A proteomic view of the *Plasmodium falciparum* life cycle

Laurence Florens*, Michael P. Washburn†, J. Dale Raine‡, Robert M. Anthony§, Munira Grainger‖, J. David Haynes§¶, J. Kathleen Moch§, Nemone Muster*, John B. Sacci§#, David L. Tabb*☆, Adam A. Witney§#, Dirk Wolters†#, Yimin Wu, Malcolm J. Gardner††, Anthony A. Holder‖, Robert E. Sinden‡, John R. Yates*† & Daniel J. Carucci§**

* *Department of Cell Biology, The Scripps Research Institute, SR-11, 10550 North Torrey Pines Road, La Jolla, California 92037, USA*
† *Department of Proteomics and Metabolomics, Torrey Mesa Research Institute, Syngenta Research & Technology, 3115 Merryfield Row, San Diego, California 92121-1125, USA*
‡ *Infection and Immunity Section, Department of Biological Sciences, Imperial College of Science, Technology & Medicine, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ, UK*
§ *Naval Medical Research Center, Malaria Program (IDD), 503 Robert Grant Avenue, Room 3A40; and ¶ Department of Immunology, Walter Reed Army Institute of Research, Silver Spring, Maryland 20910-7500, USA*
‖ *The Division of Parasitology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK*
☆ *Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA*
** *Malaria Research and Reference Reagent Resource Center, American Type Culture Collection, 10801 University Boulevard, Manassas, Virginia 20110-2209, USA*
†† *The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA*

The completion of the *Plasmodium falciparum* clone 3D7 genome provides a basis on which to conduct comparative proteomics studies of this human pathogen. Here, we applied a high-throughput proteomics approach to identify new potential drug and vaccine targets and to better understand the biology of this complex protozoan parasite. We characterized four stages of the parasite life cycle (sporozoites, merozoites, trophozoites and gametocytes) by multidimensional protein identification technology. Functional profiling of over 2,400 proteins agreed with the physiology of each stage. Unexpectedly, the antigenically variant proteins of *var* and *rif* genes, defined as molecules on the surface of infected erythrocytes, were also largely expressed in sporozoites. The detection of chromosomal clusters encoding co-expressed proteins suggested a potential mechanism for controlling gene expression.

The life cycle of *Plasmodium* is extraordinarily complex, requiring specialized protein expression for life in both invertebrate and vertebrate host environments, for intracellular and extracellular survival, for invasion of multiple cell types, and for evasion of host immune responses. Interventional strategies including anti-malarial vaccines and drugs will be most effective if targeted at specific parasite life stages and/or specific proteins expressed at these stages. The genomes of *P. falciparum*[1] and *P. yoelii yoelii*[2] are now completed and offer the promise of identifying new and effective drug and vaccine targets.

Functional genomics has fundamentally changed the traditional gene-by-gene approach of the pre-genomic era by capitalizing on the success of genome sequencing efforts. DNA microarrays have been successfully used to study differential gene expression in the abundant blood stages of the *Plasmodium* parasite[3,4]. However, transcriptional analysis by DNA microarrays generally requires microgram quantities of RNA and has been restricted to stages that can be cultivated *in vitro*, limiting current large-scale gene expression analyses to the blood stages of *P. falciparum*. As several key stages of the parasite life cycle, in particular the pre-erythrocytic stages, are not readily accessible to study, and as differential gene expression is in fact a surrogate for protein expression, global proteomic analyses offer a unique means of determining not only protein expression, but also subcellular localization and post-translational modifications.

We report here a comprehensive view of the protein complements isolated from sporozoites (the infectious form injected by the mosquito), merozoites (the invasive stage of the erythrocytes), trophozoites (the form multiplying in erythrocytes), and gametocytes (sexual stages) of the human malaria parasite *P. falciparum*. These proteomes were analysed by multidimensional protein identification technology (MudPIT), which combines in-line, high-resolution liquid chromatography and tandem mass spectrometry[5]. Two levels of control were implemented to differentiate parasite from host proteins. By using combined host–parasite sequence databases and noninfected controls, 2,415 parasite proteins were confidently identified out of thousands of host proteins; that is, 46% of all gene products were detected in four stages of the *Plasmodium* life cycle (Supplementary Table 1).

Comparative proteomics throughout the life cycle

The sporozoite proteome appeared markedly different from the other stages (Table 1). Almost half (49%) of the sporozoite proteins

Table 1 **Comparative summary of the protein lists for each stage**

Protein count	Sporozoites	Merozoites	Trophozoites	Gametocytes
152	X	X	X	X
197	–	X	X	X
53	X	–	X	X
28	X	X	–	X
36	X	X	X	–
148	–	–	X	X
73	–	X	–	X
120	X	–	–	X
84	–	X	X	–
80	X	–	X	–
65	X	X	–	–
376	–	–	–	X
286	–	–	X	–
204	–	X	–	–
513	X	–	–	–
2,415	1,049	839	1,036	1,147

Whole-cell protein lysates were obtained from, on average, 17×10^6 sporozoites, 4.5×10^9 trophozoites, 2.75×10^9 merozoites, and 6.5×10^9 gametocytes.

Present addresses: BRB 13-009, Department of Microbiology and Immunology, University of Maryland School of Medicine, 655 W. Baltimore St., Baltimore, Maryland 21201, USA (J.B.S.); Department of Medical Microbiology, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK (A.A.W.); and Ruhr-University Bochum, Institute of Analytical Chemistry, 44780 Bochum, Germany (D.W.).

were unique to this stage, which shared an average of 25% of its proteins with any other stage. On the other hand, trophozoites, merozoites and gametocytes had between 20% and 33% unique proteins, and they shared between 39% and 56% of their proteins. Consequently, only 152 proteins (6%) were common to all four stages. Those common proteins were mostly housekeeping proteins such as ribosomal proteins, transcription factors, histones and cytoskeletal proteins (Supplementary Table 1). Proteins were sorted into main functional classes based on the Munich Information Centre for Protein Sequences (MIPS) catalogue[6], with some adaptations for classes specific to the parasite, such as cell surface and apical organelle proteins (Fig. 1). When considering the annotated proteins in the database, some marked differences appeared between sporozoites and blood stages (Fig. 1). Although great care was taken to ensure that the results reflect the state of the parasite in the host, a portion of the data set may reflect the parasite's response to different purification treatments. However, the stage-specific detection of known protein markers at each stage established the relevance of our data set.

The merozoite proteome

Merozoites are released from an infected erythrocyte, and after a short period in the plasma, bind to and invade new erythrocytes. Proteins on the surface and in the apical organelles of the merozoite mediate cell recognition and invasion in an active process involving an actin-myosin motor. Four putative components of the invasion motor[7], merozoite cap protein-1 (MCP1), actin, myosin A, and myosin A tail domain interacting protein (MTIP), were abundant merozoite proteins (Supplementary Table 2). Abundant merozoite surface proteins (MSPs) such as MSP1 and MSP2 are linked by a glycosylphosphatidyl (GPI) anchor to the membrane, and both have been implicated in immune evasion (reviewed in ref. 8). A second family of peripheral membrane proteins, represented by MSP3 and MSP6, was also detected (Fig. 2a), although these proteins are largely soluble proteins of the parasitophorous vacuole, which are released on schizont rupture. Other vacuolar proteins, such as the acidic basic repeat antigen (ABRA) and serine repeat antigen (SERA), were detected in the merozoite fraction, but some such as S-antigen[9] were not (Supplementary Table 2). Notably, MSP8 and a related MSP8-like protein were only identified in sporozoites (Fig. 2a). Some MSPs are diverse in sequence and may be extensively modified by proteolysis; these features, together with the association of a variety of peripheral and soluble proteins, provide for a complex surface architecture.

Many apical organellar proteins, in the micronemes and rhoptries, have a single transmembrane domain. Among these proteins, apical membrane antigen 1 (AMA1) and MAEBL were found in both sporozoite and merozoite preparations (Fig. 2a). Erythrocyte-binding antigens (EBA), such as EBA 175 and EBA 140/BAEBL, were found only in the merozoite and trophozoite fractions. Of note, the reticulocyte-binding protein (PfRH) family (PFD0110w, MAL13P1.176, PF13_01998, PFL2520w and PFD1150c), which has similarity with the Py235 family of *P. y. yoelii* rhoptry proteins and the *Plasmodium vivax* reticulocyte-binding proteins, was not detected in the merozoite fraction. Some PfRH proteins were, however, detected in sporozoites (Fig. 2a), including RH3, which is a transcribed pseudogene in blood stages[10]. Components of the low molecular mass rhoptry complex, the rhoptry-associated proteins (RAP) 1, 2 and 3, were all found in merozoites. RAP1 was also detected in sporozoites. The high molecular mass rhoptry protein complex (RhopH), together with ring-infected erythrocyte surface antigen (RESA), which is a component of dense granules, is transferred intact to new erythrocytes at or after invasion and may contribute to the host cell remodelling process. RhopH1, RhopH2 (PFI1445w; Ling, I. T., *et al.*, unpublished data) and RhopH3 were found in the merozoite proteome. RhopH1 (PFC0120w/PFC0110w) has been shown to be a member of the cyto-adherence linked asexual gene family (CLAG)[11]; however, the presence of CLAG9 in the merozoite fraction (Fig. 2a) suggests that CLAG9 may also be a RhopH protein, casting some doubt on the proposed role for this protein in cyto-adherence[12].

The trophozoite proteome

After erythrocyte invasion the parasite modifies the host cell. The principal modifications during the initial trophozoite phase (lasting about 30 h) allow the parasite to transport molecules in and out of the cell, to prepare the surface of the red blood cell to mediate cyto-adherence, and to digest the cytoplasmic contents, particularly haemoglobin, in its food vacuole. In the next phase of schizogony (the final ~18 h of the asexual development in the blood cell), nuclear division is followed by merozoite formation and release.

Knob-associated histidine-rich protein (KAHRP) and erythrocyte membrane proteins 2 and 3 (EMP2 and -3) bind to the erythrocyte cytoskeleton (Fig. 2a). Of the proteins of the parasitophorous vacuole and the tubovesicular membrane structure extending into the cytoplasm of the red blood cell, three (the skeleton-binding protein 1, and exported proteins EXP1 and EXP2) were represented by peptides (Fig. 2a); although a fourth (Sar1 homologue, small GTP-binding protein; PFD0810w) was not. It is likely that one or more of the hypothetical proteins detected only in the trophozoite sample are involved in these unusual structures.

Digestion of haemoglobin is a major parasite catabolic process[13]. Members of the plasmepsin family (aspartic proteinases; PF14_0075 to PF14_0078)[14], falcipain family (cysteine proteinases; PF11_0161,

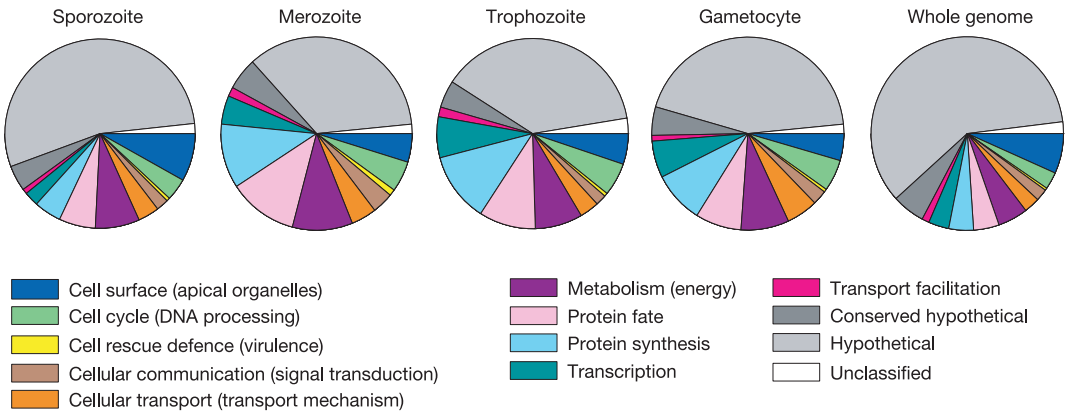

Figure 1 Functional profiles of expressed proteins. Proteins identified in each stage are plotted as a function of their broad functional classification as defined by the MIPS catalogue[6]. To avoid redundancy, only one class was assigned per protein. The complete protein list is given in Supplementary Table 1.

PF11_0162 and PF11_0165)[15], and falcilysin (a metallopeptidase; PF13_0322)[16] implicated in this process were all clearly identified (Supplementary Table 1). Several proteases expressed in the merozoite and trophozoite fractions, and not involved in haemoglobin digestion, may be important in parasite release at the end of schizogony, invasion of the new cell, or merozoite protein processing. Possible candidates for this mechanism include cysteine proteinases of the falcipain and SERA families, or subtilisins such as SUB1 and SUB2, both located in apical organelles (Fig. 2a).

The gametocyte proteome

Stage V gametocytes are dimorphic, with a male:female ratio of 1:4. They are arrested in the cell cycle until they enter the mosquito where development is induced within minutes to form the male and female gametes. Gametocyte structure reflects these ensuing fates; that is, the female has abundant ribosomes and endoplasmic reticulum/vesicular network to re-initiate translation, whereas the male is largely devoid of ribosomes and is terminally differentiated[17].

Gametocyte-specific transcription factors, RNA-binding proteins, and gametocyte-specific proteins involved in the regulation of messenger RNA processing (particularly splicing factors, RNA helicases, RNA-binding proteins, ribonucleoproteins (RNPs) and small nuclear ribonucleoprotein particles (snRNPS)) were highly represented in the gametocyte proteome (Supplementary Table 1). Transcription in the terminally differentiated gametocytes is 'suppressed', but the female gametocytes contain mRNAs encoding gamete/zygote/ookinete surface antigens (for example, P25/28)

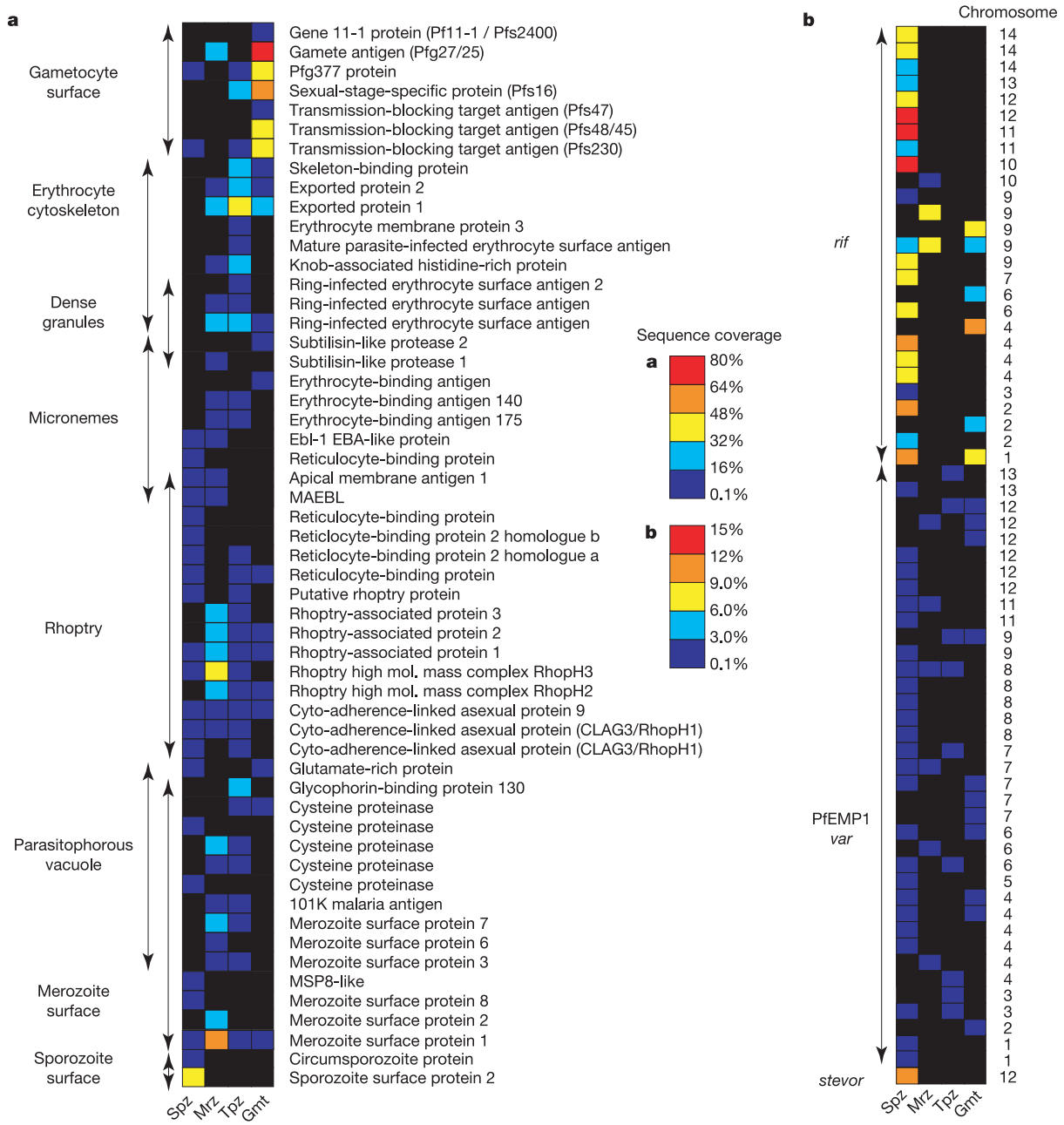

Figure 2 Expression patterns of known stage-specific proteins. **a**, Cell surface, organelle, and secreted proteins are plotted as a function of their known subcellular localization. **b**, *stevor*, *var* and *rif* polymorphic surface variants are plotted as a function of the chromosome encoding their genes. The matrices are colour-coded by sequence coverage measured in each stage (proteins not detected in a stage are represented by black squares). Locus names associated with these proteins are listed in Supplementary Table 2. Spz, sporozoite; mrz, merozoite; tpz, trophozoite; gmt, gametocyte.

that are subject to post-transcriptional control; this control is released rapidly during gamete development[17]. Ribosomal proteins were largely represented: 82% of known small subunit (SSU) proteins and 69% of known large subunit (LSU) proteins were detected in gametocytes compared to 94% and 82%, respectively, from all stages examined (Supplementary Table 1). We suggest that this reflects the accumulation of ribosomes in the female gametocyte to accommodate for the sudden increase in protein synthesis required during gametogenesis and early zygote development.

Other protein groupings highly represented in the gametocyte were in the cell cycle/DNA processing and energy classes (Fig. 1). The former is consistent with the biological observation that the mature gametocyte is arrested in G0 of the cell cycle and will require a full complement of pre-existing cell cycle regulatory cascades to respond, within seconds, to the gametogenesis stimuli (that is, xanthurenic acid and a drop in temperature)[18]. Metabolic pathways of the malaria parasite may be stage-specific, with asexual blood stage parasites dependent on glycolysis and conversion of pyruvate to lactate (L-lactate dehydrogenase) for energy. In the gametocyte and sporozoite preparations, peptides from enzymes involved in the mitochondrial tricarboxylic acid (TCA) cycle and oxidative phosphorylation were identified (Table 2). This observation suggests that gametocytes have fully functional mitochondria as a pre-adaptation to life in the mosquito, as suggested by morphological and biochemical studies[19] and their sensitivity to anti-malarials attacking respiration (primaquine and artimesinin-based products)[17]. It will be interesting to observe whether other mosquito and liver stages, which show similar drug sensitivities, express the same metabolic proteome.

Cell surface proteins (Fig. 1) included most of the known surface antigens (Fig. 2a and Supplementary Table 2). However, Pfs35 and a sexual stage-specific kinase (PF13_0258) were not detected. Nevertheless the cultured gametocytes analysed in this study expressed a specific repertoire of rifin and PfEMP1 proteins (Fig. 2b and Supplementary Table 2). Together these observations suggest that the gametocyte, which is very long-lived in the red blood cell (that is, 9–12 days compared with 2 days for the pathogenic asexual parasites), expresses a limited repertoire of the highly polymorphic families of surface antigens so widely represented in the asexual parasites.

The sporozoite proteome

Sporozoites are injected by the mosquito during ingestion of a blood meal. Although, they are in the blood stream for only minutes, sporozoites probably require mechanisms to evade the host humoral immune system in order for at least a fraction of the thousands of sporozoites injected by the mosquito to survive the hostile environment in the blood and successfully invade hepatocytes.

The main class of annotated sporozoite proteins identified was cell surface and organelle proteins (Fig. 1). Sporozoites are an invasive stage and possess the apical complex machinery involved in host cell invasion. As observed in the analysis of the *P. y. yoelii* sporozoite transcriptome[20], actin and myosin were found in the motile sporozoites (Supplementary Table 2). Many proteins associated with rhoptry, micronemes and dense granules were detected (Fig. 2a). Among the proteins found were known markers of the sporozoite stage, such as the circumsporozoite protein (CSP) and sporozoite surface protein 2 (SSP2; also known as TRAP), both present in large quantities at the sporozoite surface (Fig. 2a). Peptides derived from CTRP (circumsporozoite protein and thrombospondin-related adhesive protein (TRAP)-related protein), an ookinete cell surface protein involved in recognition and/or motility[21], were detected in the sporozoite fractions (Supplementary Table 1).

Most surprisingly, peptides derived from multiple *var* (coding for PfEMP1) and *rif* genes were identified in the sporozoite samples. PfEMP1 and rifins are coded for by large multigene families (*var* and *rif*)[22,23] and are present on the surface of the infected red blood cell. No peptides derived from *rif* genes were identified in the trophozoite sample, whereas sporozoites expressed 21 different rifins and 25 PfEMP1 isoforms (Fig. 2b); that is, a total of 14% of the *rif* genes and 33% of the *var* genes encoded by the genome. Furthermore, very little overlap was observed between stages: only ten PfEMP1 and two rifin isoforms expressed in sporozoites were found in other stages. Whereas in the blood stream the asexual stage parasites undergo asexual multiplication and therefore have an opportunity to undergo antigenic 'switching' of the variant antigen genes, the non-replicative sporozoites may not have this opportunity. Expressing such a polymorphic array of *var* (PfEMP1) and *rif* genes could be part of a sporozoite survival mechanism.

Chromosomal clusters encoding co-expressed proteins

The distinct proteomes of each stage of the *Plasmodium* life cycle suggested that there is a highly coordinated expression of *Plasmodium* genes involved in common processes. Co-expression groups are a widespread phenomenon in eukaryotes, where mRNA array analyses have been used to establish gene expression profiles. Analysis of co-regulated gene groups facilitates both searching for regulatory motifs common to co-regulated genes, and predicting protein function on the basis of the 'guilt by association' model. Furthermore, mRNA analyses in *Saccharomyces cerevisiae*[24] and *Homo sapiens*[25,26] have demonstrated that co-regulated genes do not map to random locations in the genome but are in fact

Table 2 **Examples on enzymes in stage-specific metabolic pathways**

Locus	Stage				Enzyme	EC number†	Reaction catalysed
	Spz*	Mrz*	Tpz*	Gmt*			
End of glycolysis							
PF10_0363	1.2	–	2.4	–	Pyruvate kinase	2.7.1.40	*P*-enolpyruvate to pyruvate
MAL6P1.160	8.6	66.9	18.8	14.7	Pyruvate kinase		
PF13_0141	46.2	83.9	70.9	78.8	L-lactate dehydrogenase	1.1.1.27	Pyruvate to lactate
TCA cycle and oxidative phosphorylation							
PF10_0218	12.3	–	–	–	Citrate synthase	4.1.3.7	Acetyl coA + oxaloacetate to citrate
PF13_0242	3.2	–	16.9	8.8	Isocitrate dehydrogenase (NADP)	1.1.1.41	Isocitrate to 2-oxoglutarate + CO_2
PF08_0045	2.9	–	2.2	23.1	2-Oxoglutarate dehydrogenase e1 component	1.2.4.2	2-Oxoglutarate to succinyl CoA
PF10_0334	–	–	3.5	27.7	Flavoprotein subunit of succinate dehydrogenase	1.3.5.1	Succinate to fumarate
PFL0630w	3.7	–	–	12.1	Iron-sulphur subunit of succinate dehydrogenase		
PF14_0373	–	–	–	12.7	Ubiquinol cytochrome oxidoreductase	1.10.2.2.	Ubiquinol to cytochrome *c* reductase in electron transport
PFB0795w	–	–	–	14.2	ATP synthase F1, α-subunit		
PFI1365w	–	–	–	8.8	Cytochrome *c* oxidase subunit	1.9.3.1	
PFI1340w	–	–	–	8.8	Fumarate hydratase	4.2.1.2	Fumarate to malate
MAL6P1.242	30.4	–	–	40.9	Malate dehydrogenase	1.1.1.37	Malate to oxaloacetate

Plasmodium metabolic pathways can be found at http://www.sites.huji.ac.il/malaria/. Spz, sporozoite; mrz, merozoite; tpz, trophozoite; gmt, gametocyte.
*The sequence coverage (that is, the percentage of the protein sequence covered by identified peptides) measured in each stage is reported.
†Enzyme Commission (EC) numbers are reported for each protein.

frequently organized into gene clusters on a chromosome. Gene clustering in *Plasmodium* species has been demonstrated. Ordered arrays of genes involved in virulence and antigenic variation (for example, *var*, *vir* and *rif* genes) are located in the subtelomeric regions of the chromosomes[27,28].

To determine whether gene clustering exists along the entire *P. falciparum* genome, genes whose protein products were detected in our analysis were mapped onto all 14 chromosomes in a stage-dependent manner (Fig. 3a). The 2,415 proteins identified represented an average of 45% of the open reading frames (ORFs) predicted per chromosome. The number of protein hits by chromosome was similar for all stages: sporozoite, merozoite, trophozoite and gametocyte protein lists constituting 19.7%, 15.8%, 19.5% and 21.6% of the predicted ORFs per chromosomes, respectively. Groups of three or more consecutive loci whose protein products were detected in a particular stage were defined as chromosomal clusters encoding co-expressed proteins (Fig. 3b). On the basis of this definition a total of 98 clusters containing 3 loci, 32 clusters containing 4 loci, 5 clusters containing 5 loci, and 3 clusters containing 6 loci were identified (Supplementary Table 3). For each chromosome, the frequency of finding clusters encoding co-expressed proteins containing 3–6 adjacent loci markedly exceeded the probability of finding such clusters by chance (see the footnote of Supplementary Table 3 for details on the probability calculation). Therefore, chromosomal clusters encoding co-expressed proteins were prevalent in the *P. falciparum* genome.

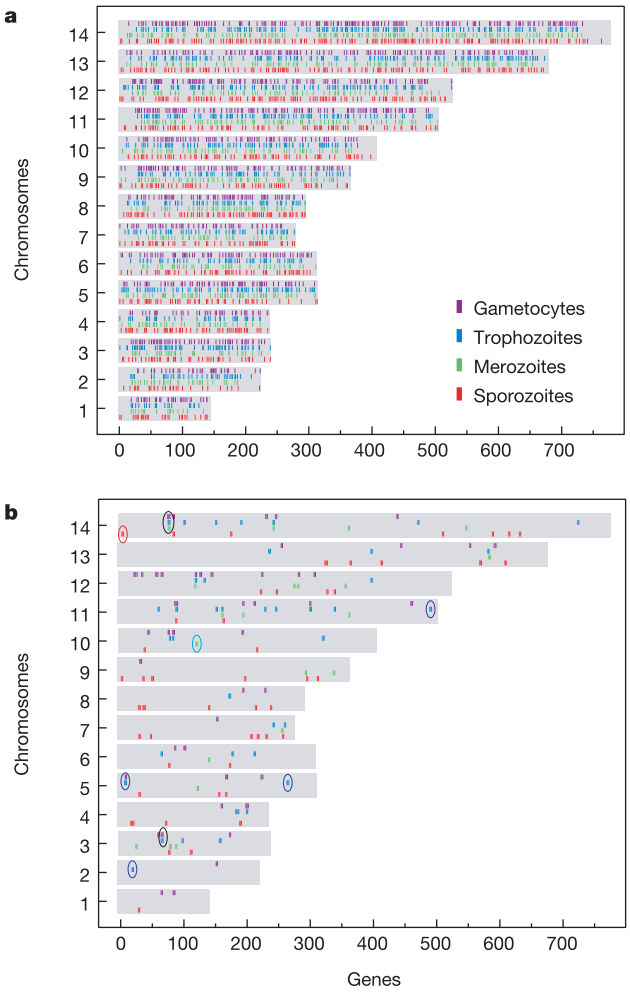

Figure 3 Distribution of expressed proteins by chromosome. **a**, For each stage, genes whose products were detected (coloured vertical bars) are plotted in the order they appear on their chromosome (grey boxes). **b**, Groups of at least three consecutive expressed genes are defined as chromosomal clusters of co-expressed proteins. Examples of such clusters, circled in **b**, are specified in Table 3 and the complete description of the 138 clusters can be found in Supplementary Table 3.

Functionally related genes have been shown to cluster in the *S. cerevisiae*[24] and human genomes[26]. This phenomenon also occurs in *P. falciparum*. A total of 138 clusters encoding co-expressed proteins were identified and 67 of them (49%) contained at least two loci that have been functionally annotated. Of these 67 clusters, 30 contained at least two loci whose annotation clearly indicates that the proteins are functionally related. For example, clusters on chromosomes 3, 5 and 10 contained ribosomal proteins, proteins involved in protein modification, and proteins involved in nucleotide metabolism, respectively (Table 3). Chromosome 14 contained a cluster of four aspartic proteases co-expressed in all of the blood stages (Table 3). This cluster was not detected in sporozoites, where no haemoglobin degradation is expected to occur. Interestingly, whereas the falcipain gene cluster on chromosome 11 appeared in our analysis as a cluster of co-expressed proteins (Supplementary Table 3), the SERA gene cluster on chromosome 2, coding for proteins that share a papain-like sequence motif[29], did not. Of the ten sporozoite-specific clusters, five involved *var* and *rif* genes, such as the *rif* cluster located in the subtelomeric domain of chromosome 14 (Table 3). On the basis of their presence in clusters encoding co-expressed proteins, we were able to suggest functional roles for 24 proteins annotated as hypothetical in the *P. falciparum* genome (Supplementary Table 3). For example, a gametocyte-specific cluster on chromosome 13 encoded two transmission-blocking antigens (Pfs48/45 and Pfs47) and a hypothetical protein, PF13_0246, which might be a gametocyte surface protein. Two clusters on chromosomes 2 and 11 were highly specific to the trophozoite stage (Table 3). Each of these clusters contained well-known secreted and surface proteins, namely KAHRP, PfEMP3, antigen 332, and RESA, all of which have been implicated in knob formation. The highly coordinated expression of these genes makes the three hypothetical proteins listed in these trophozoite-specific gene clusters possible candidates for involvement in cyto-adherence.

Discussion

Although sample handling is a principal consideration when studying pathogens, the expression of large numbers of previously identified proteins was consistent with their published expression profiles, validating our data set as a meaningful sampling of each stage's proteome. This is a particularly important aspect of our analysis as 65% of the 5,276 genes encoded by the *P. falciparum* genome are annotated as hypothetical[1], and of the 2,415 expressed proteins we identified, 51% are hypothetical proteins (Supplementary Table 1). Our results confirmed that these hypothetical ORFs predicted by gene modelling algorithms were indeed coding regions. Furthermore, from all four stages analysed, we identified 439 proteins predicted to have at least one transmembrane segment or a GPI addition signal (18% of the data set) and 304 soluble proteins with a signal sequence; that is, potentially secreted or located to organelles. Well over half of the secreted proteins and integral membrane proteins detected were annotated as hypothetical (Supplementary Table 4). The obvious interest in this class of proteins is that, with no homology to known proteins, they represent potential *Plasmodium*-specific proteins and may provide targets for new drug and vaccine development.

Our comprehensive large-scale analysis of protein expression showed that most surface proteins are more widely expressed than initially thought. In particular, the *var* and *rif* genes, which were thought to be involved in immune evasion only in the blood stage, have now been shown to be expressed in apparently large and varied numbers at the sporozoite stage. These surface proteins might be involved in general interaction processes with host cells and/or immune evasion. An alternative hypothesis is that stage-specific

Table 3 **Examples of chromosomal gene clusters encoding co-expressed proteins**

Chromosome	ID	Locus	Stage				Description	Class	SP	TM
			Spz	Mrz	Tpz	Gmt				
3	64	PFC0285c	2.1	12.7	33.2	18.7	T-complex protein β-subunit	Protein fate	0	0
3	65	PFC0290w	8.3	–	33.8	18.6	40S ribosomal protein S23	Protein synthesis	0	0
3	66	PFC0295c	–	14.9	52.5	21.3	40S ribosomal protein S12	Protein synthesis	0	0
3	67	PFC0300c	–	12.1	30.4	17.9	60S ribosomal protein L7	Protein synthesis	0	0
5	263	PFE1345c	–	–	1.9	1.6	Minichromosome maintenance protein 3	Cell transport	0	0
5	264	PFE1350c	–	–	22.4	–	Ubiquitin-conjugating enzyme	Protein fate	0	0
5	265	PFE1355	–	4.8	2.6	2.6	Ubiquitin carboxy-terminal hydrolase	Protein fate	0	0
5	266	PFE1360c	–	–	7.7	–	Methionine aminopeptidase	Protein fate	0	0
10	119	PF10_0121	10.8	74.5	29	–	Hypoxanthine phosphoribosyltransferase	Metabolism	0	0
10	120	PF10_0122	5.4	6.1	–	6.1	Phosphoglucomutase	Metabolism	0	0
10	121	PF10_0123	–	11.7	–	–	GMP synthetase	Metabolism	0	0
10	122	PF10_0124	0.9	1.8	–	–	Hypothetical protein		0	0
14	74	PF14_0074	26.6	–	–	4.9	Hypothetical protein		0	0
14	75	PF14_0075	–	26.5	43.2	47.4	Plasmepsin	Protein fate	1	0
14	76	PF14_0076	–	6.6	35.2	10	Plasmepsin 1	Protein fate	1	0
14	77	PF14_0077	–	21.2	43	11.5	Plasmepsin 2	Protein fate	1	0
14	78	PF14_0078	–	14.2	52.8	29.9	HAP protein	Protein fate	1	0
14	2	PF14_0002	3.5	–	–	–	Rifin	Surface or organelles	0	1
14	3	PF14_0003	7.9	–	–	–	Rifin	Surface or organelles	1	2
14	4	PF14_0004	6.5	–	–	–	Rifin	Surface or organelles	1	2
2	18	PFB0090c	–	–	3	–	Hypothetical protein, conserved		0	0
2	19	PFB0095c	–	–	3.4	–	Erythrocyte membrane protein 3	Surface or organelles	1	0
2	20	PFB0100c	–	1.5	24.8	–	Knob-associated histidine-rich protein	Surface or organelles	1	0
11	489	PF11_0506	–	–	6.3	4.4	Hypothetical protein		0	1
11	490	PF11_0507	–	–	0.8	–	Antigen 332	Surface or organelles	0	0
11	491	PF11_0508	–	–	3.3	–	Hypothetical protein		0	0
11	492	PF11_0509	–	6.4	3	–	RESA	Surface or organelles	0	0
13	443	PF13_0246	4.5	–	–	8.6	Hypothetical protein		0	0
13	444	PF13_0247	–	–	–	32.4	Transmission-blocking target antigen precursor (Pfs48/45)	Surface or organelles	1	1
13	445	PF13_0248	–	–	–	7.1	Transmission-blocking target antigen precursor (Pfs47)	Surface or organelles	1	1

Clusters of at least three consecutive genes encoding co-expressed proteins are reported with their position (ID) on the chromosome, the sequence coverage measured for these proteins in each stage (%), their current annotation and functional class, and the predicted presence of signal peptide (SP) or transmembrane domains (TM) (based on the TMHMM[43], a transmembrane (TM) helices prediction method based on a hidden Markov model (HMM), big-PI Predictor[44] and SignalP[45] algorithms).

regulation is not as exact as previously thought.

One mechanism of protein expression control that contributes to stage specificity in *P. falciparum* arises from the chromosomal clustering of genes encoding co-expressed proteins. The clusters described in this study demonstrate a widespread high order of chromosomal organization in *P. falciparum* and probably correspond to regions of open chromatin allowing for co-regulated gene expression. The high (A + T) content of the *P. falciparum* genome makes the identification of regulatory sequences such as promoters and enhancers challenging[31,32]. Focusing analyses on stage-specific and multi-stage clusters will facilitate finding stage-specific and general *cis*-acting sequences in the *Plasmodium* genome and will help decipher gene expression regulation during the parasite life cycle.

The malaria parasite is a complex multi-stage organism, which has co-evolved in mosquitoes and vertebrates for millions of years. Designing drugs or vaccines that substantially and persistently interrupt the life cycle of this complex parasite will require a comprehensive understanding of its biology. The *P. falciparum* genome sequence and comparative proteomics approaches may initiate new strategies for controlling the devastating disease caused by this parasite. □

Methods

Parasite material

Plasmodium falciparum clone 3D7 (Oxford) was used throughout. Sporozoites were initially isolated from the salivary glands of *Anopheles stephansi* mosquitoes, 14 days after infection, by centrifugation in a Renograffin 60 gradient, as described[33]. Four sporozoite samples were used as is. A fifth sample underwent an additional purification step on Dynabeads M-450 Epoxy coupled to NFS1 (an anti-*P. falciparum* CS protein monoclonal antibody)[34] according to the manufacturer's instructions (Dynal). Trophozoite-infected erythrocytes from synchronized cultures were purified on 70% Percoll-alanine[30], and the trophozoites released from the erythrocytes[35]. Of the of 260 parasitized erythrocytes counted by Giemsa-stained thin-blood film, 100% were identified as trophozoites. Merozoites were prepared essentially as described in ref. 36, using highly synchronized schizonts and purifying the merozoites by passage through membrane filters. Starting with synchronized asexual parasites grown in suspension culture as described[37,38], gametocytes were prepared by daily media changes of static cultures at 37 °C. When there were very few mature asexual stages present, gametocyte-infected erythrocytes were collected from the 52.5%/45% and 45%/30% interfaces of a Percoll gradient[39]. The gametocytes consisted mostly of stage IV and V parasites with minor contamination (<3%) from mixed asexual stage parasites. Finally, cellular debris from the upper bodies of parasite-free *A. stephansi* and non-infected human erythrocytes were used as controls for sporozoites and blood-stage parasites, respectively. Every effort was made to minimize enzymatic activity and protein degradation during sampling, and the subsequent isolation of the parasites; however, we cannot exclude that some of the differences in protein profiles that we observe between the different life-cycle stages may be a consequence of the sample-handling procedures.

Cell lysis

Five sporozoite, four merozoite, four trophozoite and three gametocyte preparations were lysed, digested and analysed independently. Cell pellets were first diluted ten times in 100 mM Tris-HCl pH 8.5, and incubated in ice for 1 h. After centrifugation at 18,000 *g* for 30 min, supernatants were set aside and microsomal membrane pellets were washed in 0.1 M sodium carbonate, pH 11.6. Soluble and insoluble protein fractions were separated by centrifugation at 18,000 *g* for 30 min. Supernatants obtained from both centrifugation steps were either combined (sporozoites, trophozoites and merozoites) or digested and analysed independently (gametocytes).

Peptide generation and analysis

The method follows that of Washburn *et al.*[5], with the exception that Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl; Pierce) was used to reduce urea-denatured proteins. Peptide mixtures were analysed through MudPIT as described[5].

Protein sequence databases

The *P. falciparum* database contained 5,283 protein sequences. Spectra resulting from contaminant mosquito and erythrocyte peptides had to be taken into account in the sporozoite and blood-stage samples, respectively. Tandem mass spectrometry (MS/MS) data sets from blood stages were therefore searched against a database containing both *P. falciparum* protein sequences and 24,006 ORFs from the human, mouse and rat RefSeq NCBI databases. At the date of the searches, the *Anopheles gambiae* genome was not available. The NCBI database contained 922 *Anopheles* and 313 *Aedes* proteins, which were combined to the 14,335 ORFs of the NCBI *Drosophila melanogaster*[40] database to create a control diptera database. Finally, these databases were complemented with a set of 172 known protein contaminants, such as proteases, bovine serum albumin and human keratins.

MS/MS data set analysis

The SEQUEST algorithm was used to match MS/MS spectra to peptides in the sequence databases[41]. To account for carboxyamidomethylation, MS/MS data sets were searched with a relative molecular mass of 57,000 (M_r 57K) added to the average molecular mass of cysteines. Peptide hits were filtered and sorted with DTASelect[42]. Spectra/peptide matches were only retained if they were at least half-tryptic (Lys or Arg at either end of the identified peptide) and with minimum cross-correlation scores (XCorr) of 1.8 for +1, 2.5 for +2, and 3.5 for +3 spectra and DeltaCn (top match's XCorr minus the second-best match's XCorr divided by the top match's XCorr) of 0.08. Peptide hits were deemed unambiguous only if they were not found in non-infected controls and were uniquely assigned to parasite proteins by searching against combined parasite–host databases. Finally, for low coverage loci, peptide/spectrum matches were visually assessed on two main criteria: any given MS/MS spectrum had to be clearly above the baseline noise, and both *b* and *y* ion series had to show continuity. The Contrast tool[42] was used to compare and merge protein lists from replicate sample runs and to compare the proteomes established for the four stages.

Received 31 July; accepted 9 September 2002; doi:10.1038/nature01107.

1. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419,** 498–511 (2002).
2. Carlton, J. M. *et al.* Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii*. *Nature* **419,** 512–519 (2002).
3. Ben Mamoun, C. *et al.* Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis. *Mol. Microbiol.* **39,** 26–36 (2001).
4. Hayward, R. E. *et al.* Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Mol. Microbiol.* **35,** 6–14 (2000).
5. Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19,** 242–247 (2001).
6. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **30,** 31–34 (2002).
7. Pinder, J. C. *et al.* Actomyosin motor in the merozoite of the malaria parasite, *Plasmodium falciparum*: implications for red cell invasion. *J. Cell Sci.* **111,** 1831–1839 (1998).
8. Holder, A. A. *Malaria Vaccine Development: a Multi-immune Response and Multi-stage Perspective* (ed. Hoffman, S. L.) 77–104 (ASM Press, Washington, 1996).
9. Coppel, R. L. *et al.* Isolate-specific S-antigen of *Plasmodium falciparum* contains a repeated sequence of eleven amino acids. *Nature* **306,** 751–756 (1983).
10. Taylor, H. M. *et al. Plasmodium falciparum* homologue of the genes for *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins, which is transcribed but not translated. *Infect. Immun.* **69,** 3635–3645 (2001).
11. Kaneko, O. *et al.* The high molecular mass rhoptry protein, RhopH1, is encoded by members of the clag multigene family in *Plasmodium falciparum* and *Plasmodium yoelii*. *Mol. Biochem. Parasitol.* **118,** 223–231 (2001).
12. 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).
13. Klemba, M. & Goldberg, D. E. Biological roles of proteases in parasitic protozoa. *Annu. Rev. Biochem.* **71,** 275–305 (2002).
14. Banerjee, R. *et al.* Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine. *Proc. Natl Acad. Sci. USA* **99,** 990–995 (2002).
15. Rosenthal, P. J., Sijwali, P. S., Singh, A. & Shenai, B. R. Cysteine proteases of malaria parasites: targets for chemotherapy. *Curr. Pharm. Des.* **8,** 1659–1672 (2002).
16. Eggleson, K. K., Duffin, K. L. & Goldberg, D. E. Identification and characterization of falcilysin, a metallopeptidase involved in hemoglobin catabolism within the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **274,** 32411–32417 (1999).
17. Sinden, R. E., Butcher, G. A., Billker, O. & Fleck, S. L. Regulation of infectivity of *Plasmodium* to the mosquito vector. *Adv. Parasitol.* **38,** 53–117 (1996).
18. Billker, O., Shaw, M. K., Margo, G. & Sinden, R. E. Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature* **392,** 289–292 (1998).
19. Krungkrai, J., Prapunwattana, P. & Krungkrai, S. R. Ultrastructure and function of mitochondria in gametocytic stage of *Plasmodium falciparum*. *Parasite* **7,** 19–26 (2000).
20. Kappe, S. H. *et al.* Exploring the transcriptome of the malaria sporozoite stage. *Proc. Natl Acad. Sci. USA* **98,** 9895–9900 (2001).
21. Dessens, J. T. *et al.* CTRP is essential for mosquito infection by malaria ookinetes. *EMBO J.* **18,** 6221–6227 (1999).
22. Deitsch, K. W. & Wellems, T. E. Membrane modifications in erythrocytes parasitized by *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **76,** 1–10 (1996).
23. 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).
24. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26,** 183–186 (2000).
25. Caron, H. *et al.* The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291,** 1289–1292 (2001).
26. Lercher, M. J., Urrutia, A. O. & Hurst, L. D. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genet.* **31,** 180–183 (2002).
27. Hernandez-Rivas, R. *et al.* Expressed var genes are found in *Plasmodium falciparum* subtelomeric regions. *Mol. Cell Biol.* **17,** 604–611 (1997).
28. del Portillo, H. A. *et al.* A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* **410,** 839–842 (2001).
29. Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282,** 1126–1132 (1998).
30. Kanaani, J. & Ginsburg, H. Metabolic interconnection between the human malarial parasite *Plasmodium falciparum* and its host erythrocyte. *J. Biol. Chem.* **264,** 3194–3199 (1989).
31. 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).
32. Lockhart, D. J. & Winzeler, E. A. Genomics, gene expression and DNA arrays. *Nature* **405,** 827–836 (2000).
33. Pacheco, N. D., Strome, C. P., Mitchell, F., Bawden, M. P. & Beaudoin, R. L. Rapid, large-scale isolation of *Plasmodium berghei* sporozoites from infected mosquitoes. *J. Parasitol.* **65,** 414–417 (1979).
34. Mellouk, S. *et al.* Evaluation of an *in vitro* assay aimed at measuring protective antibodies against sporozoites. *Bull. World Health Organ.* **68** Suppl., 52–59 (1990).
35. Rabilloud, T. *et al.* Analysis of membrane proteins by two-dimensional electrophoresis: comparison of the proteins extracted from normal or *Plasmodium falciparum*-infected erythrocyte ghosts. *Electrophoresis* **20,** 3603–3610 (1999).
36. Blackman, M. J. Purification of *Plasmodium falciparum* merozoites for analysis of the processing of merozoite surface protein-1. *Methods Cell Biol.* **45,** 213–220 (1994).
37. Haynes, J. D. & Moch, J. K. Automated synchronization of *Plasmodium falciparum* parasites by culture in a temperature-cycling incubator. *Methods Mol. Med.* **72,** 489–497 (2002).
38. Haynes, J. D., Moch, J. K. & Smoot, D. S. Erythrocytic malaria growth or invasion inhibition assays with emphasis on suspension culture GIA. *Methods Mol. Med.* **72,** 535–554 (2002).
39. Carter, R., Ranford-Cartwright, L. & Alano, P. The culture and preparation of gametocytes of *Plasmodium falciparum* for immunochemical, molecular, and mosquito infectivity studies. *Methods Mol. Biol.* **21,** 67–88 (1993).
40. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287,** 2185–2195 (2000).
41. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989 (1994).
42. Tabb, D. L., McDonald, W. H. & Yates, J. R. DTASelect and contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1,** 21–26 (2002).
43. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).
44. Eisenhaber, B., Bork, P. & Eisenhaber, F. Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. *Protein Eng.* **11,** 1155–1161 (1998).
45. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6 (1997).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements

We are grateful to J. Graumann, R. Sadygov, G. Chukkapalli, A. Majumdar and R. Sinkovits for computer programming; C. Deciu for the probability calculations; and C. Delahunty and C. Vieille for critical reading of the manuscript. The authors acknowledge the support of the Office of Naval Research, the US Army Medical Research and Material Command, and the National Institutes of Health (to J.R.Y.). J.D.R. is funded by a Wellcome Trust Prize Studentship. We thank the scientists and funding agencies comprising the international Malaria Genome Project for making sequence data from the genome of *P. falciparum* clone 3D7 public before publication of the completed sequence. The opinions expressed are those of the authors and do not reflect the official policy of the Department of the Navy, Department of Defense, or the US government.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.Y. (e-mail: jyates@scripps.edu).