Malaria Genomics and Post-Genomics in Leiden

Andy Waters, Taco Kooij, Marianna Karras, Shahid Khan

Date: December 2002
Updated:

- General Introduction

- Genome structure of *P. berghei* and comparative genomics (Genome databases and bioinformatics)
    - Introduction
    - Genome structure of *P. berghei*
    - GST and EST library at the University of Florida
    - *P. berghei* sequencing project, Sanger Institute
    - *P. yoelii* sequencing project, TIGR
    - Comparative genomics in Leiden
    - Genome conservation and synteny between rodent *Plasmodium* species
    - Genome conservation and synteny between human, simian and rodent *Plasmodium* species
    - The 48/45 multigene family: an extended and well-conserved gene family in *Plasmodium*
    - The B9-locus: an example of conservation of a complex and gene dense locus
    - Genes encoding variant (erythrocyte) surface antigens (VSA's): an example of species specificity
    - References

-Transcription profiling in *P. berghei* (microarrays, transcriptomics)
    - Introduction
    - *P. berghei* microarrays made in Leiden
    - Other *P. berghei* microarrays
    - Microarray hybridisations
    - Micoarray data
    - Comparative microarray analysis
    - References

- Proteome analysis of *P. berghei* (proteomics)
    - Introduction
    - Life-cycle stages and subcellular fractions analysed
    - Data to be released
    - References

- Background information on the genome and post-genome projects
    - *P. berghei* genomic DNA library (mungbean-nuclease digested)
    - *P. berghei* cDNA library

General Introduction

Genomes of malaria parasites

The publication of the nearly complete annotated genome sequence of the human malaria parasite *P. falciparum* combined with substantial amounts of assembled genome sequence from the rodent malaria parasite, *P. yoelii* (5.6x coverage) marked the beginning of what is being popularly termed the post genomic "era" in malaria research.

In addition, these datasets are supported by additional publicly accessible and searchable databases of substantial amounts of (pre-publication) genome sequence information from a further two rodent malaria parasites, *P. chabaudi* (2x coverage) and *P. berghei* (3x coverage), as well as the primate malaria parasite, *P. knowlesi* (5x coverage).

All of these unpublished projects are sufficiently well advanced that we can expect that >95% of the genes of the parasite genomes will be identifiable. Much of this data has been integrated in another public database PlasmoDB that uses the assembled *P. falciparum* genome and its 14 chromosomes as its framework.

In the following sections, we will describe the continuing genome work that the Leiden Malaria Group is involved in concerning the fuller description of the genome of rodent malaria parasite *P. berghei* in particular and rodent malaria genomes in general. Furthermore, we will outline the post-genomic initiatives active in the laboratory that the well-developed *P. berghei* model is particularly suited to facilitate.

We have the genomes and the annotation so what can we do?

Immediate exploitation of the existing genome-data is possible since basic annotation will predict proteins that are expected to be expressed upon the surface of the parasite and therefore simplistically might be considered to be candidate immunogens to produce protective immune responses (possible vaccine components).

In addition, annotation of a complete genome reveals the metabolic processes of an organism pinpointing parasite specific processes that might be exploited for drug design. For example, the discovery of the apicoplast organelle in malaria parasites and the requirement of imported proteins to have a specialised bipartite leader sequence permitted the bioinformatic identification of the repertoire of genes encoding apicoplast proteins and the realisation that this organelle can serve as an organisation centre for metabolic pathways associated with, for example, fatty acid biosynthesis (Gardner *et al.*, 2002).

We have the genomes and the annotation so why do more?

The comprehensive annotated genome assemblies for *Plasmodium* immediately permitted the initiation of comparative genomics. The initial genome-wide comparison between the human parasite *P. falciparum* and the rodent parasite *P. yoelii* (Carlton *et al.*, 2002) confirmed the conclusions of earlier studies and demonstrated long range synteny (extent of relative positional conservation of gene content) between the two genomes. The gene content of the sub-telomeric regions appeared to be species-specific encoding proteins associated with antigenic variation.

More detailed syntenic comparisons are worthwhile, as it has already been demonstrated for *Plasmodium* that this can lead to gene discovery and insight into the intron/exon structure of split genes (Lin *et al.*, 2000). Therefore, an improved genome annotation can result from such studies.

What should we do?

In principle, the release of an annotated complete genome sequence merely facilitates the next phase of experimentation, which is to confirm, correct and expand the annotation, and ascribe function to genes. Since the "core" *Plasmodium* genome (excluding gene families distributed at the telomeres) is highly conserved (Carlton *et al*, 2002), studies on stage specific gene expression at the transcriptional and

translational levels performed on rodent malaria parasites will be informative for human parasites. All model species of malaria have their particular strengths and given that *P. berghei* can allow access to all points in the parasite life cycle then it is a most suitable substrate for transcriptome (stage specific mRNA) and proteome (stage specific protein) analysis.

Post genome analyses are likely to consist of two phases, global followed by the more specific. At present, the imperative is to identify the stage specificity of gene expression e*n masse* in terms of the mRNA produced and the protein products. Although significant progress has been reported on the latter (Florens *et al.*, 2002; Lasonder *et al.*, 2002) and can be anticipated for the transcriptome, much more different and complete datasets are required. Some global strategies may require sub-cellular fractionation in order to localise protein repertoires as well as improve sensitivity of detection. Furthermore, the global transcriptional response to an altered environment (*e.g.* drug exposure, immune challenge) will be of continued interest. However, once the when of *Plasmodium* gene expression is known it is most likely that the global analyses will be superceded by studies similar to those in the pre-genomic era on single or small numbers of genes seeking to assign or confirm predicted function. The difference will be that the investigator always has recourse to global analyses and the background information they have provided. The requirement for individual gene analysis and functional assignment guarantees the genetic manipulation of malaria parasites will continue to play a pivotal role in investigations of parasite vaccine validation, functional analysis, cell and developmental biology and the rodent parasite *P. berghei* will continue to offer much in this work area.

The following sections will describe the studies in progress in this laboratory (and those of our collaborators) concerning the post-genomic analysis of genome sequence, structure and conservation ("comparative genomics") as well as transcriptome and proteome analyses. These pages will be updated as progress is made and the site will in the future serve as a portal to published (raw) data from this laboratory and derived from global analyses that can then be reanalysed according to user requirements.

References

M.J. Gardner, *et al.*, "Genome sequence of the human malaria parasite Plasmodium falciparum", Nature, 419, 6905, p498-511, (2002).

J.M. Carlton, *et al.*, "Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*", Nature, 419, 6905, p512-519, (2002).

L.H. van Lin, *et al.*, "The conserved genome organisation of non-*falciparum* malaria species: the need to know more", Int. J. Parasitol., 30, 4, p357-370 (2000).

L. Florens, *et al.*, "A proteomic view of the *Plasmodium falciparum* life cycle", Nature, 419, 6906, p520-526 (2002).

E. Lasonder *et al.*, "Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry", Nature, 419, 6906, p537-542 (2002).

Genome structure of *P. berghei* and comparative genomics (Genome databases and bioinformatics)

Introduction

The rodent malaria parasite *P. berghei* is widely used as a research model for investigation of the developmental biology of malaria parasites and to identify and analyse targets for vaccines and drugs (see Research Model). Availability of the genome sequence greatly facilitates these kinds of investigations and enhances the value of this model in malaria research.

The rodent parasite *P. yoelii* is closely related to *P. berghei* and for this species substantial amounts of assembled genome sequence (5.6x coverage) exist. For *P. berghei*, a sequencing project has been performed at the Sanger Institute (3x coverage). See below for more information on the different sequencing projects and links to the relevant sites.

Our group has been involved in studies on the genome structure of *P. berghei* and comparative genomics in *Plasmodium*. These studies show a high level of conservation of genome organisation (gene location and order) between human and rodent malaria parasites. This level of conservation

allows the discovery of orthologues and paralogues of less well-conserved genes in the different *Plasmodium* species and helps addressing questions of conservation, evolution and structure of multi-gene families. In addition, comparative genomics helps to elucidate gene content of complex loci and improves the annotation of the genome.

Genome structure of *P. berghei* (see also Research Model)

By comparison of the size chromosomes, separated by of pulsed field gel electrophoresis, the genome size has been estimated at 23-24 Mb. This genome size of *P. berghei* is comparable to the genome size of 22,9 Mb of *P. falciparum* and 23.1 Mb of *P. yoelii*. The genome is organised into 14 chromosomes in the size range of 0.61-3.8 Mb (0.64-3.3 Mb in the case of *P. falciparum*). The nuclear DNA of *P. berghei* has a high overall A+T content of about 82%.

P. berghei has two extra-nuclear DNA elements (organellar genomes) comparable to *P. falciparum*: the mitochondrial DNA and the plastid DNA. The plastid genome is 30.7 kb in size (35 kb in *P. falciparum*). Partial DNA sequence analysis revealed 69.9-95.5% homology to plastid sequences of *P. falciparum*. Arrangement of the genes (rRNA, rpo-B and tRNA) on the *P. berghei* circle is similar to that found in *P. falciparum*. All *Plasmodium* species analysed so far contain a 6 kb tandemly repeated mitochondrial (mt) genome. The mt genome of *P. berghei* has not been sequenced completely. The mtDNA of *P. yoelii* shows a remarkable 90% conservation compared with the mitochondrial genome of human parasites (for a review see Lin *et al*, 2000).

No functional information is yet available about centromeres and replication origins in *P. berghei*. Telomeres with the repeat sequence of CCCTA(G)AA have been characterised in *P. berghei* and this sequence is similar in all *Plasmodium* species analysed so far. The total length of a telomere is about 1-1.2 kb (Pace *et al*, 2000).

Size differences between homologous chromosomes of up to 0.5 Mb have been detected in parasites from different strains or clones of *P. berghei*. Size differences result from (large-scale) chromosomal rearrangements. These rearrangements mainly occur in the subtelomeric regions, while the internal parts of the chromosomes are more conserved. The order and location of genes on different chromosomes is stable in different strains of *P. berghei*, despite the extensive size polymorphisms between homologous chromosomes. A high level of conservation of gene linkage groups (synteny: gene location and order on chromosomes) exists between the four species of rodent parasites, also emphasizing a low frequency of internal large-scale rearrangements. Although the level of synteny of genes is lower when the genomes of rodent and human species of *Plasmodium* are compared, significant conservation of genome organization has been observed between human and rodent parasites (see below).

GSS and EST library at the University of Florida

In 1998, a project, funded by the National Institute of Allergy and Infectious Diseases (USA), to create a library of 10,000 partially sequenced gene tags of *P. berghei* and the human malaria *P. vivax* was initiated at the University of Florida. Jane Carlton produced almost 5,500 gene sequence tags (GSS), as well as a bit over 5,500 expressed sequence tags (EST), from the *P. berghei* ANKA strain (Carlton & Dame, 2000). DNA and cDNA fragments with sizes between 500 and 2,000 bp were cloned and partially sequenced. (see below for more information about the construction of these libraries).

The data can be searched at the custom BLAST site of the National Center for Biotechnology Information (NCBI) or at the site of the South African National Bioinformatics Institute (SANBI). The clones can be obtained from the Malaria Research and Reference Reagent Resource Center (MR4), a unique initiative to provide researchers from the malaria field with quality controlled, malaria-related reagents including parasites, mosquito vectors, antigens, antibodies, molecular probes, gene libraries and clones.

***P. berghei* sequencing project, Sanger Institute**

In our laboratory genomic DNA was isolated from asynchronous blood stages of *P. berghei (*clone 15cy1 of the ANKA strain) and subsequently cloned in the pUC18 vector. The library thus constructed was

used for the partial shotgun-sequencing project (3x coverage), performed at the Sanger Institute Pathogen Sequencing Unit, Cambridge, UK (for information contact Neil Hall, nh1@sanger.ac.uk), in collaboration with others and funded by the Wellcome Trust. The best way to search these data is at the Sanger website, where you can search the assembled contigs with an average size of 6 kb (ranging from 1 to 50 kb). At the project's main page, there is also a link to an FTP site, where you can download all sequences.

P. yoelii sequencing project, TIGR

P. yoelii is another model rodent malaria parasite that is closely related to *P. berghei*. As a complementary effort to the human malaria genome project, a whole genome-sequencing project of this model parasite was started at The Institute for Genomic Research (TIGR), resulting in a 5.6x coverage of its genome. Data from this project can be searched on both the TIGR project's page, as well as on the PlasmoDB site. Here most malaria genome data have been gathered and is easily accessible and navigable. Contigs range in size from 1 to 50 kb with an average size of 10 kb. The data have been published in Nature in the *P. falciparum* genome issue, in a paper comparing the extensive data on *P. yoelii* with the human malaria completed genome (Carlton *et al*, 2002). Our laboratory contributed to this paper by enabling an even closer comparison based on a detailed physical map of chromosome 5 of *P. berghei* constructed by Leo van Lin (Lin *et al*, 1997).

Comparative genomics in Leiden

The various sequencing projects have greatly benefited several ongoing projects here in Leiden. Below you will find a few examples of the strength of comparative genomics in *Plasmodium*. You will see how comparative genomics can uncover the coding potential of related genomes, how, from comparing multiple sets of linear DNA sequence, even putative open reading frames (ORFs) can be discovered. Further sequencing of other *Plasmodium* genomes and comparing those can result in more insight into the evolutionary history of the different species and the forces driving the evolution of the genomes. We can begin to better understand the biology of the parasite, and its interactions with the mammalian host and mosquito vector ultimately leading to a more rational approach that identifies and evaluates new targets for anti-malarial drug and vaccine development (for reviews see Lin *et al*, 2000 and Thompson *et al*, 2001).

Genome conservation and synteny between different rodent _Plasmodium_ species

Between different strains of *P. berghei*, the location and order of genes on the 14 chromosomes showed high conservation (C.J. Janse, unpublished observations). The number of chromosomes and the chromosomal location and linkage of more than 50 probes, mainly of genes, have been established in four species of *Plasmodium*, which infect African murine rodents (*P. berghei*, *P, chabaudi*, *P. vinkei*, *P. yoelii*) (Janse *et al*, 1994). All four species contained 14 chromosomes, ranging in size between 0.5 and 3.8 Mb, which showed extensive size polymorphisms. However, chromosomal location and linkage of the genes was highly conserved and nearly identical between these species. These results demonstrate that size polymorphisms of the chromosomes are caused by size-variation of the subtelomeric regions that contain species specific genes and non-coding repeat sequences.

Genome conservation and synteny between human, simian and rodent _Plasmodium_ species

The degree of linkage conservation between different species of the malaria parasite *Plasmodium* has been investigated. Initially, the chromosome locations of 42 homologous genes were established in parasites from a rodent malaria species and the human malaria parasite *P. falciparum*. Of these genes, 26 appeared to be conserved within ten synteny groups between the two genomes. Several synteny groups were analysed further by long-range restriction mapping of digested chromosomes. Finally, a fine restriction map of a small linkage group was made from the rodent malaria parasites *P. berghei* and from *P. falciparum* and from the simian malaria parasite *P. knowlesi*. The fine-scale organisation of this linkage group appears to have remained intact among the three species, despite the evolutionary distance between them (Carlton *et al*, 1998). This high level of conservation between the different

malaria parasites may allow the discovery of orthologues and paralogues of less well conserved genes in the different *Plasmodium* species and would be helpful in addressing questions of conservation, evolution and structure of multi-gene families.

The 48/45 multigene family: an extended and well-conserved gene family in *Plasmodium*

One example of a multigene family in *Plasmodium* and how genomics influenced its characterisation came from the investigation of surface proteins of gametes and zygotes of *P. falciparum.* These studies initially demonstrated that three predominant proteins are present: a 40 kDa protein, Pfs230 and Pfs48/45. The latter two have a similar modular structure consisting of loosely defined domains of about 120aa containing six cysteine residues (6-Cys domains). By expression gene cloning, an additional gene with a 6-Cys domain (*Pf12*) was discovered. Since the discovery of the first members of this *Pfs48/45* gene family, we have had to wait for the advent of the genome sequencing projects to uncover the true extent of the 6-Cys domain family of genes. Different research-groups were able to identify more genes encoding proteins with 6-Cys domains through continued analysis of growing amount of sequence data provided by the Genome Projects. The 10th and last member was only revealed when the annotation of the full *P. falciparum* genome was published.

Examination of the raw data in the publicly available databases has shown that all ten members are present in the *P. yoelii* genome. By both database searches and experimental methods, our group has shown that all ten genes also exist in the *P. berghei* genome. The function of P48/45 has been studied in *P. berghei*. Disruption of the gene encoding P48/45 strongly reduced the ability of the male parasites to fertilise (Dijk *et al*, 2001). *P. falciparum Pfs48/45*ko parasites were generated for comparison and showed similar, strong reduction in fertilization capacity. These studies have validated the use of *P. berghei* model as a relatively accessible system in which to study the function of sexual and mosquito-stage parasite proteins, such as members of the *Pfs48/45* gene family, during parasite transmission and fertilisation (for a more complete review with full references see Thompson *et al*, 2001).

The B9-locus: an example of conservation of a complex and gene dense locus

Direct comparison of syntenic regions of the genomes of two species of *Plasmodium* is also useful. A 13.6 kb contig of chromosome 5 of *P. berghei*, a rodent malaria parasite, was sequenced and analysed for its coding potential (Lin *et al*, 2001). Assembly and comparison of this genomic locus with the orthologous locus on chromosome 10 of the human malaria *P. falciparum* revealed an unexpectedly high level of conservation of the gene organisation and complexity, only partially predicted by current gene-finder algorithms and resulting in the identification of at least one gene that was completely missed by the algorithm. In fact, the comparison was able to highlight exon structure within the gene and assist the design of oligonucleotides for subsequent RT-PCR experiments that proved the existence and structure of the gene. 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 within the current models of 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.

Genes encoding variant (eryhrocyte) surface antigens (VSA's): an example of species specificity

A number of gene families are not conserved between rodent and human parasites, others are not even conserved between the different human *Plasmodium* species. The most obvious and important examples are genes involved in antigenic variation, which are mainly located in the subtelomeric regions of the chromosomes. The human parasite *P. falciparum* has a few different gene-families involved in

antigenic variation, predominantly located in the subtelomeric domains of the chromosomes. These include the *stevor* and *rifin* genes (Cheng *et al*, 1998), as well as an estimated 50 *var* genes, encoding the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) (Baruch *et al*, 1995; Su *et al*, 1995; Smith *et al*, 1995). These genes are thus far unique to *P. falciparum* and no homologues have been identified in the genome databases of any of the other *Plasmodium* species.

In the human parasite *P. vivax* and three rodent malarias (*P. berghei*, *P. chabaudi* and *P. yoelii*) (Janssen *et al*, 2002), a gene family encoding putative surface antigens has been identified, with no homology to any *P. falciparum* gene (Del Portillo *et al*, 2001). These subtelomerically located genes were named *vir*, *bir*, *cir* and *yir* genes for the respective *Plasmodium* species and these genes consist of a highly conserved three-exon structure.

Discovery of orthologues genes: pfs16 and LSA1 in rodent parasites

References

Van Lin L.H. *et al*. The conserved genome organisation of non-*falciparum* malaria species: the need to know more. Int. J.Parasitol., 30, 4, p357-370 (2000).

Pace T. *et al.* Targeted terminal deletions as a tool infunctional genomics and position effect studies in *Plasmodium*.Genome Res., 10, 9, p1414-1420 (2000).

Carlton J.M. and Dame J.B. The *Plasmodium vivax* and *P.berghei* gene sequence tag projects. Parasitol. Today, 16, 10,p409 (2000).

Carlton J.M. *et al.* Genome sequence and comparative analysis ofthe model rodent malaria parasite *Plasmodium yoelii yoelii*. Nature,419, 6905, p512-519, (2002).

Van Lin L.H. *et al.* A long range restriction map of chromosome 5of *Plasmodium berghei* demonstrates a chromosome specific symmetricalsubtelomeric organisation. Mol. Biochem. Parasitol., 86, 1, p111-115(1997).

Thompson J. *et al.* Comparative genomics in *Plasmodium*: atool for the identification of genes and functional analysis. Mol. Biochem.Parasitol., 118, 2, p147-154 (2001).

Janse C.J. *et al.* Conserved location of genes on polymorphicchromosomes of four species of malaria parasites. Mol. Biochem. Parasitol.,68, 2, p285-296 (1994).

Carlton J.M. *et al.* Gene synteny in species of *Plasmodium*.Mol. Biochem. Parasitol., 93, 2, p285-294 (1998).

Van Dijk M.R. *et al.* A central role for P48/45 in malariaparasite male gamete fertility. Cell, 104, 1, p153-164 (2001).

Van Lin L.H. *et al.* Interspecies conservation of gene order andintron-exon structure in a genomic locus of high gene density andcomplexity in *Plasmodium*. Nucleic Acids Res., 29, 10,p2059-2068 (2001).

Cheng Q. *et al. Stevor* and *rif* are*Plasmodium falciparum* multicopy gene families which potentiallyencode variant antigens. Mol. Biochem. Parasitol., 97, 1-2, p161-176(1998).

Baruch D.I. *et al.* Cloning the *P. falciparum* gene encodingPfEMP1, a malarial variant antigen and adherence receptor on the surface ofparasitized human erythrocytes. Cell, 82, 1, p77-87 (1995).

Su X.Z. *et al.* The large diverse gene family *var* encodesproteins involved in cytoadherence and antigenic variation of *Plasmodiumfalciparum*-infected erythrocytes. Cell, 82, 1, p89-100(1995).

Smith J.D. *et al.* Switches in expression of *Plasmodiumfalciparum var* genes correlate with changes in antigenic andcytoadherent phenotypes of infected erythrocytes. Cell, 82, 1,p101-110 (1995).

Del Portillo H.A. *et al.* A superfamily of variant genes encodedin the subtelomeric region of *Plasmodium vivax*. Nature, 410,6830, p839-842 (2001).

Janssen C.S. *et al.* A large gene family for putative variantantigens shared by human and rodent malaria parasites. Proc. R. Soc. Lond.B. Biol. Sci., 269,1489, p431-436 (2002).

Transcription profiling in *P. berghei* (DNA microarrays, transcriptomics)

Introduction

DNA microarrays are a powerful technology, which allows the genome and transcriptome of an organism to be screened in a single experiment. The array may consist of cDNA or gDNA sequences that have been individually amplified from a library (otherwise called an amplicon array) or of closely arrayed long oligonucleotides drawn from knowledge of the genome sequence. Both forms of array have been made for the study of transcription in the human malaria parasite (Ben Mamoun *et al*., 2000; Hayward *et al*., 2000; Le Roch *et al*., 2002). We are exploring the transcription profile of *P. berghei* using a DNA microarray based on a library of mung bean nuclease digested *P. berghei* DNA.

P. berghei microarrays made in Leiden

a) Mung bean genomic P. berghei library (Pb GSS)

The *P. berghei* DNA microarray consists of a genomic *P. berghei* library, supplied courtesy of J. Dame and J. Carlton (University of Florida). Genomic DNA was obtained from asynchronous blood stage forms of the cloned ANKA isolate of *P. berghei* (clone 15cy1) grown in laboratory Swiss white mice. Purified DNA was digested with mung bean nuclease, as described (Vernick KD *et al*., 1988). This nuclease digestion should generate fragments of DNA that contain intact functional genes and not intergenic regions thereby enriching for genes and reducing the complexity of the library. Fragments of size in the range of 500-2000 bp were selected, ends were polished and ligated into EcoR V-cleaved pBluescript SK(+) vector. Recombinant plasmids were used to transform *E. coli* XL10-Gold host cells. The library consists of 5482 clones. Inserts are amplified from the library by standard PCR using universal primers, purified through NucleoSpin columns, and resuspended in spotting buffer (ArrayIt Microspotting solution, Telechem International). Several known *P. berghei* genes but also mouse genes are amplified from genomic DNA to be used as controls on the microarray.

Microarrays are created by spotting the GSS library and the control genes in duplicate on aminosilane-coated glass slides, using the Omnigrid microarray spotter (GeneMachines). The microarray production is done in collaboration with Prof. F.C. Kafatos' group at the EMBL in Heidelberg, Germany. To account for sample heterogeneity, each gene was spotted in duplicate on the same slide. DNA is cross-linked onto the glass slides by baking 3 hours at 60°C and 10 minutes at 100°C. Microarrays are stored at room temperature and are stable for six months.

(b) GSS annotation and representation of the genome

The *P. berghei* GSS library was partially sequenced using the M13 -20 forward primer. Sequences of around 500 bp were obtained and used to retrieve the full gene sequence from the *P. berghei* contigs. BLAST search was then carried out against the annotated *P. falciparum* genome (Sanger) and 8.5% and 33% of the GSS clones were respectively homologous to known and to predicted *P. falciparum* hypothetical proteins. The *P. berghei* mung bean library is 24% redundant and represents 70% of the genome. Characterisation of the clones indicated that the DNA is "overdigested" and contains both full-length genes of appropriate size and shards of genes larger than the 2 kb upper cut-off boundary of the size fractionation. 800 sequencing reactions were not valid and the sequences of the DNA spots on the microarray are therefore not completely characterised. These "gaps" are currently being filled in and will be made publicly available through this website and others.

Other _P. berghei_ microarrays

A collaborative effort to design and produce an oligonucleotide array based on the genome of rodent malaria species, under the scheme of the Post-Genomic Initiative is supervised by the the Wellcome Trust Sanger Centre. Oligonucleotide arrays will be designed on sequence information alone and synthesised *in situ* by photosensitive photolithography allowing extremely high-density printing; hundreds of thousands of different oligonucleotides of a length that is currently being optimised (expected range 40-60mer) can be placed on a very small glass surface. The probe redundancy improves the accuracy of RNA quantitation and reduces the rate of false positives whereas the high homology between the rodent malarial parasites will allow the use of the oligonucleotide array for identifying genotypic and species-specific differences.

Microarray hybridisations

Highly synchronised cultures of blood stage parasites are grown *in vitro* and parasites harvested at distinct time points. RNA is extracted and complementary cDNA probes are synthesized and labelled with Cy3-dUTP and Cy5-dUTP fluorescent nucleotide analogs, in a random primed first strand synthesis reverse transcription reaction. Two differentially labelled probes are combined, lyophilised, resuspended in a formamide based hybridisation buffer and hybridised on the array (Dimopoulos *et al*., 2002).

Arrays are scanned, intensity of the fluorescent dyes is measured for each individual gene and ratios of the intensities of the two samples are built. The arrays are scanned using a Genetics Microsystems

scanner through the facilities offered from the Leiden Genome Technology Center. Data analysis is carried out using GenePix Pro 4.1 and Cluster and Treeview softwares. Hierarchical clustering analysis of the data allows us to arrange genes according to similarity in pattern of gene expression throughout experiments (Eisen *et al*., 1998).

The arrays and the *P. berghei* model are currently being combined in the laboratory to investigate the following:

(a) Gene expression during the asexual blood stage development

Using a mutant gametocyte-non-producer parasite clone (HPEcy1m50 cl1 of the ANKA strain) we are able to isolate asexual blood stage parasites which cannot produce gametocytes (commit to gametocytogenesis). Synchronized parasites are harvested during the asexual blood stage development at 5 time points spanning 5 to 24 h of the 24h cycle, and the stage specific RNA samples are used in competitive hybridisations. Hence information obtained from these hybridisations, allow us to establish the gene expression pattern throughout the 24-hour cycle (G1, S and M phases of the cell cycle) of asexual development of *P. berghei.*

(b) Gene expression during gametocyte development and identification of gametocyte-specific genes

Synchronized blood stages of the 'wild type' gametocyte-producer parasite clone (Clone 15cy1 of the ANKA strain) are cultured in parallel with the HPE clone mentioned above and stage specific samples are collected at the same time points during the 24 cycle of blood stage development. Furthermore, purified gametocytes are collected as described in chapter 11 of 'The Plasmodium berghei research model of malaria'.

We are in the process of comparing the transcriptome of gametocytes to that of asexual blood stage parasites. In addition we compare the transcriptome from the non-gametocyte producer clone and that of the producer clone at identical time points during blood stage development.

(c) Differences in gene expression between virulent and avirulent lines of P. yoelii

As has been described earlier, the genomes of *P. berghei* and *P. yoelii* are highly similar both at the level of synteny and sequence similarity. Indeed, the vast majority of genes within the two organisms are over 90% identical at the DNA level. Consequently, we are now performing microarray analysis on the *P. berghei* arrays with RNA extracted from blood stages of two lines of *P. yoelii*. This further illustrates the usefulness of the arrays as a substrate for analysis in other, closely related, *Plasmodium* species.

Microarray data (to be released)

Comparative microarray analysis

(a) Comparison with P. falciparum arrays

We have compared our results on transcriptome data of the rodent parasite *P. berghei* with *P. falciparum* transcriptome data. Selected asexual stage specific genes were retrieved from different microarray approaches; expressed sequence tag (EST) array (Ben Mamoun *et al*., 2000), mung bean genomic DNA library array (Hayward *et al*., 2000) and oligonucleotide array (Le Roch *et al*., 2002). By comparing the different datasets we are able to compare and contrast between different experimental approaches but also between a rodent and a human malaria parasite.

(b) Comparison with proteomic data

We are now in the fortunate position to be able to compare our transcriptome data with the recently published *P. falciparum* proteomics data (Lasonder *et al*., 2002; Florens *et al*., 2002). In addition, we plan to generate *P. berghei* proteomic data corresponding to different stages of the lifecycle (see below). Combination and analysis of the transcriptome and proteome data will provide us with a better understanding of the function and interaction of individual proteins.

References

Proteome analysis of *P. berghei* (proteomics)

Introduction.

Proteomics is the large-scale study of proteins, usually by biochemical methods. The word proteomics has been associated with the analysis of a large number of tryptic fragments of proteins by mass spectrometry (single or tandem) from a given cell line or organism after two-dimensional polyacrylamide gel fractionation of intact proteins and more recently, liquid chromatography fractionation of the tryptic digests of total protein extracts. There is no strict linear relationship between genes, or indeed with the transcriptome, and the protein complement or 'proteome' of a cell. Proteomics offers the advantage over the other large-scale initiatives of being able to define the proteins present not only within the cell but the sub-cellular compartments within it. It is our intention to generate proteomic data from the various life cycle stages of *Plasmodium berghei*. Recently two large-scale, high-throughput mass spectrometric analyses of *Plasmodium falciparum* have been conducted on a variety of *P. falciparum* life cycle stages (Lasonder *et al*., 2002; Florens *et al.*, 2002). We are collaborating with one of the groups responsible for this research, Prof. Mann's group at the University of Southern Denmark, Odense, to perform similar high-accuracy high-throughput analysis of asexual and asexual blood stages of *P. berghei*.

Life-cycle stages and sub-cellular fractions analysed.

Data to be released

References

Ben Mamoun C. *et al.* Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis. Mol Microbiol. 2001;(1):26-36.

Dimopoulos G. *et al.* Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge, and malaria infection. PNAS. 2002;99(13):8814-9.

Eisen M.B. *et al.* Cluster analysis and display of genome-wide expression patterns.PNAS. 1998;95(25):14863-8.

Florens L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle.Nature. 2002; 419(6906):520-6.

Hayward R.E. *et al.* Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malariaMol. Microbiol. 2000;35(1):6-14.

Lasonder E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. Nature. 2002; 419(6906):537-42.

Le Roch K.G. *et al.* Monitoring the chromosome 2 intraerythrocytic transcriptome of *Plasmodium falciparum* using oligonucleotide arrays. Am J Trop Med Hyg. 2002; 67(3):233-43.

Vernick K.D. *et al.* Mung bean nuclease exhibits a generalized gene-excision activity upon purified *Plasmodium falciparum* genomic DNA. Nucleic Acids Res. 1988;16(14B):6883-96.

Background information on the genome and post-genome projects

1) *P. berghei* genomic DNA library (mungbean nuclease-digested)
Library name: Pb MBN #21, Jane Carlton and Charles Yowell
Isolate: P. berghei ANKA clone 15cy1 (clone of the ANKA 8417 clone)
Vector: pBluescript SK(+)
Host: E. coli XL10-Gold
Ligation site: EcoRV
Sequencing Primer: M13 -20 forward
DNA: Genomic DNA was prepared from asynchronous blood stage forms of the cloned ANKA isolate of *P. berghei* grown in laboratory Swiss white mice. The DNA was purified from contaminating host DNA by Hoechst Dye 33258-CsCl ultracentrifugation and precipitated. Purified DNA was digested with mung bean nuclease in the presence of 36-38% formamide at 50 C, as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the digestion fragments were polished using T4 DNA polymerase, and the fragments size selected in the range 500-2000 bp. These were ligated into the EcoRV-cleaved and dephosphorylated pBluescript SK(+) vector. Recombinant plasmids were used to transform E. coli XL10-Gold host cells.

2) *P berghei* cDNA Library
For plates numbered 1-14, 22-40, 42-61:
Library name: Pb cDNA #17 (T. Pace *et al* 1998, Mol. Biochem. Parasitol. 97, 45-53.)
Isolate: P. berghei ANKA clone HP (gametocyte producer)
Vector: Lambda ZAP II
Host: E. coli XL1-Blue MRF'
Ligation site: EcoRI/XhoI
Sequencing primer: T3
RNA: Total RNA was extracted from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei, grown in Wistar rats to 30% parasitemia and 2-5% gametocytemia. Contaminating host white cells had previously been removed and final host cell contamination estimated to be approximately 5%. PolyA+ RNA was extracted and reverse transcribed using an oligo dT-Xho I primer (Lambda ZAP II cDNA cloning kit, Stratagene). Second strand cDNA was made following the manufacturer's protocol. EcoRI adaptors were ligated to the cDNA, and fragments were ligated into EcoRI /XhoI digested vector.

For plates numbered 15-21 and 63-69:
Library name: Pb cDNA #20, Charles Yowell and Jane Carlton
Isolate: P. berghei ANKA clone 15cy1 (clone of the ANKA 8417 clone)
Vector: Lambda ZAP II
Host: E. coli XL1-Blue MRF'
Ligation site: EcoRI/XhoI
Sequencing primer: T3 (M13 -20 forward plates 63, 66, 67)
RNA: Total RNA was extracted from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei grown in laboratory Swiss white mice. Contaminating host white cells had previously been removed using a novel biomagnetic bead protocol [Carlton JM, Yowell CA, Sturrock KA, Dame JB, Biomagnetic separation of contaminating host leukocytes from plasmodium-infected erythrocytes. Exp Parasitol. 2001 Feb;97(2):111-4]. PolyA+ RNA was extracted and reverse transcribed using an oligo dT-XhoI primer. Second strand cDNA was prepared using RNase H and DNA polymerase I. EcoRI adaptors were ligated to the cDNA, and it was digested with XhoI. Fragments were size selected, and those between 1-5 kb ligated into EcoRI /XhoI digested vector.