

Plasmodium post-genomics: better the bug you know?

Taco W. A. Kooij*, Chris J. Janse and Andrew P. Waters

Abstract | Since the publication of the sequence of the genome of *Plasmodium falciparum*, the major causative agent of human malaria, many post-genomic studies have been completed. Invaluably, these data can now be analysed comparatively owing to the availability of a significant amount of genome-sequence data from several closely related model species of *Plasmodium* and accompanying global proteome and transcriptome studies. This review summarizes our current knowledge and how this has already been — and will continue to be — exploited in the search for vaccines and drugs against this most significant infectious disease of the tropics.

Expressed sequence tags (ESTs). Short sequences (several hundred base pairs) of cDNA clones that are produced by reverse transcription of mRNA. ESTs contain (spliced) exons and the 5' and 3' untranslated regions. ESTs allow rapid identification ('tagging') of genes expressed in the organism, tissue or stage from which cDNA has been extracted and can expedite DNA marker (single nucleotide polymorphism) development.

Malaria Research Group,
Department of Parasitology,
Centre for Infectious
Diseases, Leiden University
Medical Centre, Albinusdreef
2, 2333 ZA, Leiden,
The Netherlands.

*Current address: Nuffield
Department of Clinical
Laboratory Sciences,
University of Oxford and
Blood Research Laboratory,
National Blood Service,
John Radcliffe Hospital,
Headington, Oxford
OX3 9DU, UK.

Correspondence to A.P.W.
e-mail: Waters@lumc.nl
doi:10.1038/nrmicro1392

It cannot be restated too often that malaria, principally through infection by the protozoan apicomplexan parasite *Plasmodium falciparum*, is responsible for more than one million deaths each year. A recent survey has shown that, in 2002, roughly 2.2 billion people were at risk of contracting *P. falciparum* infection, with 515 million a conservative estimate of the number of individuals who became infected¹.

Clearly a disease of poverty, the dwindling effectiveness of frontline affordable drugs and the continuing lack of a vaccine mean that malaria poses a greater problem now than at any time since the failure of the WHO eradication drive in the 1950s. It is in this context that malaria research must act. Basic biological investigation of malaria parasites has always offered the promise of the development of new vaccines and drugs. Although biologists initially worked on single genes of therapeutic interest, the landmark publication² of the complete *P. falciparum* genome with first-pass annotation in late 2002 irrevocably changed the face and practice of malaria research. This comprehensive data set, combined with the availability of substantial sequence tracts from the rodent malaria parasite (RMP) *Plasmodium yoelii*, has made it possible to embrace the latest global-genome-survey technologies³.

Thanks to the admirable real-time release policy of the three sequencing centres involved, this embrace was immediate, and two detailed proteomic studies accompanied the genome into press^{4,5}. Here, we review the current status of *Plasmodium* genomic and post-genomic research, indicating trends and the future requirements necessary to build a detailed 'virtual parasite' in which we are fully informed of the molecular biology that underpins its biology and interactions with both host and vector (BOX 1).

The *P. falciparum* genome

The major pre-genomic milestones in *Plasmodium* research are summarized in the accompanying TIMELINE, and the complex life cycle of a *Plasmodium* parasite is described in BOX 2.

The *P. falciparum* genome comprises 14 linear chromosomes ranging in size from 0.64–3.3 Mb and two non-nuclear genomes: a compact mitochondrial genome of 6 kb and a plastid-like, 35-kb circular genome that resides in an organelle now known as the apicoplast⁶. The genome of the 3D7 strain of *P. falciparum* was the first parasite genome to be sequenced to completion². A chromosome-by-chromosome shotgun-sequencing strategy was used based on pulsed-field gel electrophoresis (PFGE)-separated chromosomes. Mainly due to the extreme AT bias (80%) of the genome, it took almost seven years before the 22.8-Mb nuclear genome was complete, revealing 5,268 predicted protein-encoding genes at an average frequency of one gene every 4.3 kb. Matching expressed sequence tags (ESTs) and proteomic analyses were used to experimentally validate approximately 49% and 52% of these genes, respectively². At the time of publication, there were still 79 gaps in the sequence. The largest contiguous DNA sequence (contig) was chromosome 12 (2.3 Mb). Now, just three chromosomes (7, 8 and 13) are still awaiting closure, with only six gaps remaining (M. Berriman, personal communication).

Comparison of the first-pass annotation of the *P. falciparum* nuclear genome with other genomes showed that 60% of the predicted genes could not be assigned functions. The products of at least 1.3% of the *P. falciparum* genes are known to be involved in cell-to-cell adhesion or invasion of host cells, and a further 3.9% are postulated to have a role in evasion of the host

Box 1 | Databases: towards the 'virtual parasite'

Supplementary online data files and the general public databases such as GenBank aside, most genome projects require an associated database that is specifically tailored to the features of the organism. PlasmoDB^{101,102} is the malaria community's resource that attempts to present the chromosomes, their genes, primary annotation and all the relevant information pertaining to the deeper post-genomic characterization of the expression and function of the gene products. The format is logic-based and enables intuitive browsing through the genome, using a graphical interface representing the *P. falciparum* genome. The extensive amount of available data can also be explored through dedicated, complex queries. In addition, PlasmoDB integrates the ever-increasing amount of comparative information from the other *Plasmodium* genomes described here. The genome-sequencing centres also have their own curated databases, GeneDB from the Wellcome Trust Sanger Institute and Gene Indices from the Institute for Genomic Research (TIGR), which provide comprehensive information at the level of the individual genes, but without at present providing the more global synthesis attempted by PlasmoDB. Further integration of genomic, post-genomic and experimental data and enhancement of the algorithms to predict not only gene structure, but also expression profiles and protein structures, could ultimately facilitate the generation of a 'virtual parasite'. A list of useful links to genome, transcriptome and/or proteome databases and several other interesting malaria websites is provided below.

Databases with *Plasmodium* genome data

- PlasmoDB, the official site of the *Plasmodium* genome project: <http://plasmodb.org>
- Protozoan genomes at the Wellcome Trust Sanger Institute (WTSI): <http://www.sanger.ac.uk/Projects/Protozoa>
- GeneDB, the curated database of the WTSI: <http://www.genedb.org>
- Parasite databases at TIGR: <http://www.tigr.org/parasiteProjects.shtml>
- Gene Indices, the curated TIGR database: <http://www.tigr.org/tdb/tgi/protist.shtml>
- Malaria projects at the Stanford Genome Technology Center (SGTC): <http://sequence-www.stanford.edu/group/malaria/index.html>

Databases with *Plasmodium* transcription data

- Resources of the Winzeler Laboratory, The Scripps Research Institute: <http://www.scripps.edu/cb/winzeler/resources.htm>
- Microarray database of the DeRisi Laboratory, University of California San Francisco: <http://malaria.ucsf.edu/index.php>
- Malaria full-length cDNA database of the University of Tokyo: <http://fullmal.ims.u-tokyo.ac.jp/index.html>

Global networks and databases for malaria researchers

- NCBI malaria pages: <http://www.ncbi.nlm.nih.gov/projects/Malaria>
- WHO/TDR malaria pages: <http://www.who.int/maldb-www/who.html>
- Malaria Foundation International (MFI): <http://www.malaria.org>
- Multilateral Initiative on Malaria (MIM): <http://www.mim.su.se/english/index.asp>
- Malaria Research and Reference Reagent Resource Centre (MR4): <http://www.malaria.mr4.org>

More sites with information on the biology of *Plasmodium*

- WHO malaria pages: http://www.who.int/health_topics/malaria/en
- Biology, biochemistry & physiology of *Plasmodium* at the Hebrew University of Jerusalem: <http://sites.huji.ac.il/malaria/index.html>
- Basic cellular and molecular biological processes of *Plasmodium* at Tulane University: <http://www.tulane.edu/~wiser/malaria/cmb.html>
- Genome atlas maps of the 14 *P. falciparum* chromosomes at the Centre for Biological Sequence Analysis of the Technical University of Denmark: <http://www.cbs.dtu.dk/services/GenomeAtlas>
- Three-dimensional structures of 14 *Plasmodium* proteins at the Toronto-based Structural Genomics Consortium website: <http://www.sgc.utoronto.ca>
- Last but not least our own website, the *P. berghei* model parasite pages of the Leiden University Medical Centre (LUMC): <http://www.lumc.nl/1040/research/malaria/malaria.html>

immune response; many of these 250–300 proteins possess host-like extracellular adhesion domains. Curiously, only 8% of the *P. falciparum* genes could be assigned functions in metabolism, in contrast to 17% of the genes of the yeast *Saccharomyces cerevisiae*⁷. This suggests either that enzymes are more difficult to identify by sequence homology in *P. falciparum* owing to its great evolutionary distance from other well-studied organisms or that the *P. falciparum* genome contains fewer enzymes as a consequence of its parasitic lifestyle. The complement of transport molecules appeared to be similarly reduced compared with free-living organisms.

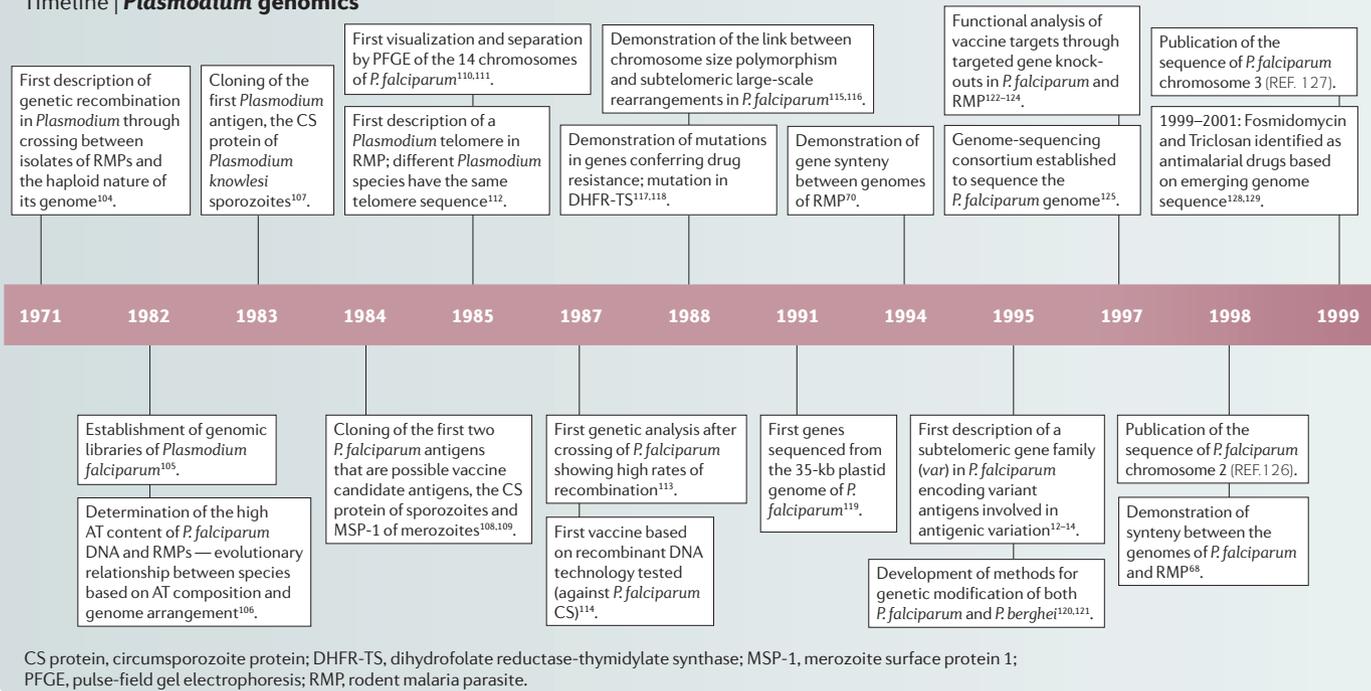
Approximately 10% of the nuclear-encoded proteins are predicted to target to the apicoplast^{2,8,9}. The apicoplast is thought to serve as an organization centre for certain metabolic pathways (isoprenoid and fatty-acid biosynthesis), was probably acquired from non-green algae by endosymbiosis, and is found in many apicomplexan parasites (for reviews, see REFS 10,11). The evolutionary origin of the apicoplast immediately suggested potential drug targets based on herbicides and antibacterial agents.

Clustering of *P. falciparum* subtelomeric genes based on hidden Markov models revealed that nearly 50% of these genes form 12 distinct *P. falciparum*-specific gene

Contig

A physical map of contiguous genomic DNA assembled using overlapping cloned segments.

Timeline | *Plasmodium* genomics



families, and of these only five subtelomeric gene families are shared with the RMPs (see **Supplementary information S1** (table)). These gene families account for 5–10% of all genes, and include gene families that are involved in immune evasion and sequestration (*var*^{12–14}, 59 members) and putative variant antigens (*rif*, 149 members, and *stevor*, 28 members^{15,16}). Although some genes that belong to the defined gene families are located in internal chromosomal regions, most are distributed in the subtelomeric regions, internal to the complex species-specific repeats that abut the telomere repeats².

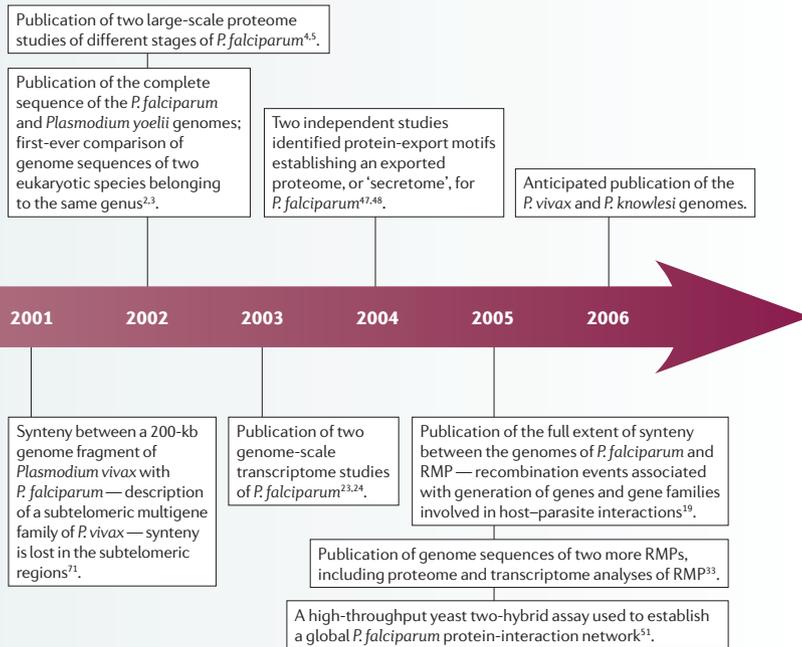
Box 2 | **The *Plasmodium* life cycle**

Plasmodium species all share the same life cycle, the stages of which are depicted in FIG. 3. The haploid sporozoites enter the bloodstream with the saliva of a feeding anopheline mosquito. After invasion of a liver cell, the parasite undergoes a period of growth (G1 phase of the cell cycle), followed by multiple genome replications and mitotic divisions (S/M phase) resulting in the production of >20,000 daughter parasites — merozoites — that invade erythrocytes after release from the hepatocyte in the circulation. After a period of intra-erythrocytic growth, several mitotic divisions result in the production of 16–32 new merozoites. The duration of the erythrocytic cycle and the number of merozoites produced are species specific. The asexual blood-stage parasites are the cause of pathology.

In the blood, a small percentage of parasites develop into sexually committed cells, the female and male gametocytes that are the precursor cells of the gametes. These cells are arrested in G0 phase and are only activated to produce gametes in the midgut when ingested as part of the bloodmeal of the female anopheline mosquito. Gamete formation is followed by fertilization, resulting in the diploid zygote. After one round of meiotic division, the zygote develops into a motile ookinete, penetrates the cells of the midgut and traverses the midgut wall to form an oocyst on the basolateral lamina. After growth and multiplication, >10,000 sporozoites develop in a single oocyst. When the motile sporozoites are released, they migrate through the haemolymph to invade the salivary glands, from which they are ready to infect a new victim and continue the cycle.

All chromosomes harbour some copies of one or more of the gene families, but the composition and order vary. *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), encoded by the *var* genes, is demonstrably associated with clonal antigenic variation, a strategy used in the erythrocytic stages to evade the adaptive host immune response. PfEMP1 also has an important role in the sequestration of infected erythrocytes in capillaries in the brain (resulting in cerebral malaria) and other tissues, including the placenta. The subtelomeric location of 60% of the *var* genes should facilitate the generation of diversity in the gene repertoire of this family^{17,18}, and the high frequency of (ectopic) recombination in these regions could also contribute to the observed size variation in the subtelomeric regions. There are also seven non-subtelomeric *var* loci containing between one and seven copies, regularly interrupted by *rif* genes (FIG. 1).

Smaller gene families that encode diverse functions are also found (see **Supplementary information S1** (table)), for example, genes encoding acyl coenzyme A synthetases (11 members) and receptor-associated protein kinases¹⁹ (21 members; FIG. 1). The genome sequence also expanded our knowledge of the extent of gene families that encode vaccine candidates, and might yet reveal new antigens for vaccine development. For example, before the sequencing initiative, the 6-Cys family, which contains transmission-blocking vaccine candidates, was thought to have three members (P48/45, P12 and P230); we now know that 10 members are expressed at different stages of the life cycle, but principally in gametocytes as five are exclusive to this stage and one (P36) is expressed in both gametocytes and sporozoites (for review, see REF. 20).



The *P. falciparum* transcriptome

Several DNA microarray studies have been published, ranging from analysis of gene transcription using random clones selected from genomic DNA libraries^{21,22} to more recent quasi-global surveys of transcription^{23,24} based on oligonucleotides designed using the emerging genome sequence. The arrays are used to probe RNA from defined parasite stages. Additionally, gene transcription in defined parasite stages has been investigated in several EST surveys^{25–28} by different techniques creating and analysing stage-specific enriched cDNA libraries^{29,30} and through serial analyses of gene expression (SAGE)^{31,32}.

Two microarray studies analysing transcription during the blood stages revealed that the parasite transcribes a large core of its genome during blood-stage development. In one study, 83% (REF. 23) of the genes present on the chip (97.2% of the annotated genes of the genome are represented on the chip) were shown to be expressed during the asexual blood stages, with 24% expressed exclusively during this part of the life cycle. Using different cut-off criteria, the other study found that at least 60% (REF. 24) of the genes in the genome are expressed during asexual blood-stage development. Furthermore, both studies found clear patterns of stage-specific gene transcription. Both studies also showed remarkable concordance in gene-transcription patterns and pinpointed genes encoding surface proteins that might serve as vaccine candidates for specific parasite stages. The study by Bozdech *et al.*²⁴ suggested a cascade of transcriptional activation during blood-stage development, where transcripts are produced in an ordered manner as and when they are required to fulfil the demands of the cell (the 'transcripts-to-go' model). This raises the hope that inhibition of a few key early transcription factors might provide a means

to arrest parasite development, a concept that remains generally valid despite the implications of the more recent discovery of translational repression in certain life-cycle stages^{33,34}. The authors of both microarray studies proposed that the groups of genes with similar transcription profiles might be involved in similar functions or cellular processes, perhaps giving insight into the role of the genes for which the annotation did not reveal a function. SAGE technology revealed an unusual feature of blood-stage transcription — the significant accumulation of antisense transcripts in a stage-specific fashion^{31,32}. At present, this is thought to represent some level of gene regulation, but the mechanism, and indeed proof of function, of antisense RNA in *Plasmodium* remains obscure at present.

By including the invasive sporozoite and merozoite stages in their analysis, Le Roch *et al.*²³ were able to identify gene groups that are associated with cell invasion, emphasizing the similarities of these stages with stage-specific modulation of gene-transcription patterns according to context. For example, one of the identified clusters contains genes transcribed in sporozoites, gametocytes and schizonts (at the onset of merozoite production) that could be essential for gliding motility and host-cell invasion, including actin and myosin.

Continued data mining of the published *P. falciparum* transcriptome, in addition to new transcriptome studies of defined developmental stages or mutant parasites, will provide a better understanding of the biology of the malaria parasite. Some examples of such studies are the analysis of the antioxidant defence system³⁵, the pentose-phosphate pathway³⁶, the transcription of variable antigens³⁷ and detailed analysis of gametocyte development^{38–40}. Not surprisingly, it seems that at least 60% of the *Plasmodium* genes are transcribed at multiple life-cycle stages, but stage- and strategy-specific transcription can also be found.

The *P. falciparum* proteome

Several detailed high-throughput mass-spectrometry studies of the *P. falciparum* proteome have been published. Reassuringly, the protein content of the different blood stages agrees well with the presence of transcripts of the genes encoding these proteins^{23,24} and, therefore, the proteome data generally support the 'transcripts-to-go' model.

Florens *et al.* characterized the proteome of four stages: sporozoites, merozoites, trophozoites and gametocytes⁴. Three additional proteomes from the ring, schizont and gamete stages were also included later in an even more global approach³⁴. Of 2,415 proteins identified in the first study (later rising to 2,904), only 6% were expressed in all four stages, whereas more than half were unique to one stage. Almost half of the sporozoite proteins were stage specific, whereas for the blood stages the numbers ranged from 20–33%. These data form a sharp contrast with the global gene-transcription study analysing the same stages, which reported that 35% of the 5,119 genes on the chip are expressed in sporozoites, gametocytes and blood stages, whereas just under 30% appeared stage specific²³. This discrepancy is most easily explained by

Sequestration

Plasmodium-infected erythrocytes can bind to endothelial cells of blood capillaries in different organs and tissues, including brain and placenta (but also to uninfected erythrocytes (rosetting), platelets (clumping) and dendritic cells). The site and degree of sequestration depend on the malaria species as well as the genes expressed on the erythrocyte surface by the particular isolate.

Clonal antigenic variation

Malaria parasites export proteins to the surface of the infected erythrocyte that can be recognized by the host immune system. An important *P. falciparum* surface protein is PfEMP1, which has a direct role in sequestration of the parasite in the capillaries through interaction with a range of endothelial-cell-surface proteins such as ICAM1 and CD36. Clonality results from the fact that each parasite expresses only one of the 59 *var* genes encoding PfEMP1; however, the variation (switch from the expression of one *var* gene to another) can occur at a rate of up to 2% per generation.

P. falciparum chromosomes I–XIV

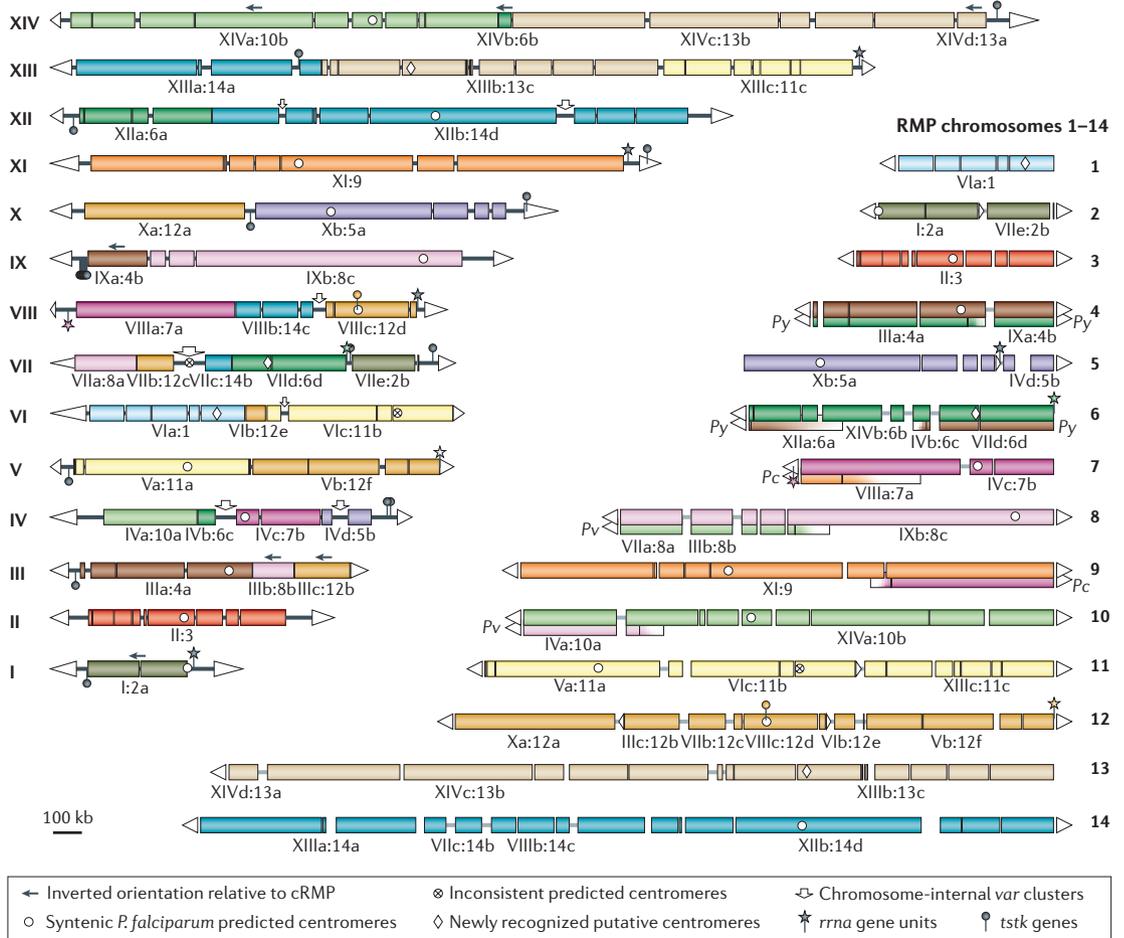


Figure 1 | A whole-genome synteny map of *Plasmodium falciparum* and three rodent malaria parasites (RMPs). A synteny map of the core regions of all chromosomes of *P. falciparum* and the RMPs, showing the 36 synteny blocks, 22 synteny breakpoints, 14 predicted centromeres, *P. falciparum*-specific indels and translocations in the RMP chromosomes. The 36 synteny blocks, coloured according to their chromosomal location in the composite RMP (cRMP) genome, are named with a Roman and an Arabic number referring to the corresponding chromosomal location in *P. falciparum* and the cRMP genome, respectively. Letters give the order in which the synteny blocks are connected. Small black arrows indicate the inverted orientation of a synteny block in *P. falciparum* relative to the cRMP genome. Indels containing *P. falciparum*-specific intrasyntenic genes are indicated by interruption of the coloured synteny blocks. Synteny blocks forming the cRMP chromosomes are linked by grey lines. Telomeres and subtelomeric linked ends are shown as white arrowheads. Bars under the cRMP chromosomes represent the differences in the organization of the synteny blocks of *Plasmodium yoelii* (Py), *Plasmodium chabaudi* (Pc) and *Plasmodium vinckei* (Pv) as a result of translocations. Colours indicate the cRMP chromosome with which recombination has taken place and colour gradients represent the ill-defined regions of the translocation breakpoints. Reproduced with permission from *PLoS Pathogens* REF. 19.

Stage-specific enriched cDNA libraries

Various techniques have been developed for large-scale generation and cloning of cDNA derived from mRNA present in different organisms, tissues or life-cycle stages to compare differences in gene expression between them. Differential display uses a combination of arbitrary primers on total cellular cDNA from which uniquely expressed PCR products are selected for further analyses. Subtractive hybridization compares two sets of cDNA from different sources (for example, parasite life-cycle stages) by hybridizing them with one in excess (the driver) and removing any hybrid sequences (and driver), leaving the unhybridized, uniquely expressed genes for further study.

Serial analyses of gene expression

(SAGE). SAGE can be used for quantitative and simultaneous analysis of a large number of transcripts without the need for a completely annotated genome. cDNA is digested with an enzyme that is expected to cleave all transcripts at least once, and linker molecules are attached. A second restriction enzyme binds to the linker sequence and cleaves away from its binding site to release short (10–14 bp) cDNA (SAGE) tags. These tags are subsequently dimerized, amplified, ligated into long concatamers and cloned. Sequencing the concatamers quantifies gene expression and is characteristic of the tissue or cell type from which the cDNA was isolated.

Limulus coagulation factor C domain

Named after the best-characterized protein containing this domain, Limulus coagulation factor C. This ~100-amino-acid domain has been proposed to be involved in lipopolysaccharide binding.

the greater sensitivity of microarray analyses and technical issues associated with proteome analysis. The same issues applied to the second study, which combined transcriptome and proteome approaches and predicted that up to 40% of the transcripts for which no proteins were detected were insoluble³⁴.

The proteome of the sexually developing parasite was described in more detail by Lasonder *et al.*⁵ Comparison of these data sets, with the help of annotation, identified candidate proteins for transmission-blocking vaccines, such as a family of genes containing Limulus coagulation factor C domains and predicted lectin domains. These proteins, initially thought to be exclusive to gametocytes and

gametes, are also expressed in ookinetes and are essential for oocyst development^{41,42}, indicating that they have a role in the interactions of the parasite with the mosquito midgut epithelium.

The annotation of the genome has benefited considerably from the proteome studies; for example, Lasonder *et al.* reported a set of peptides with significant matches in the *P. falciparum* genome for which no gene model had been predicted⁵, and further analysis of these peptides is ongoing (E. Lasonder, personal communication). Both transcriptome and proteome data revealed no tendency for the genome to be compartmentalized into regions containing genes that are coordinately expressed, ruling out

an operon-like organization. Genome-wide transcription analysis revealed just 14 clusters of three or more co-regulated genes (60 in total)²⁴, although no such clustering could be identified from the *P. falciparum* proteome⁴, probably hampered by the lower coverage of these data. However, a tandem array of five genes encoding proteins located in the Maurer's clefts (MCs) has been reported⁴³.

In addition to the reported proteomes of the whole life-cycle stages, proteome studies have focused on specific organelles and structures. For example, the proteome of gradient-purified detergent-resistant membranes of mature blood-stage parasites (late schizonts/merozoites) has been analysed⁴⁴. These membranes are greatly enriched in glycosylphosphatidylinositol-anchored proteins (GPI-APs) and their putative interacting partners. GPI-APs coat the surface of extracellular *P. falciparum* merozoites, and several are validated candidates for inclusion in a blood-stage malaria vaccine. In addition to detecting confirmed GPI-APs, this study identified new GPI-APs and several other novel, potentially GPI-AP-interacting proteins that are predicted to localize to the merozoite surface and/or apical, invasion-associated organelles (rhoptries and micronemes).

Furthermore, the proteomes of infected erythrocyte membranes and MCs of mature trophozoites and schizonts have been investigated, revealing 36 (REF. 45) and 50 (REF. 43) candidate proteins, respectively. MCs are parasite-derived membranous structures in the erythrocyte cytosol that are thought to be involved in the transport of parasite proteins to the erythrocyte surface⁴⁶. Perhaps surprisingly, the two data sets share only four proteins, which could reflect the different methods of protein-sample preparation and detection. Alternatively, this lack of overlap could suggest that the parasite proteins on the erythrocyte surface are only transiently associated with the MCs, or that proteins residing in the MCs are more easily detected. Comparison of the *P. falciparum* genes encoding MC proteins with the RMP genes using the RMP-*P. falciparum* synteny map revealed that 36 of the 50 genes (72%) are syntenic with the RMPs or belong to locally expanded gene families shared between the different species (T.W.A.K., unpublished observations). The relatively large proportion of syntenic orthologues encoding MC proteins indicates that a considerable part of the protein-export machinery is conserved between *P. falciparum* and the RMPs.

The *P. falciparum* 'secretome' and 'permeome'

Malaria parasites secrete proteins across the vacuolar membrane into the erythrocyte cytosol or to the erythrocyte membrane, inducing modifications of the erythrocyte that are necessary for parasite survival, but which are also associated with disease. Two studies have independently identified a conserved sequence motif (RXLX(E/Q)) in such secreted proteins, termed either the *Plasmodium* export element (PEXEL)⁴⁷ or the vacuolar transport signal (VTS)⁴⁸. Bioinformatics using the PEXEL/VTS signal sequence predicts a 'secretome' of 300–400 proteins for *P. falciparum* (~8% of all genes). In addition to 225 *var*, *rif* and *stevor* genes, the secretome includes 160 genes

encoding proteins that are likely to be involved in remodelling of the host erythrocyte, including heat-shock proteins, kinases, phosphatases and putative transporters⁴⁷; this vastly expands the number of potential vaccine and drug targets. The PEXEL/VTS motif seems to be distinct from known cellular-transport signals, which indicates that it might be a novel eukaryotic secretion signal associated with intracellular parasites. Interestingly, a similar sequence motif (RXLR) has been found in oomycete effector genes associated with avirulence⁴⁹. The plant-pathogenic oomycetes require host tissue for at least parts of their life cycle, and the authors hypothesized that this motif, like the *Plasmodium* PEXEL/VTS motif, mediates transport to the host cells.

In addition to the transport of parasite proteins to the erythrocyte, intra-erythrocytic parasites take up nutrients from the erythrocyte cytosol and excrete metabolic waste products. Membrane-transport proteins mediate these processes but are also implicated in antimalarial drug resistance. Furthermore, the parasites need ion channels to maintain their ion homeostasis. The initial annotation of the *P. falciparum* genome identified only a limited number of transporters and no channels. By combining different bioinformatic approaches, several putative ion channels and >100 transporters were identified, including equal numbers of known and candidate transporters for a range of organic and inorganic nutrients⁵⁰, collectively termed the *P. falciparum* 'permeome'. Most of the permeome could be organized in known classes of transporters, but the authors found that 17% (19 genes) show no sequence homology to any known family of transporters, and five of these genes were grouped as a novel family of *P. falciparum*-specific transporters. Although this doubled the amount of candidate transporters compared with the number reported in the genome paper, the transporter-gene content of the *P. falciparum* genome (~2.1%) is the lowest reported for any organism so far.

P. falciparum protein-interaction networks

Understanding the interactions between proteins can provide insights into the function of, and functional relationships between, these proteins. Recently, the first large-scale analyses of interactions between proteins during the asexual blood stages of *P. falciparum* have been published^{51,52}. Using a high-throughput yeast two-hybrid assay, 2,846 interactions were identified involving 1,312 largely uncharacterized proteins in 29 highly connected protein complexes. By combining information on protein interactions with patterns of co-expression and putative function, informed by annotation and the presence of specific domains, groups of interacting proteins were identified that have a role in chromatin modification, transcription, mRNA stability and ubiquitination, and invasion of host cells.

Improved insights into the structure–function relationships of increasing numbers of proteins might reveal new drug targets. Several groups have begun initial attempts to achieve a larger-scale protein-structure analysis by generating expression libraries of soluble proteins. The Structural Genomics Consortium has started an admirable initiative to attempt a high-throughput elucidation

Maurer's clefts

Complex parasite-derived membranous structures in the erythrocyte cytosol that are closely associated with the erythrocyte plasma membrane and are thought to have an important role in the sorting and transport of parasite proteins to the infected erythrocyte surface.

Rhoptries

Rhoptries are located at the apical end of certain invasive stages of apicomplexan parasites, such as *Plasmodium* and *Toxoplasma*, and have an important role in host-cell adhesion and invasion, as well as in the establishment of the parasitophorous vacuole. Sporozoites and merozoites each contain two of these unique secretory organelles, whereas the third invasive stage of *Plasmodium*, the ookinete, lacks rhoptries.

Micronemes

Like rhoptries, micronemes are located to the apical end of apicomplexan parasites and have an important role in gliding motility, host-cell adhesion and invasion. All three invasive stages — sporozoites, merozoites and ookinetes — contain these organelles.

Reticulocytes
Young erythrocytes.

Normocytes
Mature erythrocytes.

of three-dimensional structures of *Plasmodium* proteins (BOX 1). The structural data produced are freely available and, so far, 19 proteins from different *Plasmodium* species and other apicomplexan parasites have been resolved.

Comparative genomics

Although analysis of a single genome provides tremendous biological insights for any given organism, comparative analysis of multiple genomes can provide substantially more information on the physiology and evolution of genomes, and increases our ability to identify and assign putative functions to predicted coding regions. Orthology recognition is becoming increasingly sophisticated, and bioinformatic methods to improve *Plasmodium* annotation through the recognition of global orthologies have been developed to discover and annotate biosynthetic pathways^{25,53}. Comparative genomics can also help to identify orthologous genes or refine gene predictions through local alignments, substantially improving multi-exon gene models^{3,54}. When closely related species within a single genus are compared, this should provide additional levels of insight into, for example, the repertoire of species-specific genes that might be associated with differences in lifestyle, such as the invasion of reticulocytes versus normocytes by *Plasmodium* merozoites, and even into speciation.

Animal models of malaria, although limited, have long been established as alternative means to gain insights into many aspects of the biology underlying the parasite–host/vector interactions that cannot be obtained readily or ethically working with the human malaria parasites *P. falciparum* and *Plasmodium vivax*. In addition to several primate parasites (for example, *Plasmodium reichenowi*, a close relative of *P. falciparum* that infects chimpanzees, and *Plasmodium knowlesi*, which is more closely related to *P. vivax*) and a chicken parasite (*Plasmodium gallinaceum*, which has an intriguing

phylogenetic relationship with all four human malaria parasites^{55,56}), much work is done using RMPs as they are cheaper to maintain *in vivo* and there are fewer ethical concerns in the handling of their host organisms.

Significant amounts of genome data are available for all of the aforementioned parasites (TABLE 1), including the second most important human malaria parasite, *P. vivax*⁵⁷. Its genome sequence has almost been completed, and annotation and analyses are drawing to a close (see Gene Indices database in BOX 1; Jane Carlton, personal communication). These extensive genome data sets from different *Plasmodium* species have not only facilitated comparative genomics, but have also given rise to significant post-genomic studies, characterizing both the transcriptome and proteome of different life-cycle stages. Comparison of the *P. falciparum* genome with other genome data available in 2002 showed that 60% of the annotated *P. falciparum* genes could not be assigned functions and could therefore encode functions that are unique to *P. falciparum* or to the genus *Plasmodium*. More recently, the genomes of several other unicellular parasites have been published, allowing cross-genus genome comparisons of closely related parasites. The list of unicellular parasites for which significant amounts of genome sequence are now available includes two apicomplexan parasites infecting humans, *Cryptosporidium parvum*⁵⁸ and *Cryptosporidium hominis*⁵⁹; two apicomplexan parasites infecting cattle, *Theileria parva*⁶⁰ and *Theileria annulata*⁶¹; *Entamoeba histolytica*⁶²; and three kinetoplastid parasites, *Trypanosoma brucei*⁶³, *Trypanosoma cruzi*⁶⁴ and *Leishmania major*⁶⁵, and the genome sequence of *Toxoplasma gondii* is nearing completion.

The first comparative analysis of the genomes of two apicomplexan species, *P. falciparum* and *C. parvum*, showed that both lineages have acquired protein-adhesion domains originating from proteins of their animal hosts,

Table 1 | **Current status of the *Plasmodium* genome projects (March 2006)**

Species	Genome Size	Coverage*	Maximum contig size	Annotated full-length genes	Sequencing centre
<i>Plasmodium falciparum</i> (3D7 strain)	23.3 Mb	14.5×	3.3 Mb	5,403	SGTC, TIGR, WTSI
<i>P. falciparum</i> (Dd2 strain)	~23 Mb	8×	na	na	BI
<i>P. falciparum</i> (HB3 strain)	~23 Mb	8×	na	na	BI
<i>P. falciparum</i> (Ghanaian clinical isolate)	23 Mb	8×	na	na	WTSI
<i>P. falciparum</i> (IT strain)	23 Mb	Ongoing (1×	na	na	WTSI
<i>Plasmodium vivax</i>	26.8 Mb	10×	2.1 Mb	5,393	TIGR
<i>Plasmodium reichenowi</i>	25–27 Mb	Ongoing	na	na	WTSI
<i>Plasmodium knowlesi</i>	25 Mb	8×	na	na	WTSI
<i>Plasmodium berghei</i>	25 Mb	8×	37 kb	4,617	WTSI
<i>Plasmodium chabaudi</i>	25 Mb	8×	17 kb	4,100	WTSI
<i>Plasmodium yoelii</i>	23.1 Mb	5×	51 kb	4,034	TIGR
<i>Plasmodium gallinaceum</i>	25 Mb	3×	na	na	WTSI

*Average number of sequence reads per nucleotide. BI, Broad Institute; na, not available; SGTC, Stanford Genome Technology Center; TIGR, The Institute for Genomic Research; WTSI, Wellcome Trust Sanger Institute.

and identified at least 145 apicomplexan-specific genes⁶⁶. Initially, comparative genome analyses of *Theileria*, *Cryptosporidium* and *Plasmodium* species with other public genome databases indicated that the genomes of all three apicomplexan lineages have an unexpected paucity of specific transcription factors, despite their complex life cycles. However, a new apicomplexan family of genes with apetala 2 (AP2)-integrase DNA-binding domains was found; AP2 domains are predominantly found in plant transcription factors⁶⁷. Further discussion of (comparative) genomics of the other apicomplexan and kinetoplastid parasites is beyond the scope of this review, but it is clear that detailed comparisons of the genomes of these species will help to unravel the function of many hypothetical *Plasmodium* genes and should lead to new insights into the complex parasitic lifestyles.

Comparative genomics between species

Extensive synteny between genomes. Comparative genome analysis between *Plasmodium* species was initiated by gene-mapping studies on separated chromosomes^{68–70}, followed by more detailed analysis of small fragments of individual chromosomes^{71,72}. In general, these studies demonstrated significant conservation of gene-linkage groups (synteny) between different species. By definition, synteny is the conservation of gene association in organized blocks. Within the blocks there can be deletions and changes in gene order but the syntenic relationship of the genes remains unaltered. However, the complete picture of the degree of synteny in *Plasmodium* remained unclear until sufficient sequence data were available.

The high degree of synteny between more distantly related *Plasmodium* species was demonstrated with the publication of the extensive genome shotgun sequence of the RMP *P. yoelii*³. Contigs covering >70% of the *P. yoelii* genome could be aligned along the core regions of the 14 *P. falciparum* chromosomes. The similarity of these two *Plasmodium* genomes was not only demonstrated by the high level of synteny but also mirrored by the predicted gene content. This was the first-ever comparison of the gene content of two eukaryotic species within a single genus, and it identified >3,300 *P. yoelii* orthologues of the 5,268 predicted *P. falciparum* genes. Although the orthologues were predominantly house-keeping genes, orthologues of candidate vaccine antigens involved in parasite–host/vector interactions (for example, circumsporozoite protein (CS) and members of the merozoite surface protein (MSP) and 6-Cys families) were also identified.

Such high levels of orthology were perhaps unsurprising, given that most of the features of the life cycle are conserved between different *Plasmodium* species. However, the validation of model malaria species that is provided by their genetic similarity to those species that infect humans has emphasized the fact that structure–function studies on *P. falciparum* vaccine candidates could be carried out with the more accessible, tractable model species. Therefore, the molecular mechanisms underlying gamete fertilization⁷³ and sporozoite motility⁷⁴ could reasonably be studied in model systems. Non-primate models might be less appropriate to investigate adaptive processes of

human parasites, such as the ability of *P. falciparum* to successfully invade human erythrocytes through several independent routes, which are different from the routes of the RMPs and *P. vivax*. The complexity and flexibility of erythrocyte invasion by *P. falciparum* might have evolved as part of a selection and counter-selection ‘arms race’ that model species clearly cannot recreate.

The initial comparisons of the genomes of different *Plasmodium* species have recently been extended with the publication of two additional partial shotgun genome-sequence data sets (4× coverage each) from the RMPs *P. berghei* and *Plasmodium chabaudi*³³. The virtually complete synteny and high levels of nucleotide-sequence identity (88–92%) between the genomes of the three RMPs allowed the construction of composite RMP contigs, extending the contig size by an average of 400% (REF. 19). Combining these contigs with chromosome-mapping studies has enabled complete comparative synteny maps to be compiled for the ‘prototype’ RMP genome and that of *P. falciparum*^{19,33} (FIG. 1). These maps again showed the extensive synteny of the internal chromosome regions. Interestingly, a minimum of only 15 gross chromosomal rearrangements that reshuffle the 36 synteny blocks is needed to convert the *P. falciparum* genome into the composite RMP genome and vice versa¹⁹. Clearly, as comparative genomics is expanded to more *Plasmodium* species, it should be possible to reconstruct the organization of the minimal genome of the most recent common ancestor of the genus.

The combined sequence data from the different RMPs improved the *P. falciparum* orthologue predictions and revealed a conserved set of 2,125 genes with orthologues present in the data sets of all four species. Perhaps more telling is the fact that ~4,500 of the 5,268 *P. falciparum* full-length protein-encoding genes had an orthologue in at least one of the three RMP genome data sets³³. The discrepancy between these numbers is largely due to gaps in the sequences of the three RMPs. The 736 *P. falciparum*-specific genes without any RMP orthologue were analysed in more detail; 575 genes were located in the subtelomeric regions and sharply defined the boundaries between the subtelomeric and core regions of the chromosomes, whereas the remaining 161 *P. falciparum*-specific genes, as well as seven newly identified putative genes associated with chromosome-internal *var* clusters, were located in the core regions of the chromosomes¹⁹.

Lack of synteny in the subtelomeric regions. The subtelomeric regions of *Plasmodium* chromosomes lack synteny, which stems from their generally distinct content of gene families (for example, the *var*, *rif* and *stevor* genes in *P. falciparum* are replaced by other families in RMPs and *P. vivax*) and from the presence of large numbers of species-specific non-coding repeat sequences. However, the gene content of the subtelomeric regions of different *Plasmodium* species is not completely species specific. The RMPs and *P. vivax* share a distinct family of subtelomeric variant genes, collectively known as the *pir* (*Plasmodium* interspersed repeats) superfamily, originally described in *P. vivax* as *vir* genes⁷⁵. The *pir* superfamily is predicted to be large, with ~150–850 members in each species, and

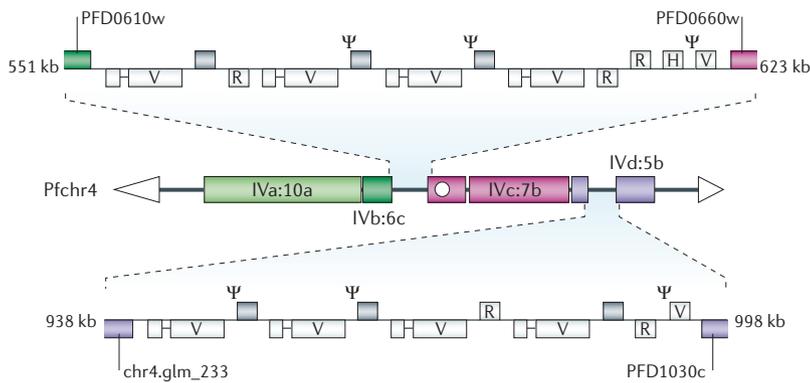


Figure 2 | Inter- and intrasyntenic var clusters of *Plasmodium falciparum* chromosome 4. Comparison of the *P. falciparum* and composite rodent malaria parasite (RMP) genomes revealed the presence of *P. falciparum*-specific gene clusters either marking synteny breakpoints or disrupting synteny blocks, termed inter- and intrasyntenic regions, respectively. Here, two chromosome-internal var clusters of *P. falciparum* chromosome 4 (Pfchr4) are shown, one linking regions with synteny to RMP chromosomes 6 (green) and 7 (pink), the other disrupting a region syntenic to RMP chromosome 5 (lavender). The annotations of the syntenic (coloured) genes flanking the regions are provided and the *P. falciparum*-specific genes (white) are designated as follows: V, var genes; R, rif genes; H, hypothetical protein. The vicar elements are shown in grey. All pseudogenes are marked with a trident.

has been proposed to include the *P. falciparum* rif family⁷⁶. The proteins encoded by certain members of the *pir* superfamily have been localized to the erythrocyte surface⁷⁷, suggesting a role in antigenic variation and immune evasion. However, *P. berghei* proteome data indicate that BIR (the related gene family in *P. berghei* is the *bir* family) proteins might have other, as-yet-unknown functions, as ~9% of the BIR proteins were exclusively detected in mosquito stages³³.

Our knowledge of the exact content and organization of subtelomeric gene families in species other than *P. falciparum* (3D7; 575 genes) remains incomplete. Although some general conservation might be anticipated between species, the true picture of multigene family diversity and organization, and the relationship to expression and function, can only emerge from increased genome sequencing. But what is clear is that the subtelomeric localization of these gene families should promote recombination, in turn generating diversity and therefore confusing synteny^{17,18}.

This tendency to diversify is exemplified by the recent analysis of members of a subtelomeric gene family present in RMPs and *P. falciparum*. These genes were first identified as two different species-specific families through BLAST analyses within the RMPs (*pyst-b*) and *P. falciparum* (*pf-fam-b*), yet could be classified as members of the same interspecies gene superfamily (renamed *pfmc-2tm*) only through shared predicted protein structure (basic proteins with two transmembrane (TM) domains), as they lacked obvious sequence similarities⁷⁸. Shared structural features of proteins encoded by subtelomeric gene families also suggest the existence of a gene superfamily within *P. falciparum* that includes both rif and *stevor* genes⁷⁸. Interestingly, this superfamily might well be extended to include the subtelomeric *pir* genes found in other

Plasmodium species, again indicating the rapid gene evolution that is one consequence of their subtelomeric location. **Supplementary information S1** (table) provides an overview of all *P. falciparum*-specific, RMP-specific and their common subtelomeric gene families.

Disruption of synteny by species-specific genes. Analysis of the 168 chromosome-internal *P. falciparum*-specific genes that are not present in the RMP genomes revealed that 126 of these genes disrupt synteny within the synteny blocks (intrasyntenic genes), with 42 located at the synteny breakpoints between synteny blocks¹⁹ (intersyntenic genes; FIG. 2). Curiously, the synteny breakpoints in the RMPs only harboured five intersyntenic genes. Most *P. falciparum* intra- (62%) and intersyntenic (88%) genes encode predicted exported proteins that are destined for the membrane surface of the merozoite or the infected erythrocyte (including 13 var and 20 rif genes in the intra- and intersyntenic regions, respectively; FIG. 2), and therefore are probably involved in parasite–host interactions. The presence of species-specific genes at synteny breakpoints suggests that gross chromosomal rearrangements might also have helped shape the species-specific gene content of the genomes of *Plasmodium* species. Evidence for the association between such gross chromosomal rearrangements and the generation of species-specific gene families has been found for a gene family encoding transforming growth factor- β (TGF- β) receptor-like serine/threonine protein kinases (*pfstk*)¹⁹, consisting of 21 copies in *P. falciparum* (and possibly *P. reichenowi*) compared with one copy for all other malaria species. This gene family is the first gene family for which a single progenitor gene in other *Plasmodium* species has been identified and which appears to have expanded relatively recently only in *P. falciparum* and *P. reichenowi*, possibly as the result of a specific adaptation to the (common ancestor of) human and chimpanzee hosts.

The analysis of the location of *P. falciparum*-specific genes using the synteny maps revealed that chromosome-internal rearrangements might have influenced the diversity and complexity of the *Plasmodium* genome, increasing the ability of the parasite to interact with its vertebrate host successfully. Furthermore, it indicates that determination of the synteny breakpoints might help to rapidly identify the species-specific gene content of future *Plasmodium* genomes.

Transcriptomes and proteomes of RMPs

Although global transcription profiles might only be correctly interpreted when a whole-genome database is available, intelligent applications of cDNA-based technologies were initiated well before the publication of homologous genome data. Several RMP EST libraries and one *P. vivax* library⁷⁹ have been produced, generating tens of thousands of sequences^{80–83} that can be compiled separately or in a common database, allowing investigations of transcript-specific features such as splicing. In addition, several stage-specific enriched suppression subtractive-hybridization libraries for RMPs throughout development in the mosquito have been generated^{41,84}. These data have not only pinpointed stage-specific

transcripts, but also confirmed the conserved nature of invasive organelles that are associated with host-cell invasion by different stages of *Plasmodium*. In keeping with their morphological similarities, certain invasive organelle proteins are expressed in more than one invasive stage⁸⁰.

DNA microarray studies covering ~70% of the genes of *P. berghei* have been performed on blood-stage parasites and generally support the 'transcripts-to-go' model²⁴. Transcription profiles of purified immature and mature gametocytes showed that these forms share many of the cellular processes common to asexual blood-stage parasites, but enter G0 (G1 arrest)³³. The switch from asexual to sexual development involves significant reprogramming of the transcriptional activity of the parasite⁷⁶ (~25% of the genes on the array were upregulated) carried out on the background of ongoing basic cellular processes³³.

An extensive high-throughput proteome survey has been carried out on five stages of the *P. berghei* life cycle, including the first survey of *Plasmodium* ookinetes and oocysts³³. The study uncovered many predicted ookinete surface proteins that can be explored for their transmission-blocking potential. In addition, this study revealed that the variable antigenic PIR proteins are not only expressed in the blood stages but are expressed as virtually non-overlapping subsets in many different life-cycle stages, such as gametocytes and sporozoites. This expression pattern is more reminiscent of RIFINs (repetitive interspersed family) in *P. falciparum*, expression of which is also not exclusive to blood stages, than of PfEMP1, suggesting that PIR and RIFIN proteins have multiple functions in their respective hosts and are not just involved in antigenic variation during the blood stages. Through comparison of the proteomes (1,836 proteins in total) of the five different stages, a dichotomous strategy of protein expression was visible: the stage-specific expression of proteins that are directly involved in the interaction between the parasite and the different host cells (43 sporozoite-, 372 asexual-blood-stage-, 127 gametocyte-, 317 ookinete- and 89 oocyst-specific proteins), coupled to more constitutive expression of 136 proteins present in at least four of the investigated stages underlying the conserved cellular machinery of the parasite in most of the different life-cycle stages. Again, this relatively low number of constitutively expressed genes is probably due to the lower coverage of proteomic techniques. Conserved elements of the cellular machinery include organelle components associated with cell invasion by the parasite and were not necessarily what might usually be considered housekeeping proteins³³. Proteomic analysis of the merozoite rhoptries of three RMPs identified 36 potential rhoptry proteins⁸⁵. Comparison of these 36 RMP rhoptry proteins with the RMP-*P. falciparum* synteny map revealed that at least four genes are located in the subtelomeric regions and are therefore species-specific (T.W.A.K., unpublished observations). When the RMP rhoptry proteome was compared with the GPI-APs reported for the late-schizont and merozoite stages of *P. falciparum*⁴⁴, only two proteins were found to be conserved. These proteins were also detected in

the MCs⁴³ and infected erythrocyte membranes⁴⁵ of mature trophozoites and schizonts, respectively, further highlighting the technical difficulties of obtaining pure subcellular parasite fractions. Indeed, five additional RMP rhoptry proteins were found in these data sets but not in the GPI-AP data set. A comparison with the genome of *P. vivax*, a parasite that also preferentially invades reticulocytes, when it is released might indicate that some of these 'species-specific' rhoptry proteins are actually conserved and can be functionally analysed for their role in determining the type of erythrocyte that the merozoite invades.

In addition to the proteomes of ookinetes and oocysts of *P. berghei*, which have not yet been reported for a human malaria parasite, the individual proteomes of the male and female gametocytes have been analysed in *P. berghei*. The two proteomes contained 36% (236 of 650) and 19% (101 of 541) sex-specific proteins, respectively⁸⁶. The protein content of the male gametocytes was the most distinct of all the proteomes reported for the life-cycle stages and shared only 69 proteins with the female gametocyte, showing the divergent features of both sexes. This proteome analysis revealed the presence of sex-specific phosphatases and protein kinases that are involved in gender-specific signalling pathways. FIGURE 3 provides a schematic overview of the available transcriptome and proteome data sets of the different *P. falciparum* and RMP life-cycle stages.

Meta-analyses and practical applications

The data sets from various malaria-parasite genome sequences and significant proteome and transcriptome surveys from at least two species provide a unique opportunity to perform comparative analyses and examine aspects of the biology of *Plasmodium* that would not be possible with data sets from a single species. Several studies have already been published that show how the use of different combinations of global *Plasmodium* databases can generate novel insights into parasite-host interactions with potential therapeutic value. Comparative post-genomic analysis of RMP genomes allowed additional detail to be teased out of the predicted protein sequences of orthologous genes. Calculation of the nonsynonymous (d_N) versus synonymous (d_S) nucleotide substitutions can reveal genes encoding more rapidly evolving proteins (high d_N/d_S values) compared with more conserved proteins (low d_N/d_S values)^{87,88}. Not surprisingly, in RMPs, proteins containing predicted signal peptide (SP) sequences and/or TM domains encoding potential secreted or surface proteins that might be exposed to the host immune response showed the highest d_N/d_S ratios³³. Analysis of the expression data generated by both transcriptome and proteome studies showed that a significantly greater number of blood-stage SP/TM proteins had high d_N/d_S ratios compared with mosquito-stage SP/TM proteins. This difference could reflect amino acid changes that have accumulated as a consequence of interactions with the host immune response and, therefore, could identify genes that are under selective immune pressure.

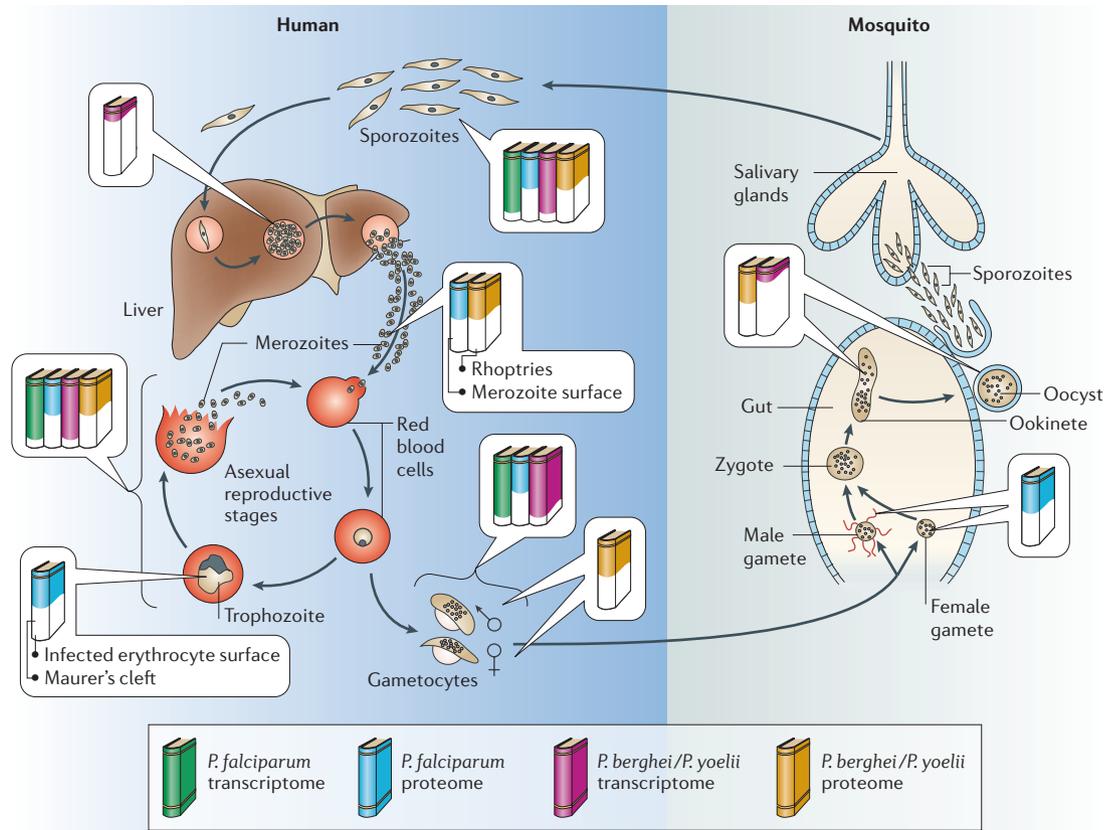


Figure 3 | **Transcriptomes and proteomes of different *Plasmodium falciparum* and rodent malaria parasite (RMP) life-cycle stages.** Transcriptome and proteome studies have covered a wide range of *Plasmodium* life-cycle stages, although no detailed studies have yet been published for the less-accessible liver stages, the short-living zygotes or oocyst/haemocoel sporozoites. The coloured books indicate stages for which *P. falciparum* and RMP transcription and expression data are available. Incomplete data sets or data sets of less pure stages are indicated by incompletely coloured books. Life cycle modified with permission from *Nature* REF. 103 © (2002) Macmillan Publishers Ltd.

Most methods to detect genes under natural selection are based on the comparative analysis of sequences within and between species. However, the single-genome-based method called codon volatility⁸⁹ defines the proportion of point mutations that result in codons encoding a different amino acid. Although this approach has been questioned^{90,91}, observations have confirmed that genes under selective pressure, such as *var* genes, contain relatively more volatile codons as opposed to genes that are under strong purifying selection to maintain their protein sequence, such as housekeeping genes⁸⁹.

An elegant method combining genome and proteome data to identify novel *P. falciparum* antigens used a strategy to mine genomic-sequence databases using epitope predictions to identify novel sporozoite antigens and epitopes recognized by experimentally vaccinated humans⁹². Such an approach could lead to the generation of an antigen map of sporozoite/liver stages ('immunosome'). Another novel method to identify antigens using genomic data has been described for *P. chabaudi*. This approach, termed linkage-group selection, is based on crossing two genetically different *Plasmodium* lines and then applying a selective pressure, in this study immune pressure, on the recombinant progeny. Subsequent analysis of the decrease in the frequency of parental alleles in

the progeny after immune pressure by using quantitative genome-wide molecular markers can identify genome loci containing genes encoding proteins that were under immune selective pressure⁹³. A third method to identify new antigens based on the genome and proteome data has been developed using *P. yoelii*. Exons of genes encoding sporozoite proteins were cloned in a DNA-immunization vector using high-throughput methods. These vectors were then used to immunize mice that were subsequently analysed for their protection against sporozoite infection⁹⁴.

Combined analysis of transcriptome and proteome data gives an insight into the regulation of transcription and protein expression. The genome of *P. falciparum* contains only a limited number of genes encoding transcription-associated proteins (one-third of the number usually found in the genomes of free-living eukaryotes⁹⁵). However, proteins containing CCCH-type zinc-finger motifs, which are often associated with modulation of mRNA decay and translation rates, are abundant⁹⁵, suggesting that post-transcriptional processes have a significant role in regulating *P. falciparum* protein levels. Bioinformatic analysis comparing mRNA-transcript and protein-abundance levels for seven different stages of the *P. falciparum* life cycle indeed implied

mechanisms of post-transcriptional control, either involving interplay between mRNA stability and degradation, gene-specific control of mRNA translation, or a combination of both³⁴. The combination of transcriptome and proteome data for *P. berghei* also demonstrated the presence of post-transcriptional control of gene expression in gametocytes through translational repression. Translational repression was known to affect the expression of two gametocyte-specific transcripts that encode vaccine-candidate antigens (P28 and P25) translated only in the zygote just after fertilization^{96,97}. Comparison of gametocyte transcriptomes and proteomes with the *P. berghei* ookinete proteome identified nine genes undergoing translational repression, and a sequence motif putatively involved in translational repression was subsequently identified in the 1-kb region downstream of these genes. This motif is not conserved in *P. falciparum* but shares a conserved sub-motif, the nanos response element (NRE), to which RNA-binding proteins of the pumilio family (PUF) can bind and which has a role in translational repression⁹⁸. Similar analysis of *P. falciparum* identified two genes that contain an NRE in their 3' untranslated region which have abundant transcripts in the gametocyte stage, whereas the proteins they encode are significantly more abundant in the gamete stage³⁴.

A detailed understanding of the specific mechanisms of transcriptional and translational control in *Plasmodium* might reveal novel therapeutic targets and strategies. For example, targeting the unlocking (derepressing) of translational repression in gametocytes circulating in the blood might lead to inappropriate expression of gametocyte-specific translationally repressed transcripts, possibly resulting in both the inhibition of further gametocyte development and exposure of their protein products (including current vaccine candidates) to the host immune system, generating transmission-blocking immune responses.

Concluding remarks

Malaria research is in a period of intense data collection, ensuring that the 'labels' on each gene in the *Plasmodium* genomes and the proteins that they encode are correct. The wealth of information on gene transcription and protein expression, in addition to the wide range of new techniques that are now available, will prove essential, as it is only through the lens of full and accurate annotation and protein characterization that we will be able to make sense of — and exploit — the genome information. One of the difficulties that malaria researchers

face is that some of the life-cycle stages are difficult to retrieve in pure form, if accessible at all. Although some stages might be accessible in other *Plasmodium* species, such as the RMPs, access to the *P. falciparum* mosquito stages and the availability of a system to study liver-cell invasion by sporozoites will be indispensable to gain a comprehensive overview of the complete *P. falciparum* life cycle. One can only hope that intriguing stages such as the *P. vivax* hypnozoites (the dormant parasites present in liver cells), which are so few in number and small in size, will one day become accessible for genome-scale analyses. Furthermore, the availability of pure subcellular organelles and structures from the different life-cycle stages, including rophtries, micronemes, dense granules and MCs, might shed light on biological processes such as host-cell invasion and modulation.

Although drug- and vaccine-discovery programmes are already (and rightly) underway as a result of the availability of the *Plasmodium* genomes, the hard choices will be at the level of inclusion or exclusion of drug targets or vaccine-candidate antigens for further development. However, the increased knowledge of structural and functional properties of a large number of *Plasmodium* proteins as well as antigenic properties of vaccine candidates will greatly benefit the decision-making process. More educated lead development might also come from large-scale gene-disruption studies. Reverse genetics is a powerful approach that is used in malaria research to specifically alter the parasite genome to explore its biology and gain new insights into gene function and expression. Recently, several reverse-genetic techniques using transposons^{99,100} have become available, but so far no genome-wide studies have been published. In a post-genomic setting, reverse genetics should be one of the principle technologies to be applied to increase our understanding of parasite biology.

HIV/AIDS is a devastating disease that the world has only known for 30 years and for which there is the prospect of combating and controlling the disease and its transmission, should the financial resources be made available to provide the drugs that have been developed. Conversely, malaria is an ancient disease that has been known for thousands of years, and the aetiological agent was first recognized more than 100 years ago, yet malaria is a steadily worsening scourge, with new therapeutics some distance away. Although the financial support for malaria research has improved, significant investment is still required at all levels of investigation, development and application to realize the potential of the *P. falciparum* genome and translate its promise into a tangible effect.

1. Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. & Hay, S. I. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* **434**, 214–217 (2005).
2. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
3. Carlton, J. M. *et al.* Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**, 512–519 (2002).
4. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
5. Lasonder, E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542 (2002).
References 2–5 form the milestone release of the first fully sequenced parasite genome, analyses of its proteome in two independent studies, and the first-ever comparison of genome sequences of two eukaryotic species belonging to the same genus.
6. Waller, R. F. & McFadden, G. I. *Malaria Parasites, Genomes and Molecular Biology* (eds Waters, A. P. & Janse, C. J.) 289–338 (Caister Academic Press, Wymondham, 2004).
7. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–567 (1996).
8. Foth, B. J. *et al.* Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science* **299**, 705–708 (2003).
This study not only shows the signals that are necessary for protein trafficking to the specialized organelle, the apicoplast, but also presents an

- algorithm to accurately predict apicoplast-targeted proteins from the genome sequences.
9. Zuegge, J., Ralph, S., Schmuker, M., McFadden, G. I. & Schneider, G. Deciphering apicoplast targeting signals — feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene* **280**, 19–26 (2001).
 10. Foth, B. J. & McFadden, G. I. The apicoplast: a plastid in *Plasmodium falciparum* and other Apicomplexan parasites. *Int. Rev. Cytol.* **224**, 57–110 (2003).
 11. Ralph, S. A. *et al.* Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nature Rev. Microbiol.* **2**, 203–216 (2004).
 12. Baruch, D. I. *et al.* Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* **82**, 77–87 (1995).
 13. 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).
 14. 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).
- References 12–14 describe for the first time the (largely subtelomeric) gene family that encodes the PfEMP1-protein repertoire that is associated with antigenic variation. The proteins are also important virulence factors, as they mediate sequestration of the infected erythrocytes in capillaries of various organs.**
15. Cheng, Q. *et al.* *stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol. Biochem. Parasitol.* **97**, 161–176 (1998).
 16. 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**, 9353–9358 (1999).
 17. Figueiredo, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marín, J. C. & Scherf, A. A central role for *Plasmodium falciparum* subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J.* **21**, 815–824 (2002).
 18. 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).
 19. Kooij, T. W. *et al.* A *Plasmodium* whole-genome synteny map: indels and synteny breakpoints as foci for species-specific genes. *PLoS Pathog.* **1**, e44 (2005).
- This study revealed the full extent of synteny between *P. falciparum* and three RMP species. The origin of a *P. falciparum*-specific gene family encoding receptor-associated protein kinases could be linked to gross chromosomal rearrangements.**
20. Thompson, J., Janse, C. J. & Waters, A. P. Comparative genomics in *Plasmodium*: a tool for the identification of genes and functional analysis. *Mol. Biochem. Parasitol.* **118**, 147–154 (2001).
 21. Hayward, R. E. *et al.* Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Mol. Microbiol.* **35**, 6–14 (2000).
 22. Mamoun, C. B. *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).
 23. Le Roch, K. G. *et al.* Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508 (2003).
 24. Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**, e5 (2003).
- Data published in references 23 and 24 describe the transcription profiles of various stages of the *P. falciparum* life cycle based on DNA microarrays. The data are easily accessible from the PlasmoDB web site.**
25. Li, L. *et al.* Gene discovery in the apicomplexa as revealed by EST sequencing and assembly of a comparative gene database. *Genome Res.* **13**, 443–454 (2003).
 26. Li, L. *et al.* ApiEST-DB: analyzing clustered EST data of the apicomplexan parasites. *Nucleic Acids Res.* **32**, D326–D328 (2004).
 27. Watanabe, J., Sasaki, M., Suzuki, Y. & Sugano, S. Analysis of transcriptomes of human malaria parasite *Plasmodium falciparum* using full-length enriched library: identification of novel genes and diverse transcription start sites of messenger RNAs. *Gene* **291**, 105–113 (2002).
 28. Watanabe, J., Suzuki, Y., Sasaki, M. & Sugano, S. Full-malaria 2004: an enlarged database for comparative studies of full-length cDNAs of malaria parasites, *Plasmodium* species. *Nucleic Acids Res.* **32**, D334–D338 (2004).
 29. Cui, L., Rzomp, K. A., Fan, Q., Martin, S. K. & Williams, J. *Plasmodium falciparum*: differential display analysis of gene expression during gametocytogenesis. *Exp. Parasitol.* **99**, 244–254 (2001).
 30. Fidock, D. A., Nguyen, T. V., Ribeiro, J. M., Valenzuela, J. G. & James, A. A. *Plasmodium falciparum*: generation of a cDNA library enriched in sporozoite-specific transcripts by directional tag subtractive hybridization. *Exp. Parasitol.* **95**, 220–225 (2000).
 31. Munasinghe, A. *et al.* Serial analysis of gene expression (SAGE) in *Plasmodium falciparum*: application of the technique to A-T rich genomes. *Mol. Biochem. Parasitol.* **113**, 23–34 (2001).
 32. Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L. M. & Wirth, D. F. Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. *Mol. Biol. Cell* **12**, 3114–3125 (2001).
 33. Hall, N. *et al.* A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* **307**, 82–86 (2005).
- Publication of the genome sequences of two additional rodent malaria parasites, *P. berghei* and *P. chabaudi*, including a comparative genome analysis with *P. yoelii* and *P. falciparum*, and transcriptome and proteome analyses of *P. berghei*.**
34. 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).
 35. Bozdech, Z. & Ginsburg, H. Antioxidant defense in *Plasmodium falciparum* — data mining of the transcriptome. *Malar. J.* **3**, 23 (2004).
 36. Bozdech, Z. & Ginsburg, H. Data mining of the transcriptome of *Plasmodium falciparum*: the pentose phosphate pathway and ancillary processes. *Malar. J.* **4**, 17 (2005).
 37. Ralph, S. A. *et al.* Transcriptome analysis of antigenic variation in *Plasmodium falciparum* — var silencing is not dependent on antisense RNA. *Genome Biol.* **6**, R93 (2005).
 38. Gissot, M. *et al.* Transcriptome of 3D7 and its gametocyte-less derivative F12 *Plasmodium falciparum* clones during erythrocytic development using a gene-specific microarray assigned to gene regulation, cell cycle and transcription factors. *Gene* **341**, 267–277 (2004).
 39. Silvestrini, F. *et al.* Genome-wide identification of genes upregulated at the onset of gametocytogenesis in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **143**, 100–110 (2005).
 40. 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).
 41. Dessens, J. T., Margos, G., Rodriguez, M. C. & Sinden, R. E. Identification of differentially regulated genes of *Plasmodium* by suppression subtractive hybridization. *Parasitol. Today* **16**, 354–356 (2000).
 42. Pradel, G. *et al.* A multidomain adhesion protein family expressed in *Plasmodium falciparum* is essential for transmission to the mosquito. *J. Exp. Med.* **199**, 1533–1544 (2004).
 43. Vincensini, L. *et al.* Proteomic analysis identifies novel proteins of the Maurer's clefts, a secretory compartment delivering *Plasmodium falciparum* proteins to the surface of its host cell. *Mol. Cell Proteomics* **4**, 582–593 (2005).
 44. Sanders, P. R. *et al.* Distinct protein classes including novel merozoite surface antigens in raft-like membranes of *Plasmodium falciparum*. *J. Biol. Chem.* **280**, 40169–40176 (2005).
 45. Florens, L. *et al.* Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol. Biochem. Parasitol.* **135**, 1–11 (2004).
 46. Przyborski, J. M., Wickert, H., Krohne, G. & Lanzer, M. Maurer's clefts — a novel secretory organelle? *Mol. Biochem. Parasitol.* **132**, 17–26 (2003).
 47. 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).
 48. Hiller, N. L. *et al.* A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).
- Two independent studies (references 47 and 48) identified a *Plasmodium* host-targeting signal, revealing the *P. falciparum* 'secretome'.**
49. Rehmany, A. P. *et al.* Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two *Arabidopsis* lines. *Plant Cell* **17**, 1859–1850 (2005).
 50. Martin, R. E., Henry, R. I., Abbey, J. L., Clements, J. D. & Kirk, K. The 'permeome' of the malaria parasite: an overview of the membrane transport proteins of *Plasmodium falciparum*. *Genome Biol.* **6**, R26 (2005).
 51. LaCount, D. J. *et al.* A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* **438**, 103–107 (2005).
- A high-throughput yeast two-hybrid method was used to provide a first insight into *P. falciparum* protein interactions.**
52. Suthram, S., Sittler, T. & Ideker, T. The *Plasmodium* protein network diverges from those of other eukaryotes. *Nature* **438**, 108–112 (2005).
 53. McConkey, G. A. *et al.* Annotating the *Plasmodium* genome and the enigma of the shikimate pathway. *Trends Parasitol.* **20**, 60–65 (2004).
 54. van Lin, L. H. *et al.* Interspecies conservation of gene order and intron–exon structure in a genomic locus of high gene density and complexity in *Plasmodium*. *Nucleic Acids Res.* **29**, 2059–2068 (2001).
 55. Waters, A. P., Higgins, D. G. & McCutchan, T. F. *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. *Proc. Natl Acad. Sci. USA* **88**, 3140–3144 (1991).
 56. Escalante, A. A. & Ayala, F. J. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc. Natl Acad. Sci. USA* **91**, 11373–11377 (1994).
 57. Carlton, J. The *Plasmodium vivax* genome sequencing project. *Trends Parasitol.* **19**, 227–231 (2003).
 58. Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445 (2004).
 59. Xu, P. *et al.* The genome of *Cryptosporidium hominis*. *Nature* **431**, 1107–1112 (2004).
 60. Gardner, M. J. *et al.* Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science* **309**, 134–137 (2005).
 61. Pain, A. *et al.* Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. *Science* **309**, 131–133 (2005).
 62. Loftus, B. *et al.* The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865–868 (2005).
 63. Berriman, M. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416–422 (2005).
 64. El Sayed, N. M. *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**, 409–415 (2005).
 65. Ivans, A. C. *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **309**, 436–442 (2005).
 66. Templeton, T. J. *et al.* Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Res.* **14**, 1686–1695 (2004).
 67. Balaji, S., Babu, M. M., Iyer, L. M. & Aravind, L. Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res.* **33**, 3994–4006 (2005).
 68. Carlton, J. M., Vinkenoog, R., Waters, A. P. & Walliker, D. Gene synteny in species of *Plasmodium*. *Mol. Biochem. Parasitol.* **93**, 285–294 (1998).
 69. Carlton, J. M., Galinski, M. R., Barnwell, J. W. & Dame, J. B. Karyotype and synteny among the chromosomes of all four species of human malaria parasite. *Mol. Biochem. Parasitol.* **101**, 23–32 (1999).
 70. Janse, C. J., Carlton, J. M., Walliker, D. & Waters, A. P. Conserved location of genes on polymorphic chromosomes of four species of malaria parasites. *Mol. Biochem. Parasitol.* **68**, 285–296 (1994).

71. Tchavtchitch, M., Fischer, K., Huestis, R. & Saul, A. The sequence of a 200 kb portion of a *Plasmodium vivax* chromosome reveals a high degree of conservation with *Plasmodium falciparum* chromosome 3. *Mol. Biochem. Parasitol.* **118**, 211–222 (2001).
72. van Lin, L. H., Pace, T., Janse, C. J., Scotti, R. & Ponzi, M. A long range restriction map of chromosome 5 of *Plasmodium berghei* demonstrates a chromosome specific symmetrical subtelomeric organisation. *Mol. Biochem. Parasitol.* **86**, 111–115 (1997).
73. van Dijk, M. R. *et al.* A central role for P48/45 in malaria parasite male gamete fertility. *Cell* **104**, 153–164 (2001).
74. Menard, R. Gliding motility and cell invasion by Apicomplexa: insights from the *Plasmodium* sporozoite. *Cell Microbiol.* **3**, 63–73 (2001).
75. del Portillo, H. A. *et al.* A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* **410**, 839–842 (2001).
76. Hayward, R. E. *Plasmodium falciparum* phosphoenolpyruvate carboxykinase is developmentally regulated in gametocytes. *Mol. Biochem. Parasitol.* **107**, 227–240 (2000).
77. Cunningham, D. A. *et al.* Host immunity modulates transcriptional changes in a multigene family (*lir*) of rodent malaria. *Mol. Microbiol.* **58**, 636–647 (2005).
78. Sam-Yellowe, T. Y. *et al.* A *Plasmodium* gene family encoding Maurer's cleft membrane proteins: structural properties and expression profiling. *Genome Res.* **14**, 1052–1059 (2004).
- Publication implying that different species of *Plasmodium* might possess related gene families that encode structurally related proteins but the interspecies similarities of which are not apparent from a simple comparison of primary sequence.**
79. Merino, E. F. *et al.* Pilot survey of expressed sequence tags (ESTs) from the asexual blood stages of *Plasmodium vivax* in human patients. *Malar. J.* **2**, 21 (2003).
80. Kappe, S. H. *et al.* Exploring the transcriptome of the malaria sporozoite stage. *Proc. Natl Acad. Sci. USA* **98**, 9895–9900 (2001).
81. Kaiser, K., Matuschewski, K., Camargo, N., Ross, J. & Kappe, S. H. Differential transcriptome profiling identifies *Plasmodium* genes encoding pre-erythrocytic stage-specific proteins. *Mol. Microbiol.* **51**, 1221–1232 (2004).
82. Abraham, E. G. *et al.* Analysis of the *Plasmodium* and *Anopheles* transcriptional repertoire during ookinete development and midgut invasion. *J. Biol. Chem.* **279**, 5573–5580 (2004).
83. Srinivasan, P. *et al.* Analysis of the *Plasmodium* and *Anopheles* transcriptomes during oocyst differentiation. *J. Biol. Chem.* **279**, 5581–5587 (2004).
84. Matuschewski, K. *et al.* Infectivity-associated changes in the transcriptional repertoire of the malaria parasite sporozoite stage. *J. Biol. Chem.* **277**, 41948–41953 (2002).
- A subtraction-hybridization study that isolated a number of interesting genes expression of which was upregulated after successful colonization of the salivary gland by the sporozoite.**
85. Sam-Yellowe, T. Y. *et al.* Proteome analysis of rhoptry-enriched fractions isolated from *Plasmodium* merozoites. *J. Proteome Res.* **3**, 995–1001 (2004).
86. Khan, S. M. *et al.* Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell* **121**, 675–687 (2005).
87. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275–276 (1977).
88. Kafatos, F. C., Efstratiadis, A., Forget, B. G. & Weissman, S. M. Molecular evolution of human and rabbit β -globin mRNAs. *Proc. Natl Acad. Sci. USA* **74**, 5618–5622 (1977).
89. Plotkin, J. B., Dushoff, J. & Fraser, H. B. Detecting selection using a single gene sequence of *M. tuberculosis* and *P. falciparum*. *Nature* **428**, 942–945 (2004).
90. Friedman, R. & Hughes, A. L. Codon volatility as an indicator of positive selection: data from eukaryotic genome comparisons. *Mol. Biol. Evol.* **22**, 542–546 (2005).
91. Sharp, P. M. Gene "volatility" is most unlikely to reveal adaptation. *Mol. Biol. Evol.* **22**, 807–809 (2005).
- An interesting approach to identify genes of pathogens that are under selective pressure, which relies upon a global comparison of the degree of variation of each predicted gene from the norm of codon usage established for the genome (reference 89), has recently been questioned in other studies (references 90 and 91).**
92. 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).
- High-throughput study profiting from proteome analyses of sporozoites to identify suitable targets for vaccine development.**
93. 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).
94. Haddad, D. *et al.* Novel antigen identification method for discovery of protective malaria antigens by rapid testing of DNA vaccines encoding exons from the parasite genome. *Infect. Immun.* **72**, 1594–1602 (2004).
95. 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).
96. Paton, M. G. *et al.* Structure and expression of a post-transcriptionally regulated malaria gene encoding a surface protein from the sexual stages of *Plasmodium berghei*. *Mol. Biochem. Parasitol.* **59**, 263–275 (1993).
97. del Carmen, R. M. *et al.* Characterisation and expression of pbs25, a sexual and sporogonic stage specific protein of *Plasmodium berghei*. *Mol. Biochem. Parasitol.* **110**, 147–159 (2000).
98. Cui, L., Fan, Q. & Li, J. The malaria parasite *Plasmodium falciparum* encodes members of the Puf RNA-binding protein family with conserved RNA binding activity. *Nucleic Acids Res.* **30**, 4607–4617 (2002).
99. Sakamoto, H. *et al.* Towards systematic identification of *Plasmodium* essential genes by transposon shuttle mutagenesis. *Nucleic Acids Res.* **33**, e174 (2005).
100. Balu, B., Shoue, D. A., Fraser, M. J. Jr & Adams, J. H. High-efficiency transformation of *Plasmodium falciparum* by the lepidopteran transposable element piggyBac. *Proc. Natl Acad. Sci. USA* **102**, 16391–16396 (2005).
101. Bahl, A. *et al.* PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Res.* **31**, 212–215 (2003).
102. Kissinger, J. C. *et al.* The *Plasmodium* genome database. *Nature* **419**, 490–492 (2002).
103. Wirth, D. F. The parasite genome: biological revelations. *Nature* **419**, 495–496 (2002).
104. Walliker, D., Carter, R. & Morgan, S. Genetic recombination in malaria parasites. *Nature* **232**, 561–562 (1971).
105. Goman, M. *et al.* The establishment of genomic DNA libraries for the human malaria parasite *Plasmodium falciparum* and identification of individual clones by hybridisation. *Mol. Biochem. Parasitol.* **5**, 391–400 (1982).
106. Pollack, Y., Katzen, A. L., Spira, D. T. & Golenser, J. The genome of *Plasmodium falciparum*. I: DNA base composition. *Nucleic Acids Res.* **10**, 539–546 (1982).
107. Ellis, J. *et al.* Cloning and expression in *E. coli* of the malarial sporozoite surface antigen gene from *Plasmodium knowlesi*. *Nature* **302**, 536–538 (1983).
108. Dame, J. B. *et al.* Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. *Science* **225**, 593–599 (1984).
109. Hall, R. *et al.* Major surface antigen gene of a human malaria parasite cloned and expressed in bacteria. *Nature* **311**, 379–382 (1984).
110. Kemp, D. J. *et al.* Size variation in chromosomes from independent cultured isolates of *Plasmodium falciparum*. *Nature* **315**, 347–350 (1985).
111. Van der Ploeg, L. H. *et al.* Chromosome-sized DNA molecules of *Plasmodium falciparum*. *Science* **229**, 658–661 (1985).
112. Ponzi, M., Pace, T., Dore, E. & Frontali, C. Identification of a telomeric DNA sequence in *Plasmodium berghei*. *EMBO J.* **4**, 2991–2995 (1985).
113. Walliker, D. *et al.* Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661–1666 (1987).
114. Ballou, W. R. *et al.* Safety and efficacy of a recombinant DNA *Plasmodium falciparum* sporozoite vaccine. *Lancet* **1**, 1277–1281 (1987).
115. Corcoran, L. M., Thompson, J. K., Walliker, D. & Kemp, D. J. Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum*. *Cell* **53**, 807–813 (1988).
116. Patarapotikul, J. & Langsley, G. Chromosome size polymorphism in *Plasmodium falciparum* can involve deletions of the subtelomeric pPFP20 sequence. *Nucleic Acids Res.* **16**, 4351–4340 (1988).
117. Cowman, A. F., Morry, M. J., Biggs, B. A., Cross, G. A. & Foote, S. J. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **85**, 9109–9113 (1988).
118. Peterson, D. S., Walliker, D. & Welles, T. E. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **85**, 9114–9118 (1988).
119. Gardner, M. J., Williamson, D. H. & Wilson, R. J. A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. *Mol. Biochem. Parasitol.* **44**, 115–123 (1991).
120. Wu, Y., Sifri, C. D., Lei, H. H., Su, X. Z. & Welles, T. E. Transfection of *Plasmodium falciparum* within human red blood cells. *Proc. Natl Acad. Sci. USA* **92**, 973–977 (1995).
121. van Dijk, M. R., Waters, A. P. & Janse, C. J. Stable transfection of malaria parasite blood stages. *Science* **268**, 1358–1362 (1995).
122. Menard, R. *et al.* Circumsporozoite protein is required for development of malaria sporozoites in mosquitoes. *Nature* **385**, 336–340 (1997).
123. Sultan, A. A. *et al.* TRAP is necessary for gliding motility and infectivity of *Plasmodium* sporozoites. *Cell* **90**, 511–522 (1997).
124. Crabb, B. S. *et al.* Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell* **89**, 287–296 (1997).
125. Hoffman, S. L. *et al.* Funding for malaria genome sequencing. *Nature* **387**, 647 (1997).
126. Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
127. Bowman, S. *et al.* The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538 (1999).
128. Jomaa, H. *et al.* Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573–1576 (1999).
129. Suroli, N. & Suroli, A. Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*. *Nature Med.* **7**, 167–173 (2001).

Acknowledgements

T.W.A.K. was supported by a Leiden University Ph.D. fellowship.

Competing interests statement

The authors declare no competing financial interests.

DATABASES

The following terms in this article are linked online to:

Entrez Genome Project: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj>
 Cryptosporidium hominis | Cryptosporidium parvum |
 Entamoeba histolytica | Leishmania major | Plasmodium berghei |
 Plasmodium chabaudi | Plasmodium falciparum | Plasmodium gallinaceum | Plasmodium knowlesi | Plasmodium reichenowi |
 Plasmodium vivax | Plasmodium yoelii | Saccharomyces cerevisiae | Theileria annulata | Theileria parva | Toxoplasma gondii | Trypanosoma brucei | Trypanosoma cruzi
 UniProtKB: <http://ca.expasy.org/sprot>
 PfEMP1 | TGF- β

FURTHER INFORMATION

Andrew P. Waters and Chris J. Janse's homepage:
<http://www.lumc.nl/1040/research/malaria/malaria.html>
 GenBank: <http://www.ncbi.nlm.nih.gov/Genbank/index.html>
 Plasmodium falciparum Project Status:
http://www.sanger.ac.uk/Projects/P_falciparum/status.shtml

SUPPLEMENTARY INFORMATION

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