Published online 28 October 2008 *Nucleic Acids Research, 2009, Vol. 37, Database issue* ***D539–D543***
doi:10.1093/nar/gkn814

PlasmoDB: a functional genomic database for malaria parasites

Cristina Aurrecoechea[1], John Brestelli[2], Brian P. Brunk[2], Jennifer Dommer[2], Steve Fischer[2], Bindu Gajria[2], Xin Gao[2], Alan Gingle[3], Greg Grant[4], Omar S. Harb[2,*], Mark Heiges[1], Frank Innamorato[2], John Iodice[2], Jessica C. Kissinger[1,5], Eileen Kraemer[6], Wei Li[2], John A. Miller[6], Vishal Nayak[2], Cary Pennington[1], Deborah F. Pinney[2], David S. Roos[7], Chris Ross[1], Christian J. Stoeckert Jr[2], Charles Treatman[2] and Haiming Wang[1]

[1]Center for Tropical & Emerging Global Diseases, University of Georgia, Athens, GA 30602, [2]Penn Center for Bioinformatics, University of Pennsylvania, Philadelphia, PA 19104, [3]Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30602, [4]School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, [5]Department of Genetics, [6]Department of Computer Science, University of Georgia, Athens, GA 30602 and [7]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

Received September 15, 2008; Accepted October 3, 2008

ABSTRACT

PlasmoDB (http://PlasmoDB.org) is a functional genomic database for *Plasmodium spp*. that provides a resource for data analysis and visualization in a gene-by-gene or genome-wide scale. PlasmoDB belongs to a family of genomic resources that are housed under the EuPathDB (http://EuPathDB.org) Bioinformatics Resource Center (BRC) umbrella. The latest release, PlasmoDB 5.5, contains numerous new data types from several broad categories—annotated genomes, evidence of transcription, proteomics evidence, protein function evidence, population biology and evolution. Data in PlasmoDB can be queried by selecting the data of interest from a query grid or drop down menus. Various results can then be combined with each other on the query history page. Search results can be downloaded with associated functional data and registered users can store their query history for future retrieval or analysis.

INTRODUCTION

Plasmodium spp. are obligate intracellular protozoan parasites of humans and animals, and are the causative agents of malaria. Transmission of these parasites to humans occurs via the *Anopheles* mosquito vector and the geographic distribution of endemic regions puts almost half of the world's population at risk to contracting malaria. This disease is a major source of morbidity and mortality worldwide, which results in 300–500 million clinical cases and 1–2 million deaths annually (1,2). While several species of *Plasmodium* cause disease in humans (including *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*), *P. falciparum* is by far the deadliest (1,3). The life cycle of the *Plasmodium* parasite takes it through multiple cell types (in the vertebrate host and arthropod vector) during which the parasite undergoes multiple developmental changes (both sexual and asexual). The different life-cycle stages are marked by specific genomic, transcriptomic, proteomic and metabolomic states. Understanding how these changes are triggered and orchestrated requires mechanisms to view and interrogate genomic and functional genomic data in a powerful and intuitive manner. Over the past 10 years, PlasmoDB has evolved into a venue that integrates such data and allows the user to perform complex queries tailored to their specific needs and interests.

UPDATED DATA CONTENT

The data available in PlasmoDB has expanded to include genomic and functional data from eight *Plasmodium* species and is summarized in Table 1 (4). The current release (PlasmoDB 5.5) contains fully sequenced and annotated genomes of *P. falciparum*, *P. vivax*, *P. yoelii*, *P. berghei*,

*To whom correspondence should be addressed. Tel: +1 215 746 7019; Fax: +1 215 573 3111; Email: oharb@pcbi.upenn.edu
Correspondence may also be addressed to Brian P. Brunk. Tel: +1 215 573 3118; Fax: +1 215 573 3111; Email: brunkb@pcbi.upenn.edu; Jessica C. Kissinger. Tel: +1 706 542 6562; Fax: +1 706 542 3582; Email: jkissing@uga.edu; David S. Roos. Tel: +1 215 898 2118; Fax: +1 215 746 6697; Email: droos@sas.upenn.edu; Christian J. Stoeckert. Tel: +1 215 573 4409; Fax: +1 215 573 3111; Email: stoeckrt@pcbi.upenn.edu

Table 1. Types of data available in PlasmoDB and example queries

Type of Data	Species for which this data is available	Example query
Genomic data		
Full sequence and annotation	*P. falciparum, P. vivax, P. yoelii, P. berghei, P. chabaudi, P. knowlesi*	Search annotations for specific keyword (see Figure 1C).
Sequence only	*P. reichenowi, P. gallinaceum*	Find sequence similarity using BLAST.
Transcript expression data		
Microarray	*P. falciparum, P. berghei, P. yoelii*	Identify genes expressed at specific life-cycle stages.
EST	*P. falciparum, P. vivax, P. berghei, P. yoelii*	Confirm gene models and alternative gene models.
SAGE	*P. falciparum*	Identify genes with transcript evidence.
Protein expression data	*P. falciparum, P. berghei, P. yoelii*	Identify genes with protein expression evidence at specific life-cycle stages.
Population biology		
SNP Microsatellite Isolate data	*P. falciparum*	Find highly polymorphic genes or distinguish isolates based on their SNP profile.
Protein interaction		
Yeast two hybrid Interactome map	*P. falciparum*	Identify possible interaction partners of a gene of interest.
Putative function		
GO annotation	*P. falciparum, P. vivax, P. yoelii, P. berghei, P. chabaudi, P. knowlesi*	Identify genes that have GO annotations.
EC numbers	*P. falciparum, P. yoelii, P. knowlesi*	Identify genes with enzymatic annotations.
Metabolic pathways	*P. falciparum*	Identify parasite-specific or missing metabolic pathways.
Evolutionary		
Orthology based	*P. falciparum, P. vivax, P. yoelii, P. berghei, P. chabaudi, P. knowlesi*	Identify genes specific to apicomplexa.
Homology based	*P. falciparum and P. yoelii*	Identify homologs of a gene or list of genes of interest.
Protein features		
Protein motifs Interpro/pfam domains Molecular weight Isoelectric point Protein structure Immune epitopes	*P. falciparum, P. vivax, P. yoelii, P. berghei, P. chabaudi, P. knowlesi*	Identify genes with specific protein attributes.
Protein localization		
Signal peptide Transmembrane domains Targeting to the RBC	*P. falciparum, P. vivax, P. yoelii, P. berghei, P. chabaudi, P. knowlesi*	Identify genes targeted to the host cell.
Apicoplast targeting	*P. falciparum*	Identify genes targeted to the apicoplast.

P. chabaudi and *P. knowlesi*. Importantly, PlasmoDB 5.5 contains results of annotation efforts from multiple sources including the recent systematic effort to update the *P. falciparum* genome that is an ongoing project started at a workshop in late 2007 co-organized by the Wellcome Trust Sanger Institute (WTSI) and EuPathDB (formerly ApiDB) teams. Reannotation data have been released in incremental steps (snapshots) in order to provide timely information to users of PlasmoDB and to solicit user comments regarding the reannotations.

Transcript expression data [microarray, expressed sequence tags (ESTs) and serial analysis of gene expression (SAGE)] available through PlasmoDB has expanded dramatically over the past few releases to include microarray data from multiple life-cycle stages, gene knock-out mutants of *P. falciparum* and *P. berghei* (5–12) and multiple stages of *P. yoelii* (mosquito, erythrocytic and liver stages) (13). Also included are EST data from over 130 libraries (*P. falciparum, P. vivax, P. berghei* and *P. yoelii*) (14,15) [dbEST (http://www.ncbi.nlm.nih.gov/dbEST/)] and SAGE data (*P. falciparum* only) (16–18). Protein expression evidence includes data from various life-cycle stages (*P. falciparum, P. berghei* and *P. yoelii*) (11,13,19–21; Leiden Malaria Group, unpublished data).

Population biology evidence (*P. falciparum* only) includes mapping of microsatellite data (22) onto the genome (available as a genome browser track), single nucleotide polymorphism (SNP) data from resequencing efforts of more than 20 *P. falciparum* strains (*P. reichenowi* is included as an out-group for comparison purposes) and data from nearly 100 *P. falciparum* isolates (23–25). OrthoMCL analyses provide ortholog determinations between the different species facilitating discovery of shared genes between lineages (26). Protein function assignments are aided by a number of additional functional data types available through PlasmoDB 5.5 including

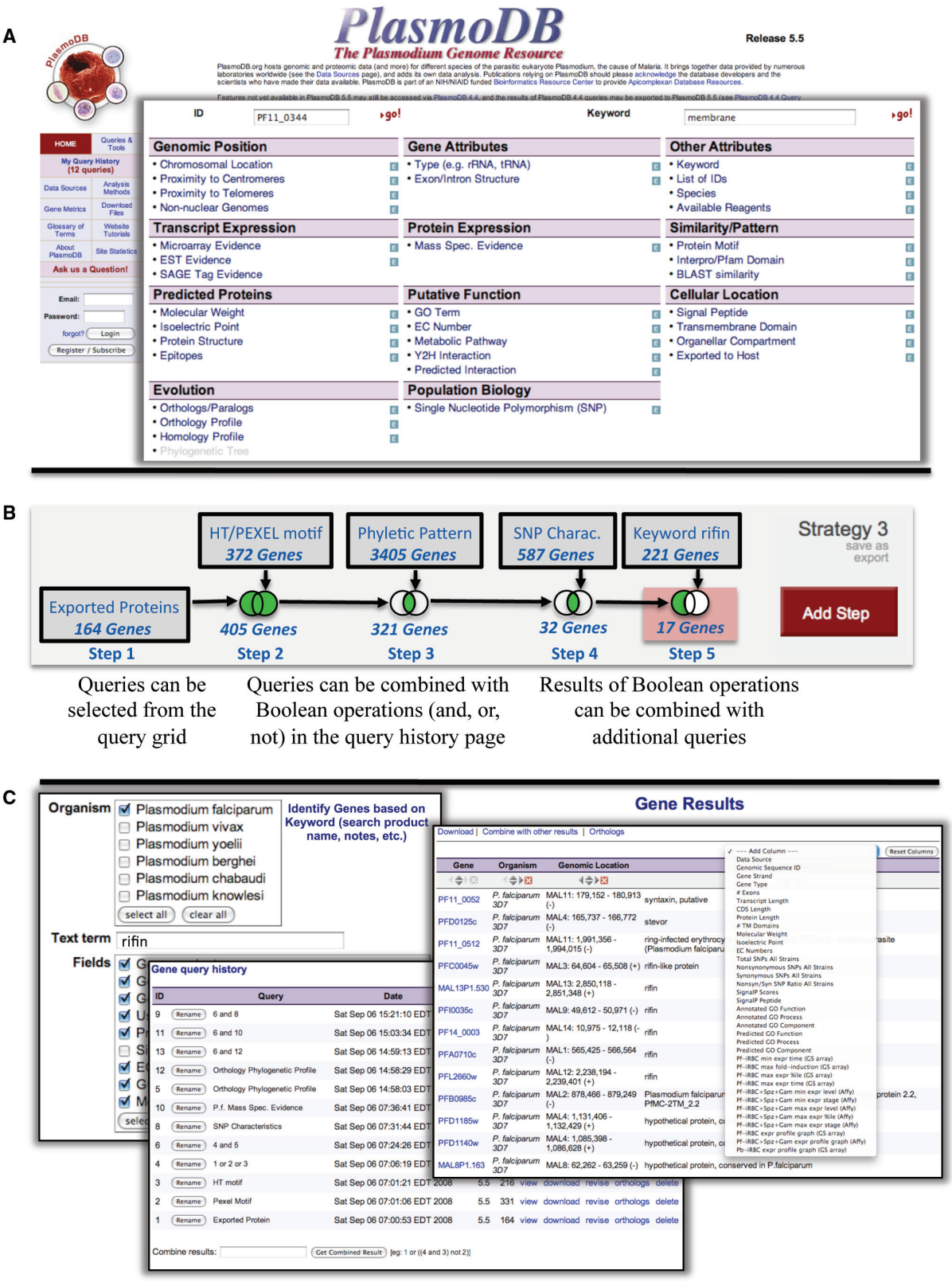

Figure 1. Screenshots from PlasmoDB 5.5 and query workflow. (**A**) The top of the screenshot shows the PlasmoDB logo. On the left side are links to various sections of PlasmoDB and a point for logging in or registering as a user (not required for using the site but useful for storing search histories. The query grid is in the center and provides an access point to all searchable data in PlasmoDB. (**B**) This is a scheme of a workflow that a user may follow when building a set of queries. Beginning at the left, queries can be performed starting from the query grid and the results can be joined using operations available through the query history page. (**C**) Screen shots of a 'key word' search page, an example gene query history and a gene results page. Note the add column feature in the results page that allows the addition of columns with additional data and the ability to sort results.

evidence of protein–protein interaction (yeast two hybrid and predicted interactome) (27,28), Genome Ontology (GO) (29) and InterPro domain (30) annotations for *P. falciparum, P. vivax, P. berghei*, *P. yoelii*, *P. knowlesi* and *P. chabaudi*, Enzyme Commission (EC) number (29) annotation for *P. falciparum*, *P. yoelii* and *P. knowlesi* (31) and metabolic pathway assignments for *P. falciparum* (31). In addition, subcellular localization of proteins is available through signal peptide (32) and transmembrane domain predictions (33) for *P. falciparum, P. vivax, P. berghei*, *P. yoelii*, *P. knowlesi* and *P. chabaudi*, and parasite-specific predictions (*P. falciparum* only) for apicoplast localization (34) and export to the host cell (35–37).

HOW TO USE PLASMODB

A visitor to PlasmoDB can use the database in two general ways: (i) To retrieve all available information associated with a particular gene of interest using a search for an exact gene ID, gene name or gene product name. (ii) To ask single questions (Table 1) and/or conduct a series of searches followed by refining the results by combining them or subtracting them from one another. Starting with the PlasmoDB home page (Figure 1A), a user can perform a quick search by entering an identifier or test term, or select a specific query from a number of drop-down menus (data not shown). Alternatively, queries may be accessed by visiting the 'Queries and Tools' section of PlasmoDB (Figure 1A), which includes a grid displaying all available queries/searches. By using the queries and tools, a user can interrogate data in PlasmoDB—the third column of Table 1 includes example data-specific questions that are available.

When conducting queries with the purpose of combining results it may be useful to visualize the searches in a workflow environment where nodes are connected using different criteria ('and', 'or', 'not') (Figure 1B). In PlasmoDB this would be accomplished by performing a number of queries and subsequently combining the results in the 'query history' section (Figure 1C, middle screen shot). For example, one may be interested in identifying a short list of possible vaccine candidates. One possible way of accomplishing this would be by identifying all proteins predicted to be exported to the host cell in *P. falciparum*. There are three exported protein datasets in PlasmoDB and a union ('or' function) of all three results retrieves 405 genes (Figure 1B, steps 1 and 2). To restrict this list further, intersecting ('and' function) these results with genes that have no orthologs in mammals reduces the results to 321 genes (Figure 1B, Step 3). Next a user may further prune this list by intersecting the results with other queries, such as genes that are nonpolymorphic between a chloroquine sensitive (3D7) and resistant strain (Dd2). This cuts the number of candidates to 32 genes (Figure 1B, Step 4 and Figure 1C, right screen shot). Alternatively, one may be interested in the genes that have protein expression evidence in a particular stage in the parasite's life cycle (the results of an intersection with genes that have proteomic evidence in gametocyte yields 27 genes). Finally, examination of the list reveals several genes encoding for rifins (a family of clonally variant proteins expressed on the surface of infected red blood cells) (38), and a user may wish to investigate genes other than rifins—this can be accomplished by excluding ('not' operation) results of a keyword query using the term 'rifin' (Figure 1B, Step 5 and Figure 1C, left most panel). A user may examine the specific gene pages for more gene-specific details, download results with their associated data or log in (if they have not done so already) to ensure that their search strategy is saved for future examination.

FUTURE DIRECTIONS

It is expected that PlasmoDB will continue its data content and tool expansion as user needs require. We anticipate the incorporation of multiple new data sets including microarray, proteomic and specific parasite isolate data. Additionally, over the next few years we look forward to incorporating sequence data from a dramatically expanded *Plasmodium spp*. sequencing effort (http://www.genome.gov/26525388). In the coming year, we will also release a new user interface that will include a workflow-based search strategy page, similar to what is shown in Figure 1B, which we anticipate will provide a more biologically intuitive and dynamic experience for scientists accessing PlasmoDB and other EuPathDB sites.

ACKNOWLEDGEMENTS

The authors wish to thank members of the *Plasmodium* research community for their willingness to share genomic-scale data sets, often prior to publication, and for numerous comments and suggestions that have helped to improve the functionality of PlasmoDB. We also wish to thank Dr Akhil Vaidya for his valuable advice and continued support to PlasmoDB. We also thank past and present staff associated with the ApiDB-BRC project, and our research laboratory colleagues whose contributions have facilitated the creation and maintenance of this database resource.

FUNDING

Federal funds from the National Institute of Allergy and Infectious Diseases; National Institutes of Health; Department of Health and Human Services, under Contract No. HHSN266200400037C. Funding to pay the Open Access publication charges for this article was provided by this contract.

Conflict of interest statement. None declared.

REFERENCES

1. Phillips,R.S. (2001) Current status of malaria and potential for control. *Clin. Microbiol. Rev.*, **14**, 208–226.
2. Snow,R.W., Guerra,C.A., Noor,A.M., Myint,H.Y. and Hay,S.I. (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*, **434**, 214–217.
3. Singh,B., Kim Sung,L., Matusop,A., Radhakrishnan,A., Shamsul,S.S., Cox-Singh,J., Thomas,A. and Conway,D.J. (2004)

A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet*, **363**, 1017–1024.
4. Bahl,A., Brunk,B., Crabtree,J., Fraunholz,M.J., Gajria,B., Grant,G.R., Ginsburg,H., Gupta,D., Kissinger,J.C., Labo,P. *et al.* (2003) PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Res.*, **31**, 212–215.
5. Bozdech,Z., Llinas,M., Pulliam,B.L., Wong,E.D., Zhu,J. and DeRisi,J.L. (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.*, **1**, E5.
6. Le Roch,K.G., Zhou,Y., Blair,P.L., Grainger,M., Moch,J.K., Haynes,J.D., De La Vega,P., Holder,A.A., Batalov,S., Carucci,D.J. *et al.* (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, **301**, 1503–1508.
7. Llinas,M., Bozdech,Z., Wong,E.D., Adai,A.T. and DeRisi,J.L. (2006) Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res.*, **34**, 1166–1173.
8. Baum,J., Maier,A.G., Good,R.T., Simpson,K.M. and Cowman,A.F. (2005) Invasion by *P. falciparum* merozoites suggests a hierarchy of molecular interactions. *PLoS Pathog.*, **1**, e37.
9. Duraisingh,M.T., Voss,T.S., Marty,A.J., Duffy,M.F., Good,R.T., Thompson,J.K., Freitas-Junior,L.H., Scherf,A., Crabb,B.S. and Cowman,A.F. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell*, **121**, 13–24.
10. Stubbs,J., Simpson,K.M., Triglia,T., Plouffe,D., Tonkin,C.J., Duraisingh,M.T., Maier,A.G., Winzeler,E.A. and Cowman,A.F. (2005) Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes. *Science*, **309**, 1384–1387.
11. Hall,N., Karras,M., Raine,J.D., Carlton,J.M., Kooij,T.W., Berriman,M., Florens,L., Janssen,C.S., Pain,A., Christophides,G.K. *et al.* (2005) A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science*, **307**, 82–86.
12. Mair,G.R., Braks,J.A., Garver,L.S., Wiegant,J.C., Hall,N., Dirks,R.W., Khan,S.M., Dimopoulos,G., Janse,C.J. and Waters,A.P. (2006) Regulation of sexual development of *Plasmodium* by translational repression. *Science*, **313**, 667–669.
13. Tarun,A.S., Peng,X., Dumpit,R.F., Ogata,Y., Silva-Rivera,H., Camargo,N., Daly,T.M., Bergman,L.W. and Kappe,S.H. (2008) A combined transcriptome and proteome survey of malaria parasite liver stages. *Proc. Natl Acad. Sci. USA*, **105**, 305–310.
14. Florent,I., Charneau,S. and Grellier,P. (2004) *Plasmodium falciparum* genes differentially expressed during merozoite morphogenesis. *Mol. Biochem. Parasitol.*, **135**, 143–148.
15. Watanabe,J., Wakaguri,H., Sasaki,M., Suzuki,Y. and Sugano,S. (2007) Comparasite: a database for comparative study of transcriptomes of parasites defined by full-length cDNAs. *Nucleic Acids Res.*, **35**, D431–D438.
16. Gunasekera,A.M., Patankar,S., Schug,J., Eisen,G., Kissinger,J., Roos,D. and Wirth,D.F. (2004) Widespread distribution of antisense transcripts in the *Plasmodium falciparum* genome. *Mol. Biochem. Parasitol.*, **136**, 35–42.
17. Gunasekera,A.M., Patankar,S., Schug,J., Eisen,G. and Wirth,D.F. (2003) Drug-induced alterations in gene expression of the asexual blood forms of *Plasmodium falciparum*. *Mol. Microbiol.*, **50**, 1229–1239.
18. Patankar,S., Munasinghe,A., Shoaibi,A., Cummings,L.M. and Wirth,D.F. (2001) Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. *Mol. Biol. Cell*, **12**, 3114–3125.
19. Florens,L., Liu,X., Wang,Y., Yang,S., Schwartz,O., Peglar,M., Carucci,D.J., Yates,J.R. III and Wub,Y. (2004) Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol. Biochem. Parasitol.*, **135**, 1–11.
20. Florens,L., Washburn,M.P., Raine,J.D., Anthony,R.M., Grainger,M., Haynes,J.D., Moch,J.K., Muster,N., Sacci,J.B., Tabb,D.L. *et al.* (2002) A proteomic view of the *Plasmodium falciparum* life cycle. *Nature*, **419**, 520–526.
21. Khan,S.M., Franke-Fayard,B., Mair,G.R., Lasonder,E., Janse,C.J., Mann,M. and Waters,A.P. (2005) Proteome analysis of separated male and female gametocytes reveals novel sex-specific. *Plasmodium Biol. Cell*, **121**, 675–687.
22. Su,X., Ferdig,M.T., Huang,Y., Huynh,C.Q., Liu,A., You,J., Wootton,J.C. and Wellems,T.E. (1999) A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science*, **286**, 1351–1353.
23. Jeffares,D.C., Pain,A., Berry,A., Cox,A.V., Stalker,J., Ingle,C.E., Thomas,A., Quail,M.A., Siebenthall,K., Uhlemann,A.C. *et al.* (2007) Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nat. Genet.*, **39**, 120–125.
24. Mu,J., Awadalla,P., Duan,J., McGee,K.M., Keebler,J., Seydel,K., McVean,G.A. and Su,X.Z. (2007) Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nat. Genet.*, **39**, 126–130.
25. Volkman,S.K., Sabeti,P.C., DeCaprio,D., Neafsey,D.E., Schaffner,S.F., Milner,D.A. Jr, Daily,J.P., Sarr,O., Ndiaye,D., Ndir,O. *et al.* (2007) A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.*, **39**, 113–119.
26. Chen,F., Mackey,A.J., Stoeckert,C.J. Jr and Roos,D.S. (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.*, **34**, D363–D368.
27. Date,S.V. and Stoeckert,C.J. Jr (2006) Computational modeling of the *Plasmodium falciparum* interactome reveals protein function on a genome-wide scale. *Genome Res.*, **16**, 542–549.
28. LaCount,D.J., Vignali,M., Chettier,R., Phansalkar,A., Bell,R., Hesselberth,J.R., Schoenfeld,L.W., Ota,I., Sahasrabudhe,S., Kurschner,C. *et al.* (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature*, **438**, 103–107.
29. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
30. Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bradley,P., Bork,P., Bucher,P., Cerutti,L. *et al.* (2005) InterPro, progress and status in 2005. *Nucleic Acids Res.*, **33**, D201–D205.
31. Ginsburg,H. (2006) Progress in in silico functional genomics: the malaria Metabolic Pathways database. *Trends Parasitol.*, **22**, 238–240.
32. Bendtsen,J.D., Nielsen,H., von Heijne,G. and Brunak,S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.*, **340**, 783–795.
33. Krogh,A., Larsson,B., von Heijne,G. and Sonnhammer,E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.
34. Foth,B.J., Ralph,S.A., Tonkin,C.J., Struck,N.S., Fraunholz,M., Roos,D.S., Cowman,A.F. and McFadden,G.I. (2003) Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science*, **299**, 705–708.
35. Hiller,N.L., Bhattacharjee,S., van Ooij,C., Liolios,K., Harrison,T., Lopez-Estrano,C. and Haldar,K. (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science*, **306**, 1934–1937.
36. Marti,M., Baum,J., Rug,M., Tilley,L. and Cowman,A.F. (2005) Signal-mediated export of proteins from the malaria parasite to the host erythrocyte. *J. Cell Biol.*, **171**, 587–592.
37. Marti,M., Good,R.T., Rug,M., Knuepfer,E. and Cowman,A.F. (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science*, **306**, 1930–1933.
38. Kyes,S.A., Rowe,J.A., Kriek,N. and Newbold,C.I. (1999) Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA*, **96**, 9333–9338.