**BMC Bioinformatics**

**RESEARCH**                                                                                      **Open Access**

# A novel subnetwork alignment approach predicts new components of the cell cycle regulatory apparatus in *Plasmodium falciparum*

Hong Cai[1†], Changjin Hong[2†], Timothy G Lilburn[3†], Armando L Rodriguez[1,4], Sheng Chen[2], Jianying Gu[5], Rui Kuang[2*], Yufeng Wang[1,6*]

## Abstract

**Background:** According to the World Health organization, half the world's population is at risk of contracting malaria. They estimated that in 2010 there were 219 million cases of malaria, resulting in 660,000 deaths and an enormous economic burden on the countries where malaria is endemic. The adoption of various high-throughput genomics-based techniques by malaria researchers has meant that new avenues to the study of this disease are being explored and new targets for controlling the disease are being developed. Here, we apply a novel neighborhood subnetwork alignment approach to identify the interacting elements that help regulate the cell cycle of the malaria parasite *Plasmodium falciparum*.

**Results:** Our novel subnetwork alignment approach was used to compare networks in *Escherichia coli* and *P. falciparum*. Some 574 *P. falciparum* proteins were revealed as functional orthologs of known cell cycle proteins in *E. coli*. Over one third of these predicted functional orthologs were annotated as "conserved *Plasmodium* proteins" or "putative uncharacterized proteins" of unknown function. The predicted functionalities included cyclins, kinases, surface antigens, transcriptional regulators and various functions related to DNA replication, repair and cell division.

**Conclusions:** The results of our analysis demonstrate the power of our subnetwork alignment approach to assign functionality to previously unannotated proteins. Here, the focus was on proteins involved in cell cycle regulation. These proteins are involved in the control of diverse aspects of the parasite lifecycle and of important aspects of pathogenesis.

## Background

Written descriptions of the symptoms of malaria have existed for over 4,000 years and evidence for the existence of the genus *Plasmodium* has been recovered from amber approximately 30 million years old [1]. Thus, the disease has probably evolved alongside its hosts since the emergence of the first humans in Africa. In 2010, it was estimate that 660,000 people died from malaria. This estimate probably represents a conservative number, as reporting of the disease is extremely variable from one region to another; generally, the regions with the highest incidence of malaria also have the weakest mechanisms for reporting and recording cases.

Malaria is caused by protozoan parasites from the Genus *Plasmodium*. Different species tend to infect different host species. Five species infect humans; the two most widespread species are *P. vivax* and *P. falciparum*. The latter species is the most lethal. *P. falciparum* has a complex life cycle that spans the arthropod vector and human host. Upon transfer from the vector to the human host, the parasite first infects the liver. After maturation in the liver, the parasite infects red blood cells. In this so-called RBC stage, the symptoms of malaria become acute.

* Correspondence: kuang@cs.umn.edu; yufeng.wang@utsa.edu
† Contributed equally
[1]Department of Biology, University of Texas at San Antonio, San Antonio, TX 78249, USA
[2]Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA
Full list of author information is available at the end of the article

A number of antimicrobial drugs have been developed over the years, notably chloroquine and artemisinin. However, in the past decades, the effectiveness of all these drugs has been significantly reduced due to the evolution of drug-resistant parasites, with the exception of artemisinin. Recently, however, evidence has emerged that resistance to artemisinin has appeared and is beginning to spread. Therefore, it is essential that new drug targets be identified and the development of new genomics-based technologies is key to this task. Genome sequences from *P. falciparum* [2] and other *Plasmodium* spp. [3-7] have been completed and these have facilitated numerous studies on, for example, parasite transcription [8-19], translation [20-29], metabolism [30-34], protein-protein interactions [35-38], and epigenetic regulation [39-42]. The data from these studies have, in turn, laid the groundwork for systems biology oriented studies of the networks associated with parasite development, survival, pathogenesis, and virulence [43-46].

Network alignment is a popular systems biology method [47-55]. However, because the malaria parasite is only distantly related to other, more completely understood model organisms, the utility of this approach may be cast in doubt. About 60% of the open reading frames in *P. falciparum* are annotated as "hypothetical proteins" [2] simply because homology transfer of information about individual proteins is not possible across extended evolutionary distances. To tackle this problem, we recently developed a neighborhood subnetwork alignment algorithm [56], which is focused on the similarities between functional modules, in other words, on the interactions among proteins rather than on individual proteins. We define a neighborhood subnetwork as the set of nodes (proteins) reachable from a central protein via a small number of edges in a protein-protein interaction (PPI) network. A proof-of-concept study predicted previously unrecognized transcriptional regulators involved in diverse facets of the parasite life cycle [43]. In this paper, we use the subnetwork alignment approach to uncover candidate proteins with roles in cell cycle regulation, several of which are potential drug targets. As our knowledge of the mechanics of the cell cycle deepens, so will our ability to influence parasite survival in the host and our ability to identify key drug targets.

## Results and discussion
### Neighborhood subnetwork alignments predicted 574 proteins that are associated with cell cycle regulation in malaria parasite
The cell cycle of the malaria parasite differs significantly from that of other model eukaryotic organisms. There is no direct correspondence between schizogony, during which the parasite undergoes multiplication, and the typical G1, S, G2 and M phases of the cell cycle in crown eukaryotes. In addition, the parasite's cell cycle features asynchronous nuclear divisions, organellar segregation, and morphogenesis of daughter merozoites. A thorough sequence similarity-based search by Doerig and Chakrabarti predicted a list of proteins that might be involved in the cell cycle [57], including cyclins, cyclin-dependent kinases, proteins critical for cell division and signal transduction. In a previous study, we used a variational Bayesian expectation maximization (VBEM) approach to reveal the dynamics of the parasite cell cycle network, and to infer regulatory relationships based on time-series transcriptomic data [58]. The results from that study exposed gaps in our cell cycle network model. Here we use our subnetwork alignment approach to try and fill these gaps.

We predicted that 574 proteins in *P. falciparum* were functional orthologs of known cell cycle proteins in *E. coli* (Additional File 1). Over 34% of these predicted functional orthologs were annotated as "conserved *Plasmodium* proteins" or "putative uncharacterized proteins" of unknown function.

### The set of functional orthologs is involved in key biological processes
Table 1 shows representative functional categories predicted for the cell cycle-associated protein set as revealed by Gene Ontology (GO) enrichment analysis. These functional categories are part of some of the most important mechanisms governing the growth and survival of the parasite. Some of the more interesting functional groups are discussed in the following sections.

#### 1. Cyclin
Our subnetwork alignment approach predicted PFL1330c to be a putative cyclin [58]. Cyclins are a family of proteins with expression levels that oscillate during the cell cycle; the synthesis and degradation of cyclins control the activity of cyclin-dependent kinases and accurate transition of key cell cycle points. Yeast two-hybrid (Y2H) experiments [37] have shown that PFL1330c has physical interaction with an apical sushi protein (ASP) (PFD0295c), which has an adhesive "sushi" domain and thought to have a role in the merozoite invasion process.

#### 2. Kinases
Signal transduction plays a key role in managing the complexity of the cell cycle [59,60]. Figure 1 shows eight kinases (in yellow) that were predicted by the subnetwork alignments and the proteins that are directly associated with them. Three protein kinases have been implicated in cell cycle regulation:

(1) PfMAP1 (PF14_0294) is a homolog of mitogen-activated protein kinase (MAPK) [61]. This kinase is believed to be a central member of the MAPKKK cascade and may be related to parasite responses to a variety of exogenous or endogenous stimuli or environmental stresses. PfMAP1 has three PPI partners: (a) a serine/threonine protein

**Table 1 Representative *P. falciparum* proteins that were predicted to be involved in cell cycle regulatory network.**

| Functional category | PlasmoDB accession number | Annotation |
| --- | --- | --- |
| Cyclin | PFL1330c | Cyclin-related protein, Pfcyc-2 |
| Cell differentiation | PFE0375w | cell differentiation protein, putative (CAF40) |
| Chromosome organization | PFE0450w | Chromosome condensation protein, putative |
|  | PF11_0062 | Histone H2B |
| Mitosis | PF13_0050 | HORMA domain protein, putative |
| DNA repair | MAL7P1.145 | Mismatch repair protein pms1 homologue, putative; |
|  | PF10_0114 | DNA repair protein RAD23, putative |
|  | PF08_0126 | DNA repair protein rad54, putative |
| DNA replication | PF07_0023 | DNA replication licensing factor mcm7 homologue, putative |
|  | PFL0580w | DNA replication licensing factor MCM5, putative |
|  | MAL7P1.21 | Origin recognition complex subunit 2, putative |
|  | PFE1345c | Minichromosome maintenance protein 3, putative |
| Regulation of cell cycle | PF07_0047 | AAA family ATPase, CDC48 subfamily (Cdc48) |
|  | PFL1925w | Cell division protein FtsH, putative |
| Protein phosphorylation | PFC0105w | Serine/threonine protein kinase, putative |
|  | MAL13P1.278 | Serine/threonine protein kinase, putative |
|  | PF14_0294 | Mitogen-activated protein kinase 1 |
|  | PFC0755c | Protein kinase, putative |
|  | PF11_0464 | Ser/Thr protein kinase, putative |
|  | PF11_0156 | Ser/Thr protein kinase |
|  | PF11_0239 | Calcium-dependent protein kinase, putative |
|  | PFL1370w | NIMA-related protein kinase, Pfnek-1 |
| Proteolysis | PF14_0517 | Peptidase, putative |
|  | MAL13P1.184 | Endopeptidase, putative |
|  | PFL1635w | Ulp1 protease, putative |
|  | PF10_0150 | Methionine aminopeptidase |
| Cytoskeleton | MAL8P1.146 | filament assembling protein, putative |
| Heat shock | PFI0875w | Heat shock protein 70 (HSP70) homologue |
|  | PFL0565w | Heat shock protein DNAJ homologue Pfj4 |
|  | PF11_0351 | Heat shock protein hsp70 homologue |
|  | PF11_0188 | Heat shock protein 90, putative |
|  | PF07_0029 | Heat shock protein 86 |
|  | PF08_0054 | Heat shock 70 kDa protein |
|  | PFB0595w | Heat shock 40 kDa protein, putative |
|  | PFI0355c | ATP-dependent heat shock protein, putative |
| Pathogenesis | PFC0005w | Erythrocyte membrane protein 1, PfEMP1 |
|  | PFI0005w | Erythrocyte membrane protein 1, PfEMP1 |
|  | PFD0005w | Erythrocyte membrane protein 1, PfEMP1 |
|  | PF08_0103 | Erythrocyte membrane protein 1, PfEMP1 |
|  | PFL0935c | Erythrocyte membrane protein 1, PfEMP1 |
|  | PFL0005w | Erythrocyte membrane protein 1, PfEMP1 |
|  | PFB1055c | Erythrocyte membrane protein 1, PfEMP1 |
|  | PFI1830c | Erythrocyte membrane protein 1, PfEMP1 |
| Microtubule cytoskeleton organization and activity | PFC0165w | Spindle pole body protein, putative |
|  | PF07_0104 | Kinesin-like protein, putative |
| Transcriptional regulation | PF10_0143 | Transcriptional coactivator ADA2 (ADA2) |
|  | PFD0985w | AP2/ERF domain-containing protein PFD0985w |
|  | PFL1085w | Transcription factor with AP2 domain, putative |
|  | PF11_0442 | Transcription factor with AP2 domain, putative |

**Table 1 Representative *P. falciparum* proteins that were predicted to be involved in cell cycle regulatory network.** (Continued)

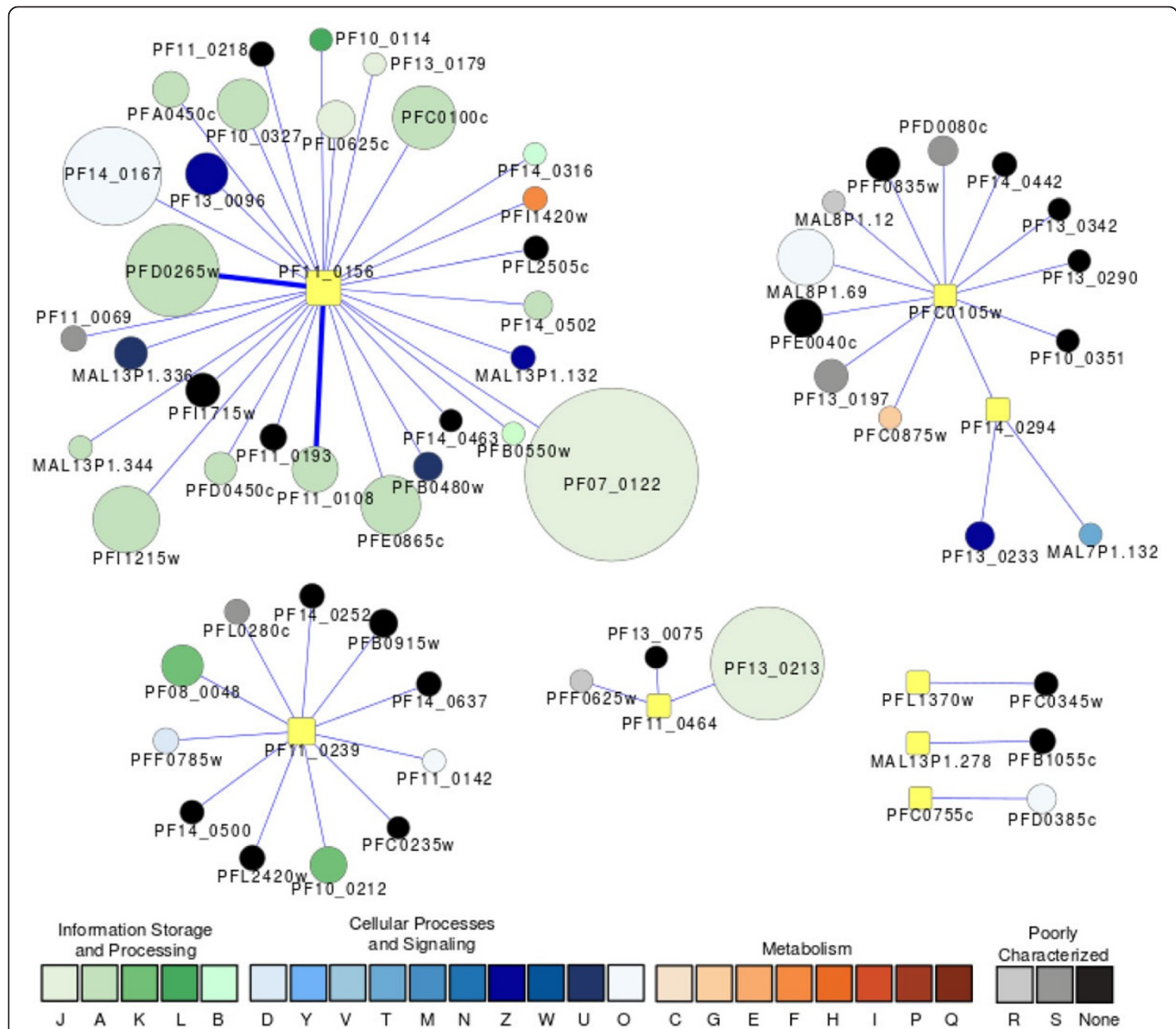| | |
|---|---|
| PFE0840c | Transcription factor with AP2 domain, putative |
| PF07_0126 | Transcription factor with AP2 domain, putative |
| PF10_0075 | Transcription factor with AP2 domain, putative |
| PFL1900w | Transcription factor with AP2 domain, putative |
| PFL0465c | Zinc finger transcription factor (Krox1) |



**Figure 1 A graph showing the proteins associated with kinases predicted to be involved in cell cycle regulation**. Square nodes represent the kinases. Node size is proportional to the degree of the connectivity of the node. Nodes are colored according to their functional classification in the eggNOG database [79]. The COG categories are [80] (J) Translation, ribosomal structure and biogenesis, (A) RNA processing and modification, (K) Transcription, (L) Replication, recombination and repair, (B) Chromatin structure and dynamics, (D) Cell cycle control, cell division, chromosome partitioning, (Y) Nuclear structure, (V) Defense mechanisms, (T) Signal transduction mechanisms, (M) Cell wall/membrane/envelope biogenesis, (N) Cell motility, (Z) Cytoskeleton, (W) Extracellular structures, (U) Intracellular trafficking, secretion, and vesicular transport, (O) Posttranslational modification, protein turnover, chaperones, (C) Energy production and conversion, (G) Carbohydrate transport and metabolism, (E) Amino acid transport and metabolism, (F) Nucleotide transport and metabolism, (H) Coenzyme transport and metabolism, (I) Lipid transport and metabolism, (P) Inorganic ion transport and metabolism, (Q) Secondary metabolites biosynthesis, transport and catabolism, (R) General function prediction only, and (S) Function unknown. Confidence scores for the interactions among the nodes (S values from STRING [81]) were divided into three groups - low (0.150-0.399), medium (0.400-0.700) and high (0.701-0.999); the groups are represented by thin, medium and heavy lines, respectively.

kinase (SRPK1) (PFC0105w). PfSRPK plays a role in mRNA splicing machinery [62]. Gene disruption of SRPK in the rodent parasite *P. berghei* suggested that it is essential during male gamete formation [63]. (b) myosin A (PF13_0233) is a component in the linear motor that promotes merozoite motility in invasion. (c) MAL7P1.132, a conserved *Plasmodium* protein of unknown function. This protein was recently annotated as a putative kinase [64].

(2) PfNek-1(PFL1370w) encodes a NIMA-related kinase and it is considered to be a potential antimalarial target. A recent study based on reverse genetics showed that it is required for the asexual cycle in red blood cells and it has sexual specificity (expression in male gametocyte) [65]. PfNek-1 is shown by yeast 2-hybrid assay to pool with a conserved hypothetical protein PFC0345w. Both proteins have abundant expression at the schizont stage.

(3) cdc2-related protein kinase 4 (CRK4) (PFC0755c) [57], was observed as a phospho-protein in the schizont stage of *P. falciparum*-infected red blood cells. Y2H showed that it has a direct interaction with an AAA family ATPase.

The most highly connected kinase predicted to be involved in the cell cycle is the serine/threonine protein kinase PfCLK-3 (PF11_0156) with 28 association partners. Ten proteins were pooled by Y2H experiments [37], including a rhoptry neck protein 3 (RON3), a splicing factor 3A subunit, eukaryotic translation initiation factor 3 subunit 10, a chloroquine resistance marker protein (CRMP), syntaxin involved in vesicle exocytosis, an export protein, and five conserved hypothetical proteins, indicating PfCLK-3's involvement in merozoite invasion, splicing, translation and trafficking. Global kinome analysis suggested that PfCLK3 is likely to be essential for parasite schizogony in RBCs [28].

A calcium-dependent protein kinase 6 (PfCDPK6) (PF11_0239) was predicted to be involved in cell cycle regulation by subnetwork alignment. Previous phenotypic analysis showed that CDPK6 plays a role in sporozoite formation and invasion of hepatocytes [66]. This kinase is associated with 11 other proteins verified by Y2H assays. Two of the association partners are likely involved in cell cycle regulation as well: a putative Ndc80 protein functions in spindle checkpoint signaling for kinetochore organization and movements, and a putative Snf2-related CBP activator (SRCAP) for base excision repair and chromosome remodeling. PfCDPK6 is also associated with PfBet1 in SNARE complex for secretion, a putative protein localized to rhoptry that might be related to merozoite invasion process, a liver-stage antigen, a ubiquitin domain containing protein, and five hypothetical proteins.

The functional roles of other predicted kinases are largely unknown. PF11_0464 is a putative serine/threonine protein kinase. A gene disruption attempt suggested that it is likely essential for the parasite RBC stage [28]. This protein is associated with two proteins required for 60S ribosomal subunit biogenesis (60S ribosomal protein L6-2 and nucleolar GTP-binding protein 1), and a pseudogene of surface-associated interspersed gene 13.1 (SURFIN13.1), which was implicated in the invasion process. MAL13P1.278 (PfArk3) is a putative serine/threonine kinase in the aurora-related kinase (ARK) family. This family of kinases has been implicated in regulation of endocytosis and of the actin skeleton [67]. PfArk3 has a weak association with an erythrocyte membrane protein 1, PfEMP1 (PFB1055) that may be related to mitotic recombination.

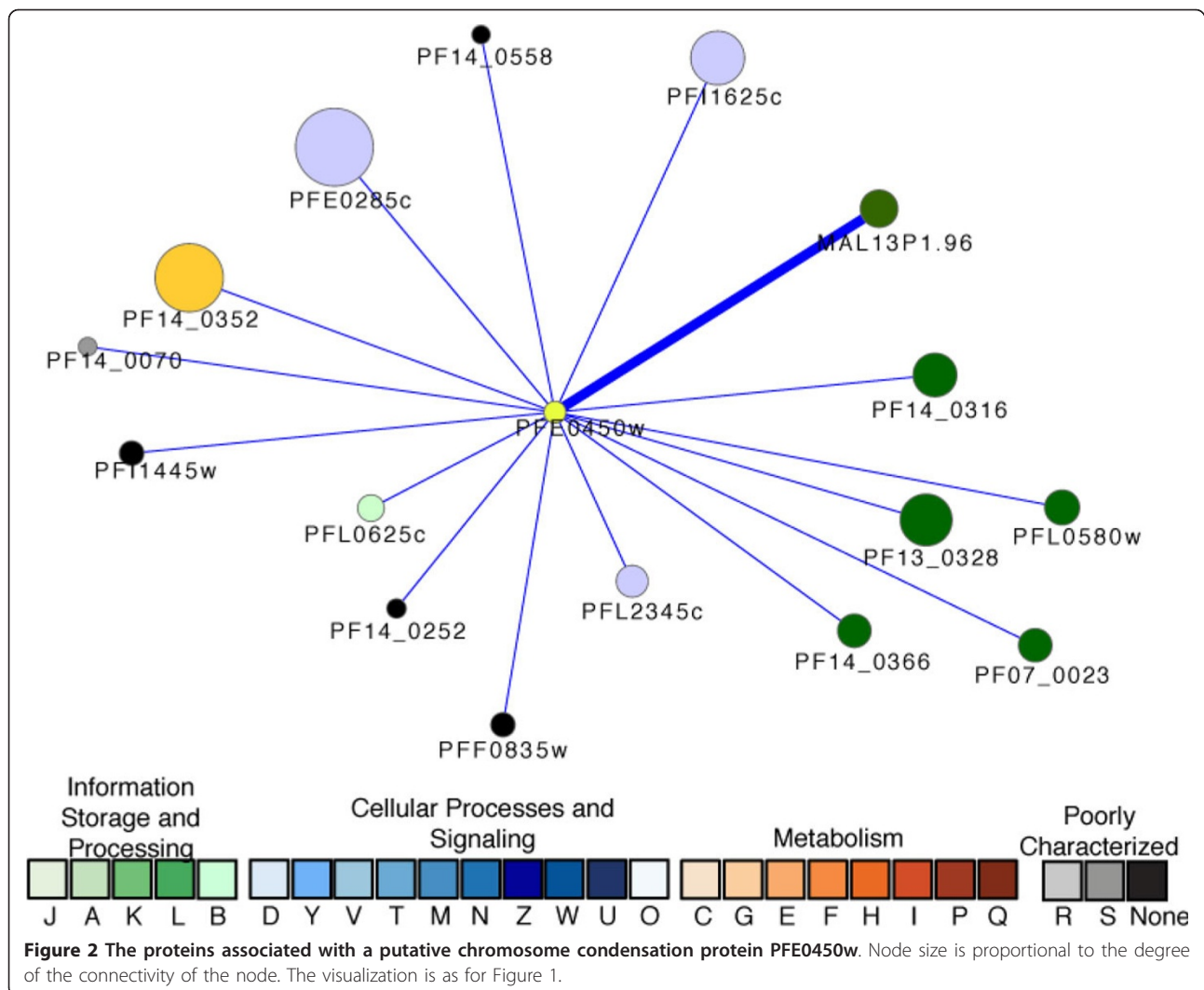### 3. Proteins implicated in cell division, chromosome organization, and DNA replication

Our analysis has implicated a number of other predicted proteins in the cell division, mitosis, chromosome organization, and DNA replication processes. PFE0450w, a putative chromosome condensation protein that forms part of the ATP-dependent chromatin remodeling complex [68], was predicted to be associated with cell cycle regulation. As shown in Figure 2, are 16 proteins associated with PFE0450w. Eight of these associations have been verified by Y2H, a set that includes two tat-binding proteins pertinent to proteasome activities, a pre-mRNA splicing factor, an eukaryotic translation initiation factor 3 subunit 10, and three conserved *Plasmodium* proteins with unknown function. Perhaps the most important association suggested by our analysis is its link with the high molecular weight rhoptry protein 2 (RhopH2). Rhop2 is localized in the rhoptries of schizonts and plays a role in cytoadherence and merozoite invasion of the red blood cell [69]. Several key components including DNA replication licensing factors and an origin recognition complex subunit were predicted by our subnetwork alignment.

### 4. DNA repair proteins

The cell cycle is also involved in involving DNA repair mechanisms that ensure genome integrity. A putative DNA repair protein RAD23 (PF10_0114) was predicted to have 92 protein-protein association partners (Figure 3), 22 of which have been demonstrated to be direct Y2H physical interactions. This protein is a member of an escort complex for proteasome-mediated degradation of non-native ER proteins. Other suggested interactors with RAD23 include heat shock chaperone proteins, ATP-dependent proteases, serine-threonine kinases, and secreted proteins that have been implicated in stress responses, signaling cascades, and protein sorting and trafficking.

### 5. Transcriptional regulators

Seven parasite-specific ApiAP2 transcription factors were predicted to have a role in cell cycle regulation, underscoring the importance of transcriptional regulation. ApiAP2 proteins are gaining recognition as attractive drug targets due to their critical roles in the parasite life cycle and their distant evolutionary relationship to the host, implying a diminished possibility of side-effects for

**Figure 2 The proteins associated with a putative chromosome condensation protein PFE0450w**. Node size is proportional to the degree of the connectivity of the node. The visualization is as for Figure 1.

humans [70]. The ApiAP2 protein with the highest degree of connectivity in the cell cycle regulatory network is PFD0985w (Figure 4). Its 17 association partners play versatile roles in epigenetic regulation, kinetochore organization, host cell entry and adhesion, secretion, and protein degradation by the ubiquitin-proteasome system [45]. The roles of another ApiAP2 protein (PF07_0126) can be inferred from its associations with 15 proteins that are related to transcriptional regulation, chromatin remodeling, replication, and repair. This protein has interactions with multiple signaling molecules including a calcium-dependent protein kinase and a ligand protein in the 14-3-3 family.

The involvement of PF10_0075 in ApiAP2 in cell cycle regulation is indicated by its Y2H interactions with another ApiAp2 protein (MAL8P1.153), a histone acetyltransferase GCN5 (PF08_0034), which is important for histone modification and chromatin remodeling [71], a DNA excision repair protein rhp16 (PFL2440w), actin (PFL2215w) and a
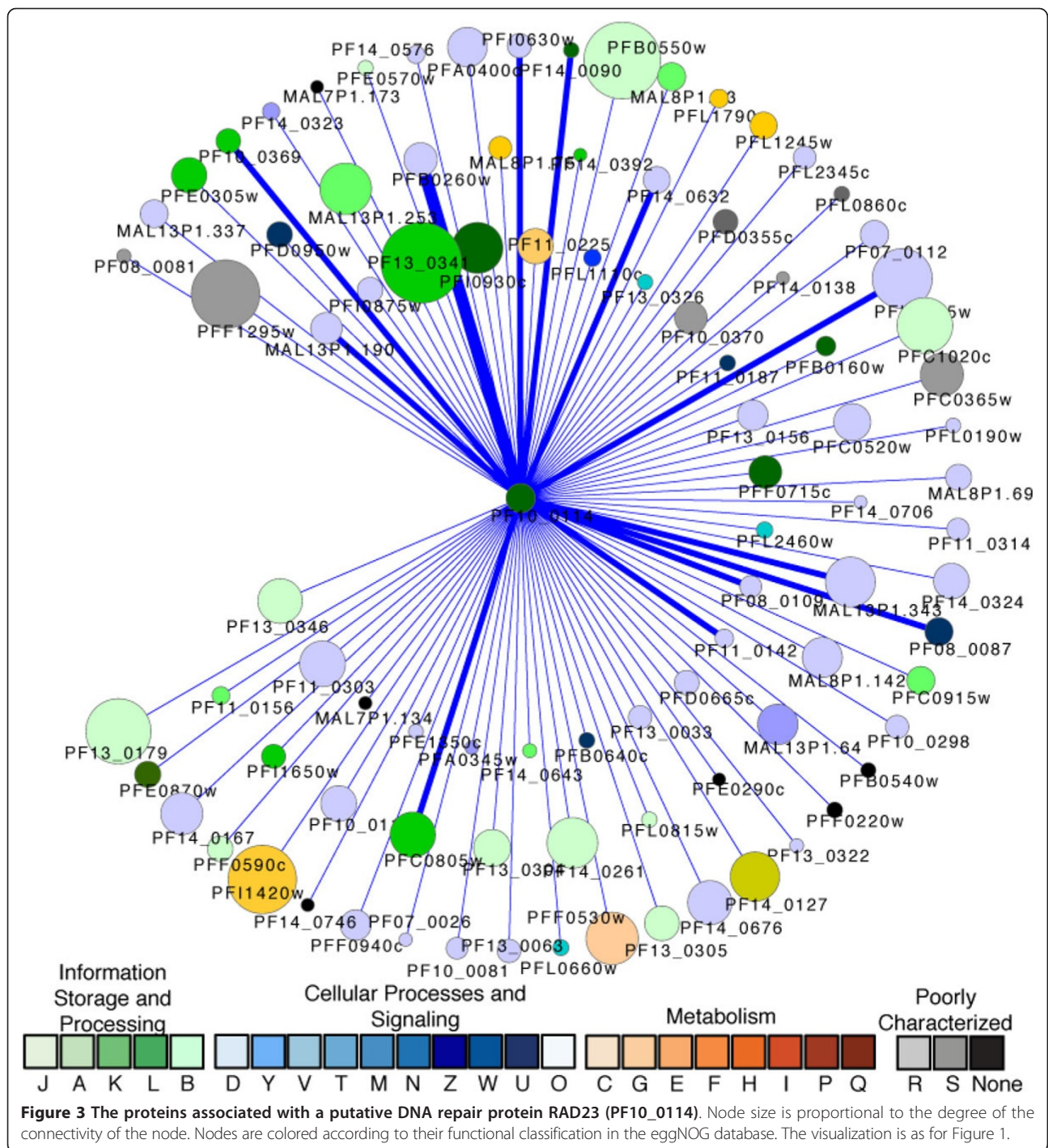
putative kelch protein whose ortholog was implicated in cytoskeletal function in Atlantic horseshoe crab, *Limulus polyphemus* [72] (Figure 4).

**6. Surface antigens**

A group of surface antigens in the *Plasmodium falciparum* erythrocyte membrane protein (PfEMP1) family (Table 1) were predicted to be associated with the cell cycle. Encoded by the var gene, PfEMP1 is one of the most abundant protein families in *P. falciparum*. Its polymorphic nature leads to antigenic variation, allowing the parasite to successfully evade the human immune systems, thus contributing to pathogenicity and virulence.

**Conclusions**

We have previously developed a neighborhood subnetwork alignment approach and here we apply this method to predict the network components involved in cell cycle regulation. The network components identified included cyclins, kinases, transcriptional regulators, and cell surface

**Figure 3 The proteins associated with a putative DNA repair protein RAD23 (PF10_0114)**. Node size is proportional to the degree of the connectivity of the node. Nodes are colored according to their functional classification in the eggNOG database. The visualization is as for Figure 1.
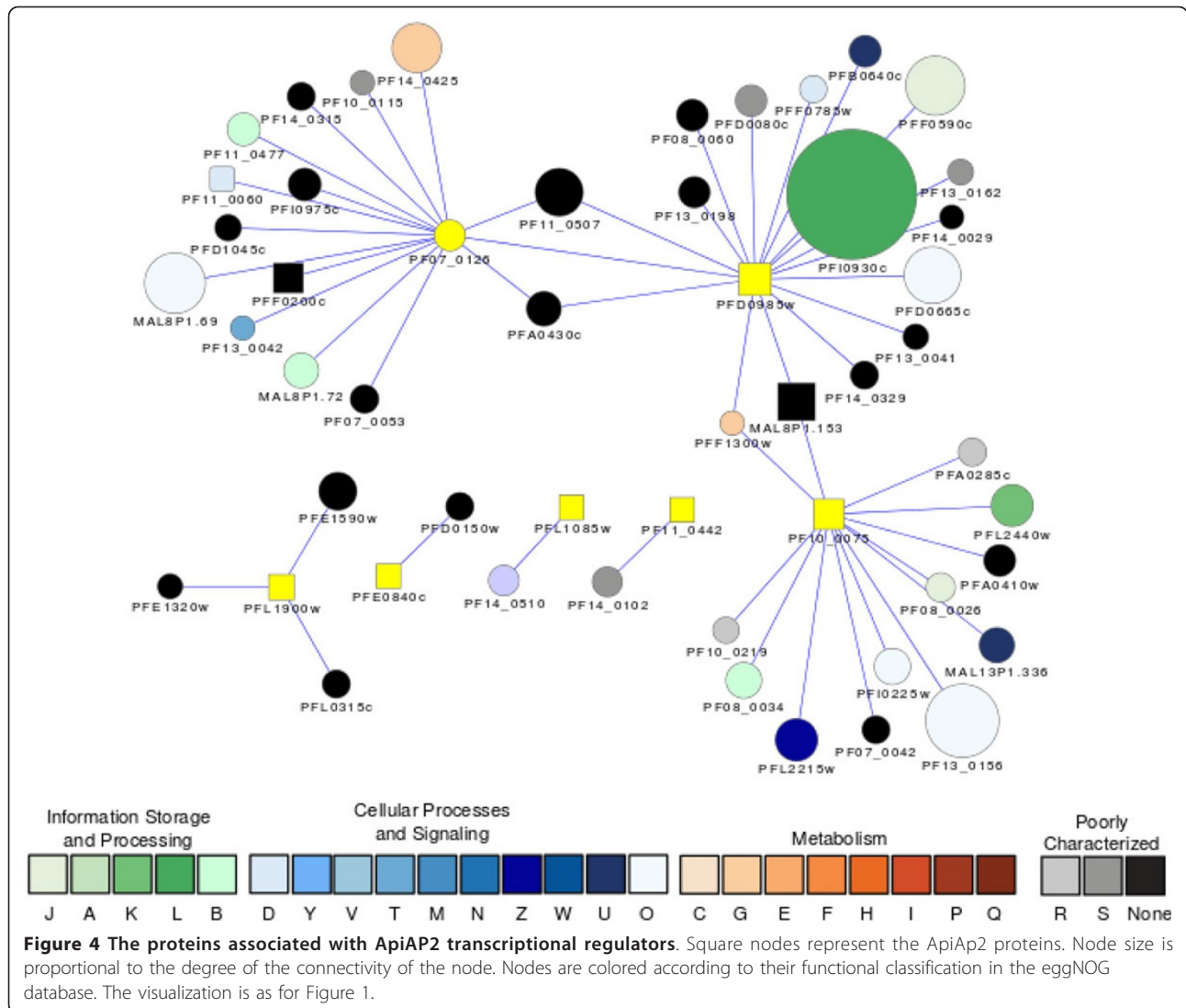
antigens, among others. Some of these are obvious and have already been confirmed by experimental approaches, such as yeast two-hybrid experiments. This validates our approach as a useful tool for *in silico* prediction of previously unrecognized interactors in cell cycle regulation and suggests that the expanded set of interactors discussed here form a new set of potential targets for drugs or therapies.

## Methods

### Subnetwork querying by neighborhood alignments

The prediction of functional orthologs for the *P. falciparum* proteins has been structured as a subnetwork querying problem. *Network Querying* is a technique that searches a large "target" network of an organism to find subnetwork regions that look similar to a given query network of another organism [73,74]. The "query" network

**Figure 4 The proteins associated with ApiAP2 transcriptional regulators**. Square nodes represent the ApiAp2 proteins. Node size is proportional to the degree of the connectivity of the node. Nodes are colored according to their functional classification in the eggNOG database. The visualization is as for Figure 1.

that we are searching against "target" network is the well-studied functional module in a model organism. Network Querying allows us to predict similar modules in the less studied target organism, providing a way to relate biological knowledge of functionality across organisms [75]. Previously, we applied a neighborhood alignment method for subnetwork querying to predict novel transcriptional regulators with versatile roles in the parasite life cycle [43]. We adopted the same method to identify proteins involved in cell cycle regulation.

First a set of proteins related to cell cycle regulation (GO:0007049: cell cycle) in *E. coli* were mapped onto the its own PPI network. For each cell cycle protein a set of "neighbors" was selected, creating a subnetwork, and by inference, a network of subnetworks in the query network. Conversely using the same technique, each *P. falciparum* protein was mapped into its own PPI network, and a sub-network of neighbors was constructed. To construct

neighborhood subnetworks of comparable size for alignment, proteins that are $k$ hops from the central were included and $k$ was chosen such that the neighbor size was under 500, unless the central protein had more than 500 neighbors.

After obtaining the neighborhood subnetworks for both the *E. coli* cell cycle proteins and the *P. falciparum* proteins, the *E. coli* subnetworks were combinatorically aligned against the *P. falciparum* subnetworks. The central protein of the best-aligned *P. falciparum* subnetwork was labeled a functional ortholog of the proteins involved in cell cycle regulation in *E. coli* .

Analysis to determine how well the *P. falciparum* neighborhood subnetworks aligned with the *E. coli* neighborhood subnetworks was done by assigning a numerical score for each alignment by a shortest-path graph kernel to measure the similarity between two labeled networks [76]. To optimize the graph kernel for this specific use case; only

paths between the central protein and other subnetwork proteins are counted. Each shortest path through the central protein characterizes the functional role of the protein in the chained molecular activities along the path. As shown in Figure 5, given two subnetworks $S_p$ with central protein $p$ and $S_q$ with central protein $q$, the shortest path similarity function is defined as follows,

$$K\left(S_q, S_p\right) = \frac{1}{|S_q| + |S_p|} \prod_{\forall (i1,i2) \in S_q} B((i1, i2) S_p)$$

Where

$$B\left((i1, i2), S_p\right) = \max_{\forall (j1,j2) \in S_p} \frac{2E(i1, j1)E(i2, j2)}{dist(i1, i2) + dist(j1, j2)},$$

$E(x, y) = exp(-\frac{Eval(x, y)}{\sigma})$ with the normalization parameter $\sigma = 10$ measures the sequence similarity between proteins × and y based on the E-value of the sequence alignment, and dist(x,y) is the length of the shortest path connecting proteins x and y in the PPI subnetwork. The computation was done on a $-\log_{10}$ scale. The method outlined here takes each pair of proteins (i1, i2) from one subnetwork and seeks the maximum ratio of sequence similarity with respect to the closeness (shortest path through the central protein) of the networks, in order to identify proteins (j1,j2) in the target subnetwork. From this algorithm, a subnetwork alignment score is obtained by, collecting the shortest paths between two neighborhood subnetworks, getting an alignment score for each pair of proteins, and totaling
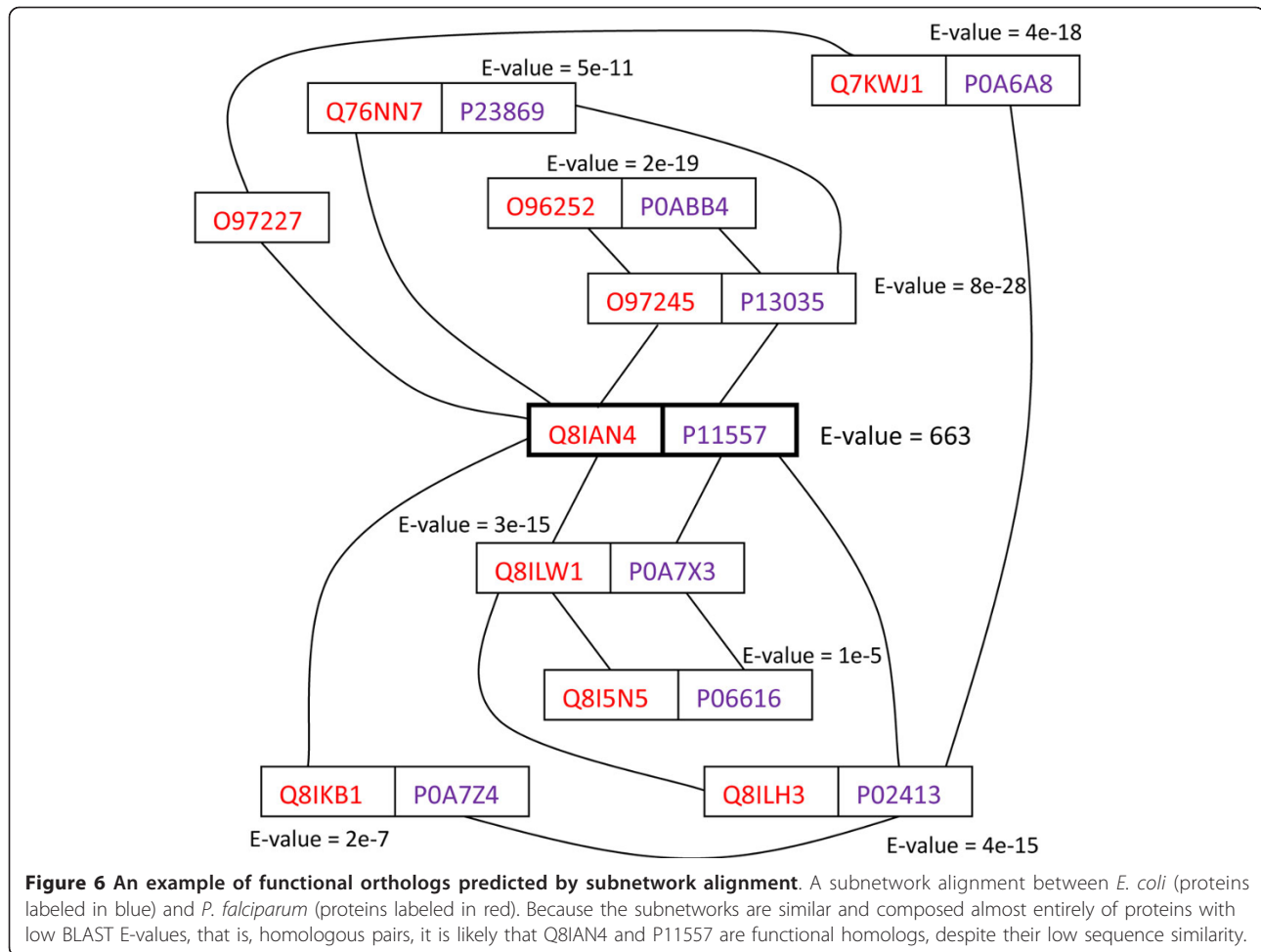
all of the alignment values. This approach allows for the summarization of the functional coherence, and distance between two central proteins, into a numerical score by way of evaluating the sequence similarity and the role of the central protein between two subnetworks.

An example of how the subnetwork alignment approach is used to predict functional orthologs is shown in Figure 6 (annotations are shown in Additional File 2). Although the *P. falciparum* protein encoded by locus PF08_0126 (Uniprot ID Q8IAN4, a putative DNA repair protein rad54) and *E. coli* protein DamX (P11557) showed no significant homology, they did share eight pairs of sequence and network orthologs when their PPI networks were aligned. DamX has been shown to directly or indirectly interfere with cell division in *E. coli* [77,78]. Despite their low sequence similarity (BLAST E-value 663), the network alignment evidence suggests that DamX and Q8IAN4 are likely to be functional orthologs.

### Data preparation and network analysis

Protein-protein interaction data for *E. coli* were downloaded from the IntAct database [44]. Protein association data for *P. falciparum* were extracted from the STRING database [45]. STRING assigns association confidence scores (S), ranging from 0.15 to 0.999, based on sequence similarity, pathway analysis [24,46], chromosome synteny, genome organization, phylogenetic reconstruction, and literature text mining. Cytoscape 2.8.3 was used for network visualization [47]. Nodes are colored according to their functional classification in the eggNOG database



**Figure 5 Subnetwork alignment**. See Methods section for the description of the algorithm.

**Figure 6 An example of functional orthologs predicted by subnetwork alignment**. A subnetwork alignment between *E. coli* (proteins labeled in blue) and *P. falciparum* (proteins labeled in red). Because the subnetworks are similar and composed almost entirely of proteins with low BLAST E-values, that is, homologous pairs, it is likely that Q8IAN4 and P11557 are functional homologs, despite their low sequence similarity.

[48]. NetworkAnalyzer was used to compute topological parameters of the networks [49], with the default settings. Gene Ontology (GO) enrichment analysis was conducted using BiNGO [50]. The hypergeometric test was used with the Benjamini and Hochberg false discovery rate (FDR) correction with a significance level of 0.05.

## Additional material

**Additional file 1: Functional orthologs involved in cell cycle regulation in *P. falciparum*.** The query genome is *P. falciparum*, and the target genome is *E. coli*. GO: Gene Ontology. BP: Biological Process. MF: Molecular Function. CC: Cellular Component.

**Additional File 2: An example of functional orthologs predicted by subnetwork alignment**. The predicted pair is shaded.

## List of abbreviations used

ARK: aurora-related kinase; ASP: apical sushi protein; CRMP: chloroquine resistance marker protein; GO: Gene Ontology; MAPK: mitogen-activated protein kinase; PPI: protein-protein interaction; RBC: red blood cell; VBEM: variational Bayesian expectation maximization; Y2H: yeast 2-hybrid.

Conference on Bioinformatics and Biomedicine 2012: Bioinformatics. The full contents of the supplement are available online at http://www.biomedcentral.com/bmcbioinformatics/supplements/14/S12.

## Authors' details

[1]Department of Biology, University of Texas at San Antonio, San Antonio, TX 78249, USA. [2]Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA. [3]Department of Bacteriology, American Type Culture Collection, Manassas, VA 20110, USA. [4]Department of Computer Science, University of Texas at San Antonio, San Antonio, TX 78249, USA. [5]Department of Biology, College of Staten Island, City University of New York, Staten Island, NY 10314, USA. [6]South Texas Center for Emerging Infectious Diseases, University of Texas at San Antonio, San Antonio, TX 78249, USA.

## References

1. Poinar G Jr: Plasmodium dominicana n. sp. (Plasmodiidae: Haemospororida) from Tertiary Dominican amber. *Systematic parasitology* 2005, **61**(1):47-52.
2. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, *et al*: Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 2002, **419**(6906):498-511.
3. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, *et al*: Comparative genomics of the neglected human malaria parasite Plasmodium vivax. *Nature* 2008, **455**(7214):757-763.
4. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, *et al*: Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. *Nature* 2002, **419**(6906):512-519.
5. Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M, Florens L, Janssen CS, Pain A, Christophides GK, *et al*: A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 2005, **307**(5706):82-86.
6. Pain A, Bohme U, Berry AE, Mungall K, Finn RD, Jackson AP, Mourier T, Mistry J, Pasini EM, Aslett MA, *et al*: The genome of the simian and human malaria parasite Plasmodium knowlesi. *Nature* 2008, **455**(7214):799-803.
7. Tachibana S, Sullivan SA, Kawai S, Nakamura S, Kim HR, Goto N, Arisue N, Palacpac NM, Honma H, Yagi M, *et al*: Plasmodium cynomolgi genome sequences provide insight into Plasmodium vivax and the monkey malaria clade. *Nature genetics* 2012, **44**(9):1051-1055.
8. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL: The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium falciparum. *PLoS Biol* 2003, **1**(1):E5.
9. Bozdech Z, Mok S, Hu G, Imwong M, Jaidee A, Russell B, Ginsburg H, Nosten F, Day NP, White NJ, *et al*: The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(42):16290-16295.
10. Dharia NV, Bright AT, Westenberger SJ, Barnes SW, Batalov S, Kuhen K, Borboa R, Federe GC, McClean CM, Vinetz JM, *et al*: Whole-genome sequencing and microarray analysis of ex vivo Plasmodium vivax reveal selective pressure on putative drug resistance genes. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(46):20045-20050.
11. Dharia NV, Sidhu AB, Cassera MB, Westenberger SJ, Bopp SE, Eastman RT, Plouffe D, Batalov S, Park DJ, Volkman SK, *et al*: Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in Plasmodium falciparum. *Genome biology* 2009, **10**(2):R21.
12. Florent I, Porcel BM, Guillaume E, Da Silva C, Artiguenave F, Marechal E, Brehelin L, Gascuel O, Charneau S, Wincker P, *et al*: A Plasmodium falciparum FcB1-schizont-EST collection providing clues to schizont specific gene structure and polymorphism. *BMC genomics* 2009, **10**:235.
13. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ, *et al*: Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 2003, **301**(5639):1503-1508.
14. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL: Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. *Nucleic acids research* 2006, **34**(4):1166-1173.
15. Lopez-Barragan MJ, Lemieux J, Quinones M, Williamson KC, Molina-Cruz A, Cui K, Barillas-Mury C, Zhao K, Su XZ: Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. *BMC genomics* 2011, **12**:587.
16. Mair GR, Braks JA, Garver LS, Wiegant JC, Hall N, Dirks RW, Khan SM, Dimopoulos G, Janse CJ, Waters AP: Regulation of sexual development of Plasmodium by translational repression. *Science* 2006, **313**(5787):667-669.
17. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Bohme U, Lemieux J, Barrell B, Pain A, Berriman M, *et al*: New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Molecular microbiology* 2010, **76**(1):12-24.
18. Vignali M, Armour CD, Chen J, Morrison R, Castle JC, Biery MC, Bouzek H, Moon W, Babak T, Fried M, *et al*: NSR-seq transcriptional profiling enables identification of a gene signature of Plasmodium falciparum parasites infecting children. *J Clin Invest* 2011, **121**(3):1119-1129.
19. Watanabe J, Wakaguri H, Sasaki M, Suzuki Y, Sugano S: Comparasite: a database for comparative study of transcriptomes of parasites defined by full-length cDNAs. *Nucleic acids research* 2007, **35**(Database):D431-438.
20. Acharya P, Kumar R, Tatu U: Chaperoning a cellular upheaval in malaria: Heat shock proteins in Plasmodium falciparum. *Molecular and biochemical parasitology* 2007, **153**(2):85-94.
21. Bowyer PW, Simon GM, Cravatt BF, Bogyo M: Global profiling of proteolysis during rupture of Plasmodium falciparum from the host erythrocyte. *Mol Cell Proteomics* 2010, **10**(5):M110 001636.
22. Florens L, Liu X, Wang YF, Yang SG, Schwartz O, Peglar M, Carucci DJ, Yates JR, Wu YM: Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Molecular and biochemical parasitology* 2004, **135**(1):1-11.
23. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL, *et al*: A proteomic view of the Plasmodium falciparum life cycle. *Nature* 2002, **419**(6906):520-526.
24. Lasonder E, Ishihama Y, Andersen JS, Vermunt AMW, Pain A, Sauerwein RW, Eling WMC, Hall N, Waters AP, Stunnenberg HG, *et al*: Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. *Nature* 2002, **419**(6906):537-542.
25. Lasonder E, Janse CJ, van Gemert GJ, Mair GR, Vermunt AM, Douradinha BG, van Noort V, Huynen MA, Luty AJ, Kroeze H, *et al*: Proteomic profiling of Plasmodium sporozoite maturation identifies new proteins essential for parasite development and infectivity. *PLoS pathogens* 2008, **4**(10):e1000195.
26. Roobsoong W, Roytrakul S, Sattabongkot J, Li J, Udomsangpetch R, Cui L: Determination of the Plasmodium vivax schizont stage proteome. *Journal of proteomics* 2011, **74**(9):1701-1710.
27. Silvestrini F, Lasonder E, Olivieri A, Camarda G, van Schaijk B, Sanchez M, Younis Younis S, Sauerwein R, Alano P: Protein export marks the early phase of gametocytogenesis of the human malaria parasite Plasmodium falciparum. *Mol Cell Proteomics* 2010, **9**(7):1437-1448.
28. Solyakov L, Halbert J, Alam MM, Semblat JP, Dorin-Semblat D, Reininger L, Bottrill AR, Mistry S, Abdi A, Fennell C, *et al*: Global kinomic and phospho-proteomic analyses of the human malaria parasite Plasmodium falciparum. *Nature communications* 2011, **2**:565.
29. Tarun AS, Peng X, Dumpit RF, Ogata Y, Silva-Rivera H, Camargo N, Daly TM, Bergman LW, Kappe SH: A combined transcriptome and proteome survey of malaria parasite liver stages. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(1):305-310.
30. Ginsburg H: Progress in in silico functional genomics: the malaria Metabolic Pathways database. *Trends in parasitology* 2006, **22**(6):238-240.
31. Ginsburg H, Tilley L: Plasmodium falciparum metabolic pathways (MPMP) project upgraded with a database of subcellular locations of gene products. *Trends Parasitol* 2011, **27**:285-286.
32. Lakshmanan V, Rhee KY, Daily JP: Metabolomics and malaria biology. *Molecular and biochemical parasitology* 2011, **175**(2):104-111.
33. Plata G, Hsiao TL, Olszewski KL, Llinas M, Vitkup D: Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. *Mol Syst Biol* 2010, **6**:408.
34. Yeh I, Hanekamp T, Tsoka S, Karp PD, Altman RB: Computational analysis of Plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome research* 2004, **14**(5):917-924.

35. Date SV, Stoeckert CJ Jr: **Computational modeling of the Plasmodium falciparum interactome reveals protein function on a genome-wide scale.** *Genome research* 2006, **16**(4):542-549.
36. LaCount DJ, Schoenfeld LW, Fields S: **Selection of yeast strains with enhanced expression of Plasmodium falciparum proteins.** *Molecular and biochemical parasitology* 2009, **163**(2):119-122.
37. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, Sahasrabudhe S, Kurschner C, et al: **A protein interaction network of the malaria parasite Plasmodium falciparum.** *Nature* 2005, **438**(7064):103-107.
38. Suthram S, Sittler T, Ideker T: **The Plasmodium protein network diverges from those of other eukaryotes.** *Nature* 2005, **438**(7064):108-112.
39. Bartfai R, Hoeijmakers WA, Salcedo-Amaya AM, Smits AH, Janssen-Megens E, Kaan A, Treeck M, Gilberger TW, Francoijs KJ, Stunnenberg HG: **H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3.** *PLoS pathogens* 2010, **6**(12):e1001223.
40. Lopez-Rubio JJ, Mancio-Silva L, Scherf A: **Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites.** *Cell host & microbe* 2009, **5**(2):179-190.
41. Ponts N, Harris EY, Prudhomme J, Wick I, Eckhardt-Ludka C, Hicks GR, Hardiman G, Lonardi S, Le Roch KG: **Nucleosome landscape and control of transcription in the human malaria parasite.** *Genome research* 2010, **20**(2):228-238.
42. Westenberger SJ, Cui L, Dharia N, Winzeler E, Cui L: **Genome-wide nucleosome mapping of Plasmodium falciparum reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes.** *BMC genomics* 2009, **10**:610.
43. Cai H, Hong C, Gu J, Lilburn TG, Kuang R, Wang Y: **Module-based subnetwork alignments reveal novel transcriptional regulators in malaria parasite Plasmodium falciparum.** *BMC Syst Biol* 2012, **6**(Suppl 3):S5.
44. Isokpehi RD, Hide WA: **Integrative analysis of intraerythrocytic differentially expressed transcripts yields novel insights into the biology of Plasmodium falciparum.** *Malaria journal* 2003, **2**(1):38.
45. Lilburn TG, Cai H, Zhou Z, Wang Y: **Protease-associated Cellular Networks in Malaria Parasite Plasmodium falciparum.** *BMC genomics* 2011, **12**(Suppl 5):S9.
46. Westenberger SJ, McClean CM, Chattopadhyay R, Dharia NV, Carlton JM, Barnwell JW, Collins WE, Hoffman SL, Zhou Y, Vinetz JM, et al: **A systems-based analysis of Plasmodium vivax lifecycle transcription from human to mosquito.** *PLoS Negl Trop Dis* 2010, **4**(4):e653.
47. Bandyopadhyay S, Sharan R, Ideker T: **Systematic identification of functional orthologs based on protein network comparison.** *Genome research* 2006, **16**(3):428-435.
48. Bruckner S, Huffner F, Karp RM, Shamir R, Sharan R: **TORQUE: topology-free querying of protein interaction networks.** *Nucleic acids research* 2009, **37**:W106-108, (Web Server issue).
49. Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T: **PathBLAST: a tool for alignment of protein interaction networks.** *Nucleic acids research* 2004, **32**:W83-88, (Web Server issue).
50. Klau GW: **A new graph-based method for pairwise global network alignment.** *BMC bioinformatics* 2009, **10**(Suppl 1):S59.
51. Koyutürk M, Kim Y, Subramaniam S, Szpankowski W, Grama A: **Detecting conserved interaction patterns in biological networks.** *J Comput Biol* 2006, **13**(7):1299-1322.
52. Kuang R, Ie E, Wang K, Wang K, Siddiqi M, Freund Y, Leslie C: **Profile-based string kernels for remote homology detection and motif extraction.** *J Bioinform Comput Biol* 2005, **3**(3):527-550.
53. Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T: **Conserved patterns of protein interaction in multiple species.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(6):1974-1979.
54. Singh R, Xu J, Berger B: **Global alignment of multiple protein interaction networks.** *Pac Symp Biocomput* 2008, 303-314.
55. Zhenping L, Zhang S, Wang Y, Zhang XS, Chen L: **Alignment of molecular networks by integer quadratic programming.** *Bioinformatics* 2007, **23**(13):1631-1639.
56. Cai H, Hong C, Gu J, Lilburn TG, Kuang R, Wang Y: **Prediction of novel systems components in cell cycle regulation in malaria parasite by subnetwork alignments.** *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on: 4-7 October 2012* 2012, 1-6.
57. Doerig C, Chakrabarti D: **Cell cycle control in Plasmodium falciparum: a genomics perspective.** In *Malaria parasites: genomes and molecular biology.* Wymondham, UK: Caister Academic Press;Waters AP, Janse CJ 2004:249-287.
58. Tienda-Luna IM, Yin Y, Carrion MC, Huang Y, Cai H, Sanchez M, Wang Y: **Inferring the skeleton cell cycle regulatory network of malaria parasite using comparative genomic and variational Bayesian approaches.** *Genetica* 2008, **132**(2):131-142.
59. Le Roch K, Sestier C, Dorin D, Waters N, Kappes B, Chakrabarti D, Meijer L, Doerig C: **Activation of a Plasmodium falciparum cdc2-related kinase by heterologous p25 and cyclin H. Functional characterization of a P. falciparum cyclin homologue.** *The Journal of biological chemistry* 2000, **275**(12):8952-8958.
60. Ward P, Equinet L, Packer J, Doerig C: **Protein kinases of the human malaria parasite Plasmodium falciparum: the kinome of a divergent eukaryote.** *BMC genomics* 2004, **5**(1):79.
61. Doerig CM, Parzy D, Langsley G, Horrocks P, Carter R, Doerig CD: **A MAP kinase homologue from the human malaria parasite, Plasmodium falciparum.** *Gene* 1996, **177**(1-2):1-6.
62. Dixit A, Singh PK, Sharma GP, Malhotra P, Sharma P: **PfSRPK1, a novel splicing-related kinase from Plasmodium falciparum.** *The Journal of biological chemistry* 2010, **285**(49):38315-38323.
63. Tewari R, Straschil U, Bateman A, Böhme U, Cherevach I, Gong P, Pain A, Billker O: **The systematic functional analysis of Plasmodium protein kinases identifies essential regulators of mosquito transmission.** *Cell Host Microbe* 2010, **8**(4):377-387.
64. Talevich E, Mirza A, Kannan N: **Structural and evolutionary divergence of eukaryotic protein kinases in Apicomplexa.** *BMC evolutionary biology* 2011, **11**:321.
65. Dorin-Semblat D, Schmitt S, Semblat JP, Sicard A, Reininger L, Goldring D, Patterson S, Quashie N, Chakrabarti D, Meijer L, et al: **Plasmodium falciparum NIMA-related kinase Pfnek-1: sex specificity and assessment of essentiality for the erythrocytic asexual cycle.** *Microbiology* 2011, **157**(Pt 10):2785-2794.
66. Coppi A, Tewari R, Bishop JR, Bennett BL, Lawrence R, Esko JD, Billker O, Sinnis P: **Heparan sulfate proteoglycans provide a signal to Plasmodium sporozoites to stop migrating and productively invade host cells.** *Cell host & microbe* 2007, **2**(5):316-327.
67. Smythe E, Ayscough KR: **The Ark1/Prk1 family of protein kinases. Regulators of endocytosis and the actin skeleton.** *EMBO reports* 2003, **4**(3):246-251.
68. Clapier CR, Cairns BR: **The biology of chromatin remodeling complexes.** *Annual review of biochemistry* 2009, **78**:273-304.
69. Vincensini L, Fall G, Berry L, Blisnick T, Braun Breton C: **The RhopH complex is transferred to the host cell cytoplasm following red blood cell invasion by Plasmodium falciparum.** *Molecular and biochemical parasitology* 2008, **160**(2):81-89.
70. Painter HJ, Campbell TL, Llinas M: **The Apicomplexan AP2 family: integral factors regulating Plasmodium development.** *Molecular and biochemical parasitology* 2011, **176**(1):1-7.
71. Lindner SE, De Silva EK, Keck JL, Llinas M: **Structural determinants of DNA binding by a P. falciparum ApiAP2 transcriptional regulator.** *Journal of molecular biology* 2010, **395**(3):558-567.
72. Way M, Sanders M, Chafel M, Tu YH, Knight A, Matsudaira P: **beta-Scruin, a homologue of the actin crosslinking protein scruin, is localized to the acrosomal vesicle of Limulus sperm.** *Journal of cell science* 1995, **108**(Pt 10):3155-3162.
73. Sharan R, Ideker T: **Modeling cellular machinery through biological network comparison.** *Nature biotechnology* 2006, **24**(4):427-433.
74. Barabasi AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nature reviews Genetics* 2004, **5**(2):101-113.
75. Sahraeian SME, Yoon BJ: **Fast Network Querying Algorithm for Searching Large-Scale Biological Networks.** *Int Conf Acoust Spee* 2011, 6008-6011.
76. Borgwardt KM, Kriegel HP, Vishwanathan SV, Schraudolph NN: **Graph kernels for disease outcome prediction from protein-protein interaction networks.** *Pac Symp Biocomput* 2007, 4-15.
77. Jonczyk P, Hines R, Smith DW: **The Escherichia coli dam gene is expressed as a distal gene of a new operon.** *Mol Gen Genet* 1989, **217**(1):85-96.
78. Lyngstadaas A, Løbner-Olesen A, Boye E: **Characterization of three genes in the dam-containing operon of Escherichia coli.** *Mol Gen Genet* 1995, **247**(5):546-554.

79.  Muller J, Szklarczyk D, Julien P, Letunic I, Roth A, Kuhn M, Powell S, von Mering C, Doerks T, Jensen LJ, *et al*: eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic acids research* 2010, **38**(Database):D190-195.
80.  Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, *et al*: The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 2003, **4**:41.
81.  Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, *et al*: The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research* 2010, **39**(Database):D561-568.