THE ROYAL SOCIETY

Published online 17 January 2002

Orthology between the genomes of *Plasmodium falciparum* and rodent malaria parasites: possible practical applications

A. P. Waters

Department of Parasitology, Leiden University Medical Centre, Postbus 9600, 2300 RC Leiden, The Netherlands (*waters@lumc.nl*)

The work of the consortium that has been formed to complete the entire sequence of the genome of a selected clone of the human malaria parasite, *Plasmodium falciparum*, is almost finished. Already huge tracts of the genome are available as fully assembled chromosomes or large contigs and the work of initial annotation is in an advanced state. Post-genomic research is in one sense the process of furthering the process of annotation, creating biological atlases and preliminary attempts to make global descriptions of gene transcription and proteome analysis are underway. Comparison between significant amounts of genome data from both closely, and more distantly related organisms, can facilitate the identification of genes themselves, coordinately regulated gene expression groups, gene function and genome organization. Models of malaria can fulfil these functions and in addition provide an experimental system wherein predictions can be tested and basic experimental investigations performed within numerous aspects of disease, pathology, parasite–host and parasite–vector interactions. Comparative genomics in *Plasmodium* has already been shown to have informative roles in the completion of annotation and the elucidation of gene function. These roles will be illustrated by example and used as the basis for a discussion of the utility of genome information and malaria models in realizing the desired product of *Plasmodium* genomics, the development of malaria therapies.

Keywords: malaria; comparative genomics; *Plasmodium falciparum*; *Plasmodium berghei*

1. INTRODUCTION

Models of malaria using non-human malaria parasites were developed to serve as convenient laboratory references for the provision of biological insight into the human forms of the disease that would otherwise either be practically, or ethically, impossible or difficult to obtain. Prior to the development of *in vitro* cultivation of *Plasmodium falciparum* (Trager & Jensen 1976), animal models of malaria were widely used and have facilitated descriptions of the biology of the parasite since the initial investigations of Ross (1898). Work on such models continues as the investigation of human malaria parasites can frequently be supplemented by valuable information gained through the use of these animal models. In some situations the models represent the only feasible avenue of detailed investigation, for example the invasion of hepatocytes by sporozoites or the development of the malaria zygote to the ookinete. Successful and widely used models include a wide range of species that infect laboratory rodents (notably *Plasmodium yoelii*, *Plasmodium berghei* and *Plasmodium chabaudi*), birds (*Plasmodium gallinaceum*) and primates (notably *Plasmodium knowlesi* and *Plasmodium cynomolgi* and, less frequently, *Plasmodium reichenowi*). All of these species reproduce many of the biological characteristics of human malaria and, as will become clear, this is, in large part, because the genus *Plasmodium* appears to be relatively well conserved at the level of genome organization and individual gene content and structure. In the modern age, model species of *Plasmodium* were the first to be genetically transformed on both a transient (Goonewardene *et al.* 1993) and stable (Van Dijk *et al.* 1995) basis, establishing one foundation for post-genomic research that will continue to refer to, and use, animal models because of the difficulties of working solely with human-infective species of malaria. Thus it is essential that the genomes of the model malaria parasites are also characterized. This paper attempts to show that comparative analysis of the genome of rodent malaria species and *P. falciparum* can provide practical insights into genome structure and content that can be exploited in gene discovery and vaccine candidate validation.

2. COMPARATIVE GENOMICS

Comparative genomics can describe relative genome composition and identify conserved and species–genus–phylum-specific gene components. The initial comparison of the first three fully sequenced eukaryotic genomes yielded numerous insights into the relative evolution of unicellular and multicellular organisms and their respective genomes (Rubin *et al.* 2000). These ranged from the organizational (the extent of gene duplication) through to the developmental (nature of genes expanded and range of predicted function maintained within the genome). At the microbial level, comparative genomics is already yielding insight into the mechanisms of pathogenesis (Kerr

DOI 10.1098/rstb.2001.1011
© 2002 The Royal Society

1999; Alm *et al.* 2000) and although it is still in its infancy, the same can be expected to be true for *Plasmodium*.

Comparative genomics in *Plasmodium* was first initiated by gene mapping studies on separated chromosomes. It was shown that between different strains of *Plasmodium* species, the gene content of chromosomes was stable, despite extensive size polymorphisms between homologous chromosomes that resulted from copy number variations in subtelomeric repeat elements. Genome rearrangements that alter gene order and content of chromosomes, such as chromosomal translocations, have only been detected infrequently. This was emphasized by the demonstration of a high level of conservation of gene linkage groups (synteny) between four species of rodent *Plasmodium* parasites (Janse *et al.* 1994). The observed level of genetic synteny was reduced as genomes of less closely related species of *Plasmodium* were compared (Carlton *et al.* 1998, 1999) indicating that reassortment of gene content of chromosomes does occur but leaving open the question of its association with the process of speciation. However, a thorough analysis of relative genome organization within *Plasmodium* would only be possible once significant amounts of assembled genome sequence from different *Plasmodium* species became available. The sequence of additional genomes from other *Apicomplexa* would also be useful in this context, as it would permit discrimination between genus- and phylum-specific genes, gene families and the possible organization of clusters of genes into coordinately expressed units. Today, we are on the threshold of being able to perform useful analyses with the existing (and anticipated) data and pave the way for meaningful post-genomic analyses that will inform attempts to generate the hoped for anti-malarial therapies.

3. THE GENOME INITIATIVES

An international consortium of genome sequencing centres and teams of malaria researchers was formed in 1996 to begin the task of sequencing the whole genome of *P. falciparum*. Sequencing of the chromosomes is proceeding at three centres: TIGR/NMRI (Rockville, USA; see http://www.tigr.org/tdb/edb2/pfa1/htmls/), the Sanger Sequencing Centre (Hinxton, UK; see http://www.sanger.ac.uk/projects/p_falciparum/) and Stanford University (Stanford, USA; see http://sequence-www.stanford.edu/group/malaria/index.html). It is anticipated that shortly (late 2001 to mid-2002) the whole genome sequence assembled into the 14 chromosomes will be available. In addition, there have been several initiatives to generate large amounts of genome information about additional parasite genomes, as well as an expression profile (through cDNA sequencing) of *P. falciparum*, *Plasmodium vivax* and *P. berghei* blood-stage parasites (Carlton & Dame 2000). There are two different types of genome sequence available. Genome Survey sequence data, consisting of a single sequencing read in one direction of a bank of clones that were isolated from libraries enriched in protein-coding segments of the genome. These are therefore sequences that cannot be linked together. However, a recent *P. yoelii* initiative characterized random DNA clones from a sheared DNA library to an extent that the entire genome should have been sequenced twice (TIGR/NMRI, USA; see http://www.tigr.org/tdb/edb2/pya1/htmls/). This data can be, and has been, assembled in short contigs and preliminarily annotated. In short, there is significant genome information available in publicly accessible databases for another human malaria (*P. vivax*), and three rodent malarias (in order of the amount of data available; *P. yoelii*, *P. berghei* and some for *P. chabaudi*). The initiatives that do not involve the main *P. falciparum* genome sequence consortium are summarized in table 1.

An important use of model malaria parasites is the further investigation of orthologues of *P. falciparum* vaccine candidate proteins. Where previously such orthologues were identified by biochemical, immunological and molecular biological approaches, the *Plasmodium* genome database resources now permit the identification *in silico*, using BLAST analyses, of the existence of orthologues and paralogues of well-characterized *P. falciparum* genes in other species of *Plasmodium*. Such analyses can also document the existence of families of homologues of vaccine candidates within *P. falciparum* and define multigene families and the extent of conservation of such families in the genus. Once the genes have been identified the biology and function of the encoded proteins can be investigated in the most appropriate *Plasmodium* species. One example of this is the gene family that has been defined by structural similarity to *Pfs48/45*, a gene that encodes a surface protein of gametocytes and gametes (Rener *et al.* 1983).

4. THE Pfs48/45 GENE SUPERFAMILY

The gene encoding Pfs48/45 was first cloned during the study of surface proteins of the gametocytes and gametes of *P. falciparum* (Kocken *et al.* 1993) although its importance as a transmission blocking vaccine candidate had already been defined (Kumar & Carter 1984; Vermeulen *et al.* 1983). Pfs48/45 has a modular structure consisting of two loosely defined domains of *ca.* 120 amino acids that contain six cysteine residues (6-Cys domains) (Templeton & Kaslow 1999). The two domains are separated by a cysteine-containing intervening region of unknown significance. The protein is predicted to have a conventional signal sequence for distribution to the surface of the cell and is glycophosphoinositol anchored at the parasite surface. Conventional gene cloning uncovered two additional genes that encoded proteins with a similar modular structure, Pf12, a blood-stage protein, which is similar in overall structure to Pfs48/45 minus the intervening 'degenerate' domain (Elliot *et al.* 1990) and Pfs230, another gametocyte–gamete surface antigen. Pfs230 contains six 6-Cys domains, a signal sequence but with no obvious means to anchor it to the membrane (Williamson *et al.* 1993), and the similarity in structure of all three proteins was noted (Carter *et al.* 1995). Subsequent physical analysis indicated that Pfs230 and Pfs48/45 form a complex (Kumar 1987) yet Pfs48/45 has no role either in the production or processing of Pfs230 (Van Dijk *et al.* 2001). The gene encoding Pfs230 was subsequently shown to be located on chromosome 2 of *P. falciparum*, immediately downstream of a shorter paralogous gene, Pf230 II, that contains five 6-Cys domains and also lacks an anchor sequence (Gardner *et al.* 1998). Based on this information Templeton & Kaslow (1999) interrogated, through BLAST analysis, the emerging *P. falciparum* database with individual 6-Cys domains and uncovered a

Table 1. Summary of genome initiatives other than the *P. falciparum* 3D7 genome-sequencing consortium.
(Superscripted M indicates clones sequenced from libraries constructed from Mung Bean Nuclease digested genomic DNA. Y indicates that a single YAC of *P. vivax* DNA containing a telomere was sequenced. Abbreviations: nt, nucleotide; pr, protein.)

	Plasmodium species						
	P. falciparum		*P. vivax*		*P. yoelii*	*P. berghei*	*P. chabaudi*
site of genome initiative[a]	University of Florida	University of Tokyo	University of Florida	Universities of Sao Paulo & Heidelberg/ Sanger Centre	TIGR/Naval Medical Research Center	University of Florida	University of Glasgow/ Sanger Centre
gDNA[b]	1726^{M} (14%)	N/A	10 237[c] ^{M}(26%)	155 711 bp (Y)	2 × genome coverage	5482^{M} (24%)	766 shotgun reads
blood-stage cDNA	1115 (15%)	2490 (40%)	N/A		N/A[e]	5582 (50%)	N/A
number of entries in GenBank[d]	nt 8981 (1.64) pr 3679 (1.63)		nt 736 (3.35) pr 796 (3.32)		nt 120 (1.85) pr 133 (1.69)	nt 98 (1.36) pr 106 (1.38)	nt 831 (15.4) pr 89 (1.48)
URL	http://parasite.etmed.ufl.edu/	http://133.11.149.55/aboutdb.html	http://parasite.vetmed.ufl.edu/	http://www.sanger.ac.uk/projects/p_vivax/	http://www.tigr.org/tdb/edb2/pya1/htmls/	http://parasite.vetmed.ufl.edu/	http://www.sanger.ac.uk/projects/p_chabaudi/

[a] Most of the information is available at the NCBI Malaria Genetics Web Page (http://www.ncbi.nlm.nih.gov/malaria/).
[b] Shotgun reads unless indicated. The figure in brackets after the number of entries is the percentage redundancy—figures kindly supplied by Dr Jane Carlton and Dr Junichi Watanabe.
[c] The total is broken down into 1925 GSS reads of *P. vivax* strain Belem and 8312 of strain Sal I (redundancy is 26%, Sal I and 25%, Belem).
[d] Status as of 6 March 2001. The figure in brackets indicates the factor by which the number of entries has increased since the databases were last reviewed (22 July 1999).
[e] N/A, not applicable.

further three genes (*Pf36*, *Pf41* and *Pf47*; nomenclature based upon predicted molecular weight of the ORF) forming what they termed a Pfs48/45 gene superfamily that was restricted to *Plasmodium* (Templeton & Kaslow 1999).

Continued searches on the maturing database for *P. falciparum* have revealed an additional two genes that bring the total number of Pfs48/45 gene family members to nine (Thompson *et al.* 2001). Taking advantage of the subsequent release of significant genome data from *P. berghei*, *P. chabaudi* (Janssen *et al.* 2001), *P. vivax* and recently *P. yoelii*, (see Web addresses in § 3 and table 1) it has been possible to document the conservation of this family within the genus. All sequences encoding the nine members of the family that have been identified to date in *P. berghei* have been independently confirmed and used to confirm the database searches (Thompson *et al.* 2001) and they appear to be present in the *P. yoelii* database. To date, and based upon sequence similarity, various orthologues of all nine members have been identified in the databases of the four additional species of *Plasmodium* mentioned above. Data from additional apicomplexan genomes is not abundant but homologues of this family outside *Plasmodium* have yet to be detected. Therefore, since it is clear that the majority of the members of this gene superfamily appear to be well represented in all species of *Plasmodium,* the family can be referred to as the P48/45 gene family.

5. ORGANIZATION OF THE P48/45 GENE FAMILY

The comparison of the complete genomes of *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster* demonstrated that gene duplications were easily detected in whole genome analyses. Indeed, as much as 48.6% of the total number of genes in the *C. elegans* genome arose as a result of duplications (Rubin *et al.* 2000). It is clear that *Plasmodium* also contains multigene families that have arisen through gene duplication events. The publication of the completed sequence of chromosome 2 of *P. falciparum* (Gardner *et al.* 1998) demonstrated the existence of numerous multigene families; the subtelomeric families encoding variant antigens (*var*, *rif*, *stevor*) chromosomal internal gene families (*sera, resa*) and simple duplicated genes (*msp4*, *msp5*). The sequence of chromosome 3 reinforced this impression (Bowman *et al.* 1999).

As noted above, the sequence of chromosome 2 also demonstrated that from the P48/45 gene superfamily, *pfs230* and its closest paralogue, *pfs230 II*, were adjacent collinear genes separated by 1.5 kb (Gardner *et al.* 1998). Examination of the contigs of the *P. falciparum* database and physical analysis of the genome has demonstrated that this is not an isolated phenomenon within the P48/45 gene superfamily. *pfs47* is a paralogue of *pfs48/45* and has been shown to lie 1.5 kb upstream and an orthologous gene arrangement exists in *P. berghei* (Van Dijk *et al.* 2001) and in *P. yoelii*. Both gene pairs exist in a head-to-tail organization. It has not yet been established whether the paralogous pair of *p230/p230 II* is also linked in other *Plasmodium* genomes. In both cases the pairs represent the most closely related members of the family, which suggests that they have arisen through local gene duplication. *pf12* and *pf41* appear to be single genes as they are not duplicated in either *P. falciparum* or *P. yoelii*. Furthermore, the conserved arrangement of P48/45 gene superfamily members and their closest paralogues in tandem pairs suggests that each has arisen as the result of a duplication event in the common ancestor of the genus. Finally, it has also been demonstrated that in addition to the known gametocyte specific transcription pattern of *pfs230* and *p48/45*, *P. berghei pb47* and *pbs230 II* are also transcribed in gametocytes (Van Dijk *et al.* 2001) and raises the possibility that all members of this gene family encode gametocyte–gamete surface proteins. Nevertheless, the data already available significantly expands the number of candidate antigens that can be assessed for their potential to generate immune responses that block transmission and are a practical demonstration of the immediate value of whole genome characterization. Research on the rodent malaria parasite genomes has already, and will continue to, provide information for decisions to pursue additional members of this gene superfamily as transmission-blocking vaccine candidate antigens. Clearly, the functional significance of the members of the P48/45 gene family can be most thoroughly and rapidly addressed in rodent malaria models (de Koning-Ward *et al.* 2000*b*).

The maintenance of pairs of genes encoding surface proteins that are close paralogues of one another is not restricted to the P48/45 gene superfamily. The P25/28 gene pair that encodes highly structured proteins found on the surface of the zygote and ookinetes of *Plasmodium* was known to be tightly linked in *P. falciparum* (Duffy & Kaslow 1997) and has recently been shown to exist head-to-tail in three diverse species of *Plasmodium* as a tandem array (Tomas *et al.* 2001). The expressed MSP encoding genes *msp4*, and *msp5* are similarly linked on chromosome 2 in *P. falciparum* (Gardner *et al.* 1998; Wu *et al.* 1999). Other examples may yet be uncovered. Lastly, it is significant that there may, of course, be species-specific duplications. A recent report demonstrates that the human parasite *Plasmodium ovale* contains three copies of the P25/28 genes involving a duplication of *pov28,* although the genomic arrangement of the genes was not clarified (Tachibana *et al.* 2001).

One can only speculate about the role of this type of gene duplication. Such duplications have been thought of as an expansion of the gene pool from which mutation and selection can lead to the generation of proteins with new functions. Alternatively, multiple receptor–ligand systems on the surface of gametes might also serve as a fail-safe mechanism to ensure accurate cell–cell contact that would have obvious advantages for fertilization. Recently it has been proposed that gene pairs may be maintained for mutual necessity. In this situation, shortly after duplication both copies accumulate mutations that do not cause the function of either protein to diverge but make both functionally less efficient. Such a situation creates a selective pressure to maintain both copies of the gene (Force *et al.* 1999).

6. PRACTICAL BENEFITS OF A COMPARISON OF *PLASMODIUM* GENOME INFORMATION AND THE NEED TO INTEGRATE MORE DATA

Information about orthology in terms of gene content of genomes of different species can be revealed by simple

comparison of randomly generated DNA sequences. However, bilateral comparisons of the genome information of different species of *Plasmodium* can gain an extra dimension when the data is assembled in contigs. With such datasets the earlier indications of synteny in *Plasmodium* can be examined in much more detail and the *P. yoelii* genome initiative will be invaluable in this respect. Independent of the whole genome sequencing projects a small-scale, but detailed characterization of a 13.6 kb locus of a non- subtelomeric region of chromosome 5 of *P. berghei* and its comparison with *P. falciparum* has foreshadowed some of what we can expect from larger scale ventures (Van Lin *et al.* 2001). This pilot project also indicated the extent of characterization of each genetic locus that it might be necessary to perform in order to fully appreciate its complexity and potential to encode proteins. Lastly, the comparison demonstrated that the degree of similarity between the two species at the level of gene order and conservation of gene structure is great and can be usefully used as a predictive tool for the identification of genes and their organization.

The locus (B9) was chosen because of evidence that there is a clustering of genes that are expressed during parasite sexual development and transmission from the host to the vector on *P. berghei* chromosome 5 (Van Lin *et al.* 1997). Therefore, the DNA sequence of the locus surrounding a novel gametocyte specific gene (ORF3, B9) on that chromosome was compiled (Van Lin *et al.* 2001). Systematic Northern blot analysis of the locus revealed numerous gene transcripts including three that were gametocyte specific. Systematic cloning and analysis of cDNA of mRNA produced by genes in the locus revealed a complex arrangement of six genes, four of which overlap, two in their coding regions. The degree of overlap is significant as the 13.6 kb locus produces 15.5 kb of mRNA. Three of the genes (ORFs1–3) generate multiple mRNA species through alternative splicing assembly with eight possible models for ORF2 (Van Lin *et al.* 2001).

The orthologous locus of *P. falciparum* (on chromosome 10) was assembled and verified, and cDNA analysis performed, to confirm that the organization of the locus was absolutely conserved. The presence of six genes in a 13.6 kb region indicated a gene density greater than the annotated average of one gene every 4.0–4.5 kb for *P. falciparum* chromosomes 2 and 3. Therefore, GlimmerM analysis was performed on both loci to assess the ability of the algorithm to make accurate predictions on characterized complex loci (figure 1). It appeared that the algorithm was not able to accurately predict the presence of the complex genes with multiple short introns contained on the B9 locus. GlimmerM analysis was more accurate using the *P. falciparum* chromosome 10 locus, perhaps reflecting the fact that the algorithm was trained on sequence data from the human parasite. However, ORF2 was not predicted by the GlimmerM algorithm in either species, yet local sequence alignment could clearly reveal the local similarity between the two genome segments and prompt a closer investigation (Van Lin *et al.* 2001).

Such local analyses will be of tremendous use as significant amounts of additional *Plasmodium* genomes become available. Indeed, the recent release of the contigs resulting from the analysis of *P. yoelii* twofold coverage shotgun-sequencing data indicates that regions annotated as barren in the published *P. falciparum* chromosomes having significant homology with *P. yoelii* contigs (L. H. M. Van Lin, personal communication). These regions can now be investigated for their protein-coding potential and demonstrate the value of local DNA sequence alignments for predicting gene structure. As has recently become apparent from studies on the human genome, a significant amount of work will still be necessary to characterize individual cDNAs from cloned populations and, in turn, couple these to proteome studies to reveal the full protein-coding potential of the minimal *Plasmodium* genome. Random sequencing of libraries of cDNA clones may not be efficient and normalized libraries could lack splice variants—thus a combination of microarray assisted global transcription surveys and directed isolation and characterization of cDNA aided by local pairwise alignments of two *Plasmodium* genomes might prove the most effective strategy for comprehensive annotation of the diversity of transcripts that can be generated by any given genomic locus.

7. THE ROLE OF ANIMAL MODELS OF MALARIA IN POST-GENOME RESEARCH

(a) *Functional analysis of gene products*

Besides offering the possibility to improve the annotation of the *P. falciparum* genome through a comparative analysis, animal models represent a potent source of information concerning protein function within the context of the infected host or vector. Several model species of *Plasmodium can* be genetically manipulated and currently offer a more flexible choice of forms of manipulation. Currently, two rodent malaria parasites, *P. berghei* (Van Dijk *et al.* 1995, 1996) and *P. yoelii* (Mota *et al.* 2001) and two primate parasites, *P. knowlesi* (Van der Wel *et al.* 1997) and *P. cynomolgi* (Kocken *et al.* 1999) can be stably transformed with positive selection for acquisition of drug resistance. Importantly, exogenous DNA can integrate site specifically via homologous recombination into the genome of both rodent species and *P. knowlesi*. Truly random integration has not been reported. Site-specific homologous recombination has been exploited to both modify and disrupt genes with examination of the subsequent phenotype. As in *P. falciparum* both rodent species support single-site insertion recombination into their genome, an event that is potentially reversible. The reversibility can be exploited to help confirm phenotype. Double-site replacement integration that is irreversible, giving stable mutant stocks, occurs in both *P. berghei* and *P. knowlesi* and has not yet been reported for *P. falciparum*. A further advantage of double-site integration is that it might be used simultaneously to knock-out more than one gene. This is particularly advantageous when paralogous genes which might share function are located adjacent to one another in the genome (Tomas *et al.* 2001; Van Dijk *et al.* 2001).

Rodent models have already shown that detailed studies of surface protein function can be obtained for a large proportion of the parasite life cycle, beginning with gametocytogenesis and continuing through to sporozoite invasion and development in cultured hepatocytes. Knock-out studies performed on surface proteins expressed on the surface of the gametes have demonstrated a sex specific role for the Pbs48/45 in gamete fertilization. Knock-out

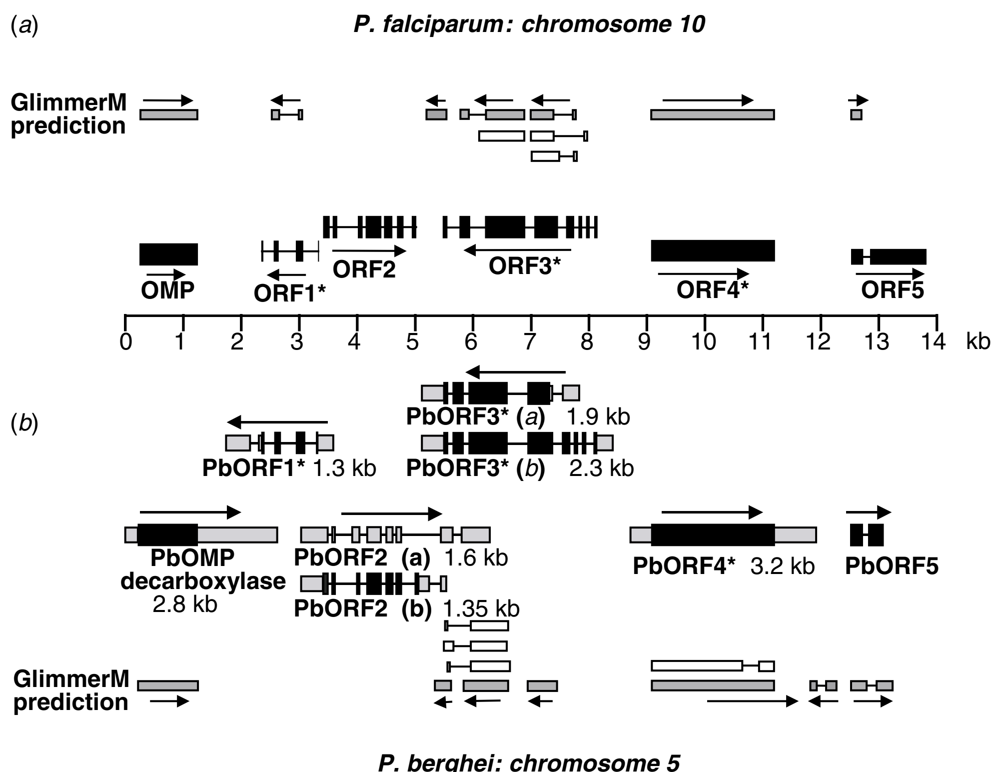

Figure 1. (*a*) A visual representation of the GlimmerM gene prediction in the B9-locus of *P. berghei* in relation to the organization of the coding regions as it became apparent during this study. The predictions with the highest probability are shown in grey; the predictions with lower probability are shown in white. An asterisk indicates the gametocyte-specific genes. No experimental evidence, or any significant sequence homology, was found for the gene predicted between ORF4 and ORF5 in *P. berghei*. (*b*) The same annotation applies for *P. falciparum*. For the second exon of ORF5 in *P. falciparum*, the ORF continues from the end of our sequence for another 1059 nucleotides (contig c10m345, TIGR database). Without this additional sequence, GlimmerM is unable to predict the exon accurately. Black boxes indicate exons; grey boxes indicate untranslated regions, where mapped lines indicate introns. Reproduced from Van Lin *et al.* (2001) with permission from *Nucleic Acids Research*, Oxford University Press.

of the Pbs48/45 gene renders the male gamete infertile, a curious finding given that the protein is also expressed in female gametes, which can still be fertilized in the Pbs48/45 knock-out lines. Other studies on the later stages of parasite development in the mosquito have demonstrated a central role for the CS protein in sporozoite development within the growing oocyst (Menard *et al.* 1997). In addition, malaria parasites apparently use a conserved mechanism of motility but involving stage-specific, surface-force transducing molecules such as TRAP (sporozoites, Sultan *et al.* 1997; Wengelnik *et al.* 1999) and CTRP (ookinetes, Yuda *et al.* 1999; Dessens *et al.* 1999) that contain so-called A-domains and thrombospondin repeats which are biologically conserved protein modules involved in cell attachment. The conservation of this mechanism of motility extends throughout apicomplexans (Kappe *et al.* 1999).

All of the proteins that have been investigated by genetic manipulation in *P. berghei* have orthologues in the human parasites that are expressed at the same stage of the life cycle as their rodent parasite counterparts. Therefore, the use of model species of *Plasmodium*, in particular, currently, *P. berghei* , has been informative to the human situation even where the structures and morphologies of parasite structures (e.g. gametocytes) were clearly comparable.

8. CONCLUDING REMARKS

Animal models of malaria will continue to play a significant role in the comprehensive investigation of the biology of malaria parasites and their interactions with both host and vector. Their ability to provide access to areas of the life cycle not easily available with human parasites will provide valuable resources for analytical material for transcriptome and proteome studies. Models will also be intrinsic to the evaluation of the effectiveness of proposed vaccine candidates as part of the process of candidate validation. Thus, whilst it is possible to elegantly explore certain features of the *Plasmodium* life cycle with the cultured *P. falciparum* system for example, the functional genetics of erythrocyte invasion and cytoadherence (Cowman *et al.* 2000), *in vivo* analysis of *P. falciparum* will remain largely unattainable. Nevertheless, the models discussed above can be used for the study of pathological states such as cerebral malaria and cytoadherence to other deep tissues (Mota *et al.* 2000), antigenic variation (Phillips *et al.* 1997) and reticulocyte invasion (Preiser *et al.* 1999). Furthermore, with the discovery of conserved gene families in *Plasmodium* that have been demonstrably associated with human parasite cytoadherence (e.g. clag, Gardner *et al.* 1998; Janssen *et al.* 2001) and large multigene gene families that resemble the *P. vivax vir* subtelom-

eric genes (del Portillo *et al.* 2001), it becomes possible to consider the use of model *Plasmodium* species to perform *in vivo* investigations into important pathological features of malaria that have a common molecular basis. Furthermore, a range of hosts and host genetic backgrounds are available that permit a unique combination of host and parasite genetics or host susceptibility to disease to be considered experimentally. Lastly, from the point of view of tissue pathology it would be useful to be able to dynamically visualize parasite distribution during infection. The recent successes in tumour visualization using cells expressing luciferase and ultrasensitive CCD cameras (Sweeney *et al.* 1999; Lavon *et al.* 2000) might prove applicable to malaria.

In-depth analysis of the biology of *Plasmodium* is already possible, and in particular, with existing technologies, multiple genetic manipulations (gene modification or knock-out) can be performed on the same parasite stock (de Koning-Ward *et al.* 2000*a*). This allows approaches such as 'add-back' experiments to knock-out parasite mutants to validate observations made with the original knock-out parasites (Sultan *et al.* 2001). However, further development of the genetic manipulation technologies for application in malaria models would be useful. One example of such a technology that has not yet been achieved in model parasites is 'hit and run' technology, which can be used to recycle the limited number of positive selectable markers currently available (two). In order to achieve this an effective negative selectable marker is required. There are several candidates for such markers, notably viral thymidine kinase and bacterial cytidine deaminase. However, we have derived transgenic populations of *P. berghei* which overexpress mutants of the herpes simplex virus thymidine kinase (Black *et al.* 1996) that have a greater affinity for poison metabolites such as ganciclovir and a reduced affinity for thymidine; however, we have been unable to demonstrate an increase in sensitivity of the transgenic parasite to either ganciclovir or acyclovir (B. Franke-Fayard, unpublished data). We are now in the process of testing additional mutants and new drugs that are available for this purpose.

The strong impression from the recent releases of genome data and emerging analyses is that, apart from the telomeric and subtelomeric regions, malaria parasite chromosomes are very well conserved between species. This conservation is reflected in gene content and the local organization of these genes if not necessarily the base composition (McCutchan *et al.* 1984). Thus, animal models of *Plasmodium* can continue to be exploited and the powerful global analytical technologies that have recently been developed, such as DNA microarrays and proteome studies can be applied to the complete life cycle. Appropriate implementation will provide a detailed analysis of stage-specific protein expression and function and hopefully yield insights that can be converted into therapies.

The author thanks members, past and present, of the malaria group in Leiden. In particular, Chris Janse for a critical reading of this manuscript and continuous scientific stimulation; Joanne Thompson, Milly van Dijk, Blandine Franke-Fayard and Leo van Lin for direct contributions that have been discussed in this article. This work has involved extensive collaboration with the malaria research groups at the University of Nijmegen and Dr Marta Ponzi and Dr Tommaso Pace at ISS in Rome. The group is financially supported by LUMC; the European Community; the Dutch Organization for Scientific Research (NOW) and the World Health Organization.

REFERENCES

Alm, R. A., Bina, J., Andrews, B. M., Doig, P., Hancock, R. E. & Trust, T. J. 2000 Comparative genomics of *Helicobacter pylori:* analysis of the outer membrane protein families. *Infect. Immun.* **68**, 4155–4168.

Black, M. E., Newcomb, T. G., Wilson, H. M. & Loeb, L. A. 1996 Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proc. Natl Acad. Sci. USA* **93**, 3525–3529.

Bowman S. (and 24 others) 1999 The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538.

Carlton, J. M.-R. & Dame, J. B. 2000 The *Plasmodium vivax* and *P. berghei* gene sequence tag projects. *Parasitol. Today* **16**, 409.

Carlton, J. M.-R., Vinkenoog, R., Waters, A. P. & Walliker, D. 1998 Gene synteny in species of *Plasmodium*. *Mol. Biochem. Parasitol.* **93**, 285–294.

Carlton, J. M.-R., Galinski, M. R., Barnwell, J. W. & Dame, J. B. 1999 Karyotype and synteny among the chromosomes of all four species of human malaria parasite. *Mol. Biochem. Parasitol.* **101**, 23–32.

Carter, R., Coulson, A., Bhatti, S., Taylor, B. J. & Elliott, J. F. 1995 Predicted disulfide-bonded structures for three uniquely related proteins of *Plasmodium falciparum*, Pfs230, Pfs48/45 and Pf12. *Mol. Biochem. Parasitol.* **71**, 203–210.

Cowman, A. F., Baldi, D. L., Healer, J., Mills, K. E., O'Donnell, R. A., Reed, M. B., Triglia, T., Wickham, M. E. & Crabb, B. S. 2000 Functional analysis of proteins involved in *Plasmodium falciparum* merozoite invasion of red blood cells. *FEBS Lett.* **476**, 84–88.

de Koning-Ward, T. F., Fidock, D., Menard, R., Thathy, V., Van Spaendonk, R. M. L., Waters, A. P. & Janse, C. J. 2000*a* Human dihydrofolate reductase: a new selectable marker that enables sequential genetic manipulation of the *Plasmodium berghei* genome. *Mol. Biochem. Parasitol.* **106**, 199–212.

de Koning-Ward, T. F., Janse, C. J. & Waters, A. P. 2000*b* The development of genetic tools for dissecting the biology of malaria parasites. *A. Rev. Microbiol.* **54**, 157–185.

del Portillo, H. A. (and 12 others) 2001 A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* **410**, 839–842.

Dessens, J. T., Beetsma, A. L., Dimopoulos, G., Wengelnik, K., Crisanti, A., Kafatos, F. C. & Sinden, R. E. 1999 CTRP is essential for mosquito infection by malaria ookinetes. *EMBO J.* **18**, 6221–6227.

Duffy, P. E. & Kaslow, D. C. 1997 A novel malaria protein, Pfs28, and Pfs25 are genetically linked and synergistic as falciparum malaria transmission-blocking vaccine. *Infect. Immun.* **65**, 1109–1113.

Elliott, J. F., Albrecht, G. R., Gilladoga, A., Handunnetti, S. M., Neequaye, J., Lallinger, G., Minjas, J. N. & Howard, R. J. 1990 Genes for *Plasmodium falciparum* surface antigens cloned by expression in COS cells. *Proc. Natl Acad. Sci. USA* **87**, 6363–6367.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. & Postlethwait, J. 1999 Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545.

Gardner, M. J. (and 25 others) 1998 Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. Science **282**, 1126–1132.

Goonewardene, R., Daily, J., Kaslow, D., Sullivan, T. J., Duffy, P., Carter, R., Mendis, K. & Wirth, D. 1993 Transfection of the malaria parasite and expression of firefly luciferase. *Proc. Natl Acad. Sci. USA* **90**, 5234–5236.

Janse, C. J., Carlton, J. M.-R., Walliker, D. & Waters, A. P. 1994 Conserved location of genes on polymorphic chromosomes of four species of malaria parasites. *Mol. Biochem. Parasitol.* **68**, 285–296.

Janssen, C. S., Barrett, M. P., Lawson, D., Quail, M. A., Harris, D., Bowman, S., Phillips, R. S. & Turner, C. M. 2001 Gene discovery in *Plasmodium chabaudi* by genome survey sequencing. *Mol. Biochem. Parasitol.* **113**, 251–260.

Kappe, S., Bruderer, T., Gantt, S., Fujioka, H., Nussenzweig, V. & Menard, R. 1999 Conservation of a gliding motility and cell invasion machinery in apicomplexan parasites. *J. Cell. Biol.* **147**, 937–944.

Kerr, J. R. 1999 Cell adhesion molecules in the pathogenesis of and host defence against microbial infection. *Mol. Pathol.* **52**, 220–230.

Kocken, C. H., Jansen, J., Kaan, A. M., Beckers, P. J., Ponnudurai, T., Kaslow, D. C., Konings, R. N. & Schoenmakers, J. G. 1993 Cloning and expression of the gene coding for the transmission blocking target antigen Pfs48/45 of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **61**, 59–68.

Kocken, C. H. M., Van der Wel, A. & Thomas, A. W. 1999 *Plasmodium cynomolgi:* transfection of blood-stage parasites using heterologous DNA constructs. *Exp. Parasitol.* **93**, 58–60.

Kumar, N. 1987 Target antigens of malaria transmission blocking immunity exist as a stable membrane bound complex. *Parasite Immunol.* **9**, 321–335.

Kumar, N. & Carter, R. 1984 Biosynthesis of the target antigens of antibodies blocking transmission of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **13**, 333–342.

Lavon, I. (and 10 others) 2000 High susceptibility to bacterial infection but no liver dysfunction in mice compromised for hepatocyte NF-kappa B activation. *Nature Med.* **6**, 573–577.

McCutchan, T. F., Dame, J. B., Miller, L. H. & Barnwell, J. 1984 Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. *Science* **225**, 808–811.

Menard, R., Sultan, A. A., Cortes, C., Altszuler, R., Van Dijk, M. R., Janse, C. J., Waters, A. P., Nussenzweig, R. S. & Nussenzweig, V. 1997 Circumsporozoite protein is required for development of malaria sporozoites in mosquitoes. *Nature* **385**, 336–340.

Mota, M. M., Jarra, W., Hirst, E., Patnaik, P. K. & Holder, A. A. 2000 *Plasmodium chabaudi*-infected erythrocytes adhere to CD36 and bind to microvascular endothelial cells in an organ-specific way. *Infect. Immun.* **68**, 4135–4144.

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

Phillips, R. S., Brannan, L. R., Balmer, P. & Neuville, P. 1997 Antigenic variation during malaria infection—the contribution from the murine parasite *Plasmodium chabaudi*. *Parasite Immunol.* **19**, 427–434.

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

Rener, J., Graves, P. M., Carter, R., Williams, J. L. & Burkot, T. R. 1983 Target antigens of transmission-blocking immunity on gametes of *Plasmodium falciparum*. *J. Exp. Med.* **158**, 976–981.

Ross, R. 1898 The role of the mosquito in the evolution of the malaria parasite. *Lancet* **ii**, 488–489.

Rubin, G. M. (and 50 others) 2000 Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215.

Sultan, A. A., Thathy, V., Frevert, U., Robson, K. J., Crisanti, A., Nussenzweig, V., Nussenzweig, R. S. & Menard, R. 1997 TRAP is necessary for gliding motility and infectivity of *Plasmodium* sporozoites. *Cell* **90**, 511–522.

Sultan, A. A., Thathy, V., de Koning-Ward, T. F. & Nussenzweig, V. 2001 Complementation of *Plasmodium berghei* TRAP knockout parasites using human dihydrofolate reductase gene as a selectable marker. *Mol. Biochem. Parasitol.* **113**, 151–156.

Sweeney, T. J., Mailander, V., Tucker, A. A., Olomu, A. B., Zhang, W., Cao, Y., Negrin, R. S. & Contag, C. H. 1999 Visualizing the kinetics of tumor-cell clearance in living animals. *Proc. Natl Acad. Sci. USA* **96**, 12 044–12 049.

Tachibana, M., Tsuboi, T., Templeton, T. J., Kaneko, O. & Torii, M. 2001 Presence of three distinct ookinete surface protein genes, Pos25, Pos28-1, and Pos28-2, in *Plasmodium ovale*. *Mol. Biochem. Parasitol.* **113**, 341–344.

Templeton, T. J. & Kaslow, D. C. 1999 Identification of additional members defines a *Plasmodium falciparum* gene superfamily which includes Pfs48/45 and Pfs230. *Mol. Biochem. Parasitol.* **101**, 223–227.

Thompson, J., Janse, C. J. & Waters, A. P. 2001 Comparative genomics in *Plasmodium*: a tool for the identification of genes and functional analysis. *Mol. Biochem. Parasitol.* **118**, 147–154.

Tomas, A. M. (and 16 others) 2001 The conserved P25/28 proteins of the malaria ookinete surface have multiple, partially redundant functions. *EMBO J.* **20**, 3975–3983.

Trager, W. & Jensen, J. B. 1976 Human malaria parasites in continuous culture. *Science* **193**, 673–675.

Van der Heyde, H. C., Bauer, P., Sun, G., Chang, W. L., Yin, L., Fuseler, J. & Granger, D. N. 2001 Assessing vascular permeability during experimental cerebral malaria by a radio-labeled monoclonal antibody technique. *Infect. Immun.* **69**, 3460–3465.

Van der Wel, A. M., Tomas, A. M., Kocken, C. H. M., Malhotra, P., Janse, C. J., Waters, A. P. & Thomas, A. W. 1997 Transfection and *in vivo* selection of the primate malaria parasite, *Plasmodium knowlesi*. *J. Exp. Med.* **185**, 1499–1503.

Van Dijk, M. R., Waters, A. P. & Janse, C. J. 1995 Stable transfection of malaria parasite blood stages. *Science* **268**, 1358–1362.

Van Dijk, M. R., Janse, C. J. & Waters, A. P. 1996 Expression of a *Plasmodium* gene introduced into subtelomeric regions of *Plasmodium berghei* chromosomes. *Science* **271**, 662–665.

Van Dijk, M. R., Janse, C. J., Thompson, J., Waters, A. P., Braks, A., Dodemont, H., Stunnenberg, H. G., Van Gemert, G.-J., Sauerwein, R. W. & Eling, W. 2001 A central role for P48/45 in malaria parasite male gamete fertility. *Cell* **104**, 153–164.

Van Lin, L. H. M., Pace, T., Janse, C. J., Scotti, R. & Ponzi, M. 1997 A long range restriction map of chromosome 5 of *Plasmodium berghei* demonstrates a chromosome specific symmetrical subtelomeric organisation. *Mol. Biochem. Parasitol.* **86**, 111–115.

Van Lin, L. H. M., Pace, T., Janse, C. J., Birago, C., Ramesar, J., Picci, L., Ponzi, M. & Waters, A. P. 2001 Interspecific 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.

Vermeulen, A. N., Ponnudurai, T., Beckers, P. J., Verhave, J. P., Smits, M. A. & Meuwissen, J. H. 1983 Sequential expression of antigens on sexual stages of *Plasmodium falciparum* accessible to transmission-blocking antibodies in the mosquito. *J. Exp. Med.* **162**, 1460–1476.

Wengelnik, K., Spaccapelo, R., Naitza, S., Robson, K. J. H., Janse, C. J., Bistoni, F., Waters, A. P. & Crisanti, A. 1999

The A-domain and the thrombospondin-related motif of *Plasmodium falciparum* TRAP are implicated in the invasion process of mosquito salivary glands. *EMBO J.* **18**, 5195–5204.

Wu, T., Black, C. G., Wang, L., Hibbs, A. R. & Coppel, R. L. 1999 Lack of sequence diversity in the gene encoding merozoite surface protein 5 of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **103**, 243–250.

Yuda, M., Sakaida, H. & Chinzei, Y. 1999 Targeted disruption of the *Plasmodium berghei* CTRP gene reveals its essential role in malaria infection of the vector mosquito. *J. Exp. Med.* **190**, 1711–1716.