ELSEVIER

Comparative genomics of malaria parasites

Neil Hall and Jane Carlton

In the past few years, the area of comparative genomics of malaria parasites has begun to come of age, with the completion of genome sequencing projects of four *Plasmodium* species, and several functional genomics studies. A picture is emerging of a parasite genome that is highly adapted to its mammalian and vector hosts, and which uses post-transcriptional gene-silencing as one method for the control of gene expression. The genome is compartmentalized into a core of conserved housekeeping genes, sandwiched between subtelomerically located genes encoding surface antigens. Species-specific gene families shape the preference of the parasite for host cells, in addition to determining interactions with the host immune-system. Recent research has led to the description of a motif that is conserved across *Plasmodium* species and which plays a central role in protein export into the host cell.

Addresses
The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20851, USA

Corresponding authors: Hall, Neil (nhall@tigr.org); Carlton, Jane (carlton@tigr.org)

Current Opinion in Genetics & Development 2005, **15**:609–613

This review comes from a themed issue on
Genomes and evolution
Edited by Stephen J O'Brien and Claire M Fraser

Available online 22nd September 2005

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2005.09.001

Introduction

Malaria infects 300–600 million people and causes more than one million deaths annually in tropical and subtropical parts of the world, making it one of the most important diseases affecting mankind [1]. The disease is caused by species of the genus *Plasmodium*, intracellular protozoan parasites that are transmitted from host to host by mosquito vectors. There are four species of *Plasmodium* that infect humans, including *Plasmodium falciparum*, which is the most virulent species, and *Plasmodium vivax*, which is the most prevalent. As *Plasmodium* parasites are host-restricted, making studies of the human-infective species difficult, there has been considerable interest in model malaria parasites that can be adapted to growth in laboratory rodents: these model parasites include *Plasmodium berghei*, *Plasmodium chabaudi* and *Plasmodium yoelii*, and species, such as *Plasmodium knowlesi* and *Plasmodium cynomolgi*, that infect non-human primates. In addition, the avian malaria parasite *Plasmodium gallinaceum* is used as a model for the study of the mosquito stages.

Genome sequencing projects of four *Plasmodium* species have now been published: the complete genome sequence of *P. falciparum* and the rodent malaria parasite *P. y. yoelii* in 2002 [2,3], and two further rodent malaria species in 2005 [4••]. Several other *Plasmodium* species are currently being sequenced (see [5•] for review). Examination of data from all of these organisms shows that the *Plasmodium* genomes are haploid, have a standard size of approximately 22–26 Mb and are distributed among 14 linear chromosomes with a size range of 0.5–3.0 Mb. Genome composition varies among species, from the extremely (A + T)-rich genome of *P. falciparum* and the rodent malaria species (~80%) to the more (G + C)-rich *P. vivax* and *P. cynomolgi* genomes (~70%), which, in addition, have an isochore structure with regions of high (G + C) content interspersed between regions of high (A + T) content [6]. Each *Plasmodium* species appears to have 5000–6000 predicted genes per genome [2,3], ~60% of which are orthologous among the species. Many of the genes unique to each species are located within subtelomeric regions, and many code for immunodominant antigens. The difference in gene number between species is caused by (i) differential gene expansion in distinct lineages; and (ii) in some species, the presence of a large, variant gene family, the *Plasmodium* interspersed repeats (PIRs) family, members of which are predicted to be involved in antigenic variation. Finally, studies involving mapping of conserved genes to separations of *Plasmodium* chromosomes [7–9], and the generation of a whole genome alignment map among the rodent malaria species and *P. falciparum* [3] have shown that gene location and order, and even exon–intron boundaries and the fine-scale organization of genes, are preserved over large regions across *Plasmodium* species. The degree of conservation of synteny is greatest when comparing genomes of more closely related species.

In this article, we review the most recent studies of comparative genomics emerging in the wake of the completion of four *Plasmodium* genome sequencing projects. These studies are in their infancy but promise to provide a wealth of data greater than that provided by analysis of the individual genomes alone.

Comparative analysis of *Plasmodium* antigens, and antigenic variation

It is now well established that the major differences in gene content among *Plasmodium* species occur between

genes involved in interaction with the host immune system. Many of these genes are located in the subtelomeric regions of the parasite genome, and sequencing of these regions in several species has identified such antigen families, some of which are conserved between species (for example, the *P. vivax* [10] and rodent malaria species [3,11••] PIR families, which are clearly related and might have shared a common ancestor with the *rifin* genes of *P. falciparum* [11••]). *P. falciparum* contains other gene families that encode proteins involved in antigenic variation and evasion of immune responses (the *var*, *rifin*, *stevor* gene families, with 60, 140 and 25 copies each, respectively; reviewed in [12]). In *P. knowlesi*, the *SICAvar* (*Schizont-infected cell agglutination variant antigen*) gene family has also been described [13]. This gene encodes a protein that is expressed on the surface of infected erythrocytes and is implicated in antigenic variation in this species. There is little homology between the *SICAvar* and the *var* genes, despite their common function. Recently, a comparative study has identified the *P. falciparum SURFIN* gene family [14•], which forms a clade with the gene encoding the *P. vivax* transmembrane protein PvSTP1 (*P. vivax* subtelomeric protein 1) [10]; together, these proteins contain features of several exported and surface-expressed proteins from human, rodent and monkey malaria species. These studies suggest that species-specific evolution of antigen genes, most probably in response to pressure from differing host immune systems, has led to the current diverse repertoire of malaria antigens found in different species.

More recent whole genome analysis studies have demonstrated that proteins which are exported to the red blood cell surface contain a motif (termed the Pexel motif) that appears to be conserved across several *Plasmodium* species [15•,16]. A search of all the *P. falciparum* proteins identified many containing the Pexel motif, indicating that many as yet uncharacterized subtelomeric proteins are exported to the infected red blood cell surface. Although this motif is present in *var*, *rifin* and *stevor* genes and in members of several hypervariable gene families of *P. yoelii*, *P. vivax* and *P. gallinaceum*, it is not present in any members of the *pir* gene family [15•], which might suggest that proteins of this family are not exported to the red blood cell surface and might, indeed, have a different role than originally inferred.

Comparative gene expression and regulation

Until recently, very little was known regarding the regulation of gene expression in *Plasmodium*. Genes are monocistronically transcribed, implying the presence of regulatory sequence elements flanking coding regions: and the few promoters that have been identified appear to conform to the standard eukaryotic promoter structure [i.e. a basal promoter regulated by upstream enhancer elements (reviewed in [17]]. Few of the identified DNA elements direct the transcription of *Plasmodium* genes, although one promising candidate, the G-box, was recently identified upstream of several *P. falciparum* heat shock protein genes, and was found to be conserved across several *Plasmodium* species [18]. Since publication of the *P. falciparum* genome sequence, several transcriptome [19,20,21•,22,23] and proteome [3,24,25] studies of various life-cycle stages have been completed. These sequences have enabled further insight into gene regulation in *Plasmodium*; for example, microarray studies have shown that steady-state RNA levels for many transcripts change throughout the parasite life-cycle, indicating transcriptional regulation at the level of RNA synthesis and/or stability. Bozdech *et al.* [22] generated microarray data for the *P. falciparum* asexual stages and suggested that a small number of transcription factors with overlapping binding site specificities could account for the mechanical character of transcriptional control. These studies have also been cross-referenced to proteomic studies of the asexual, sexual and mosquito stages of *P. falciparum* [24,25], subsequently revealing that large proportions of the genome encode proteins that are used in multiple stages of the life-cycle. This has led to theories that a complex, multilayer regulatory network is employed by the parasite for gene expression, a different mode of regulation than that observed in other eukaryotes [22].

More recently, a search for transcription-associated proteins within the complete *P. falciparum* genome found that the parasite contained far fewer than expected in comparison with those of other eukaryotes [26], leading to the conclusion that *Plasmodium* protein levels might be primarily determined by post-transcriptional mechanisms. A global analysis of transcript and protein levels in *P. falciparum* also led to a similar conclusion [21•]. As a step towards identifying motifs involved in such gene regulation, a recent study took advantage of available *P. berghei* microarray studies and proteomic data and identified a motif in the 3′UTR (untranslated region) of genes that are upregulated in gametocyte stages but whose protein products appear after transmission from the vertebrate to the invertebrate host [4••]. This enabled identification of the motif as a putative control-element involved in the translational repression of transcripts produced during the sexual stages. Downstream regions of *P. falciparum* orthologs of these genes did not contain the same motif, indicating that the regulation of gene expression could be a major contributor to *Plasmodium* host specificity and parasite diversification.

Finally, the first proteomic analysis of separated male and female gametocytes in *P. berghei* has shown that expression of sex-specific proteins is probably controlled by the 5′UTR and not by the 3′UTR nor through post-translational processes [27•]. It remains to be seen how applicable this finding is to the control of gene expression in other *Plasmodium* species.

Comparative evolutionary studies

Comparative genomic methods have been essential in attempting to understand the evolutionary history of malaria parasites in addition to the selective pressures acting upon them (for a recent review, see [28•]). For example, several studies using a limited number of genes have proposed a 'Malaria's Eve' hypothesis, which suggests that, on the basis of the sparse genetic diversity exhibited by the parasite, extant *P. falciparum* originated from a population bottleneck around 6000–10 000 years ago [29,30]. By contrast, Mu *et al.* generated a SNP (single nucleotide polymorphism) density map of *P. falciparum* chromosome 3 from five geographically distinct isolates, and based on the level of divergence, dated the origin of *P. falciparum* to 100 000–180 000 years ago [31], significantly older than Malaria's Eve. Another comparative genomics study, which somewhat reconciled these two extreme views, analyzed the genome sequence of the 6 kb mitochondrial genome from 100 *P. falciparum* isolates world-wide, and provided compelling evidence of an ancient origin for the species (50 000–100 000 years old) but with a recent expansion in the African malaria parasite population [32]. Such large-scale mitochondrial sequencing studies have been repeated using *P. vivax*, and this species, too, was found to be an ancient parasite [33••] that most probably arose by a host switch from macaque monkeys [34].

Understanding the diversity of human *Plasmodium* isolates is of serious consequence for control measures, because a genetically homogenous population is easier to control than a variable, heterogenous one through the rational use of drugs and/or vaccines. In addition to the studies mentioned above, several large-scale genomics approaches have been used to study *P. vivax* diversity. Feng *et al.* [35] sequenced a 100 kb region syntenic to *P. falciparum* chromosome 3 in five *P. vivax* isolates and compared the perceived evolutionary histories of the orthologs between species and within species. A highly diverse *P. vivax* genome was revealed, and orthologous genes between the species were found to evolve at different rates and with different mutation patterns. More recently, whole genome analysis of *P. falciparum* and its closest relative, the chimpanzee species, *Plasmodium reichenowi*, has shown that the highest level of divergence occurs within genic sequences at four-fold synonymous sites, followed by introns and then intergenic sequences [36•]. This is similar to the pattern seen in primates and suggests that the greater level of conservation in intergenic sites might be caused by conserved regulatory sequences [36•]. Comparison of evolution rates in duplicated versus non-duplicated genes in *P. falciparum* and *P. y. yoelii* has demonstrated that duplicated genes are evolving more rapidly at the nucleotide level and have accelerated rates of intron gain and loss [37]. This supports the theory that paralogous gene family expansion and diversification is playing a major role in the evolution of malaria parasites.

Finally, identification of rapidly evolving genes in *Plasmodium* species is of considerable interest because such genes might be interacting with the host immune system. A genome-wide analysis of selective constraints in the genomes of *P. berghei* and *P. chabaudi* demonstrated that putative surface-proteins that are expressed in the vertebrae host are evolving more rapidly than those expressed in the mosquito vector [4••]. As the genome sequences of more *Plasmodium* species are completed, this comparative analysis will become more powerful and promises to become a useful method for identifying host-interacting proteins.

Conclusion

Notwithstanding the recent tremendous advances in understanding *Plasmodium* biology that have been facilitated by comparative genomics of *Plasmodium* parasites, there remains much more that can be done. To all intents, the whole genome comparisons completed to date have been with only two species of *Plasmodium*, because the three rodent malaria genomes sequenced are too closely related to provide more than a single reference point. With the completion of the genome sequence of a second human species (*P. vivax*) and a second model species genome (*P. knowlesi*), expected in Spring 2006, novel analyses such as determining 'original synteny' and dissecting the evolutionary pathway of gene expression regulation might be possible. The landscape of comparative genomics of malaria parasites is set to change significantly over the next few years.

Acknowledgements

The authors would like to acknowledge the Burroughs Wellcome Fund, National Institute of Allergy and Infectious Diseases, US Department of Defense and the Wellcome Trust for their support of the *Plasmodium* genome sequencing initiatives.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI: **The global distribution of clinical episodes of *Plasmodium falciparum* malaria**. *Nature* 2005, **434**:214-217.

2. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S *et al.*: **Genome sequence of the human malaria parasite *Plasmodium falciparum***. *Nature* 2002, **419**:498-511.

3. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL *et al.*: **Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii***. *Nature* 2002, **419**:512-519.

4. •• Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M, Florens L, Janssen CS, Pain A, Christophides GK *et al.*: **A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses**. *Science* 2005, **307**:82-86.

This study includes comparative genome sequence analysis across four species, plus a microarray and proteomic study of *P. berghei*. Integration of these datasets enabled the authors to study evolution rates of genes

expressed in different stages, and to identify a sequence involved in post-transcriptional repression.

5. Carlton J, Silva J, Hall N: **The genome of model malaria**
• **parasites, and comparative genomics**. *Curr Issues Mol Biol* 2005, **7**:23-37.
This is a particularly useful review describing the structure of *Plasmodium* genomes in detail for readers without a genomic background. Although it predates the publication of the *P. berghei* and *P. chabaudi* genomes it provides broad comparisons across the most studied species, and detailed comparisons of *P. falciparum* with *P. yoelii*

6. McCutchan TF, Dame JB, Miller LH, Barnwell J: **Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA**. *Science* 1984, **225**:808-811.

7. Janse CJ, Carlton JM, Walliker D, Waters AP: **Conserved location of genes on polymorphic chromosomes of four species of malaria parasites**. *Mol Biochem Parasitol* 1994, **68**:285-296.

8. Carlton JM, Galinski MR, Barnwell JW, Dame JB: **Karyotype and synteny among the chromosomes of all four species of human malaria parasite**. *Mol Biochem Parasitol* 1999, **101**:23-32.

9. Carlton JM, Vinkenoog R, Waters AP, Walliker D: **Gene synteny in species of *Plasmodium***. *Mol Biochem Parasitol* 1998, **93**:285-294.

10. del Portillo HA, Fernandez-Becerra C, Bowman S, Oliver K, Preuss M, Sanchez CP, Schneider NK, Villalobos JM, Rajandream MA, Harris D *et al.*: **A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax***. *Nature* 2001, **410**:839-842.

11. Janssen CS, Phillips RS, Turner CM, Barrett MP: ***Plasmodium***
•• **interspersed repeats: the major multigene superfamily of malaria parasites**. *Nucleic Acids Res* 2004, **32**:5712-5720.
This is an excellent bioinformatics study of the *pir* gene family of *Plasmodium* species. It describes the structure and diversity of the genes and, most importantly, suggests that the *rifin* family is in fact a member of the *pir* superfamily, which is found in all other sequenced species.

12. Rasti N, Wahlgren M, Chen Q: **Molecular aspects of malaria pathogenesis**. *FEMS Immunol Med Microbiol* 2004, **41**:9-26.

13. al-Khedery B, Barnwell JW, Galinski MR: **Antigenic variation in malaria: a 3′ genomic alteration associated with the expression of a *P. knowlesi* variant antigen**. *Mol Cell* 1999, **3**:131-141.

14. Winter G, Kawai S, Haeggstrom M, Kaneko O, von Euler A,
• Kawazu S, Palm D, Fernandez V, Wahlgren M: **SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes**. *J Exp Med* 2005, **201**:1853-1863.
This study identifies a new family of proteins in *P. falciparum*, members of which are directed to the surface of the red blood cell. Comparison to other species suggests several conserved motifs with other gene families such as *vir* and *SICAvar*.

15. Knuepfer E, Rug M, Klonis N, Tilley L, Cowman AF: **Trafficking**
• **of the major virulence factor to the surface of transfected *P. falciparum*-infected erythrocytes**. *Blood* 2005, **105**:4078-4087.
This study identifies and validates a motif involved in targeting *Plasmodium* proteins to the surface of the red blood cell. The authors demonstrate conservation of this motif across several *Plasmodium* species.

16. Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, Lopez-Estrano C, Haldar K: **A host-targeting signal in virulence proteins reveals a secretome in malarial infection**. *Science* 2004, **306**:1934-1937.

17. Horrocks P, Dechering K, Lanzer M: **Control of gene expression in *Plasmodium falciparum***. *Mol Biochem Parasitol* 1998, **95**:171-181.

18. Militello KT, Dodge M, Bethke L, Wirth DF: **Identification of regulatory elements in the *Plasmodium falciparum* genome**. *Mol Biochem Parasitol* 2004, **134**:75-88.

19. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ *et al.*: **Discovery of gene function by expression profiling of the malaria parasite life cycle**. *Science* 2003, **301**:1503-1508.

20. Le Roch KG, Zhou Y, Batalov S, Winzeler EA: **Monitoring the chromosome 2 intraerythrocytic transcriptome of *Plasmodium falciparum* using oligonucleotide arrays**. *Am J Trop Med Hyg* 2002, **67**:233-243.

21. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A,
• Grainger M, Yan SF, Williamson KC, Holder AA, Carucci DJ *et al.*: **Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle**. *Genome Res* 2004, **14**:2308-2318.
The authors compare mRNA and protein abundance from a number of studies to quantify the role of post-transcriptional regulation in parasite development. Much of the abundance data between mRNA and proteins are in agreement, suggesting that mRNA regulation is the predominant form of expression regulation. The authors did find that the gametocyte transcriptome was most similar to the gamete proteome, which suggested that post-transcriptional regulation was also important.

22. Bozdech Z, Zhu J, Joachimiak MP, Cohen FE, Pulliam B, DeRisi JL: **Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray**. *Genome Biol* 2003, **4**:R9.

23. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL: **The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum***. *PLoS Biol* 2003, **1**:E5.

24. Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW, Eling WM, Hall N, Waters AP, Stunnenberg HG *et al.*: **Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry**. *Nature* 2002, **419**:537-542.

25. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL *et al.*: **A proteomic view of the *Plasmodium falciparum* life cycle**. *Nature* 2002, **419**:520-526.

26. Coulson RM, Hall N, Ouzounis CA: **Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum***. *Genome Res* 2004, **14**:1548-1554.

27. Khan SM, Franke-Fayard B, Mair GR, Lasonder E, Janse CJ,
• Mann M, Waters AP: **Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology**. *Cell* 2005, **121**:675-687.
This proteomic study of male and female gametocytes in *P. berghei* identified male- and female-specific proteins. Regulatory sequences are validated using green fluorescent protein (GFP) tagging, and gene disruptions are carried out to validate putative signal-transduction proteins, enabling the identification of new signal-transduction mechanisms.

28. Hartl DL: **The origin of malaria: mixed messages from genetic**
• **diversity**. *Nat Rev Microbiol* 2004, **2**:15-22.
This review gives an excellent overview of the differing data either supporting or refuting the 'Malaria's Eve' hypothesis.

29. Volkman SK, Barry AE, Lyons EJ, Nielsen KM, Thomas SM, Choi M, Thakore SS, Day KP, Wirth DF, Hartl DL: **Recent origin of *Plasmodium falciparum* from a single progenitor**. *Science* 2001, **293**:482-484.

30. Rich SM, Licht MC, Hudson RR, Ayala FJ: **Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum***. *Proc Natl Acad Sci USA* 1998, **95**:4425-4430.

31. Mu J, Duan J, Makova KD, Joy DA, Huynh CQ, Branch OH, Li WH, Su XZ: **Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum***. *Nature* 2002, **418**:323-326.

32. Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E *et al.*: **Early origin and recent expansion of *Plasmodium falciparum***. *Science* 2003, **300**:318-321.

33. Jongwutiwes S, Putaporntip C, Iwasaki T, Ferreira MU, Kanbara H,
•• Hughes AL: **Mitochondrial genome sequences support ancient population expansion in *Plasmodium vivax***. *Mol Biol Evol* 2005, **22**:1733-1739.
This study of 106 mitochondrial genomes dates the emergence of *P. vivax* to ~30 000 years ago, suggesting that it emerged with the host species. A similar finding was reached in a study of mitochondrial genome polymorphisms in *P. falciparum* [32].

34. Mu J, Joy DA, Duan J, Huang Y, Carlton J, Walker J, Barnwell J, Beerli P, Charleston MA, Pybus OG *et al.*: **Host switch leads to emergence of *Plasmodium vivax* malaria in humans**. *Mol Biol Evol* 2005, **22**:1686-1693.

35. Feng X, Carlton JM, Joy DA, Mu J, Furuya T, Suh BB, Wang Y, Barnwell JW, Su XZ: **Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax***. *Proc Natl Acad Sci USA* 2003, **100**:8502-8507.

36. • Neafsey DE, Hartl DL, Berriman M: **Evolution of noncoding and silent coding sites in the *Plasmodium falciparum* and *P. reichenowi* genomes**. *Mol Biol Evol* 2005, **22**:1621-1626.

This study, although based on an early and fragmented assembly of *P. reichenowi*, measures the evolution rates of different silent sites in the *P. falciparum* genome, using comparative analysis. The authors demonstrate that the intergenic sites are not mutating as rapidly as intron or fourfold synonymous coding sites. They conclude that there is evidence of selection on the intergenic sites, probably as a result of regulatory sequences.

37. Castillo-Davis CI, Bedford TB, Hartl DL: **Accelerated rates of intron gain/loss and protein evolution in duplicate genes in human and mouse malaria parasites**. *Mol Biol Evol* 2004, **21**:1422-1427.