Update TRENDS in Parasitology Vol.21 No.9 September 2005

Research Focus

Systems biology in malaria research

Martin J. Fraunholz

Junior Group Applied Bioinformatics and Cell Biology, Institute of Microbiology, Ernst Moritz Arndt University, Friedrich-Ludwig-Jahn Strasse 15, D-17487 Greifswald, Germany

A recent publication of genome and expression analyses of the murine parasites *Plasmodium chabaudi chabaudi* and *Plasmodium berghei* presents the state of the art in *Plasmodium* systems biology. By integrating genomics, transcriptomics and proteomics, the authors can classify and annotate genes by their expression profiles and can even detect evidence of posttranscriptional gene silencing in the murine malaria species.

Integrating genomics and postgenomics of malaria research

Hall *et al.* recently described 'a comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic and proteomic analyses' [1]. The analysis of the draft genome sequences of the two murine *Plasmodium* species *Plasmodium chabaudi chabaudi* AS and *Plasmodium berghei* ANKA was integrated with genome-wide transcriptomics and proteomics analyses that covered almost all stages of the *P. berghei* life cycle; RNA was prepared from seven blood stages, and protein samples were harvested from asexual blood stages, gametocytes, ookinetes, oocysts and salivary gland sporozoites. The integration of sequence, DNA microarray and proteomics [multidimensional protein identification technology (MudPIT)] data revealed a battery of interesting information.

Expression profiles classify proteins

Orthology assignments by reciprocal analysis with the basic local alignment search tool (BLAST) and gene synteny detected a universal rodent malaria gene set of ~4400 genes and indicated that 15–20% of the genes might be specific to *Plasmodium*, although this estimate will change with each genome sequence that is sequenced. Annotation of such organism-specific genes is extremely difficult because it normally involves heterologous expression of the gene product and subsequent biochemical characterization of the protein. Because proteins in the same pathway must be present at the same time, expression profiles can be used to deduce the cellular role of proteins. Proteins that are present in all or most stages of the life cycle can be labeled as 'housekeeping' proteins; Hall *et al.* [1] put 136 gene products in this category. By contrast, MudPIT identified 948 proteins that were present in only one of the examined developmental stages; these have, therefore, been termed 'stage-specific' proteins. Furthermore, so-called 'strategy-specific' genes are expressed differentially in either asexual replicative stages (472 proteins) or invasive stages (234 proteins). 'Host-specific' genes are present in either the mosquito (229 proteins) or the mammalian host (417 proteins). As an example, the authors show that the expression of genes involved in the tricarboxylic acid cycle (TCA) (Figure 1) and other mitochondrial functions starts in gametocytes and increases throughout the vector stages. This is in agreement with the long-known fact that a complete TCA cycle is absent from asexual blood stages [2]. Overall, this relatively crude classification can be refined further by using more-quantitative data such as DNA microarrays (Box 1). However, because of mechanisms such as posttranscriptional gene silencing – which has been observed in *Plasmodium* [3,4] – mRNA levels do not necessarily correlate with protein levels.

Posttranscriptional gene silencing

The real power of the systems biology approach lies in the integration of data from different experimental sources and the subsequent discovery of novel information contained within the datasets. By comparing the gametocyte DNA array results with the gametocyte and ookinete proteomics data, Hall *et al.* identified nine cases in which message was abundant in gametocytes but protein was produced only in ookinetes [1]. These results indicate that translational repression, first described by Paton *et al.* [3] and analogous to findings from genome-wide studies in *Plasmodium falciparum* [4], is silencing the messages in gametocytes of murine malaria species. In 2/3 of the detected mRNA in this phenotype, Hall *et al.* detected a 47-nucleotide-sequence motif in the 3′-UTR. A subsequent genome-wide motif scan revealed a total of 29 candidates for translational repression. The transcription of most of these candidates is upregulated throughout the gametocyte stage, whereas MudPIT did not identify any protein. It will be interesting to investigate further these candidates and their 22 *P. falciparum* homologs to determine whether posttranscriptional gene silencing is a common regulatory mechanism in the genus *Plasmodium*.

The *pir* gene family

The *Plasmodium* interspersed repeats (*pir*) gene family proposed by Hall *et al.* [1] shares an evolutionary relationship and incorporates the *vir*, *yir*, *cir* and *bir* gene families (for *vivax*, *yoelii*, *chabaudi* and *berghei* interspersed repeats, respectively). *pir* genes are unrelated to the *P. falciparum var* genes. However, like the *var*

Corresponding author: Fraunholz, M.J. (martin.fraunholz@uni-greifswald.de).
Available online 25 July 2005

(a)

Description	Locus	Protein sequence coverage (%) ABS	Gct	Okn	Ocy	Spz
Glycolysis						
Hexokinase	PB000727.00.0	40.5	15	32.6		16.6
Glucose-6-phosphate isomerase	PB001005.02.0	28.8	20.7	28.2		
6-phosphofructokinase	PB000520.00.0	38	8.6	16.1		7.5
Triose-phosphate isomerase	PB300820.00.0	50	39.5	57.7		
Triose-phosphate isomerase	PB001162.00.0	50	39.5	57.7		
Phosphoglycerate kinase	PB000318.00.0	63.1	26.3	60.7	11.6	
Phosphoglycerate mutase	PB000139.01.0	79.2	48.4	76	14.4	
Phosphoglycerate mutase	PB000556.01.0	10.1	12.1	10.4		
Enolase	PB000456.03.0	75.1	59	63.7	26.5	14.6
Pyruvate kinase	PB000521.02.0	68.5	37.4	70.5		9.6
L-lactate dehydrogenase	PB000185.00.0	93.7	64.8	83.8	24.1	42.5
Tricarboxylic acid cycle						
Citrate synthase	PB001560.02.0		4.4	16.6		9.2
Aconitase/IRP-like protein	PB001120.02.0	6.7	4.5	12.3		
Isocitrate dehydrogenase	PB000917.03.0		7.7	20.9		
Branched-chain α-keto acid dehydrogenase	PB001062.02.0	15.9	6.3	41.8	3.5	
Succinyl-CoA synthetase α subunit	PB001002.01.0	10.9	14	43.3		
Succinyl-CoA synthetase β subunit	PB000281.02.0		2.6	36		
Iron–sulfur subunit of succinate dehydrogenase	PB000191.02.0		5.5	6.8		
Flavoprotein subunit of succinate dehydrogenase	PB001230.02.0			11.9		
Flavoprotein subunit of succinate dehydrogenase	PB300700.00.0			8.6		
Fumarate hydratase	PB000657.03.0			7.8		
Malate dehydrogenase	PB000128.01.0	41.7	30.8	50.6	16	14.7

Key: 95% 40% 30% 20% 10% 0.1% None

(b)

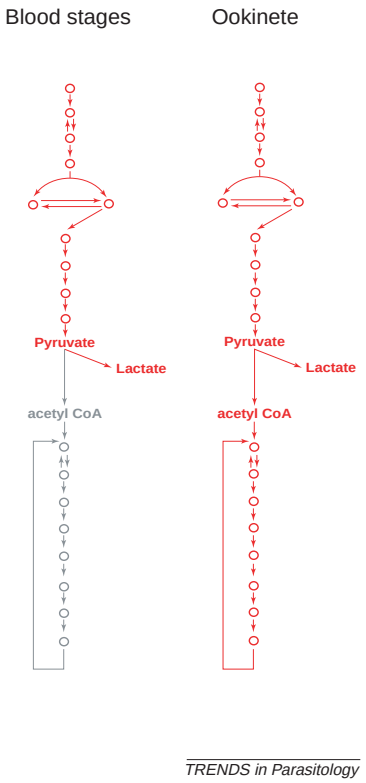

Figure 1. Metabolic pathway reconstruction by expression analysis. **(a)** Table showing representation of protein expression from the glycolytic and TCA pathways, color coded by protein sequence coverage identified through proteomics experiments (see key). The sequence coverage values in the table serve as a crude measurement of protein abundance. Values are from Ref. [1]. During asexual blood stages, the glycolytic pathway is active, resulting in the production of lactate, whereas there is no evidence of a complete TCA cycle during these stages. **(b)** Pathway glyph representation of the activity of metabolic pathways, as determined by their proteomics expression profiles. Red indicates active pathways and gray indicates inactive pathways.

gene family, *pir* is expanded into multiple copies, the number of which varies between species. The assessment of the ratio of synonymous to non-synonymous substitutions indicates diversifying selection of *pir*, outlining its function as a 'smoke screen' for the adaptive immune system in the mammalian host. However, the authors detected that ~1/10 of the bir proteins in *P. berghei* is present exclusively in mosquito stages, which indicates roles that are alternative to evasion of the mammalian immune system. As invertebrates, mosquitoes lack an adaptive immune system and rely completely on innate immunity. Therefore, the insect defense against microbial pathogens is limited to phagocytosis, antimicrobial peptides, melanization or encapsulation (for review, see Ref. [5]). The exclusive expression of some *var* genes in *P. falciparum* salivary gland sporozoites [6] and of *bir* genes in *P. berghei* mosquito stages [1] raises the question of whether these variant gene families also function to avoid elimination by the innate mosquito defenses.

Negatively selected genes in subtelomeric regions

Besides the *pir* family, multiple genes that are involved in immune evasion have been identified. Candidate transmembrane proteins have been selected by intersecting sets of genes that have a signal peptide and mammalian-host-specific expression with genes that show signs of diversifying pressure acting on them.

Overall, possible immune-evasive proteins and species-specific genes are positioned in the subtelomeric regions of chromosomes, whereas conserved genes are located preferentially in the central portions. This is congruent with findings in other *Plasmodium* species [7–9].

The roadmap to the transparent parasite

Hall *et al.* [1] demonstrate the feasibility and maturity of genomics and postgenomics methods for malaria research. The published draft genome sequences of *P. chabaudi chabaudi* AS and *P. berghei* ANKA line up with the previously published sequences of the murine malaria parasite *Plasmodium yoelii yoelii* [9] and the human malaria parasite *P. falciparum* 3D7 [10]. This genome coverage renders *Plasmodium* the eukaryotic genus with the most species sequenced to date. With many more species and isolates in the sequencing pipeline, this status will be emphasized even more in the future and will make the genus extremely interesting for, among other things, comparative genomics, studies of genome evolution and analysis of host–pathogen interactions.

The published postgenomics data become more than the sum of their parts only when direct comparisons of them are possible: in gametocytes, where datasets exist for both experiments, using DNA arrays and MudPIT. This outlines the infrastructural necessities for the future of high-throughput malaria research: multi-laboratory cooperations must begin at the point of planning the

Box 1. A selection of *Plasmodium* data-mining tools

- The *Plasmodium* Genome Resource (http://PlasmoDB.org) contains sequence data from multiple murine and human malaria species integrated with microarray and proteomics expression data. The site enables the construction of complex queries to identify candidate genes of interest.
- The Sanger Institute pathogen genome database GeneDB (http://www.genedb.org) is based on the same engine as PlasmoDB. It also incorporates other pathogen genome sequences.
- The NCBI malaria database (http://www.ncbi.nlm.nih.gov/projects/Malaria/) integrates genomic information about *Plasmodium falciparum* and *Plasmodium yoelii* with linkage maps, markers, epidemiological data, and literature and GenBank searches.
- The Genomics Institute of the Novartis Research Foundation (http://chemlims.com:8080/OPI20/MServlet.ChemInfo) provides a database of *P. falciparum* gene function prediction that uses ontology-based pattern identification based on Affymetrix expression data.
- The DeRisi Laboratory Malaria Transcriptome Database (http://malaria.ucsf.edu) is a malaria transcriptome database based on high-resolution time-point sampling of blood stages and glass-slide DNA array analyses.
- The official *Plasmodium vivax* genome-sequencing project at the Institute for Genomic Research [TIGR (http://www.tigr.org/tdb/e2k1/pva1/)] currently offers only licensed sequence download and BLAST searches. Contains links to the other TIGR *Plasmodium*-sequencing sites and to TIGR gene indices that provide information about, for example, cDNA abundance and predicted alternative splice sites.
- Full-malaria (http://fullmal.ims.u-tokyo.ac.jp) is a database containing full-length *Plasmodium* cDNAs.
- Malariabase (http://malariabase.org) is a *Plasmodium* genome database that also presents *Anopheles* data. 'Binary signatures' (accessible at http://matrices.malariabase.org) indicate the abundance of annotation features for *Plasmodium* genes.
- PlasmoCyc (http://plasmocyc.stanford.edu) is a malaria metabolic pathway database and genome browser.
- Malaria Parasite (http://sites.huji.ac.il/malaria) is a manually curated website with pathway maps and literature references.
- The database of the Structural Genomics of Pathogenic Protozoa (http://www.sgpp.org) project contains protein crystal structures of malaria and other pathogenic protozoa.

experiments. Laboratories with expertise in growing and isolating synchronized *Plasmodium* blood stages *in vitro*, capable of laser dissecting enough liver schizonts from infected mice, passing the parasites through mosquito vectors or isolating salivary gland sporozoites and ookinetes will be able to provide enough material to overcome the sensitivity issues of MudPIT or enable multiple repetitions of array experiments to increase statistical support. Protein and RNA samples must be isolated from the same culture at the same time points to ensure that the results are truly comparable with each other to elucidate to what extent posttranscriptional gene silencing and, possibly, other regulatory mechanisms function within *Plasmodium*. It is hoped that, in future experiments, the liver stages of malaria parasites will be included to make full use of the murine model. Then, researchers could finally obtain information about the elusive early stages of the infection and go 'full circle' by using a genome-wide account of mRNA and protein quantities throughout the complete life cycle of malaria parasites.

Quo vadis, Plasmodium research?

New technologies for cell biology manipulation and postgenomics analysis define members of the genus *Plasmodium* as model organisms for intracellular eukaryotic pathogens. The fully sequenced or highly sampled genomes enable, for example, genome-wide analyses of the levels of transcripts and proteins, identification of single nucleotide polymorphisms and genotyping of uncultured isolates. Classical methods of expression analysis such as expressed sequence tag (EST) and serial analysis of gene expression (SAGE) are quickly becoming superseded as slide scanners and Affymetrix (http://www.affymetrix.com/index.affx) chip readers become standard laboratory equipment. The expertise and machinery for modern mass-spectrometric methods such as multidimensional chromatographic separation of peptides, coupled with mass-spectrometric detection and identification, are – or will be – concentrated in central facilities at universities or companies, enabling submission of samples as known from current sequencing services. However, the advent of high-throughput methods does not mean the demise of traditional experimentation but, instead, will enable investigators to sift through giant datasets in seconds using computers and databases, thereby reducing the dataset from thousands of candidate genes in a *Plasmodium* genome to mere tens. These, in turn, can be investigated more thoroughly by classical biochemical and cell biology methods using gene-by-gene approaches. Box 1 lists some of the available resources for online data mining.

References

1 Hall, N. *et al*. (2005) A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 307, 82–86

2 Sherman, I.W. (1998) Carbohydrate metabolism of asexual stages. In *Malaria: Parasite Biology, Pathogenesis, and Protection* (Sherman, I.W., ed.), pp. 93–120, ASM Press

3 Paton, M.G. *et al*. (1993) Structure and expression of a post-transcriptionally regulated malaria gene encoding a surface protein from the sexual stages of *Plasmodium berghei*. *Mol. Biochem. Parasitol.* 59, 263–275

4 Le Roch, K.G. *et al*. (2004) Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res.* 14, 2308–2318

5 Osta, M.A. *et al*. (2004) Innate immunity in the malaria vector *Anopheles gambiae*: comparative and functional genomics. *J. Exp. Biol.* 207, 2551–2563

6 Florens, L. *et al*. (2002) A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 419, 520–526

7 Lanzer, M. *et al*. (1994) Organization of chromosomes in *Plasmodium falciparum*: a model for generating karyotypic diversity. *Parasitol. Today* 10, 114–117

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

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 Gardner, M.J. *et al*. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511

1471-4922/$ - see front matter
doi:10.1016/j.pt.2005.07.007