partners might be an indication that a protein is essential for viability, the identities of a protein's interacting partners might provide clues about the protein's function. For instance, if a protein interacts with a known gene that is involved in invasion, has an expression pattern that mimics other genes involved in invasion, and carries promoter motifs that indicate that it might be co-regulated with other invasion genes, it is probable that this protein is involved, directly or indirectly, in invasion.

   The power of systems biology in malaria is that it allows accurate predictions of gene function and serves as a complement for powerful traditional genetic methods that have proven so useful in model systems. However, tools are needed that can assign probabilities to predictions and that can flag interactions that are probably spurious.

Blocking transmission and development

In response to a molecular signal that has not yet been identified, some blood-stage parasites exit the asexual erythrocytic cycle and undergo sexual development, differentiating into male and female gametocytes, which are taken up by the mosquito in a blood meal. These sexual forms of *Plasmodium* are responsible for transmission of the disease from one person to the next. An ideal antimalarial drug or drug combination would both kill rapidly multiplying asexual parasites and interfere with sexual development to block the transmission cycle. Signalling and developmental cascades such as those that might occur during sexual development or in the vector stages of the *Plasmodium* life cycle are common points of therapeutic intervention in humans. Because classic forward-genetic strategies cannot easily be applied to *P. falciparum*, identification of genes with important roles in *Plasmodium* developmental biology has been difficult, and compared with model organisms we know little about the identity of proteins that regulate growth and development. Knowing the identity of such proteins could provide new strategies for malaria control. For example, G-protein-coupled receptors and nuclear receptors, which bind small molecules and ultimately change transcriptional patterns, are some of the most important classes of drug targets in humans.

It is likely that expression data from *P. falciparum* can be used to find novel transcription-factor-binding sites (motifs) that are associated with the transcriptional changes that occur during sexual development or other life-cycle stages. Sequences bearing these motifs could then be used to biochemically purify potential regulatory proteins. Using gene-expression information to find sequence motifs that are involved in controlling transcription has worked well in many other microorganisms, including bacteria[51] and *S. cerevisiae*[52], in which most (but not all) known transcription-factor-binding sites can be rediscovered. A comprehensive review and assessment of these techniques has recently been published[53]. Genes are first grouped by their common expression pattern[52] or by their common function[54]. Then, the regions upstream of the start codon are systematically searched for all overrepresented sequences (such as ATGGAC) or by deterministic methods, such as MEME[10]. In *Plasmodium*, site-directed mutagenesis of promoter sequences is considerably more difficult than in yeast. Therefore, expression profiling combined with bioinformatics analysis and comparative genomics (BOX 1) is probably the most efficient way to identify specific motifs that can be used to construct stage-dependent reporters of gene activity or to identify the transcriptional regulatory protein that binds to the motif. A search of the 1,000-base-pair regions upstream of the start codons of genes that are expressed during sexual development in *P. falciparum* identified a statistically over-represented novel palindromic sequence (TGTANNTACA)[55]. When compared with the frequency of the motif in all other upstream regions in the genome, the probability of this enrichment by chance is roughly 1 in 10^{24}. The motif is also statistically enriched upstream of homologous genes in *P. berghei*, *Plasmodium yoelii* and *Plasmodium chabaudi*, and it is preferentially found about 400–600 nucleotides upstream of the ATG in co-regulated genes but shows a random distribution in unregulated genes. Previous random exonuclease deletion mapping of a promoter carrying this motif[56] showed that its disruption almost completely abolished promoter activity, providing strong evidence of its functional significance and the power of the bioinformatic approach. The next goal is to find the protein that binds to this or other motifs. If the activity of this transcription factor is controlled by the binding of a small molecule or a metabolite,

an analogue of the small molecule could be used as a transmission-blocking drug. If other uncharacterized motifs are found to control the transcriptional regulation of known drug targets, other members of the regulatory pathway might be discovered among uncharacterized genes if they also bear the motif.

Motifs can also be found by examining the proteome. Although quantitative proteomic methods that rely on enzymatic or metabolic labelling have not gained widespread use in malaria parasites, estimates of protein abundance can be determined by counting the number of peptide spectra that are detected for a protein using mass spectrometry for a particular life-cycle stage. Two studies have sought to compare protein and transcript levels in *P. falciparum* and *P. berghei*[35,43]. Both studies found a delay between the time a transcript appeared and the time a cognate protein appeared for some life-cycle stages, and concluded that post-transcriptional gene silencing might have a role in control of translation during sexual development. Furthermore, analysis of the upstream untranslated regions of *P. berghei* genes that are subject to post-transcriptional gene silencing revealed the presence of a 47-nucleotide sequence that was overrepresented among these genes[35]. The hypothesis that there might be a translational delay makes sense because the parasite must undergo a rapid morphological transition once it leaves the mammalian host and enters the mosquito. Therefore, the parasite might store transcripts in preparation for this transition. If a drug was identified that could relieve translational repression in maturing sexual stages, it might effectively sterilize these parasites, leaving them unable to complete their life cycle and preventing transmission.

Systems biology and vaccine development

Although discovering appropriate formulations and adjuvants will probably have a key role in the development of a successful vaccine, most research has tended to focus on a fairly small class of historical antigens (FIG. 1). The acquisition and analysis of genome-scale data sets through systems biology might assist in the identification of novel vaccine targets, which could theoretically be more protective. The study of gene-expression and proteomic data sets can reveal which of the 5,300 *Plasmodium* proteins are expressed in the invasive stages of the *Plasmodium* life cycle (sporozoites or merozoites) and might be the

most promising antigens. Doolan *et al.*[57] used proteomic and genomic data to identify 27 potential sporozoite-stage antigens. To determine whether these proteins were antigenic, peripheral blood mononuclear cells (PBMCs) were obtained from volunteers that had previously been immunized with an irradiated-sporozoite vaccine and subsequently challenged with infectious mosquitoes. The antigens were tested for their ability to induce and recall an *ex vivo* interferon-γ response in the PBMCs of these volunteers. Of 27 proteins, 16 were antigenic using this assay — in some cases, the proportion of volunteers who showed a response to the antigen was higher than for antigens used in clinical trials, such as the CSP[57]. PBMCs of all the immunized volunteers reacted with one uncharacterized protein, PFL0800c, discovered by this genomic approach[57]. Both transcription[39] and proteomic[42] data indicate that this protein is one of the most highly expressed in the sporozoite stages.

Although data on stage-specific expression and the levels of expression in proteomic and transcription data allow the selection of novel proteins for further characterization as vaccine candidates, further information is provided by studying the frequency of polymorphisms in different genes across *Plasmodium* strains. *Plasmodium* is known to vary the presentation of antigens in successive generations[58,59]. Some of the antigenic variation is probably the result of mitotic recombination in members of multigene families, such as the mostly subtelomeric *var* genes, which encode versions of the *P. falciparum* erythrocyte membrane protein-1 (PfEMP1)[60]. Most of these proteins are weakly expressed, with a selective silencing mechanism in place to ensure that only one member of the family is fully transcribed at any time[61]. However, other *P. falciparum* immunogens, such as the CSP, the merozoite surface protein-1 (MSP1) or the apical membrane antigen-1 (AMA1), are not members of multigene families, are centromere proximal and are expressed at high levels. Levels of genetic variability in the introns of housekeeping genes are low in *P. falciparum*[62] but in these immunogens, variability is exceptionally high[63–65], implying that these genes are under intense selection pressure from the host. These and similar highly variable, uncharacterized proteins might be the targets of human naturally acquired immunity. Indeed, in *P. chabaudi*, mice that have been infected with one MSP1

variant cannot be reinfected by a different strain that carries the original MSP1 allele[66]. Genome-wide hybridization methods have been used to rapidly and inexpensively characterize genetic variability in yeast[67], and this method also works in *P. falciparum*[68], potentially providing an easy way to identify genes under selection across strains.

Outlook

Although systems biology holds significant promise, challenges remain. It is theoretically feasible to create, at least in rodent parasites, lines that express epitope-tagged or green-fluorescent-protein-tagged versions of all parasite proteins, and this could allow the rapid characterization of protein complexes or permit systematic, informative protein-localization studies. Indeed, by studying the localization and expression patterns of a handful of genes, and by comparing their protein sequences, a short amino-acid motif that directs a nascent protein out of the parasite to the surface of the red blood cell was discovered[69,70].

The sequential disruption of genes is time consuming, and thought should be given to devising strategies for systematically creating and phenotyping mutant *P. falciparum* strains. Because *Plasmodium* genes are more difficult to clone and overexpress than genes from model organisms, there might be exceptional difficulties associated with creating arrays of *Plasmodium* recombinant proteins that can be used to examine kinase specificity or other assays of protein–protein interaction. However, the Gateway recombinational cloning system, which has been used in several genome-scale cloning efforts in humans, has recently been adapted to *P. falciparum*[71].

Ultimately, all of the gene-identification and gene-expression data will need to be assembled in a repository where it can be easily retrieved. The malaria-research community is fortunate to have a dedicated repository — PlasmoDB — for malaria sequence and functional-genomics data sets and, therefore, this is unlikely to be a large hurdle. As systems-biology data are integrated, malariologists will probably use PlasmoDB to determine the probability that a gene is involved in a defined process based on expression data, proteomic data, protein-interaction and protein-localization data. Comparative sequencing or comparative genome hybridizations might identify subsets of genes that are under selection pressure from the host or from drugs. Discovering novel host alleles that result in increased or reduced

susceptibility to malaria might reveal new interactions between parasite and host genes. Many of the approaches suggested here are not specific to *P. falciparum* but could be applied to a range of microbial pathogens. Researchers must keep in mind that more data is better, because the quality of a data set with 5,000 protein–protein interactions can be more readily evaluated with statistical methods than one that contains only 10 interactions. Support for this type of research, which might not involve testing a specific hypothesis, needs to be encouraged by the funding agencies. It might take several years before the impact of systems biology on malaria research can be fully realized because researchers will need to test the hypotheses using traditional methods and this is likely to be time consuming. Although every discovery might not lead directly to a new drug or a vaccine, the basic findings that are made through integrative biology will contribute to our understanding of *Plasmodium* biology, and this will serve as a foundation that will ultimately increase the speed at which new therapies can be developed.

Elizabeth A. Winzeler is at the Department of Cell Biology, ICND202, The Scripps Research Institute, La Jolla, California 92037, USA.
e-mail: winzeler@scripps.edu

1. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
2. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
3. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
4. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
5. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
6. Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
7. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
8. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
9. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
10. Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76 (2003).
11. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
12. Kelley, B. P. *et al.* Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl Acad. Sci. USA* **100**, 11394–11399 (2003).
13. Jansen, R. *et al.* A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science* **302**, 449–453 (2003).
14. Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88–93 (2004).
15. Greenwood, B. M., Bojang, K., Whitty, C. J. & Targett, G. A. Malaria. *Lancet* **365**, 1487–1498 (2005).
16. Sachs, J. & Malaney, P. The economic and social burden of malaria. *Nature* **415**, 680–685 (2002).
17. Good, M. F., Xu, H., Wykes, M. & Engwerda, C. R. Development and regulation of cell-mediated immune responses to the blood stages of malaria: implications for vaccine research. *Annu. Rev. Immunol.* **23**, 69–99 (2005).
18. Doolan, D. L. *et al.* Utilization of genomic sequence information to develop malaria vaccines. *J. Exp. Biol.* **206**, 3789–3802 (2003).
19. Nussenzweig, R. S., Vanderberg, J., Most, H. & Orton, C. Protective immunity produced by the injection of X-irradiated sporozoites of *Plasmodium berghei*. *Nature* **216**, 160–162 (1967).
20. Clyde, D. F., Most, H., McCarthy, V. C. & Vanderberg, J. P. Immunization of man against sporozite-induced falciparum malaria. *Am. J. Med. Sci.* **266**, 169–177 (1973).
21. Cohen, S., Mc, G. I. & Carrington, S. Gamma-globulin and acquired immunity to human malaria. *Nature* **192**, 733–737 (1961).
22. Alonso, P. L. *et al.* Efficacy of the RTS, S/AS02A vaccine against *Plasmodium falciparum* infection and disease in young African children: randomised controlled trial. *Lancet* **364**, 1411–1420 (2004).
23. Richie, T. L. & Saul, A. Progress and challenges for malaria vaccines. *Nature* **415**, 694–701 (2002).
24. Baird, J. K. Effectiveness of antimalarial drugs. *N. Engl. J. Med.* **352**, 1565–1577 (2005).
25. Trager, W. & Jensen, J. B. Human malaria parasites in continuous culture. *Science* **193**, 673–675 (1976).
26. Al-Olayan, E. M., Beetsma, A. L., Butcher, G. A., Sinden, R. E. & Hurd, H. Complete development of mosquito phases of the malaria parasite *in vitro*. *Science* **295**, 677–679 (2002).
27. Golenda, C. F., Li, J. & Rosenberg, R. Continuous *in vitro* propagation of the malaria parasite *Plasmodium vivax*. *Proc. Natl Acad. Sci. USA* **94**, 6786–6791 (1997).
28. Goonewardene, R. *et al.* Transfection of the malaria parasite and expression of firefly luciferase. *Proc. Natl Acad. Sci. USA* **90**, 5234–5236 (1993).
29. van Dijk, M. R., Waters, A. P. & Janse, C. J. Stable transfection of malaria parasite blood stages. *Science* **268**, 1358–1362 (1995).
30. Crabb, B. S. & Cowman, A. F. Characterization of promoters and stable transfection by homologous and nonhomologous recombination in *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **93**, 7289–7294 (1996).
31. Wu, Y., Kirkman, L. A. & Wellems, T. E. Transformation of *Plasmodium falciparum* malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *Proc. Natl Acad. Sci. USA* **93**, 1130–1134 (1996).
32. de Koning-Ward, T. F., Janse, C. J. & Waters, A. P. The development of genetic tools for dissecting the biology of malaria parasites. *Annu. Rev. Microbiol.* **54**, 157–185 (2000).
33. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
34. Carlton, J. M. *et al.* Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**, 512–519 (2002).
35. Hall, N. *et al.* A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* **307**, 82–86 (2005).
36. Jomaa, H. *et al.* Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573–1576 (1999).
37. Callebaut, I., Prat, K., Meurice, E., Mornon, J. P. & Tomavo, S. Prediction of the general transcription factors associated with RNA polymerase II in *Plasmodium falciparum*: conserved features and differences relative to other eukaryotes. *BMC Genomics* **6**, 100 (2005).
38. Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**, E5 (2003).
39. Le Roch, K. G. *et al.* Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508 (2003).
40. Wirth, D. F. Biological revelations. *Nature* **419**, 495–496 (2002).
41. Coulson, R. M., Hall, N. & Ouzounis, C. A. Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. *Genome Res.* **14**, 1548–1554 (2004).
42. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
43. Le Roch, K. G. *et al.* Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res.* **14**, 2308–2318 (2004).
44. Greenbaum, D. C. *et al.* A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* **298**, 2002–2006 (2002).
45. Lasonder, E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542 (2002).
46. Khan, S. M. *et al.* Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell* **121**, 675–687 (2005).
47. Sam-Yellowe, T. Y. *et al.* Proteome analysis of rhoptry-enriched fractions isolated from *Plasmodium* merozoites. *J. Proteome Res.* **3**, 995–1001 (2004).
48. Meissner, M. *et al.* Tetracycline analogue-regulated transgene expression in *Plasmodium falciparum* blood stages using *Toxoplasma gondii* transactivators. *Proc. Natl Acad. Sci. USA* **102**, 2980–2985 (2005).
49. Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. & Feldman, M. W. Evolutionary rate in the protein interaction network. *Science* **296**, 750–752 (2002).
50. LaCount, D. J. *et al.* A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* **438**, 103–107 (2005).
51. McGuire, A. M., Hughes, J. D. & Church, G. M. Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res.* **10**, 744–757 (2000).
52. Roth, F. P., Hughes, J. D., Estep, P. W. & Church, G. M. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nature Biotechnol.* **16**, 939–945 (1998).
53. Tompa, M. *et al.* Assessing computational tools for the discovery of transcription factor binding sites. *Nature Biotechnol.* **23**, 137–144 (2005).
54. Hughes, J. D., Estep, P. W., Tavazoie, S. & Church, G. M. Computational identification of *cis*-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **296**, 1205–1214 (2000).
55. Young, J. A. *et al.* The *Plasmodium falciparum* sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol. Biochem. Parasitol.* **143**, 67–79 (2005).
56. Dechering, K. J. *et al.* Isolation and functional characterization of two distinct sexual-stage-specific promoters of the human malaria parasite *Plasmodium falciparum*. *Mol. Cell. Biol.* **19**, 967–978 (1999).
57. Doolan, D. L. *et al.* Identification of *Plasmodium falciparum* antigens by antigenic analysis of genomic and proteomic data. *Proc. Natl Acad. Sci. USA* **100**, 9952–9957 (2003).
58. Smith, J. D. *et al.* Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101–110 (1995).
59. 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).
60. Freitas-Junior, L. H. *et al.* Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407**, 1018–1022 (2000).
61. Scherf, A. *et al.* Antigenic variation in malaria: *in situ* switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. *EMBO J.* **17**, 5418–5426 (1998).
62. Volkman, S. K. *et al.* Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* **293**, 482–484 (2001).
63. Polley, S. D. & Conway, D. J. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics* **158**, 1505–1512 (2001).
64. Miller, L. H., Roberts, T., Shahabuddin, M. & McCutchan, T. F. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). *Mol. Biochem. Parasitol.* **59**, 1–14 (1993).
65. de la Cruz, V. F., Lal, A. A. & McCutchan, T. F. Sequence variation in putative functional domains of the circumsporozoite protein of *Plasmodium falciparum*. Implications for vaccine development. *J. Biol. Chem.* **262**, 11935–11939 (1987).
66. Martinelli, A. *et al.* A genetic approach to the *de novo* identification of targets of strain-specific immunity in malaria parasites. *Proc. Natl Acad. Sci. USA* **102**, 814–819 (2005).

67. Winzeler, E. A. *et al.* Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
68. Volkman, S. K. *et al.* Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum*. *Science* **298**, 216–218 (2002).
69. Marti, M., Good, R. T., Rug, M., Knuepfer, E. & Cowman, A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933 (2004).
70. Hiller, N. L. *et al.* A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).
71. Aguiar, J. C. *et al.* High-throughput generation of *P. falciparum* functional molecules by recombinational cloning. *Genome Res.* **14**, 2076–2082 (2004).
72. Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255 (2003).
73. Janse, C. J. *et al.* High efficiency transfection of *Plasmodium berghei* facilitates novel selection procedures. *Mol. Biochem. Parasitol.* 7 Oct 2005 [epub ahead of print].
74. Roos, D. S. Genetics. Themes and variations in apicomplexan parasite biology. *Science* **309**, 72–73 (2005).
75. Tonkin, C. J. *et al.* Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. *Mol. Biochem. Parasitol.* **137**, 13–21 (2004).
76. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
77. Ross-Macdonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).

DATABASES
The following terms in this article are linked online to:
Entrez: http://www.ncbi.nlm.nih.gov/Entrez
Escherichia coli | *Plasmodium berghei* | *Plasmodium chabaudi* | *Plasmodium falciparum* | *Plasmodium gallinaceum* | *Plasmodium knowlesi* | *Plasmodium reichenowi* | *Plasmodium vivax* | *Plasmodium yoelii* | *Saccharomyces cerevisiae* | *Toxoplasma gondii*
UniProtKB: http://us.expasy.org/uniprot
AMA1 | CSP | MSP1 | PfEMP1 | PFL0800c

FURTHER INFORMATION
Elizabeth A. Winzeler's homepage:
http://www.scripps.edu/cb/winzeler
The Medicines for Malaria Venture:
http://www.mmv.org
MEME:
http://meme.sdsc.edu/meme/meme.html
PlasmoDB:
http://www.PlasmoDB.org
Protozoan genomes at the Sanger Institute:
http://www.sanger.ac.uk/Projects/Protozoa
The Institute for Genomic Research (TIGR):
http://www.tigr.org

SUPPLEMENTARY INFORMATION
See online article: S1 (table)
Access to this links box is available online.

OPINION

Helicobacter pylori phase variation, immune modulation and gastric autoimmunity

Mathijs Bergman, Gianfranco Del Prete, Yvette van Kooyk and Ben Appelmelk

Abstract | *Helicobacter pylori* can be regarded as a model pathogen for studying persistent colonization of humans. Phase-variable expression of Lewis blood-group antigens by *H. pylori* allows this microorganism to modulate the host T-helper-1-cell versus T-helper-2-cell response. We describe a model in which interactions between host lectins and pathogen carbohydrates facilitate asymptomatic persistence of *H. pylori*. This delicate balance, favourable for both the pathogen and the host, could lead to gastric autoimmunity in genetically susceptible individuals.

Helicobacter pylori is a human pathogen that persistently colonizes the stomach of approximately half of the world's population, for as long as the lifetime of its host. Colonization of the gastric mucosa typically occurs during childhood, and ~10% of those infected with *H. pylori* ultimately develop disease, which ranges from gastritis to peptic-ulcer disease to mucosa-associated lymphoid tissue (MALT) lymphoma or gastric cancer[1]. However, in most cases,

H. pylori persists without inducing clinical disease in its host, indicating that, at the gastric mucosal interface, there is a host–pathogen equilibrium that is beneficial for both organisms. Indeed, more recent data indicate that, in a small subpopulation of infected individuals, infection with *H. pylori* during childhood could protect against the later development of severe gastric-reflux disease, Barrett's oesophagus and adenocarcinoma of the oesophagus[2].

H. pylori is equipped with an impressive range of mechanisms that facilitate persistent colonization of its host[3], and because of the severity of *H. pylori*-associated diseases, the virulence factors of *H. pylori* have been studied extensively[4]. Infection with *H. pylori* results in vigorous innate and acquired immune responses by the host, as manifested by release of cytokines by epithelial cells and infiltration of the gastric mucosa by neutrophils, macrophages and lymphocytes, as well as by induction of specific humoral responses[5–10]. *H. pylori* triggers the innate immune system through interaction with Toll-like receptor 2 (TLR2)[11] and increases the influx of neutrophils and mononuclear cells to the mucosa through expression of neutrophil-activating protein[12]. In addition, *H. pylori* uses several mechanisms to evade or downregulate both innate and adaptive host immune responses. Lipopolysaccharide (LPS, see Glossary) expressed by *H. pylori* has low endotoxic and immunobiological activity compared with LPS of other bacteria[13,14] and can antagonize TLR4 signalling[11]. In addition, *H. pylori* can evade interaction with the host receptor TLR5 (REF. 15), a property that could contribute to its persistence at the mucosal surface. Arginase expressed by *H. pylori* downregulates nitric-oxide production by macrophages[16]. Efficient phagocytosis and killing of *H. pylori* is prevented by the presence of the *cag* pathogenicity island, which encodes a type IV secretion system[17,18]. Finally, *H. pylori* vacuolating cytotoxin A (VacA) inhibits the activation and function of T cells[19,20].

Helicobacter spp. are proposed to be native inhabitants of the stomach, and substantial evidence supports the idea of co-evolution of *H. pylori* and humans[21]. It is thought that, during this proposed co-evolution, bacteria were selected for their ability to induce sufficient epithelial damage to free nutrients but not to threaten the viability of the host. Insight into the host–pathogen interactions that are involved in *H. pylori* persistence should increase our understanding of molecular mechanisms that are involved in persistence of other, less well-known, human pathogens. In this article, we focus on a recently discovered role for *H. pylori* LPS in the modulation of the host immune response towards a local inflammatory environment that facilitates persistence. Following initial colonization by *H. pylori*, most infected individuals remain asymptomatic for decades; however, a small number of individuals can develop full-blown gastric autoimmunity as a result of persistence of this microorganism.