ELSEVIER

Vivax series

The *Plasmodium vivax* genome sequencing project

Jane Carlton

The Institute for Genomic Research 9712 Medical Center Drive Rockville, MD 20850, USA.

With the successful completion of the project to sequence the *Plasmodium falciparum* genome, researchers are now turning their attention to other malaria parasite species. Here, an update on the *Plasmodium vivax* genome sequencing project is presented, as part of the *Trends in Parasitology* series of reviews expanding on various aspects of *P. vivax* research.

Of the four species of malaria parasite known to infect humans, *Plasmodium vivax* has achieved the widest global distribution, and is responsible for >50% of malaria cases in Central and South America, Asia and the Indian subcontinent [1,2]. Over one billion people are at risk from infection, with 70 million to 80 million cases reported annually. Historically, *P. vivax* malaria or 'benign tertian' malaria is known to have been prevalent in temperate regions, such as parts of continental USA [3] and the south of England [4], through the latter part of the past millennium. Sporadic cases of local transmission in some temperate areas still occur today [3–6]. Despite its widespread prevalence, studies concerning the parasite's biology and disease manifestation have been limited, in part due to the attention directed towards understanding the biology of the greater menace, *Plasmodium falciparum*, but also because of the lack of a continuous *in vitro* culture system for the parasite.

A landmark in the fight against malaria was passed recently with the publication of the complete genome sequence of *P. falciparum* [7], in combination with that of the mosquito vector *Anopheles gambiae* [8]. Concurrently, the PARTIAL SHOTGUN COVERAGE (see Glossary) of a model rodent malaria parasite *Plasmodium yoelii yoelii* and comparative analysis with the *P. falciparum* sequence [9] hinted at the wealth of possibilities that could result from comparative analyses of multiple malaria genomes. To this end, several initiatives to provide sequence data from other species of malaria parasite are currently underway (Table 1). The project to sequence the complete genome of *P. vivax* is in progress at The Institute for Genomic Research (TIGR), Rockville, MD, USA, with the goal of producing a FINISHED SEQUENCE at least as good as the sequence of *P. falciparum*. Surplus funds from the US Department of Defense and the National Institute of Allergy and Infectious Diseases, National Institutes of Health, which supported part of the *P. falciparum* genome sequencing project, are being used to finance the project.

Genome characteristics

What is known concerning the characteristics of the *P. vivax* nuclear genome? Previous studies have identified 12–14 linear chromosomes that range in size from 1.2 Mb to 3.5 Mb [10]. Size differences between homologous chromosomes have been detected in field isolates of *P. vivax* [11], a phenomenon first described in laboratory clones and isolates of *P. falciparum*, and due to recombination between chromatids during meiosis and copy number variations in subtelomeric repeats during mitosis (reviewed in Ref. [12]). Initial estimates put the genome

Glossary

Contig: Contiguous DNA sequence produced from joining overlapping raw sequence reads

Expressed sequence tag (EST): Generated by sequencing one end of a recombinant clone from a complementary DNA library. ESTs are single-pass reads and are therefore prone to contain sequence errors.

Finished sequence: Complete sequence of a genome with no gaps and an accuracy of >99.9%.

Full shotgun coverage: Genome coverage in a random raw sequence that is required to produce finished sequence – usually eightfold to tenfold.

Genome coverage: Average number of times a nucleotide is represented by a high-quality base in a random raw sequence.

Genome survey sequence (GSS): Generated by sequencing one end of a recombinant clone from a genomic DNA library. The genomic DNA library can in some instances be enriched for the presence of coding regions, for example, through use of mung bean nuclease digestion of genomic DNA before cloning.

Homolog: A general term used to describe genes containing significant sequence similarity to each other.

Open reading frame (ORF): Stretches of codons in the same reading frame uninterrupted by stop codons and calculated from a six-frame translation of DNA sequence.

Ortholog: Homologous genes generated by speciation (i.e. related to each other by vertical descent)

Paired read: Sequence reads determined from both ends of a cloned insert in a recombinant clone

Partial shotgun coverage: Typically fourfold to sixfold random coverage of a genome which produces sequence data of sufficient quality to enable gene identification, but which is not sufficient to produce a finished genome sequence.

Raw sequence: Unassembled sequence reads produced from sequencing of inserts from individual recombinant clones of a genomic DNA library.

Scaffold: A group of ordered and orientated contigs known to be physically linked to each other by paired read information.

Sequence tagged site (STS): A unique 100–1000 bp sequence used to identify clones to generate overlapping inserts for a physical map.

Synteny: Strictly, this refers to the presence of two or more genes on the same chromosome in the same species. However, it is frequently used to mean the conservation of gene location between species (i.e. the presence of orthologous genes that are syntenic in one species and also located on the same chromosome in a second species, without regard to their order).

Corresponding author: Jane Carlton (carlton@tigr.org).

http://parasites.trends.com 1471-4922/03/$ - see front matter doi:10.1016/S1471-4922(03)00066-7

Table 1. Current *Plasmodium* genome sequencing projects[a]

Species	Sequencing center	Funding agency	Project type	Status	Ref. or website
P. falciparum	TIGR, SG, ST	NIAID, DoD, BWF, WT	Finished sequence	Complete	[7], http://plasmodb.org
P. vivax	TIGR	NIAID, DoD	Finished sequence	5X	http://www.tigr.org/tdb/e2k1/pva1/
P. y. yoelii	TIGR	DoD	Partial	5X	[9], http://www.tigr.org/tdb/e2k1/pya1/
P. knowlesi	SG	WT	Partial	5X	http://www.sanger.ac.uk/Projects/P_knowlesi/
P. c. chabaudi	SG	WT	Partial	3X	http://www.sanger.ac.uk/Projects/P_chabaudi/
P. berghei	SG	WT	Partial	3X	http://www.sanger.ac.uk/Projects/P_berghei/
P. reichenowi	SG	WT	Partial	LC	http://www.sanger.ac.uk/Projects/ (under construction)
P. gallinaceum	SG	WT	Partial	LC	http://www.sanger.ac.uk/Projects/ (under construction)
P. falciparum	SG	WT	Partial	LC	http://www.sanger.ac.uk/Projects/ (under construction)

[a]Abbreviations: 3X, threefold coverage; 5X, fivefold coverage; BWF, Burroughs Wellcome Fund; DoD, US Department of Defense; NIAID, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA; LC, library construction; SG, The Sanger Center, Cambridge, UK; ST, Stanford University, CA, USA; TIGR, The Institute for Genomic Research, Rockville, MD, USA; WT, Wellcome Trust.

size at 35–40 Mb based on pulsed-field gel electrophoresis (PFGE) and separation of *P. vivax* chromosomes, but this is likely to be an overestimate based on current genome data for other *Plasmodium* spp. [9] and the identification of large contaminating DNA molecules from the host [13]. Analysis of several *Plasmodium* genomes using ultra-centrifugation through Hoechst dye caesium chloride gradients [14] established the *P. vivax* nuclear genome as having an isochore structure (i.e. areas of the genome that exhibit a marked bias towards AT nucleotide composition). Such isochores represent a higher order structuring of the genome, which are of unknown significance. In the simian parasite *Plasmodium cynomolgi*, which has a similar isochore organization [14], protein-coding genes have been located to GC-rich isochores, whereas chromosome ends containing the telomeres were located in AT-rich isochores [15]. Evidence to support a similar organization of the *P. vivax* genome comes from sequence analysis of two regions of the genome, an internal chromosomal region of 200 kb [16] and a subtelomeric region of 150 kb [17]. The GC content of the 200 kb sequence increases from 25% at the telomere proximal end to 48% for the remaining 160 kb, whereas the 150 kb subtelomeric sequence has an overall GC composition of ~21% that varies only by 5–10% in coding versus non-coding regions. Thus, the chromosome ends of *P. vivax* parasites appear to consist of AT-rich sequences, whereas telomere-distal regions consist of GC-rich sequences. By contrast, *P. falciparum* has a uniform GC distribution of 23% throughout the genome, with the exception of 2–3 kb regions of >97% AT content on each chromosome which are thought to contain the centromeres [18]. The evolution and function of isochores remains in dispute (reviewed in Ref. [19]), but it is tempting to speculate that the isochores could encode genes that mediate phenomena specific to the pathophysiology of vivax malaria [14].

A large-scale sampling of ~20% of the *P. vivax* genome [11 000 GENOME SURVEY SEQUENCES (GSSs)], through partial sequencing of random clones from mung bean nuclease-digested genomic DNA libraries, has given some insight into the coding potential of the parasite [20]. For example, ORTHOLOGS of previously identified *Plasmodium* genes were identified, and similar numbers of proteins were found to be common between the *P. vivax* proteome and proteomes of *P. falciparum* and *Plasmodium berghei*, a rodent malaria species. Functional analysis of the proteins using the gene ontology (GO) classification system [21] also suggested that similar proportions of proteins exist in each of the GO categories between the three species. A large multi-gene family of 600–1 000 variant proteins termed the *vir* family has recently been identified in *P. vivax* [17], HOMOLOGS of which have been found in several rodent malaria species [9,22]. It seems probable that these proteins, which could play a role in antigenic variation of the parasite, will be found to make up a large proportion of the *P. vivax* proteome.

Sequencing the *P. vivax* genome

The Salvador I strain of *P. vivax*, isolated from a naturally acquired infection of a patient from the La Paz region of El Salvador [23], has been chosen for sequencing. This strain has been passaged through human volunteers and *Aotus* (owl) and *Saimiri* (squirrel) monkeys by mosquito and blood infection [23]. It has been the subject of drug susceptibility and relapse activity studies (e.g. Ref. [24]), and has been used to test the immunogenicity and protective efficacy of recombinant antigen constructs (e.g. Refs [25,26]). Salvador I chromosomes can be separated by PFGE for karyotype and physical mapping studies [10], and >7000 GSSs have been generated for this strain [20]. Thus, similar to the 3D7 clone of *P. falciparum* [27], it is often regarded as the standard reference strain for *P. vivax*. Genomic DNA for the project was provided by John Barnwell at the Centers for Disease Control, Atlanta, GA, USA. Briefly, the parasite is grown in splenectomized *Saimiri boliviensis boliviensis* monkeys, infected blood collected, and filtration techniques used to deplete the blood of monkey leukocytes. Parasite DNA is then prepared using conventional methods.

A whole genome shotgun (WGS) strategy, applied successfully to sequence the first eukaryotic genome [28], is being used to sequence the *P. vivax* genome. This technique follows a series of standard steps. First, genomic DNA of an organism is sheared into pieces of varying length and cloned into plasmid vectors. Random clones are selected, DNA extracted, and the inserts are sequenced at both 5′ and 3′ ends, producing so-called PAIRED READS. Once a sufficient number of RAW SEQUENCE reads have

Table 2. Current status of the *Plasmodium vivax* whole genome shotgun project[a]

Genome size (Mb)	Total no. of sequences	Genome coverage	No. of contigs	Mean contig length (kb)	No. of scaffolds	Mean scaffold span (kb)	Total scaffold span (Mb)	GC content (%)	No. of *Saimiri* sequences[b]	No. of BLASTX hits[c]
~24	194 000	5X	5126	4.4	479	48.5	23.2	40	44	1906

[a]All preliminary sequence data have been released and is accessed at http://www.tigr.org/tdb/e2k1/pva1/ with reference to TIGR's data release policy. Abbreviations: 5X, fivefold; TIGR, The Institute for Genomic Research, Rockville, MD, USA.
[b]Identified through sequence similarity analysis, with significance cut-off at >90% identity over >100 bases
[c]Identified through sequence similarity analysis, with cut-off values set at expect >1e − 15 and score >100 for significance.

been obtained, the sequences are assembled by use of a computer algorithm into contiguous DNA sequences (CONTIGS). The WGS approach is slightly different from that used to sequence the genome of *P. falciparum*, where a chromosome-by-chromosome strategy was chosen: each of the 14 chromosomes being purified by PFGE and sequenced individually using shotgun methodology. In addition, physical map data comprising SEQUENCE TAGGED SITES (STSs), microsatellite markers and optical restriction sites were used to map groups of linked contigs to individual chromosomes (reviewed in Ref. [29]).

Mass sequencing of *P. vivax* genomic libraries started in the spring of 2002, and details concerning the current status of the fivefold GENOME COVERAGE are shown in Table 2. All preliminary sequence data have been released and can be accessed at http://www.tigr.org/tdb/e2k1/pva1/ with reference to TIGR's data release policy. At FULL SHOTGUN COVERAGE, work will begin to close the gaps between the contigs by identifying clones that span sequence gaps, and by using PCR amplification and transposon-insertion methods for physical gaps. The length of this closure period is the most difficult to predict for any sequencing project because it depends upon individual genome characteristics such as the number and size of repeats, and random nature of the libraries used for sequencing. The closure process took many years for several *P. falciparum* chromosomes [18,30,31], and is still continuing for several others. A notable difference between the *P. falciparum* and *P. vivax* projects is the ease to which sequencing has progressed for *P. vivax*. This is in large part due to the more-moderate AT content of the *P. vivax* genome, which has enabled the construction of large-insert libraries that are invaluable for the linking of contigs into SCAFFOLDS, and also because of the fewer number of poly(A) and poly(T) regions, which proved to be extremely difficult to sequence in *P. falciparum*. In addition, considerable advances in high-throughput robotics and sequencing technology have been made since the *P. falciparum* genome project began [29].

A concern with certain genome sequencing projects, such as those of parasitic or symbiotic organisms, is the presence of contaminating genomic DNA of the host. As an example, 10% of the sequence data for the *P. y. yoelii* rodent malaria genome was found to be of mouse origin [9], a consequence of growing the parasite in its laboratory host, *Mus musculus*. Several filtration and magnetic methods exist to purify parasite from host material [32–34], but because the genome size of eukaryotic hosts in particular can be many times greater than that of the parasite, the probability of significant host contamination of parasite DNA is high. In some cases, such as the *P. y. yoelii–M. musculus* model system, a significant difference in genome composition between the two genomes (23% GC for *P. y. yoelii* versus 42% GC for *M. musculus*) and the availability of the complete mouse genome sequence means that bioinformatic analysis is sufficient to weed out contaminating sequences. In the case of the *P. vivax* project, however, the genome composition of the parasite and its laboratory host are very similar (40% GC for *P. vivax* versus 49% GC for coding regions of *S. b. boliviensis*), and little information is available on the monkey sequence. As a preliminary test of the degree of monkey contamination among the *P. vivax* sequence data, *P. vivax* contigs have been searched against: (1) all available *Saimiri* sequences; and (2) all human transcripts. Less than 1% of the sequences were identified as monkey contaminants (Table 2), indicating successful elimination of monkey leukocytes during preparation of the *P. vivax* genomic DNA.

Comparative malaria genomics

At fivefold coverage of the *P. vivax* genome, what is the sequence data starting to reveal? A preliminary comparison of the *P. vivax* OPEN READING FRAMES (ORFs) against public protein databases reveals that 37% of putative genes have homology to identified proteins, and 78% of those are to *P. falciparum* proteins (Table 2). The remaining 63% of proteins have no apparent homology to any known protein, a figure similar to that found with analysis of the *P. falciparum* proteome [7]. Once the random phase of the genome sequencing is complete, it will be possible to identify all orthologs between the two species, using 'reciprocal best match' sequence similarity searching in combination with gene SYNTENY. Identification in one species of the ortholog of a candidate gene in a second species is important for cross-species comparison of gene function and comparison of molecular mechanisms associated with a common phenotype. For example, identification of the exact ortholog of the *P. falciparum* chloroquine-resistance gene, *Pfcrt*, in *P. vivax* enabled a thorough investigation into whether the molecular mechanism of resistance to chloroquine (CQ) in *P. vivax* parasites is different from that in *P. falciparum* parasites [35]. Identification of orthologous genes between *P. falciparum* and *P. vivax* will also bring to light unique genes in the *P. vivax* genome. *Plasmodium vivax* has many distinct biological features, such as the ability to form a dormant phase (the hypnozoite) in the liver, and a preference for invasion and replication within reticulocytes, which are not shared with *P. falciparum*. Although stage-specific gene expression studies are limited in *P. vivax* because of the problems of obtaining biological material from distinct life-stages [for example, there are no *P. vivax* EXPRESSED SEQUENCE TAGS (ESTs)], identification of unique *P. vivax*

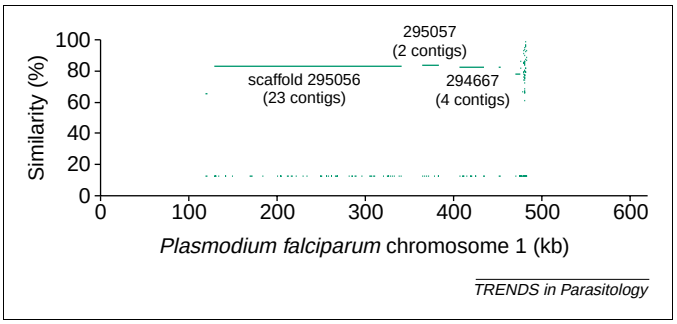

Fig. 1. Alignment of several *Plasmodium vivax* scaffolds with *Plasmodium falciparum* chromosome 1. The 643 Mb chromosome is represented along the x-axis, with percentage similarity plotted along the y-axis. Maximal unique matches (MUMs) of a minimum of five amino acids, which have been clustered and extended, are plotted immediately above the *x*-axis, whereas scaffolds that contain these MUMs are plotted according to their percent similarity. Note the presence of a repeat region on *P. falciparum* chromosome 1 that has resulted in a 'pile-up' of short *P. vivax* scaffolds at this point.

genes could aid in addressing questions pertinent to *P. vivax* biology. It is now apparent from several gene-mapping studies [10,36,37], targeted sequencing projects [16,35,38] and through the construction of a genome-wide synteny map [9] that conservation of gene synteny exists at a high level within the *Plasmodium* genus. Not only are gene content and order highly conserved across hundreds of kilobases of DNA, but the fine structure of genes is also highly conserved [9,38]. For example, in a 200 kb segment of one *P. vivax* chromosome, 36 contiguous genes are in the same order and orientation, and have the same structure as the genes on the orthologous segment of *P. falciparum* chromosome 3 [16]. The degree of synteny between *Plasmodium* spp. decreases as the phylogenetic distance between them increases. A preliminary comparison of the *P. vivax* scaffolds with *P. falciparum* chromosomes is starting to reveal the degree of overall synteny between these two genomes. Using the software MUMmer2, which identifies clusters of maximal unique matches (MUMs) from six-frame translations of two input sequences, and is therefore independent of gene prediction data [39], a tiling path of *P. vivax* scaffolds along each *P. falciparum* chromosome is being created. Figure 1 shows an example of three large scaffolds comprising 23, two and four contigs, respectively, which map to the internal region of *P. falciparum* chromosome 1. The telomeric and subtelomeric regions of the chromosome are devoid of matches, in accord with previous findings [9,36] that *Plasmodium* chromosomes consist of a conserved body flanked at either end with subtelomeric regions containing species-specific genes, many of which are involved in evading the host's immune system. The construction of a synteny map between *P. falciparum* and *P. vivax* will enable: (1) evolutionary mapping of the *Plasmodium* genus, through identification of the chromosome 'splitting' and 'joining' events that led to the creation of different species; (2) refinement of gene predictions through simultaneous annotation of both genomes; and (3) identification of conserved non-coding motifs (phylogenetic footprints) between the two species which could play a role in gene regulation.

Population genetics and evolutionary history

Understanding the population structure of the malaria parasite is of great practical relevance to developing control measures to contain the disease. For example, predicting the flow of drug resistance mutations through populations of parasites has consequences for the type and range of drug therapies made available from healthcare facilities in endemic countries. The genetic structure of the *P. falciparum* parasite has been studied in much greater detail than that of *P. vivax* [40] primarily due to the lack of genome data for *P. vivax*. Numerous microsatellite markers [41], restriction fragment length polymorphism (RFLP) markers [42] and single nucleotide polymorphism (SNP) markers [43] have been described and used in studies of *P. falciparum*. A few genome markers, mostly orthologs of previously identified *P. falciparum* antigen genes, have been used for population studies of *P. vivax* [44,45]. The lack of genetic markers for the *P. vivax* genome has severely hampered an in-depth analysis of the population structure and evolutionary history of the parasite, and prevented efforts to map determinants contributing to important parasite phenotypes such as antimalarial drug resistance [46,47] and patterns of relapse [48]. The complete sequence of the *P. vivax* genome promises to provide a wealth of data from which candidate markers for polymorphism studies can be identified. Preliminary studies from previously published sequences are already uncovering microsatellite markers for genotyping isolates from Papua New Guinea, Indonesia and Guyana (Peter Zimmerman, pers. commun.), although the frequency of microsatellites in the *P. vivax* genome appear to be much less than in the *P. falciparum* genome. A dense SNP map over a 100-kb region of one *P. vivax* chromosome of several strains from geographically diverse areas also promises to provide insight into the diversity and evolutionary history of the parasite (Xin-Zhuan Su, pers. commun.). In the light of these studies and the speed of the *P. vivax* genome sequencing effort, the future looks promising for diversity studies of this neglected species.

The future of *P. vivax* research

The genome sequence of *P. falciparum* has provided scientists with the means to jump-start biological experimentation into varied aspects of malaria research. Many reports have been published in which use of the sequence was acknowledged, most notably in the area of novel antimalarial development [29]. Can we expect the same to happen with release of the *P. vivax* genome data? An essential difference between research in the two species remains in the limited availability of *P. vivax* biological materials such as DNA, RNA, and whole parasite preparations of different life-stages. Few malaria laboratories are equipped with the insectaries and nonhuman primates necessary for completion of the life cycle, and fewer still can provide a constant supply of material to other laboratories. At a recent meeting convened by the Multilateral Initiative on Malaria [49], steps recommended to facilitate *P. vivax* research included the development of standardized *P. vivax* reagents that could be maintained and distributed through the Malaria Repository, MR4 [50], and a concerted effort to develop a continuous *in vitro* culture

system for the growth of blood stages of the parasite. With the expected completion of the *P. vivax* genome sequencing project in 2004, it is vital that researchers have access to reagents so that the *P. vivax* genome sequence can be as much of a valuable resource as the *P. falciparum* genome sequence has proven to be.

Acknowledgements

The *P. vivax* genome sequencing project is funded by the US Department of Defense through cooperative agreement with the US Army Medical Research and Materiel Command and by the Naval Medical Research Center, and by the National Institute of Allergy and Infectious Diseases, National Institutes of Health. Thanks to Adam Phillippy for MUMmer2 analysis, and Malcolm Gardner and Steven Sullivan for critical reading of the manuscript.

References

1 Carter, R. and Mendis, K.N. (2002) Evolutionary and historical aspects of the burden of malaria. *Clin. Microbiol. Rev.* 15, 564–594
2 Mendis, K. *et al.* (2001) The neglected burden of *Plasmodium vivax* malaria. *Am. J. Trop. Med. Hyg.* 64, 97–106
3 Zucker, J.R. (1996) Changing patterns of autochthonous malaria transmission in the United States: a review of recent outbreaks. *Emerg. Infect. Dis.* 2, 37–43
4 Reiter, P. (2000) From Shakespeare to Defoe: malaria in England in the Little Ice Age. *Emerg. Infect. Dis.* 6, 1–11
5 Local transmission of *Plasmodium vivax* malaria–Virginia, 2002. *Morb. Mortal. Wkly Rep.* 51, 921–923
6 MacArthur, J.R. *et al.* (2001) Probable locally acquired mosquito-transmitted malaria in Georgia. *1999. Clin. Infect. Dis.* 32, E124–E128
7 Gardner, M.J. *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511
8 Holt, R.A. *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129–149
9 Carlton, J.M. *et al.* (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 419, 512–519
10 Carlton, J.M-R. *et al.* (1999) Karyotype and synteny among the chromosomes of all four species of human malaria parasite. *Mol. Biochem. Parasitol.* 101, 23–32
11 Langsley, G. *et al.* (1988) *Plasmodium vivax*: karyotype polymorphism of field isolates. *Exp. Parasitol.* 67, 301–306
12 Janse, C.J. (1993) Chromosome size polymorphism and DNA rearrangements in *Plasmodium*. *Parasitol. Today* 9, 19–22
13 Deitsch, K.W. *et al.* (2001) Host sequences in *Plasmodium falciparum* and *Plasmodium vivax* genomic DNA: horizontal transfer or contamination artifact? *FEBS Lett.* 491, 164–165
14 McCutchan, T.F. *et al.* (1984) Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. *Science* 225, 808–811
15 McCutchan, T.F. *et al.* (1988) The genome of *Plasmodium cynomolgi* is partitioned into separable domains which appear to differ in sequence stability. *Nucleic Acids Res.* 16, 4499–4510
16 Tchavtchitch, M. *et al.* (2001) The sequence of a 200 kb portion of a *Plasmodium vivax* chromosome reveals a high degree of conservation with *Plasmodium falciparum* chromosome 3. *Mol. Biochem. Parasitol.* 118, 211–222
17 del Portillo, H.A. *et al.* (2001) A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* 410, 839–842
18 Hall, N. *et al.* (2002) Sequence of *Plasmodium falciparum* chromosomes 1, 3-9 and 13. *Nature* 419, 527–531
19 Eyre-Walker, A. and Hurst, L.D. (2001) The evolution of isochores. *Nat. Rev. Genet.* 2, 549–555
20 Carlton, J.M. *et al.* (2001) Profiling the malaria genome: a gene survey of three species of malaria parasite with comparison to other apicomplexan species. *Mol. Biochem. Parasitol.* 118, 201–210
21 Consortium, T.G.O. (2001) Creating the gene ontology resource: design and implementation. *Genome Res.* 11, 1425–1433
22 Janssen, C.S. *et al.* (2001) Gene discovery in *Plasmodium chabaudi* by genome survey sequencing. *Mol. Biochem. Parasitol.* 113, 251–260
23 Collins, W.E. *et al.* (1972) Transmission of four central american strains of *Plasmodium vivax* from monkey to man. *J. Parasitol.* 58, 332–335
24 Contacos, P.G. *et al.* (1972) Studies on the characterization of *Plasmodium vivax* strains from central America. *Am. J. Trop. Med. Hyg.* 21, 707–712
25 Collins, W.E. *et al.* (1997) Protective immunity induced in squirrel monkeys with a multiple antigen construct against the circumsporozoite protein of *Plasmodium vivax*. *Am. J. Trop. Med. Hyg.* 56, 200–210
26 Yang, C. *et al.* (1997) Induction of protective antibodies in *Saimiri* monkeys by immunization with a multiple antigen construct (MAC) containing the *Plasmodium vivax* circumsporozoite protein repeat region and a universal T helper epitope of tetanus toxin. *Vaccine* 15, 377–386
27 Walliker, D. *et al.* (1987) Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* 236, 1661–1666
28 Myers, E.W. *et al.* (2000) A whole-genome assembly of *Drosophila*. *Science* 287, 2196–2204
29 Gardner, M.J. (2001) A status report on the sequencing and annotation of the *P. falciparum* genome. *Mol. Biochem. Parasitol.* 118, 133–138
30 Hyman, R.W. *et al.* (2002) Sequence of *Plasmodium falciparum* chromosome 12. *Nature* 419, 534–537
31 Gardner, M.J. *et al.* (2002) Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature* 419, 531–534
32 Janse, C.J. *et al.* (1994) Removal of leucocytes from *Plasmodium vivax*-infected blood. *Ann. Trop. Med. Parasitol.* 88, 213–216
33 Carlton, J.M.R. *et al.* (2001) Biomagnetic separation of contaminating host leukocytes from *Plasmodium*-infected erythrocytes. *Exp. Parasitol.* 97, 111–114
34 Homewood, C.A. (1976) A comparison of methods used for the removal of white cells from malaria-infected blood. *Ann. Trop. Med. Parasitol.* 70, 249–251
35 Nomura, T. *et al.* (2001) Evidence for different mechanisms of chloroquine resistance in two *Plasmodium* species that cause human malaria. *J. Infect. Dis.* 183, 1653–1661
36 Janse, C.J. *et al.* (1994) Conserved location of genes on polymorphic chromosomes of four species of malaria parasites. *Mol. Biochem. Parasitol.* 68, 285–296
37 Carlton, J.M.R. *et al.* (1998) Gene synteny in species of *Plasmodium*. *Mol. Biochem. Parasitol.* 93, 285–294
38 van Lin, L.H. *et al.* (2001) Interspecies conservation of gene order and intron-exon structure in a genomic locus of high gene density and complexity in *Plasmodium*. *Nucleic Acids Res.* 29, 2059–2068
39 Delcher, A.L. *et al.* (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 30, 2478–2483
40 Walliker, D. *et al.* (1998) The genetic structure of malaria parasite populations. In *Malaria: Parasite Biology, Pathogenesis and Protection* (Sherman, I.W. *et al.*, eds), pp. 235–252, ASM Press
41 Su, X-Z. and Wellems, T.E. (1996) Towards a high resolution *Plasmodium falciparum* linkage map: polymorphic markers from hundreds of simple sequence repeats. *Genomics* 33, 430–444
42 Walker-Jonah, A. *et al.* (1992) An RFLP map of the *Plasmodium falciparum* genome, recombination rates and favored linkage groups in a genetic cross. *Mol. Biochem. Parasitol.* 51, 313–320
43 Mu, J. *et al.* (2002) Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. *Nature* 418, 323–326
44 Figtree, M. *et al.* (2000) *Plasmodium vivax* synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes. *Mol. Biochem. Parasitol.* 108, 53–66
45 Tsuboi, T. *et al.* (1998) Sequence polymorphism in two novel *Plasmodium vivax* ookinete surface proteins, Pvs25 and Pvs28, that are malaria transmission-blocking vaccine candidates. *Mol. Med.* 4, 772–782
46 Collins, W.E. and Jeffery, G.M. (1996) Primaquine resistance in *Plasmodium vivax*. *Am. J. Trop. Med. Hyg.* 55, 243–249
47 Rieckmann, K.H. *et al.* (1989) *Plasmodium vivax* resistance to chloroquine. *Lancet* 2, 1183–1184
48 Krotoski, W.A. (1989) The hypnozoite and malarial relapse. *Prog. Clin. Parasitol.* 1, 1–19
49 Sina, B. (2002) Focus on *Plasmodium vivax*. *Trends Parasitol.* 18, 287–289
50 Adams, J.H. *et al.* (2000) Malaria research and reference reagent resource center. *Parasitol. Today* 16, 89