

Interspecies conservation of gene order and intron–exon structure in a genomic locus of high gene density and complexity in *Plasmodium*

Leonard H. M. van Lin, Tomasino Pace¹, Chris J. Janse, Cecilia Birago¹, Jai Ramesar, Leonardo Picci¹, Marta Ponzi¹ and Andrew P. Waters*

Department of Parasitology, Leiden University Medical Centre, PO Box 9600, 2300 RC Leiden, The Netherlands and ¹Laboratorio di Biologia Cellulare, Istituto Superiore di Sanità, Rome, Italy

Received January 29, 2001; Revised and Accepted March 19, 2001

DDBJ/EMBL/GenBank accession no. AJ278826

ABSTRACT

A 13.6 kb contig of chromosome 5 of *Plasmodium berghei*, a rodent malaria parasite, has been sequenced and analysed for its coding potential. Assembly and comparison of this genomic locus with the orthologous locus on chromosome 10 of the human malaria *Plasmodium falciparum* revealed an unexpectedly high level of conservation of the gene organisation and complexity, only partially predicted by current gene-finder algorithms. Adjacent putative genes, transcribed from complementary strands, overlap in their untranslated regions, introns and exons, resulting in a tight clustering of both regulatory and coding sequences, which is unprecedented for genome organisation of *Plasmodium*. In total, six putative genes were identified, three of which are transcribed in gametocytes, the precursor cells of gametes. At least in the case of two multiple exon genes, alternative splicing and alternative transcription initiation sites contribute to a flexible use of the dense information content of this locus. The data of the small sample presented here indicate the value of a comparative approach for *Plasmodium* to elucidate structure, organisation and gene content of complex genomic loci and emphasise the need to integrate biological data of all *Plasmodium* species into the *P.falciparum* genome database and associated projects such as PlasmodB to further improve their annotation.

INTRODUCTION

In 1996 an international *Plasmodium falciparum* genome sequencing consortium was established to sequence and annotate the entire genome of *P.falciparum* (1); it is expected to complete its task before the end of 2001. To date, the full sequences of two of the 14 chromosomes have been published

(2,3) and preliminary sequence data is available for the remaining chromosomes from the web sites of the sequencing centres. It is a great challenge to identify novel genes from the bulk of genomic sequence data, an approach that relies to a large extent upon the ability of computer algorithms to predict open reading frames (ORFs) in the compiled DNA sequence.

From experience with organisms whose genomes have been completely sequenced it has been shown that comparative genomics can be highly informative (4). Such comparisons allow inferences to be drawn about one genome and its coding potential from the properties of another related genome, as well as about the evolutionary forces that influenced genome organisation (5). The insight that comes from cross-species sequence comparisons directly assists in the identification of coding regions, regulatory signals and other functional elements of a genome and may even be helpful in deducing complexities such as alternative splicing (6). Genome comparisons of distantly related organisms can reveal global biological patterns, such as the existence and extent of conserved gene families (7,8). In addition, comparisons of more closely related species or different strains from a single species may contribute to a better insight into species- or strain-specific traits involved in pathogenicity and virulence (9).

Rodent malaria parasites have served as useful comparative tools that reproduce much of the biology of human malaria. Like *P.falciparum*, the haploid genomes of rodent parasites are organised in 14 chromosomes (10,11) that are compartmentalised into relatively conserved internal regions flanked by highly polymorphic terminal regions (12). Comparative mapping of genes on chromosomes of human and rodent parasites has shown, within the limits of the number of probes used, that both linkage and gene order appear to be well conserved (13–15). Such mapping studies may have practical use in positional cloning of less well conserved genes.

Previously we have shown a relative clustering of genes that are expressed during sexual development on chromosome 5 of the rodent malaria parasite *Plasmodium berghei* (16). In this study we characterise a 13.6 kb internal region of chromosome 5 to which the sex-specific marker B9 had been mapped (16) and we compare this genomic region with its orthologous locus

*To whom correspondence should be addressed. Tel: +31 71 5265069; Fax: +31 71 5266907; Email: waters@lumc.nl

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

of *P.falciparum* on chromosome 10. In both the rodent and human malaria parasite we demonstrate that six genes are present in the 13.6 kb genomic locus, three of which are specifically transcribed in gametocytes. Sequence comparison of the orthologous loci of *P.berghei* and *P.falciparum* reveals a high level of conservation of the complex gene organisation, consisting of overlapping genes with an intricate intron–exon structure. The comparative approach taken was invaluable in elucidation of the unexpected high gene density and complexity, not predicted by the current gene-finder software due to difficulties in predicting small exon genes. If the observed underestimation of gene complexity and density applies to other genomic regions of *P.falciparum* this will justify a critical view of the prediction of the total number of genes encoded by the genome of the malaria parasite.

MATERIALS AND METHODS

Sequences reported in this paper have EMBL accession no. AJ278826 (*P.berghei* B9 locus).

Parasite stage-specific RNA and genomic DNA

For *P.berghei* the gametocyte producing clone 8417HP (11) and the gametocyte non-producing clone 233 (17) of the ANKA strain have been used, as well as the gametocyte non-producing clone 1 of the K173 strain (18). Collection, purification and cultivation of blood and mosquito stages of *P.berghei* were performed as described (19,20). Stage-specific RNA was obtained from ring forms, trophozoites, gametocytes and ookinetes and isolated as described before (21) or with an Rneasy kit according to the manufacturer's instructions (Qiagen). For *P.falciparum*, parasites of clone 3D7 were grown *in vitro* (22) and asexual blood stages and gametocytes were obtained (23).

Southern and northern analysis

For northern blotting total RNA was size fractionated on 1% agarose gels under denaturing conditions and transferred to Amersham Hybond-C extra or to Schleicher & Schuell Nytran SuPerCharge membranes. All experiments involving the manipulation of DNA by Southern blot hybridisation were carried out using standard procedures (24). Chromosomes of *P.falciparum* were separated by pulsed field gel electrophoresis using contour-clamped homogeneous electric field gel electrophoresis (CHEF) as described before (25).

Contig assembly of the *P.berghei* B9 locus

A partial *Sau3AI* genomic library of *P.berghei* clone 8417HP (26) was screened with a photobiotinylated subtracted cDNA probe enriched for gametocyte-specific sequences. This probe was prepared by hybridisation of cDNA of gametocytes of clone 8417HP to mRNA of strain K173 using the Subtractor kit according to the manufacturer's instructions (Invitrogen). A genomic clone referred to as B9 was selected and used as the starting point in three successive steps of chromosome walking. Newly identified clones were purified and subcloned with a Rapid Excision Kit according to the manufacturer's instructions (Stratagene) and mapped according to their restriction pattern and assembled in a contig. The contig will be referred to as the B9 locus and its sequence has been generated according to the method described below.

Table 1. Genomic clones of *P.falciparum*

Clone	Position
f-OMP37 ^a	367–2444
f-B9-4 ^a	7332–9574
f-B9-99 ^a	7725–10487
PCR clone 1	2178–4160
PCR clone 2	5875–6312
PCR clone 3	7024–8280
PCR clone OMP	347–1112

^aClones from the *P.falciparum* genomic library (27).

Contig assembly of the B9 locus of *P.falciparum*

Based on the sequence of the B9 locus of *P.berghei*, an attempt was made to assemble the homologous locus for *P.falciparum* from preliminary sequence data of the TIGR shotgun database. As the starting point the putative PbOMP decarboxylase gene, located in the B9 locus of *P.berghei*, was chosen. Degenerate PCR primers (L512 and L513, Table 2) were designed based on amino acid homology between the gene of *P.berghei* and its homologous counterpart in *Trypanosoma cruzi* (accession no. T30520), exploiting the biased codon usage of *P.falciparum*. After PCR amplification the PCR product was cloned and sequenced as described before. This sequence, together with sequences from putative coding regions of ORFs in the B9 locus of *P.berghei* identified by ongoing gene characterisation, were used to blast the TIGR database. The preliminary sequence data for *P.falciparum* chromosomes 10 and 11 were obtained from The Institute for Genomic Research (TIGR) web site (www.tigr.org). Sequencing of chromosomes 10 and 11 was part of the International Malaria Genome Sequencing Project and was supported by an award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. From the collected data a partial contig of *P.falciparum* was assembled and is considered the orthologue of the B9 locus of *P.berghei*. Gaps in the contig of *P.falciparum* were closed by PCR on genomic DNA. In addition, a genomic library of *P.falciparum* (27) was screened to identify clones that covered remaining gaps. All genomic clones used in the process of assembly are listed in Table 1. This assembled sequence subsequently proved identical to that contained in contig c10m345 (61 138 bp) which is available in the *P.falciparum* chromosome 10 database from TIGR.

RT-PCR, RACE and screening of cDNA libraries

RT-PCR. RT-PCR was performed on total RNA obtained from three different sources using a Superscript One-Step RT-PCR System kit according to the manufacturer's instructions (Gibco BRL): for *P.berghei* asynchronous blood stage parasites and a gametocyte-enriched parasite population; for *P.falciparum* a gametocyte-enriched parasite population from the 3D7 clone. Oligonucleotide primers for RT-PCR spanned multiple exons of individual ORFs (Table 2). For subsequent sequencing RT-PCR products were cloned into either the pGEM-T vector (Promega) or the PCR II-TOPO vector using a

Table 2. Oligonucleotide primers used in this study

Gene	Primer code	Nucleotide sequence (5'→3')	Start position
<i>Plasmodium berghei</i>			
PbORF1	RT1-a	TCCCAAATAAGCGCTTTGCC	2042
	RT1-b	CATGCAATATTGGCAAATG	2299
PbORF2	RT2-a	GCATTGTGTGAATAATCAGA	3987
	RT2-b	GACATATACTCTAGGGTCACTAC	4791
	L619	AATTATAAATATGGGAGACGC	3054
	L625	TACTTGACATATACTCTAGGG	4796
PbORF3	L556	TTGCTTATCGCTTTTTCTATCC	5677
	L551	AAATGAAGAACCCAATATTGCC	7247
PbORF4	L503	TTCAAAAAGCAGTAACTTCAAGG	9063
	L504	TATATTCGAGCGTCTATAGGG	11 155
PbORF5	L547	TAAATATGATTTTGCAAATGTTCC	12 479
	L549	ACAACCTTCAAAAATATGTATCCC	12 965
<i>Plasmodium falciparum</i>			
PfOMP decarboxylase	L512	ACiTiTTT(A/G)TGTATTGGiTT(A/G)GATCC ^a	347
	L513	CATTTTGTGC(A/T)CC(A/T)ATiCC(A/T)GG ^a	1112
PfORF1	F-ORF1-1	CCTTATCCATATTTGTATGC	1879
	F-ORF1-10	CCAACTAAAGTATTGAATTT	3647
PfORF2	F-B9-2A	ATGGAACTTATTTAGTAGA	3421
	F-B9-2B	CCAGTCATCTTCAAGTTGTC	4656
PfORF3	F-ORF3-3	GCATACATAGATTTGTCTAC	5272
	F-B9-3A	TGAAGATATTAGCTCCAAG	6903
	F-B9-3D	GGAAAAAATTGTGTCTTC	6618
	F-ORF3-1	CACCTTAAGAACGTACAAAC	8259
PfORF4	F-B9-1	GAAGGTCGTACGATGAGCTTTG	9206
	F-B9-7	TCTAGAATCATCTAATACATATCTA	9973

^aThe degenerate sequences in parentheses have been synthesised in equimolar ratios. i, inosine.

TOPO TA Cloning Kit (Invitrogen) according to the manufacturer's low melting point agarose method.

cDNA. To obtain cDNA clones of ORFs located in the B9 locus of *P.berghei* we screened a cDNA library constructed from asynchronous blood stage mRNA of clone 8417HP (28) with probes representing the coding region of the PbOMP decarboxylase gene or an additional part of the *P.berghei* B9 locus (positions 4006–13 629). Individual clones from the former were subcloned using a Rapid Excision Kit as described before, whereas clones from the latter were subcloned after PCR amplification of the insert using a universal vector oligonucleotide primer combined with a *Plasmodium*-specific internal primer. Exons of the individual ORFs located in the B9 locus of *P.falciparum* were identified by PCR amplification using the gametocyte cDNA library of clone 3D7 of *P.falciparum* as template (25). The positions of the oligonucleotide primers were chosen in genomic regions

that showed homology at the amino acid level to the *P.berghei* contig (Table 2). All PCR products were cloned for subsequent sequencing as described before.

5'-RACE. The 5'-RACE System for Rapid Amplification of cDNA Ends v2.0 (Gibco BRL) was used to identify the transcription start sites of PfORF1, PbORF2, PbORF3 and PbORF4 using gametocyte-enriched total RNA of *P.berghei* and *P.falciparum* as templates.

Sequencing and sequence quality

About 60% of the sequencing reactions were commercially performed on automatic fluorescent DNA sequencers (Base-Clear). This involved sequencing with universal M13 vector primers on a LI-COR4200 sequencer using a SequiTherm EXCEL II Kit-LC (Epicentre) and sequencing with *Plasmodium*-specific primers on an ABI PRISM 310 or ABI PRISM 377 DNA sequencer using an ABI PRISM BigDye Terminator

Cycle Sequencing Kit (Applied Biosystems). The remaining 40% were sequenced in-house using a Sequenase v2.0 DNA Sequencing Kit (Amersham Pharmacia Biotech). To ensure the quality of the sequence of *P.berghei* the genomic non-coding regions in the contig were sequenced at least twice and a mean redundancy reading for coding regions of 4.6 per bp was obtained.

Computer-assisted analysis

Contig assembly was performed using the assembly program from the GCG package v10.0 (Genetics Computer Group, Madison, WI), ContigExpress from the Vector Nti package v1.0 and ClustalW (29). To detect protein homology BLAST and FASTA searches were done against the SWISS-PROT, EMBL and GenBank databases. The TBLASTN program on the TIGR web site for the chromosomes 10 and 11 project was used to find cryptic exons using putative coding regions of the *P.berghei* locus derived from all six reading frames as queries. ORF predictions for the complete contigs of both *Plasmodium* species were kindly performed by Steven Salzberg at TIGR with GlimmerM, a program especially trained for *P.falciparum* (30). Additional sequence analysis was done on the EMBL and NCBI servers (www.embl-heidelberg.de/Services/index.html and www.ncbi.nlm.nih.gov/Tools/index.html, respectively).

RESULTS

Contig assembly of the B9 locus in *P.berghei*

Previously we reported a long-range restriction map of chromosome 5 of *P.berghei* showing the position of a set of chromosome-specific markers (16). One of these markers, a PCR fragment referred to as marker B9, hybridised exclusively to gametocyte RNA in northern blot analysis and has been mapped to the more conserved internal region of chromosome 5 as opposed to the polymorphic sub-telomeric regions. Originally marker B9 was amplified from a genomic clone B9 that had been obtained after screening a partial *Sau3AI* genomic DNA library with a subtracted cDNA probe enriched for genes expressed in gametocytes (Materials and Methods). To investigate the possibility of clustering of sex-specific genes in the internal regions of chromosome 5 we characterised the B9 locus in more detail. To isolate and characterise genes in close proximity to clone B9 a bi-directional chromosome walk was performed resulting in the isolation of 23 genomic DNA clones. After restriction mapping, 14 of these appeared to be non-redundant, overlapping clones with an insert ranging in size from 2.3 to 5.9 kb. In Figure 1A a tiling path is shown of clone B9 and five additional clones that were used to assemble the consensus sequence of 13 629 bp for the B9 locus.

Gene identification in the B9 locus of *P.berghei* and *P.falciparum*

To examine the B9 locus for the presence of putative genes the consensus sequence was analysed using the NCBI program ORF Finder. Two ORFs were identified as putative single exon genes. One was identified as the homologue of the gene encoding orotidine 5'-monophosphate decarboxylase (OMP decarboxylase) and the second will be referred to as PbORF4 (Fig. 1C). In this study unknown genes were named ORFs 1–5

as a working nomenclature and will be referred to as ORF or gene in the text.

A better representation of the putative genes in the B9 locus was obtained after GlimmerM analysis, the gene-finding program specially trained on *P.falciparum*. This program predicted two additional ORFs: PbORF3 and PbORF5 (Fig. 1C). In Figure 2A a visual representation of the GlimmerM prediction of the B9 locus of *P.berghei* is shown in relation to the gene organisation as it became apparent during this study.

Gene expression in the B9 locus was studied by selecting nine adjacent genomic DNA probes along the 13.6 kb sequence of the locus (probes a–k, Fig. 1B). Northern analysis revealed that at least five genes were present in this locus, three of which were predominantly transcribed in gametocytes (Fig. 1D).

An additional valuable source of sequence data became available with the orthologous B9 locus of *P.falciparum* of 13 797 bp (Materials and Methods). GlimmerM analysis of this sequence is shown in Figure 2B and revealed a partial gene structure, PfORF1, not predicted in *P.berghei*. For the orthologous region in *P.berghei* we had shown by northern analysis the expression of a gametocyte-specific transcript of 1.2 kb and constitutively expressed transcripts of 1.4, 1.7 and 2.1 kb (probes b, RT-1 and RT-2, Fig. 1D), while gene-finding software had been unable to predict any statistical gene model. Comparative sequence analysis of both *Plasmodium* species revealed regions with a significant degree of conservation at the amino acid level as well as at the nucleotide level, located in the region of the newly predicted ORF1 with probe b of Figure 1B. In *P.berghei* the presence of ORF1 could be confirmed by RT-PCR based on the conserved regions. In addition, in both *Plasmodium* species a previously unpredicted ORF, ORF2, became apparent (Fig. 1C). RT-PCR probes from several genes (Fig. 1B) were generated (primers listed in Table 2) for subsequent northern blot analyses (Fig. 1D). PbORF5 was not detectable by northern analysis, although a RT-PCR product with a correctly spliced intron could be amplified from gametocyte-enriched cDNA.

Gene organisation in the B9 locus of *P.berghei*

To characterise the structure of the putative genes in the B9 locus in more detail a cDNA library of *P.berghei* was screened and 5'-RACE was performed. Seven cDNA clones were isolated, two of which were characterised as PbOMP decarboxylase, four as ORF2 and one as ORF3. Together with the RT-PCR results, the sequence data for these cDNA clones and the results of the 5'-RACE experiments showed that the B9 locus is gene dense and displays a complex gene organisation. At least six genes are contained in <14 kb and four of these have (multiple) introns. Untranslated regions (UTRs) overlap with neighbouring UTRs and even with coding regions of adjacent genes; PbOMP decarboxylase with PbORF1, PbORF1 with PbORF2 and PbORF2 with PbORF3. Because the UTRs of PbORF5 remain undetermined, overlap between PbORF4 and PbORF5 has not been demonstrated.

The cDNA sequence data not only revealed a complex organisation and tight clustering of the putative genes relative to each other, but also variation within the transcripts of the individual genes. Alternative sites for transcription initiation and polyadenylation were observed and two of the genes appear to be subjected to alternative splicing. These mechanisms contribute

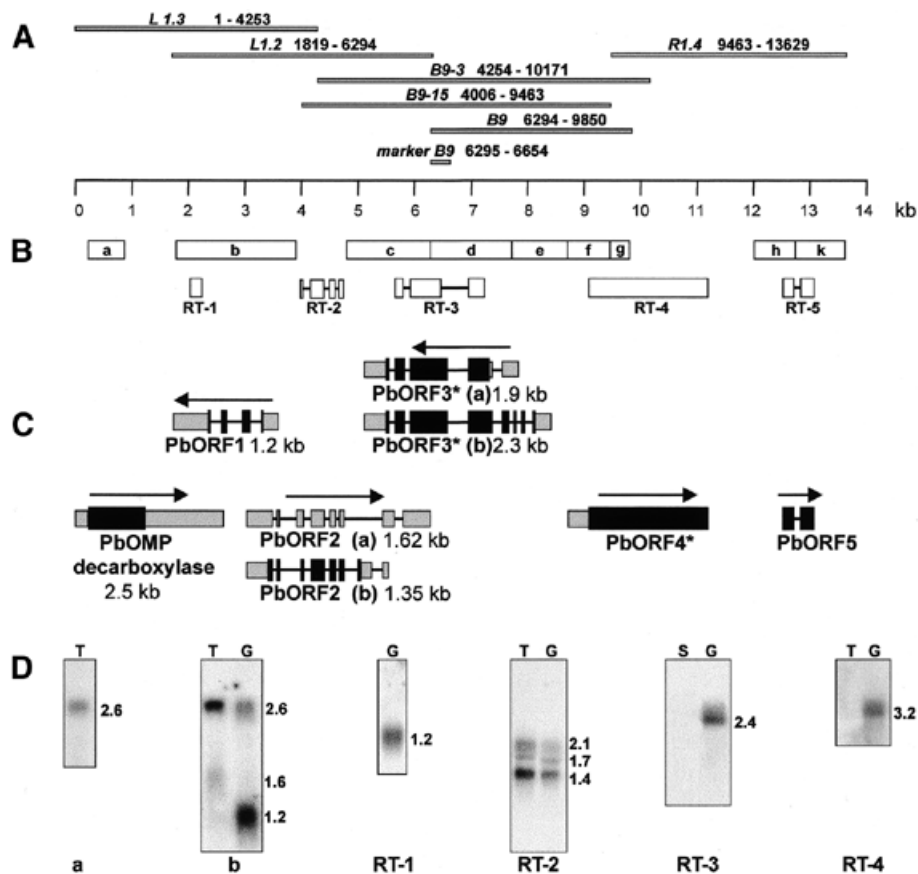


Figure 1. (A) Tiling path of the genomic DNA clones used to assemble the B9 locus in addition to marker B9 (16). (B) Molecular probes used in northern blot analysis relative to their position in the B9 locus. Letters in alphabetic order represent names of the genomic probes and RT numbers refer to RT-PCR probes. (C) Gene organisation in the B9 locus, confirmed by cDNA clone, RT-PCR and 5'-RACE analyses as described. Black and grey boxes represent coding and UTR regions, respectively, and lines the introns. Arrows indicate the direction of transcription and asterisks mark the gametocyte-specific genes. Models (a) and (b) for PbORF2 and PbORF3 represent the two most diverse cDNA clones and 5'-RACE models, respectively. The coding region in model (a) of PbORF2 remains undetermined due to a frameshift and/or internal stop codons as a result of alternative splicing. (D) Compilation of results after northern blot analysis with the probes from (B). T, trophozoite stage; G, gametocyte stage; S, schizont stage. Probe b shows the expression of PbOMP decarboxylase, PbORF1 and PbORF2 relative to each other. Probe b shows a very faint band in the region of 1.6 kb in the G lane compared to the T lane due to unequal amounts of parasite RNA. Probe RT-2 shows three clear bands in the region of 1.6 kb, with equal amounts of parasite RNA loaded in the G and T lanes. The pattern of multiple bands for this probe is most likely a representation of the different alternative splice transcripts.

to variability in transcript sizes, resulting in predicted variant C- and N-termini or even truncated proteins. In Figure 1C two alternative models for both PbORF3 and PbORF2 have been included [models (a) and (b)] based on 5'-RACE experiments and sequencing of cDNA clones, respectively.

Multiple polyadenylation sites. For PbOMP decarboxylase two isolated cDNA clones showed a polyadenylation site at position 2643, in addition to a site at position 2695 found by PCR on the cDNA library. For PbORF2 the variation was more prominent, because the sites were separated by an intervening intron. In Figure 1C model (b) of PbORF2 indicates a polyadenylation site in exon 8 at position 5478 and model (a) at position 6240 in exon 9.

Alternative transcription start sites. For PbORF3 transcription start sites were found by 5'-RACE at positions 8393 and 7672. As shown in Figure 1C, this variability affects the coding

potential of the transcripts of models (a) and (b), leading to an alternative N-terminus of the putative protein.

Alternative splicing. For PbORF1 and PbORF2 differential or alternative splicing was shown while all intron-exon junctions matched the consensus sequence of splicing: GT at the 5'-end and AG at the 3'-end of the intron. For PbORF1 differential splicing of an intron in the 3'-UTR has been observed as a rare event (one out of 10 RT-PCR clones) at positions 2259-2331 (not shown in Fig. 1C). In addition, splicing in the 3'-UTR of the homologous gene of *P.falciparum* has not yet been detected. For PbORF2 a complex pattern of splicing was found. In Figure 1C the two models for PbORF2 represent the two most diverse cDNA clones out of four clones isolated. Not only differential splicing of the whole of exon 7 was evident, but also alternative use of different splice positions for a given exon. The 3'-end of exon 1 in model (a) was positioned 4 nt farther downstream compared to model (b), at positions 3446 and 3442, respectively. Because this nucleotide shift is not a

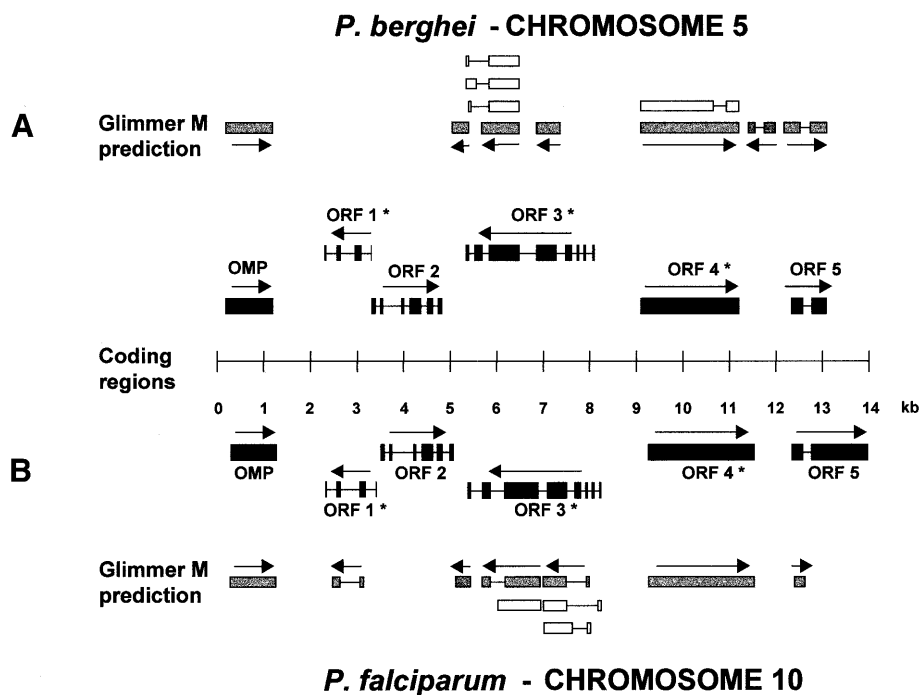


Figure 2. (A) A visual representation of GlimmerM gene prediction in the B9 locus of *P.berghei* in relation to the organisation of the coding regions as it became apparent during this study. The predictions with the highest probability are shown in grey; the predictions with lower probability in white. Asterisks indicate 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 nt (contig c10m345, TIGR database). Without this additional sequence GlimmerM is unable to predict the exon accurately.

multiple of three, the reading frame was altered, giving rise to an alternative start of the coding region or a truncated protein. For exon 3 the 5'-end varied between positions 3898 and 3987, as shown by models (a) and (b), respectively. RT-PCR using primers L619 and L625 detected additional splice variants. Exon 4 showed variation in its 5'-end (positions 4154 and 4185) and an additional intron in the 5'-UTR was found (positions 3189–3291) (not shown in Fig. 1C). In total we found five non-redundant transcript variants for PbORF2.

High level of conservation in orthologous 'B9 loci' of *P.berghei* and *P.falciparum*

By hybridising genomic DNA probes to chromosomes of *P.falciparum* separated by CHEF electrophoresis we confirmed the location of the B9 locus on chromosome 10 (not shown). At the time of writing our complete assembled contig of *P.falciparum* is present in the TIGR database as part of a 61.1 kb contig (c10m345). Comparison of the B9 locus of *P.berghei* with the orthologous locus of *P.falciparum* reveals a high level of conservation. All six putative genes found in *P.berghei* are present in *P.falciparum* as a cluster, in the same order and orientation (Fig. 2). RT-PCR probes for PfOMP decarboxylase and PfORF1–PfORF4 were generated with specific primers (Table 2) to hybridise to RNA from asexual trophozoites and gametocytes of *P.falciparum*. PfORF1, PfORF3 and PfORF4 were expressed only in gametocytes, while a transcript of PfOMP decarboxylase was also detected in trophozoites (data not shown). PfORF2 could not be detected by northern blot analysis, probably due to the

relatively low expression of this gene (similar to PbORF2), but its presence was confirmed by RT-PCR. This pattern of expression is identical to that of the orthologous genes of *P.berghei*. For PfORF5 we have not been able to show expression, although an EST from asexual blood stages (accession no. T02495) was found, representing the 3'-end of this gene. Sequence comparisons of both *Plasmodium* species combined with RT-PCR analysis of cDNA from *P.falciparum* revealed a highly conserved exon–intron organisation for all putative genes (Fig. 2B). The intron–exon boundaries found in the coding regions of *P.berghei* appear to be conserved in *P.falciparum* and all match the consensus for splicing. For PfORF2 homology was found only in the first six exons of PbORF2 and exons in the 3'-UTR remained undetected, even with additional RT-PCR experiments.

Characterisation of the ORFs in the B9 locus

The main characteristics of the six genes in the B9 locus are described below complete with additional bioinformatics analyses (summary in Table 3). During the preparation of this manuscript, preliminary sequence and preliminary annotated sequence data for the *Plasmodium yoelii* genome were made available by the Institute for Genomic Research (www.tigr.org). The complete orthologous B9 locus of *P.yoelii* could be assembled from this, with the exception of a minor gap of a few nucleotides within exon 6 of ORF3 and a gap of ~800 bp from the middle of exon 5 to the beginning of exon 2 of that same ORF. In order of appearance in the B9 locus, from OMP decarboxylase to ORF5, the references to the contigs in the database of

Table 3. Summary of characteristics of the ORFs in the 'B9 locus' of *P.berghei* and *P.falciparum*

Gene (expression) ^a	Exon no.	mRNA <i>P.berghei</i> <i>(P.falciparum)</i>	Putative protein <i>P.berghei</i> <i>(P.falciparum)</i>	Nucleotide identity ^b coding region (%)	Amino acid identity (%) ^b <i>(P.yoelii, P.vivax)</i>	Amino acid similarity (%) ^b <i>(P.yoelii, P.vivax)</i>
OMP decarboxylase (AS, S)	1	2.6 kb (2.6 kb)	334 aa (323 aa)	75 75	67.2 (94, 62)	82 (97, 79)
ORF1 (S)	4/5 ^c	1.2 kb (1.5 kb)	79 aa (79 aa)	85 85	84.8 (98, 73)	95 (99, 89)
ORF2 (AS, S)	9 ^d	2.1/1.7/1.4 kb (unknown)	193 aa ^e (185 aa)	73.5 ^f 73.5 ^f	62.2 (97, na)	84 (97, na)
ORF3 (S)	8	2.4 kb (2.0 kb)	492/512 aa ^g (512 aa)	71 71	57.8 (94, na)	72 (96, na)
ORF4 (S)	1	3.2 kb (3.0 kb)	699 aa (758 aa)	63.2 63.2	47.9 (88, 44)	63 (94, 62)
ORF5 (BS)	2 ^h	unknown (unknown)	184 aa (757 aa)	ns ns	50.0 (85, 41) ⁱ	71 (91, 68) ⁱ

ns, not significant; na, not available.

^aExpression of the gene in blood stage parasites [AS, asexual blood stage; S, sexual (gametocyte); BS, blood stage parasites but precise stage undefined].

^bSequence comparison between *P.berghei* and *P.falciparum*. Values given in parentheses are the comparisons of *P.berghei* with first *P.yoelii* and second *P.vivax*, respectively.

^cA differentially spliced intron in the 3'-UTR was found only in *P.berghei*.

^dTotal number of exons is nine (*P.berghei*), but the number per transcript varies due to alternative splicing.

^eBased on model (b) of Figure 1C.

^fBased only on the first six exons.

^gBased on models (a) and (b) of Figure 1C.

^hNumber of exons based on RT-PCR results; additional exons cannot be excluded.

ⁱNot including the region of exon 2 containing the repeat motifs unique to *P.falciparum*.

P.yoelii are 755, 3329, 5703 and 1035. Although we lack experimental data to prove conservation of the genome organisation in this rodent malaria species, the homology to the coding regions in the B9 locus of *P.berghei* suggests a high level of conservation (see below). In addition, conserved synteny has been shown before for six genetic markers located on chromosome 5 of *P.berghei* and the orthologous chromosome of *P.yoelii*, as well as for 47 other markers located on the other chromosomes (13). Genome survey sequences (GSS) of *Plasmodium vivax* with homology to coding regions in the B9 locus of *P.berghei* were found in the database with preliminary sequences from a mung bean nuclease-digested genomic DNA library of *P.vivax* (<http://parasite.vetmed.ufl.edu>) and will be referred to in comparisons of sequence homology.

OMP decarboxylase. OMP decarboxylase catalyses the final step in pyrimidine biosynthesis. The *P.berghei* OMP decarboxylase gene consists of one continuous exon including a remarkably long 3'-UTR of 1438 bp. Several conserved domains are present in the proteins of *P.berghei* and *P.falciparum*, including the catalytic site (31). Highest homology in GenBank was observed with the orthologous genes from the protozoan parasites *T.cruzi* (32) and *Leishmania mexicana* (accession no. BAA94297). In the databases orthologous genes were identified for both *P.yoelii* and *P.vivax* (GSS 318PvC04, Table 3).

ORF1. This gene consists of five exons with a rarely spliced intron in the 3'-UTR. The presence of the gametocyte-specific 1.2 kb transcript is shown in Figure 1D by probes b and RT-1. Protein structure analysis gave indications for the presence of a signal sequence and a transmembrane domain. Orthologues were observed in the databases for *P.yoelii* and *P.vivax* (273PvD07 encoding exon 3 plus flanking introns, Table 3).

ORF2. This gene consists of nine exons with a complex pattern of alternative splicing (see Supplementary Material). Constitutive expression of the 1.4/1.7/2.1 kb transcripts during blood stage development was demonstrated by probes b and RT-2 in Figure 1D. The predicted protein of model (b) (Fig. 1C) revealed homology to the human MAD2B (mitotic arrest-deficient) (33) and MAD2-like proteins that are required for implementation of the mitotic and meiotic spindle checkpoint during cell division (34) (see Supplementary Material). Protein sequence alignment of a group of MAD-like proteins of different species showed a conserved motif in the first half of the amino acid sequence (Supplementary Material). An orthologue was found in *P.yoelii*.

ORF3. The sequence and further analysis of this gene revealed the presence of eight exons. A minor variant of mRNA was detected that lacks the first three 5'-exons, resulting in a different N-terminus of the predicted protein. Expression of a gametocyte-specific 2.4 kb transcript could be detected by

probe RT-3 (Fig. 1D). Coiled-coil domains, potentially involved in protein-protein interactions, and a discernible nuclear localisation signal (NLS) were predicted for the proteins of *P.berghei* and *P.falciparum*. An orthologue was found in *P.yoelii* (missing exons 2–4 and the 5'-start of exon 5 and with a small gap in exon 6, Table 3).

ORF4. This gene consists of a single exon and encodes a cysteine-rich protein of which all 16 cysteines are conserved between *P.berghei* and *P.falciparum*. In Figure 1D expression of a gametocyte-specific 3.2 kb transcript was detected with probe RT-4. In all three *Plasmodium* species a transmembrane domain has been predicted in the 3'-end of the putative protein, but no signal sequence. Orthologues were found in *P.yoelii* and *P.vivax* (226PvD02 and 230PvB03, Table 3). All cysteine residues are conserved in all four proteins.

ORF5. This gene with two exons in *P.falciparum* and possibly three in *P.berghei* is the least conserved among the six genes in the B9 locus. The position of the first intron is confirmed experimentally in *P.berghei* and shown by comparison to be conserved in *P.falciparum*. The second exon in *P.falciparum* is 685 residues long in contig c10m345 (TIGR), compared to 103 in *P.berghei*. The deduced protein of this long exon contains two blocks of repeat motifs, one of six and one of nine amino acid residues, that are lacking in *P.berghei*. In *P.yoelii* a possible third exon was found at the end of the B9 locus, coding for 424 amino acids. This third exon has no significant homology to exon 2 of *P.falciparum*, except for a regular spacing of the relatively abundant glutamine residues. In all three *Plasmodium* species several NLS were detected in the putative protein. Although PbORF5 was amplified from gametocyte-enriched cDNA of *P.berghei* we have no further evidence to confirm that it is gametocyte specific. ORF5 is conserved between the two rodent malaria species (Table 2) and for *P.vivax* a GSS was found (300PvD03) that contained exon 1 with the position of the splice site conserved (Table 1).

DISCUSSION

It is evident from the analysis of organisms whose genomes have been completely sequenced that comparative genomics can be highly informative in identifying novel genes, gene classes or gene functions with the availability of phylogenetically closely related species (4). In *Plasmodium* this comparative approach can still be considered to be in its infancy given the relatively small number of genes present in GenBank for different *Plasmodium* species and the preliminary status of most of the sequencing data for various *Plasmodium* species released by the sequencing centres involved in the malaria genome sequencing project.

Initial results from studies focusing on evolutionary conservation among these species hint that there are significant similarities in the genome organisation of all malaria parasites. A high degree of conserved clustering of several genes within the genome (synteny) has been demonstrated among rodent malaria species (13), to a lesser degree between rodent malaria parasites and *P.falciparum* (14) and among different human malaria parasites (15). As an example of a conserved multi-gene family among different *Plasmodium* species preliminary indications (J.Thompson, personal communication) suggest

the presence of orthologues of members of the Cys₆ domain gene superfamily in *P.berghei*, *P.vivax* and *Plasmodium chabaudi*, including the gametocyte-specific genes Pfs48/45 and Pfs230 (35). In addition, indications of a high level of conservation of complex gene structure among different *Plasmodium* species were found in the organisation of the intron-exon structure of individual genes, e.g. the *cdc2*-related kinase 2 gene (36), the α -tubulin I and II genes (J.Renz and A.P.Waters, unpublished results) and the guanylyl cyclase gene (preliminary data from www.ncbi.nlm.nih.gov/Malaria/plasmodiumblcus.html; J.Thompson, personal communication).

This study presents a detailed analysis of a 13.6 kb region, the B9 locus, located in an internal position on chromosome 5 of *P.berghei*. The *P.falciparum* counterpart identified through a comparative analysis of *P.falciparum* genome sequence databases maps to chromosome 10. The gene number, organisation and expression pattern of this region is highly conserved between the two species. Three out of six genes identified within this region are transcribed exclusively in sexual stages. Four of the six genes have introns with canonical splice junctions that, in two of these genes, are differentially or alternatively spliced. Physical analysis revealed significant overlap between the genes, including partial overlap of the coding regions with gene structures of adjacent genes, a phenomenon not previously described in *Plasmodium*. This compression of the B9 locus results in an observed gene density of one gene per 2.2 kb, which is greater than observed for chromosomes 2 (one per 4.8 kb; 2) and 3 (one per 4.5 kb; 3). The relatively high gene density of the B9 locus and extent of transcript overlap on the two strands are emphasised by its transcript capacity of 15.5 kb within the 13.6 kb genomic sequence (DNA:RNA ratio 1:1.14). An intricate organisation of the genome, as shown here for the B9 locus, might have a favourable influence on the conservation of synteny and gene order in *Plasmodium*; spontaneous large-scale recombination events like translocations are unlikely to be neutral, but would probably have a significant effect on gene expression of at least one gene.

In this study the identification of genes was significantly aided by direct comparison of the genomic DNA sequence of the B9 locus from *P.berghei* and *P.falciparum*. This allowed initial identification of the intron-exon structure of the genes consisting of multiple small exons and specific design of primers for cDNA synthesis. It was of crucial importance to the identification of ORF1 in *P.berghei* and ORF2, which was not predicted by the GlimmerM algorithm, in both *Plasmodium* species. In addition, direct sequence comparison and subsequent cDNA analysis complemented the prediction of ORF3 in *P.berghei*. Since GlimmerM was only trained on a subset of *P.falciparum* genes it is likely that the accuracy of prediction in *P.berghei* is not as good as in *P.falciparum*, although in general both genomes can be considered to be very similar in their A+T content (~82%) (37). In the B9 locus we have found overall A+T contents of 77 and 81% for *P.berghei* and *P.falciparum*, respectively. A lower A+T content of the introns and other non-coding regions of *P.berghei* (78%) compared to the average number found for *P.falciparum* (~85%) (3) might be an additional complicating factor in gene structure prediction. The acknowledged difficulty of accurately predicting the complex nature of genes consisting of multiple small exons in the absence of similarity data makes it reasonable that this class of genes in the current genome annotation of *P.falciparum* is

being missed or incorrectly predicted (3,38,39). However, it must be emphasised that the sample of *P.berghei* reported here is extremely small and that generalised conclusions might be somewhat misleading. This caveat notwithstanding, a relative underestimation of split genes can have practical consequences, as we have shown, and lead to a failure to detect genes in important loci, e.g. the *P.falciparum* gene PfCRT, encoding a vacuolar transport protein, in which mutations have been reported to be associated with resistance to chloroquine (40). Although additional algorithms will play a role in improving gene identification (41), experimental data is crucial to the accurate assignation of gene extent and exon structure (3,39). In this study the demonstration of split genes consisting of multiple small exons, alternative splicing and gene overlap merely emphasises both this need and the fact that it might take a considerable amount of work to generate sufficient appropriate data.

Although an ability to alternatively select different combinations of exons during RNA splicing has been reported for *Plasmodium* (42), an additional feature of alternative splicing so far unique to *Plasmodium* is found for ORF2: differential use of canonical splice sites for the same exon. In total at least five variant transcripts are present, reflected in multiple bands in northern blot analyses. The coding potential of several transcript variants remains speculative, but truncated protein isoforms, non-coding antisense regulatory RNA, translation read-through through an internal stop codon (43) and simple transcription errors can all be hypothesised. In higher eukaryotes alternative splicing can act as a powerful regulatory mechanism that can affect quantitative control of gene expression and functional diversification of proteins (44,45). We note that both the ORF2 gene, which has homology to the human spindle checkpoint gene MAD2, and the PfPK6 gene of *P.falciparum*, whose encoded kinase is also involved in cell cycle progression, produce multiple mRNA species (46).

The locus described here, containing several genes that use alternative transcription start and stop sites, may represent a small but interesting database in which to look for structural features relevant to transcription promotion, stage specificity and mRNA stability. For example, the region between ORF3 and ORF4 is a good candidate as a bi-directional gametocyte-specific promoter element. In addition, the information on gene organisation in the B9 locus will be a valuable set of data to improve the gene-finding algorithm of GlimmerM (47).

In conclusion, characterisation of complex genes and alternative splicing will rely on direct physical analysis of the coding potential of genomes aided by comparison with a closely related genome. Thus, it would be most useful to incorporate sequence information for all *Plasmodium* species and related apicomplexan parasites into the *P.falciparum* genome database to improve its current annotation. Short homologies of protein coding DNA or regulatory elements would therefore become evident even in A+T-rich regions, giving a framework for the discovery and accurate prediction of novel genes and their structural features. Furthermore, we anticipate that direct genome comparison will have additional value for the identification of genes that are expressed in parasite forms for which RNA and cDNA are not readily available for analysis due to the complex life cycle of the parasite.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge Pietro Alano and Francesco Silvestrini for *P.falciparum* cultures and RNA analysis and for providing the pJFE14DAF gametocyte cDNA library of clone 3D7 of *P.falciparum*; Clara Frontali and Joanne Thompson for critical reading of the manuscript and helpful suggestions; and the latter also for communication of unpublished findings; Steven Salzberg, Leda Cummings and Malcolm Gardner for their help with GlimmerM analysis. This work received support in the form of a grant from the European Community INCO-DC programme of the Fourth Framework of DG XII (grant IC18-CT96-0052). Preliminary sequence data from *P.falciparum* chromosomes 10 and 11 and the *P.yoelii* genome was obtained from The Institute for Genomic Research web site (www.tigr.org). Sequencing of chromosomes 10 and 11 was part of the International Malaria Genome Sequencing Project and was supported by an award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. The *P.yoelii* sequencing program is being carried on in collaboration with the Naval Medical Research Center and is supported by the US Department of Defense. Sequence data for *P.vivax* were obtained from the University of Florida Gene Sequence Tag Project web site at <http://parasite.vetmed.ufl.edu>. Funding was provided by the National Institute of Allergy and Infectious Diseases.

REFERENCES

- Hoffman, S.L., Bancroft, W.H., Gottlieb, M., James, S.L., Burroughs, E.C., Stephenson, J.R. and Morgan, M.J. (1997) Funding for malaria genome sequencing. *Nature*, **387**, 647.
- Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V., Shallom, S., Mason, T., Yu, K., Fujii, C. *et al.* (1998) Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science*, **282**, 1126–1132.
- Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Feltwell, T. *et al.* (1999) The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature*, **400**, 532–538.
- Mushegian, A.R., Garey, J.R., Martin, J. and Liu, L.X. (1998) Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode and yeast genomes. *Genome Res.*, **8**, 590–598.
- Doolittle, R.F. (1997) A bug with excess gastric avidity. *Nature*, **388**, 515–516.
- Ansari-Lari, M.A., Oeltjen, J.C., Schwartz, S., Zhang, Z., Muzny, D.M., Lu, J., Gorrell, J.H., Chinault, A.C., Belmont, J.W., Miller, W. *et al.* (1998) Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6. *Genome Res.*, **8**, 29–40.
- Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W. *et al.* (2000) Comparative genomics of the eukaryotes. *Science*, **287**, 2204–2215.
- Chervitz, S.A., Aravind, L., Sherlock, G., Ball, C.A., Koonin, E.V., Dwight, S.S., Harris, M.A., Dolinski, K., Mohr, S., Smith, T. *et al.* (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science*, **282**, 2022–2028.
- Field, D., Hood, D. and Moxon, R. (1999) Contribution of genomics to bacterial pathogenesis. *Curr. Opin. Genet. Dev.*, **9**, 700–703.
- Sheppard, M., Thompson, J.K., Anders, R.F., Kemp, D.J. and Lew, A.M. (1989) Molecular karyotyping of the rodent malarial *Plasmodium chabaudi*, *Plasmodium berghei* and *Plasmodium vinckei*. *Mol. Biochem. Parasitol.*, **34**, 45–52.

11. Janse, C.J., Boorsma, E.G., Ramesar, J., van Vianen, P., van der Meer, R., Zenobi, P., Casaglia, O., Mons, B. and van der Berg, F.M. (1989) *Plasmodium berghei*: gametocyte production, DNA content and chromosome-size polymorphisms during asexual multiplication *in vivo*. *Exp. Parasitol.*, **68**, 274–282.
12. Janse, C.J. (1993) Chromosome size polymorphism and DNA rearrangements in *Plasmodium*. *Parasitol. Today*, **9**, 19–22.
13. Janse, C.J., Carlton, J.M., Walliker, D. and Waters, A.P. (1994) Conserved location of genes on polymorphic chromosomes of four species of malaria parasites. *Mol. Biochem. Parasitol.*, **68**, 285–296.
14. Carlton, J.M., Vinkenoog, R., Waters, A.P. and Walliker, D. (1998) Gene synteny in species of *Plasmodium*. *Mol. Biochem. Parasitol.*, **93**, 285–294.
15. Carlton, J.M., Galinski, M.R., Barnwell, J.W. and Dame, J.B. (1999) Karyotype and synteny among the chromosomes of all four species of human malaria parasite. *Mol. Biochem. Parasitol.*, **101**, 23–32.
16. van Lin, L.H.M., Pace, T., Janse, C.J., Scotti, R. and 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.
17. Dearsly, A.L., Sinden, R.E. and Self, I.A. (1990) Sexual development in malarial parasites: gametocyte production, fertility and infectivity to the mosquito vector. *Parasitology*, **100**, 359–368.
18. Janse, C.J., Ramesar, J., van den Berg, F.M. and Mons, B. (1992) *Plasmodium berghei*: *in vivo* generation and selection of karyotype mutants and non-gametocyte producer mutants. *Exp. Parasitol.*, **74**, 1–10.
19. Mons, B., Janse, C.J., Boorsma, E.G. and Van der Kaay, H.J. (1985) Synchronized erythrocytic schizogony and gametocytogenesis of *Plasmodium berghei in vivo* and *in vitro*. *Parasitology*, **91**, 423–430.
20. Janse, C.J. and Waters, A.P. (1995) *Plasmodium berghei*: the application of cultivation and purification techniques to molecular studies of malaria parasites. *Parasitol. Today*, **11**, 138–143.
21. Paton, M.G., Barker, G.C., Matsuoka, H., Ramesar, J., Janse, C.J., Waters, A.P. and Sinden, R.E. (1993) 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.
22. Alano, P., Silvestrini, F. and Roca, L. (1996) Structure and polymorphism of the upstream region of the pfg27/25 gene, transcriptionally regulated in gametocytogenesis of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.*, **79**, 207–217.
23. Carter, R., Graves, P.M., Creasey, A., Byrne, K., Read, D., Alano, P. and Fenton, B. (1989) *Plasmodium falciparum*: an abundant stage-specific protein expressed during early gametocyte development. *Exp. Parasitol.*, **69**, 140–149.
24. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. Alano, P., Read, D., Bruce, M., Aikawa, M., Kaido, T., Tegoshi, T., Bhatti, S., Smith, D.K., Luo, C., Hansra, S. *et al.* (1995) COS cell expression cloning of Pfg377, a *Plasmodium falciparum* gametocyte antigen associated with osmiophilic bodies. *Mol. Biochem. Parasitol.*, **74**, 143–156.
26. Birago, C., Pace, T., Barca, S., Picci, L. and Ponzi, M. (1996) A chromatin-associated protein is encoded in a genomic region highly conserved in the *Plasmodium* genus. *Mol. Biochem. Parasitol.*, **80**, 193–202.
27. Pace, T., Ponzi, M., Scotti, R. and Frontali, C. (1995) Structure and superstructure of *Plasmodium falciparum* subtelomeric regions. *Mol. Biochem. Parasitol.*, **69**, 257–268.
28. Birago, C., Pace, T., Picci, L., Pizzi, E., Scotti, R. and Ponzi, M. (1999) The putative gene for the first enzyme of glutathione biosynthesis in *Plasmodium berghei* and *Plasmodium falciparum*. *Mol. Biochem. Parasitol.*, **99**, 33–40.
29. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
30. Salzberg, S.L., Pertea, M., Delcher, A.L., Gardner, M.J. and Tettelin, H. (1999) Interpolated Markov models for eukaryotic gene finding. *Genomics*, **59**, 24–31.
31. Traut, T.W. and Temple, B.R. (2000) The chemistry of the reaction determines the invariant amino acids during the evolution and divergence of orotidine 5'-monophosphate decarboxylase. *J. Biol. Chem.*, **275**, 28675–28681.
32. Gao, G., Nara, T., Nakajima-Shimada, J. and Aoki, T. (1999) Novel organization and sequences of five genes encoding all six enzymes for *de novo* pyrimidine biosynthesis in *Trypanosoma cruzi*. *J. Mol. Biol.*, **285**, 149–161.
33. Cahill, D.P., da Costa, L.T., Carson-Walter, E.B., Kinzler, K.W., Vogelstein, B. and Lengauer, C. (1999) Characterization of MAD2B and other mitotic spindle checkpoint genes. *Genomics*, **58**, 181–187.
34. Shonn, M.A., McCarroll, R. and Murray, A.W. (2000) Requirement of the spindle checkpoint for proper chromosome segregation in budding yeast meiosis. *Science*, **289**, 300–303.
35. Templeton, T.J. and Kaslow, D.C. (1999) Identification of additional members define a *Plasmodium falciparum* gene superfamily which includes Pfs48/45 and Pfs230. *Mol. Biochem. Parasitol.*, **101**, 223–227.
36. Vinkenoog, R., Veldhuisen, B., Speranca, M.A., del Portillo, H.A., Janse, C. and Waters, A.P. (1995) Comparison of introns in a cdc2-homologous gene within a number of *Plasmodium* species. *Mol. Biochem. Parasitol.*, **71**, 233–241.
37. McCutchan, T.F., Dame, J.B., Miller, L.H. and Barnwell, J. (1984) Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. *Science*, **225**, 808–811.
38. Pertea, M., Salzberg, S.L. and Gardner, M.J. (2000) Finding genes in *Plasmodium falciparum*. *Nature*, **404**, 34.
39. Lawson, D., Bowman, S. and Barrell, B. (2000) Finding genes in *Plasmodium falciparum*. *Nature*, **404**, 34–35.
40. Fidock, D.A., Nomura, T., Talley, A.K., Cooper, R.A., Dzekunov, S.M., Ferdig, M.T., Ursos, L.M., bir Singh Sidhu, A., Naude, B., Deitsch, K.W. *et al.* (2000) Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell*, **6**, 861–871.
41. Yeramian, E. (2000) The physics of DNA and the annotation of the *Plasmodium falciparum* genome. *Gene*, **255**, 151–168.
42. Knapp, B., Nau, U., Hundt, E. and Kupper, H.A. (1991) Demonstration of alternative splicing of a pre-mRNA expressed in the blood stage form of *Plasmodium falciparum*. *J. Biol. Chem.*, **266**, 7148–7154.
43. Bischoff, E., Guillotte, M., Mercereau-Puijalon, O. and Bonnefoy, S. (2000) A member of the *Plasmodium falciparum* Pf60 multigene family codes for a nuclear protein expressed by readthrough of an internal stop codon. *Mol. Microbiol.*, **35**, 1005–1016.
44. Sharp, P.A. (1994) Split genes and RNA splicing. *Cell*, **77**, 805–815.
45. Lopez, A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.*, **32**, 279–305.
46. Bracchi-Ricard, V., Barik, S., Delvecchio, C., Doerig, C., Chakrabarti, R. and Chakrabarti, D. (2000) PfPK6, a novel cyclin-dependent kinase/mitogen-activated protein kinase-related protein kinase from *Plasmodium falciparum*. *Biochem. J.*, **347**, 255–263.
47. van Lin, L.H.M., Janse, C.J. and Waters, A.P. (2000) Genome organisation of non-Falciparum malaria species: the need to know more. *Int. J. Parasitol.*, **30**, 357–370.