

RESEARCH ARTICLE

Open Access

# Quantitative trait loci mapping reveals candidate pathways regulating cell cycle duration in *Plasmodium falciparum*

Heather B Reilly Ayala<sup>1,2</sup>, Mark A Wacker<sup>2</sup>, Geoffrey Siwo<sup>2</sup>, Michael T Ferdig<sup>2\*</sup>

## Abstract

**Background:** Elevated parasite biomass in the human red blood cells can lead to increased malaria morbidity. The genes and mechanisms regulating growth and development of *Plasmodium falciparum* through its erythrocytic cycle are not well understood. We previously showed that strains HB3 and Dd2 diverge in their proliferation rates, and here use quantitative trait loci mapping in 34 progeny from a cross between these parent clones along with integrative bioinformatics to identify genetic loci and candidate genes that control divergences in cell cycle duration.

**Results:** Genetic mapping of cell cycle duration revealed a four-locus genetic model, including a major genetic effect on chromosome 12, which accounts for 75% of the inherited phenotype variation. These QTL span 165 genes, the majority of which have no predicted function based on homology. We present a method to systematically prioritize candidate genes using the extensive sequence and transcriptional information available for the parent lines. Putative functions were assigned to the prioritized genes based on protein interaction networks and expression eQTL from our earlier study. DNA metabolism or antigenic variation functional categories were enriched among our prioritized candidate genes. Genes were then analyzed to determine if they interact with cyclins or other proteins known to be involved in the regulation of cell cycle.

**Conclusions:** We show that the divergent proliferation rate between a drug resistant and drug sensitive parent clone is under genetic regulation and is segregating as a complex trait in 34 progeny. We map a major locus along with additional secondary effects, and use the wealth of genome data to identify key candidate genes. Of particular interest are a nucleosome assembly protein (PFL0185c), a Zinc finger transcription factor (PFL0465c) both on chromosome 12 and a ribosomal protein L7Ae-related on chromosome 4 (PFD0960c).

## Background

Malaria is one of the deadliest infectious diseases in the world with the most lethal form, *Plasmodium falciparum*, infecting more than 500 million people each year, two to three million of whom die [1]. The characteristic malarial fevers occur in multiples of 24 hr due to synchronous parasite development and proliferation in the host's red blood cells (RBC), corresponding to cell lysis and massive liberation of new parasites and toxins into the host's bloodstream [2,3]. Clinical studies in South east Asia have demonstrated that parasite lines

which proliferate at an increased rate in RBC are more virulent than those with low multiplication rates, indicating a relationship between parasite growth and disease severity [4,5]. The molecular mechanisms directing the rate of parasite growth and development in the erythrocytic cycle are not well understood, underscoring the need to identify candidate genes regulating these processes.

The parasite erythrocytic cycle involves invasion of RBC by a merozoite, followed by a 'ring' stage that begins to ingest haemoglobin. Digestive vesicles merge into a larger digestive vacuole characteristic of the metabolically active trophozoite stage that is active for DNA replication, transcription and translation functions [6-8]. Unlike other eukaryotic organisms, *Plasmodium* spp.

\* Correspondence: ferdig.1@nd.edu

<sup>2</sup>Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA

Full list of author information is available at the end of the article

does not undergo cytokinesis after each successive round of DNA replication. Instead, DNA replication and mitosis occur multiple times within the same cell body - a process known as endomitosis resulting in the schizont containing 8-32 merozoites [9]. Progression of *P. falciparum* through the erythrocytic cycle takes approximately 48 hours [10]; however, we previously observed a shortened cell cycle in Dd2 compared to HB3 due to a shortened time in the ring and trophozoite stages [11]. These stages correspond to the G1 phase of the cell cycle and make up the majority of the parasites erythrocytic cycle. This observation is consistent with the progression of *Toxoplasma gondii*, another Apicomplexan parasite, through its tachyzoite cell cycle [12].

The development through the erythrocytic cycle requires coordinated expression of distinct sets of genes. Based on expectations of homologous functions with yeast, progression through the malaria parasite cycle is directed by cyclins and cyclin-dependent kinases (CDKs) [13]. Five candidate CDKs have been described in *P. falciparum*; PfPK5 a homologue to CDK1 and CDK5, PfPK6 a homologue to CDK1 and MAPKs, Pfmrk a homologue to CDK7, Pfcrk-1 a homologue to *cdc2* [14], and most recently Pfcrk-3 a homologue of CDK-related kinase 3 [15]. Both PfPK5 and Pfmrk have cyclin-dependent activity, whereas PfPK6 does not [16,17]. PfPK5 is active during the erythrocytic cycle and has been implicated in regulation of nuclear division [18,19]. Knock-out studies in *P. berghei* with the orthologue for Pfcrk-1 (Pbcrk-1) indicate this cyclin is essential for the completion of the erythrocytic cycle in Plasmodium [20]. Likewise, studies on Pfcrk-3 demonstrate that it is also necessary for the development of the erythrocytic parasite and most likely plays a role in chromatin modification [15]. The Myb-related transcription factors also participate in regulating the expression of genes involved in growth control and cell differentiation [21]. Gissot et al. demonstrated that a decrease in the *P. falciparum* transcription factor PfMyb1 resulted in a 40% parasite growth inhibition altering the transcription of a handful of genes implicated in cell cycle regulation, DNA repair and replication. These included genes such as PfPK5, a proliferating cell nuclear antigen (PCNA) shown to play a role in DNA replication, and a PP1-like phosphatase [22,23]. Additionally, a study by Janse et al. demonstrated that removal of one of two copies of the eukaryotic elongation factor 1 $\alpha$  (*ef1 $\alpha$* ) from *P. berghei* increased the cell cycle length by extending the G1 phase compared to the parent parasite line [24].

Quantitative trait loci (QTL) mapping combined with the resolution of whole-genome technologies offers unbiased access to genes, biological processes, and mutations throughout a genome and provides a point of entry into the parasite genome from which a list of

candidate genes can be refined and narrowed down in an unbiased manner. The goal of QTL mapping is to determine the genetic architecture, i.e. the number, position, and types of genetic contributions, for a given phenotype. Quantitative traits generally involve input from multiple genes acting in ways which may be additive, dominant, epistatic, or interactive and therefore will be associated with multiple QTLs [25,26].

QTL mapping in thirty-four progeny derived from a genetic cross between Dd2, a chloroquine resistant (CQR) parasite from Indochina, and HB3, a chloroquine sensitive (CQS) parasite from Honduras has been used to identify the gene regulating chloroquine (CQ) resistance, *Plasmodium falciparum* chloroquine resistance transporter (*pfcr1*) [27], and more recently the loci involved in quinine (QN) susceptibility [28]. The Dd2  $\times$  HB3 cross is characterized by high recombination resolution and cultured haploid progeny lines that can be repeatedly assayed for genome-wide sequence and transcription data, making it an excellent system for identifying genes and mechanisms that have diverged between the parental lines [29,30]. Here we characterize the genetic architecture of erythrocytic cell cycle duration in progeny from the Dd2  $\times$  HB3 cross and pinpoint candidate genes regulating this cycle time (CT) in *P. falciparum* [11]. We identify a major QTL on chromosome 12 that interacts epistatically with a locus on chromosome 4 and two additional additive loci to account for nearly 75% of the phenotype variation inherited among the progeny. We draw on existing genome-wide datasets including sequence polymorphisms [31], transcriptional polymorphisms and expression correlations between progeny expression and CT [30,32] and interaction network [33,34] to prioritize candidate genes and their interactions.

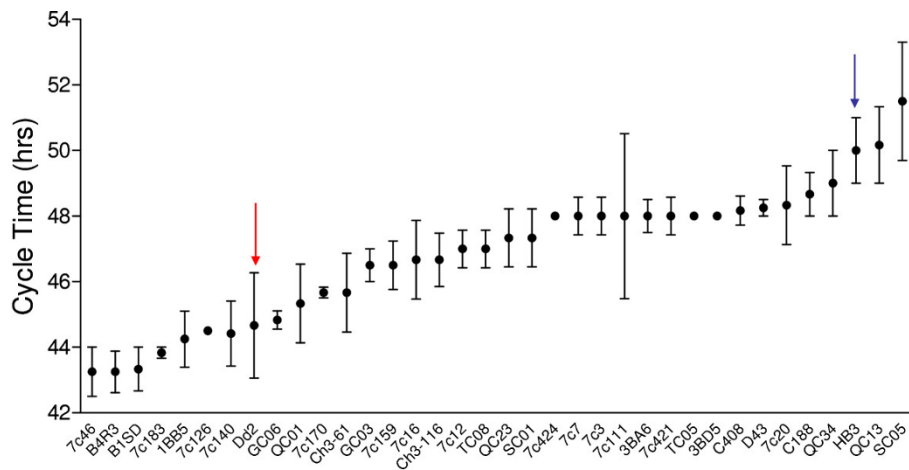
## Results

### Progeny inherit quantitative differences in cycle time

Previous work identified a statistically significant difference in CT duration in the parent lines (Dd2, 44.1 h; HB3, 49.6 h) [11]. The greatest deviation in cycle time duration was observed in the ring and early trophozoite stages whereas the late trophozoite and schizont stages showed insignificant deviations from each other. Cycle time was measured with biological replication in all 34 progeny of the Dd2  $\times$  HB3 cross, revealing a continuous distribution of this trait across the progeny set, indicating that CT is a complex, quantitative trait controlled by multiple genes (Figure 1).

### Quantitative trait loci (QTL) mapping of cycle time

QTL mapping was used to identify associations between CT and microsatellite (MS) markers distributed throughout the genome to identify chromosome regions



**Figure 1 Cycle Time phenotypes in progeny of Dd2 x HB3 cross.** Progeny are arranged in ascending order of cycle time duration. Dots represent means of a minimum of 3 replicates, error bars are SEM. Parents marked by arrows, Dd2 in red and HB3 in blue.

carrying polymorphisms that impact CT. QTL scans locate most of the genetic determination of the phenotype variation (43.76%) at one main locus on chromosome 12 at marker C1M48 (Table 1) with a highly significant LOD score ( $p < 0.01$  using 1000 permutations) (Figure 2A). In addition to the main effect, several minor peaks on chromosomes 3, 8, 13 and 14 were evident but not significant in the context of the chromosome 12 effect. Consequently, the remaining unassigned fraction of CT variation was analyzed using secondary scans of the residual variation after statistically removing the effect from the chromosome 12 locus; subsequent genome-wide scans identified markers C14M96 and C14M68 on chromosome 14 (LOD scores of 2.388 and 2.392 respectively) (Figure 2B). Inheritance of a D allele at any of these three markers produces progeny with a significantly faster CT than progeny inheriting an H allele (Figure 3A, B). Marker C1M48 exhibits an additive effect with marker C14M96 on chromosome 14 (Figure 4, 5A) such that progeny inheriting a D allele at both

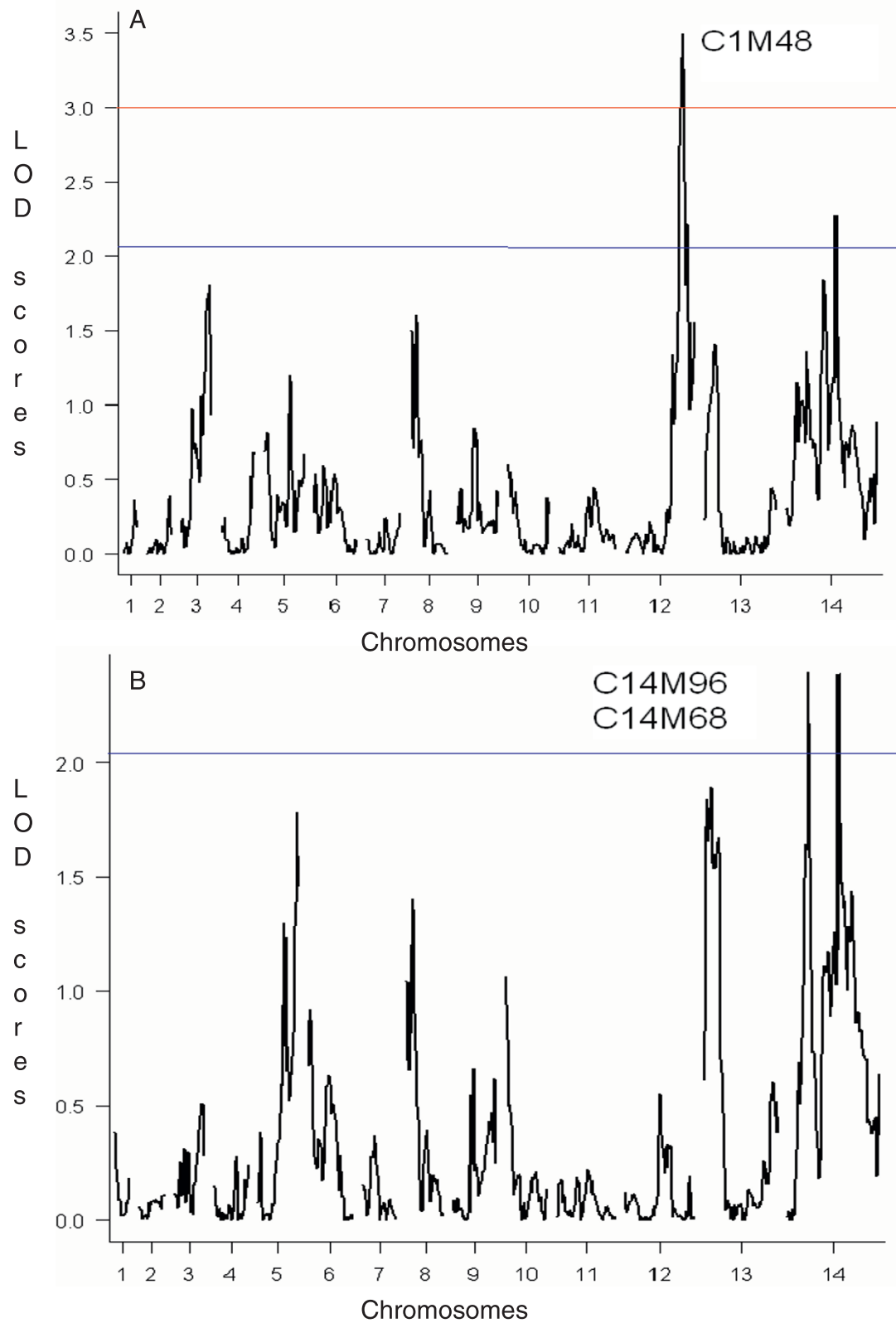
markers have a significantly ( $P < 0.0001$ ) faster CT (45.06 hr) than those with an H allele at both markers (49.14 hr) (Figure 5A).

Next, we ran pairwise scans to search for epistatic (non-additive) interactions between loci; an epistatic interaction is non-additive, such that the allelic contribution of one locus is observable only in the context of allele at the second locus. An interaction between C1M48 and a fourth marker on chromosome 4, C4M81 was identified (Figure 4). No effect was detected for C4M81 independently. Inheritance of the H allele at marker C4M81 and a D allele at C1M48 results in parasites with the fastest CT (44.09 hr) (Figure 5B). Progeny inheriting an H allele at both markers have a significantly ( $p < 0.0001$ ) slower CT (48.71 hr). When all three markers are analyzed together, progeny inheriting the HHH (C4M81, C1M48, C14M96) allele combination have the slowest CT (49.35 hr) while HDD progeny exhibit the fastest CT (44.15 hr) (Figure 5C). Together the multiple QTL model for CT shows four loci, one on

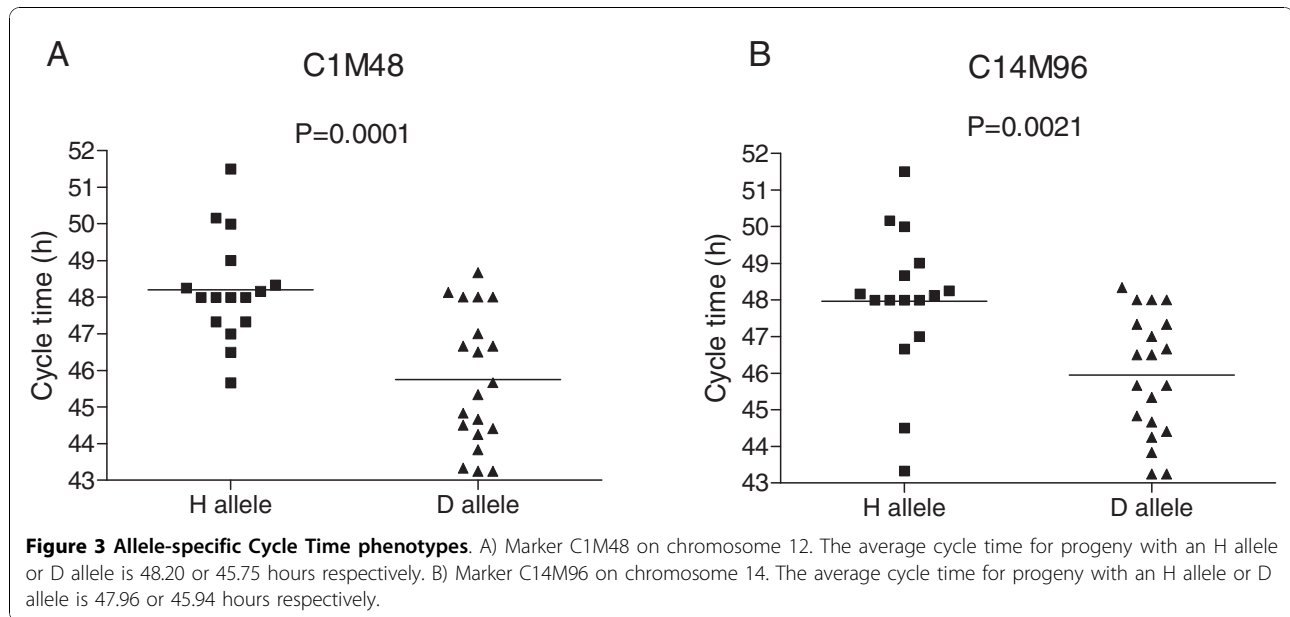
**Table 1 Summary of QTL effects for Cycle Time**

Nearest Marker	Effect	Chr	Position (cM)	LOD	Sig.	df	SS	Adj. SS	% variance	F value	P value <sup>a</sup>
C1M48	Main	12	140.1	3.84	99%	1	57.83	63.85	43.76	25.40	4.23E-07
C14M96	Residual	14	124.3	2.39	66%	1	23.84	11.64	7.98	9.26	0.00495
C14M68	Residual	14	50.0	2.39	66%	1	11.78	5.65	3.88	4.50	0.0426
C4M81	Interaction	4	50.0	0.06	n.s.	1	1.13	15.34	10.52	6.10	0.00614
C1M48 x C4M81						1	14.85	14.85	10.18	11.82	0.0018
Error						29	36.46				
Total						34	145.89		75.01		

Definitions of terms: Sig. significance values were determined by 1000 permutations of data, df - degree of freedom associated with each term in the model, SS - associated sum of squares, Adj. SS - adjusted sums of squares (type III sum of squares adjusting for all other terms in the model), % variance - total variance explained by the adjusted sums of squares, F statistics are based on Adj SS, P values are based on the F distribution.



**Figure 2 QTL scans of Cycle Time.** Red line represents significant threshold at 95% and blue line is suggestive at 66% based on 1000 permutations of the data. A) Main scan of cycle time. B) Residual scan of cycle time. Peaks on chromosome 14 from left to right are C14M96 and C14M68.

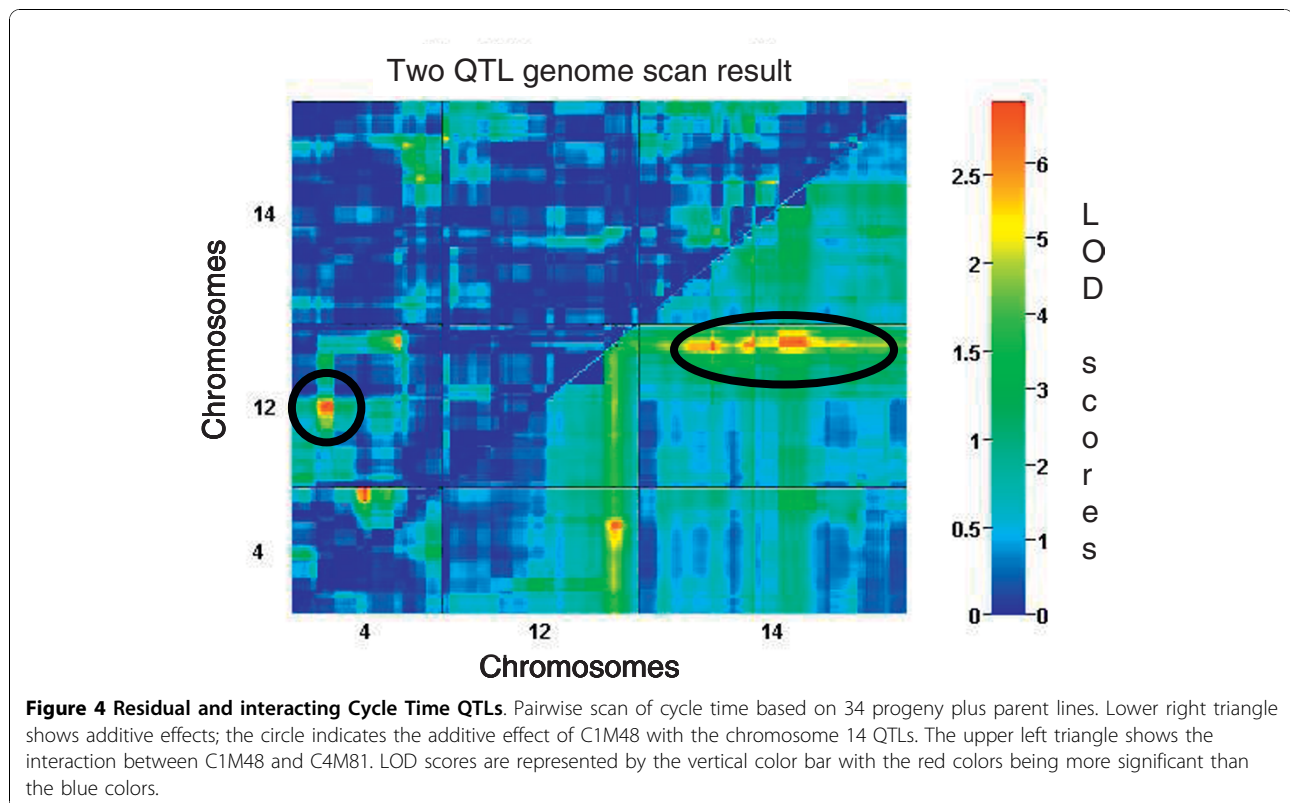


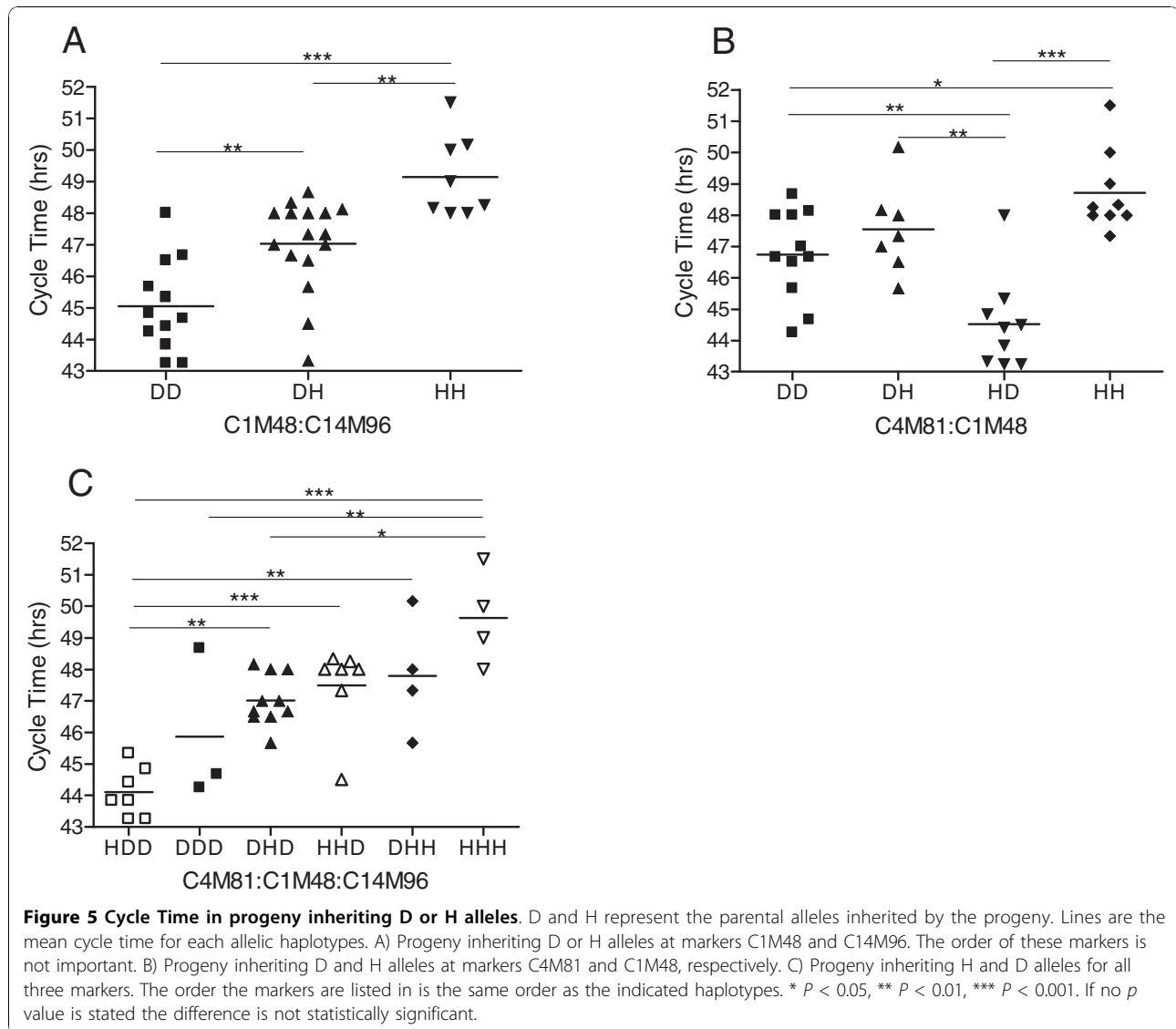
chromosome 12, two on chromosome 14 and one on chromosome 4 contributing to 75.01% of the observed variation (Table 1).

**Bioinformatic filtering of CT candidate gene list**

The four CT QTLs cover approximately 3.0% of the parasite's 23 Mb genome. A preliminary cumulative list

of 165 candidate genes was compiled for these loci (Figure 6A). The comprehensive gene list in these loci, including gene IDs and ontology (GO) annotations, is provided in the supplemental data (See additional file 1). As an initial step, we examine the genes in this locus for functionally relevant candidates that could plausibly be tied to cell cycle regulatory events or DNA replication.

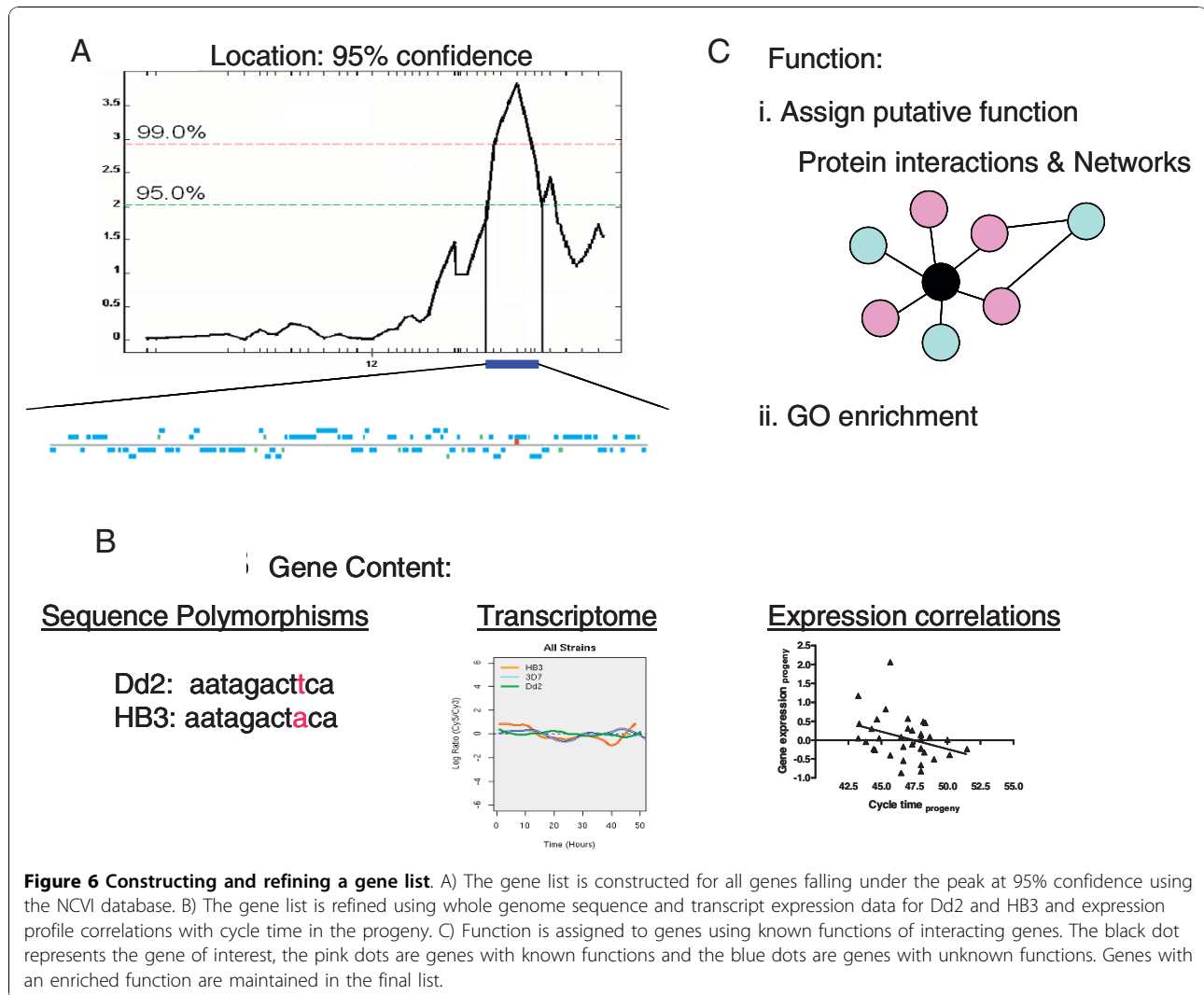




Some such candidates include two minichromosome maintenance proteins involved in DNA replication (PFL0560c and PFL0580w) in the chromosome 12 QTL. The chromosome 14 locus contains two serine/threonine protein kinases that could participate in cell cycle control (PF14\_0408 and PF14\_0392). Marker C14M39 is near a PP1-like protein serine/threonine phosphatase (PF14\_0224). The chromosome 4 region contains a Ran binding protein implicated in DNA replication (PFD0950w). Ran proteins are nuclear GTPases thought to be involved in nucleocytoplasmic transport and to play a role in mRNA processing and cell cycle regulation. Notably, genes with functional annotations are in the minority and may not necessarily be involved in the measured phenotype; nearly 68% of the genes encode conserved hypothetical proteins. While functional candidates may merit further interest, we also established

unbiased biological criteria to generate comprehensive gene lists irrespective of annotation.

Three criteria were used to refine the gene list: SNP density between parental lines, correlations between transcriptome profiles of the parents, and correlation of the phenotype with gene expression in the progeny (Figure 6B). Given the continued growth of the genome-wide information, e.g. metabolites, protein interactions, chromatin modifications, these filtering criteria will become increasingly expansive in future studies. Progeny are likely to inherit structural or expressional polymorphisms in genes which differ between parent lines thereby contributing to observed differences in CT. The top 25% of the values from each criterion was used for cross-filtering. (See additional file 2). SNP density refers to a comparison of the parent lines to determine the number of SNPs (synonymous or nonsynonymous)



**Figure 6 Constructing and refining a gene list.** A) The gene list is constructed for all genes falling under the peak at 95% confidence using the NCVI database. B) The gene list is refined using whole genome sequence and transcript expression data for Dd2 and HB3 and expression profile correlations with cycle time in the progeny. C) Function is assigned to genes using known functions of interacting genes. The black dot represents the gene of interest, the pink dots are genes with known functions and the blue dots are genes with unknown functions. Genes with an enriched function are maintained in the final list.

found within a genomic region. For SNP density the cutoff was set at 1.4 (Additional file 2, Figure s1A). Because not all polymorphic differences are due to structural changes in the gene itself, we also used expression data from the intraerythrocytic developmental cycle. The values used are a correlation of the expression profiles between the parents. In this case the cutoff value was assigned as the lowest 25% of the correlation values at 0.69 (Additional file 2, Figure s1B). Finally, because correlating expression traits with phenotypes can highlight pathways, we evaluated expression profiles for correlation with CT variation within the progeny. The cutoff value was assigned as the top 25% of the expression correlation values at 0.18 (Additional file 2, Figures 1C). These three screening criteria were used in parallel each resulting in an approximate 75% reduction of genes. There was some overlap between the genes which were eliminated resulting in 57% reduction in the gene list. Genes meeting any one of these criteria

were kept, reducing the list to 71 genes of which 51 are of unknown biological function (71.8%).

To include the large percentage of hypothetical genes in our analytical framework, we incorporated data from protein interactions and gene networks to assign likely functions to hypothetical proteins. Twenty of 71 genes in the candidate list were included in the yeast two hybrid network containing 1267 proteins [33]. In addition, PlasmoMAP, a protein network generated with computational and functional genomics data, provides information for 3667 genes [34]. For our analysis, once a putative function is assigned, genes containing an enriched function are flagged as possibly contributing to CT. Functional enrichment is defined as an over-representation of a particular gene ontology (GO) annotation in a subset of genes whose expression profiles across the progeny correlate with CT. The most enriched GO functions in the genes correlating with CT are: DNA metabolic process (GO: 0006259), antigenic variation (GO: 0020033), protein amino acid

**Table 2 Final list of prioritized candidate Cycle Time genes**

Chr	Gene ID	Gene Name	Proposed function	Data set <sup>a</sup>
12	PFL0170w	hypothetical protein	DNA metabolic process	PM
	PFL0185c	nucleosome assembly protein 1, putative	DNA metabolic process	Y2
	PFL0210c	eukaryotic initiation factor 5a, putative	DNA metabolic process	PM
	PFL0250w	hypothetical protein	DNA metabolic process	PM
	PFL0265w	hypothetical protein	DNA metabolic process	PM
	PFL0285w	glyoxalase II family protein, putative	DNA metabolic process	PM
	PFL0295c	hypothetical protein, conserved	antigenic variation	PM
	PFL0315c	hypothetical protein	DNA metabolic process	PM
	PFL0325w	hypothetical protein	antigenic variation	PM
	PFL0435w	hypothetical protein	DNA metabolic process	PM
	PFL0465c	Zinc finger transcription factor (krox1)	DNA metabolic process	Y2
	PFL0500w	50S ribosomal protein L1, putative	DNA metabolic process	Y2, PM
	PFL0575w	hypothetical protein	antigenic variation	Y2
	PFL0585w	PfpUB <i>Plasmodium falciparum</i> polyubiquitin	DNA metabolic process	Y2, PM
	PFL0630w	iron-sulfur subunit of succinate dehydrogenase	DNA metabolic process	PM
14-1	PF14_0217	hypothetical protein	antigenic variation	PM
	PF14_0218	actin, putative	antigenic variation	PM
14-2	PF14_0385	hypothetical protein	antigenic variation	PM
	PF14_0396	hypothetical protein	DNA metabolic process	PM
	PF14_0397	hypothetical protein, conserved	DNA metabolic process	PM
	PF14_0404	hypothetical protein	antigenic variation	PM
	PF14_0407	hypothetical protein	DNA metabolic process	PM
4	PFD0950w	ran binding protein 1	DNA metabolic process	GO
	PFD0960c	ribosomal protein L7Ae-related protein, putative	DNA metabolic process	PM

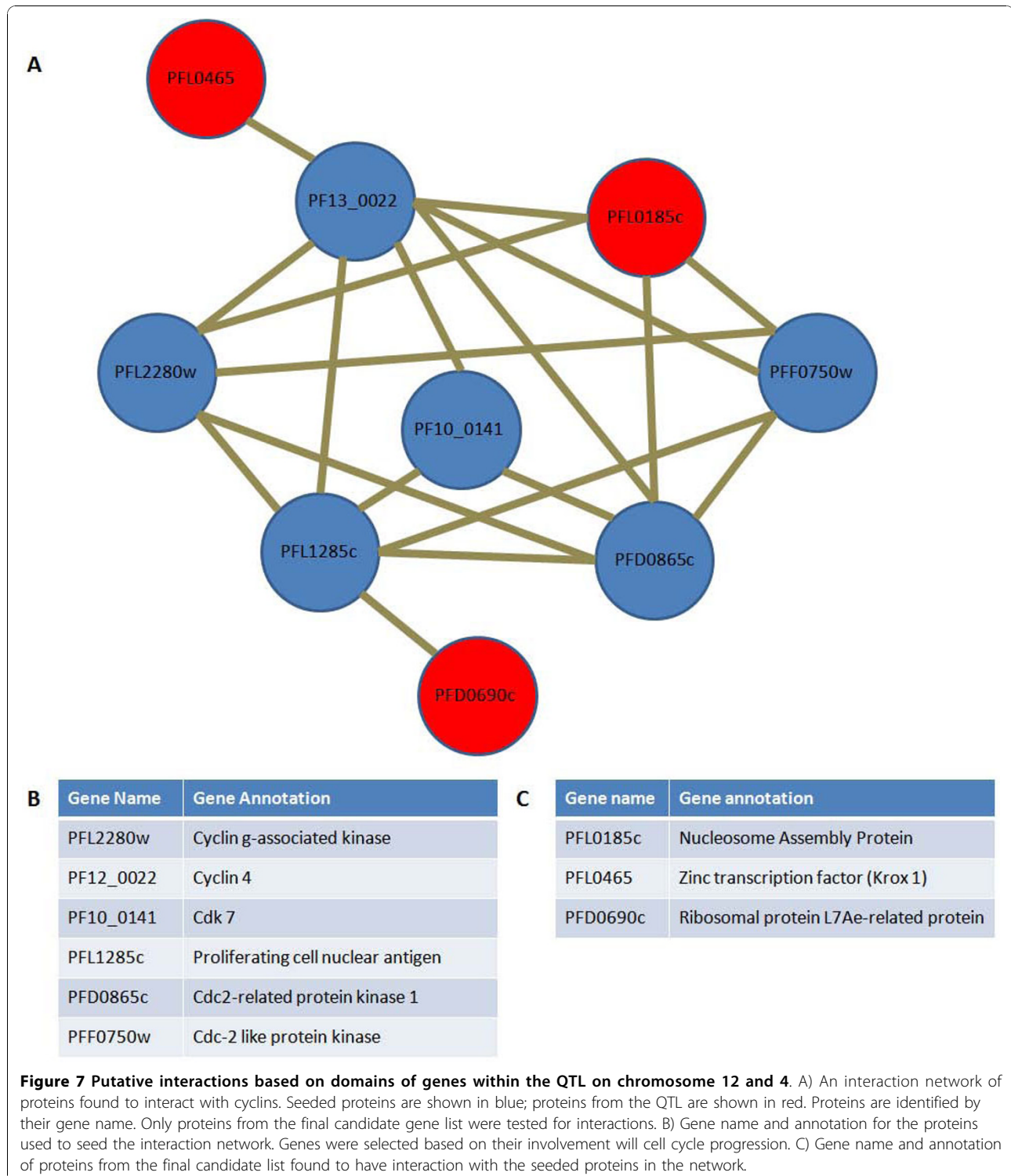
<sup>a</sup>Data set refers to the interaction data used to imply the function of the protein. Data from PlasmoMAP is denoted by PM and data from the yeast two-hybrid is denoted by Y2. Genes with a known function from the gene ontology are indicated by GO.

dephosphorylation (GO: 0006470) and cell-cell adhesion and cytoadherence to microvasculature (GO: 0016337). Candidate genes involved in any of these functions is given preference as being a gene regulating CT differences in *P. falciparum*. This step effectively eliminated another two thirds of the genes in the initial list of 167 narrowing it down to 24 candidate CT genes (Table 2).

Given the well-described role of cyclins in regulating cell cycle progression, we ascertained the potential direct physical interactions with cyclins among our candidate genes. Using domain information for proteins encoded by candidate genes within the QTL, we built an interaction network between the proteins using cyclins as a seed using a previously described algorithm for predicting domain-domain interactions given proteins with identifiable domains [35]. Only genes annotated as coding for cyclins were considered (Figure 7). A protein-protein interaction network in which three of the genes from the list of 24 candidates interact with the seeded proteins was produced by this procedure (Figure 7).

We extended this network to test possible interactions involving the additional 104 genes from the initial candidate gene list, but no additional interactions were found. Several proteins are connected to the network, including nucleosome assembly protein (PFL0185c) and Zinc transcription factor Krox 1 (PFL0465), both found at the main QTL on chromosome 12, and the protein ribosomal protein 7LAe-related protein (PFD0960c) located at the interacting locus on chromosome 4. The nucleosome assembly protein (PFL0185c) was selected based on its relatively weak correlation between Dd2 and HB3 transcription profiles, however, SNP density and phenotype-expression correlation highlight differences. In contrast, the zinc finger transcription factor (PFL0465c) had a high SNP density between Dd2 and HB3 exhibiting 3 nonsynonymous and 2 synonymous SNPs. In addition to having a high SNP density, the ribosomal protein also showed a strong correlation between the cycle time phenotype and expression data across the progeny.





## Discussion

We showed previously that the CQR and multi-drug resistant parasite, Dd2, proliferates faster than its CQS counterpart, HB3, and that one component of this faster growth rate is due to a shorter G1 phase in the

erythrocytic cycle of Dd2 [11]. Here we demonstrate that progeny inherit CT as a range of values indicative of a complex trait influenced by multiple genes. Our multi-locus genetic model of the determinants of this trait includes four loci accounting for 75.01% of the

variance in CT; the main effect on chromosome 12, two additive interactions on chromosome 14, and a locus on chromosome 4 that interacts epistatically with the large chromosome 12 effect (Table 1). Progeny inheriting a D allele at the markers on chromosomes 12 and 14 exhibit significantly faster CTs compared to progeny inheriting an H allele (Figure 3).

This difference in CT length between parasites appears to be a result of a shortened or elongated amount of time in the ring and trophozoite stages [11,24]. It is during these stages that the DNA replication, transcription and translation occurs [6,7]. To narrow the list of genes falling within QTL regions, we developed a method to draw on the existing databases to prioritize candidate genes using genome-wide data from independent sources and to indicate functions of candidates based on their relationship with other known genes, even in the absence of an annotation.

*Plasmodium* spp. have unique features of cell cycle regulation; consequently standard homology-based assumptions may miss important components. By utilizing an unbiased approach that relies on both genetics and networks of biological interactions we can retain hypothetical genes in our model, while also drawing on the rich data available from model organisms. Interestingly, all three of the genes predicted by our protein interaction network were also contained in the final gene list produced by the unbiased filtering (Figure 7, Table 2).

Functional analysis of genes correlating with the CT trait identified an enrichment of genes involved in DNA metabolism (GO:0006259). According to the gene ontology database <http://amigo.geneontology.org/cgi-bin/amigo/go.cgi>, the GO annotation of DNA metabolism includes a suite of molecular functions including: DNA replication, elongation, ligation, maintenance, mutagenesis, repair, modification and packaging. Mutations within genes involved in the DNA metabolic process would affect the efficiency of DNA replication and therefore the rate at which it is completed. Assuming *P. falciparum*, like other eukaryotic organisms, has a DNA checkpoint in its cell cycle, a delay in DNA replication would have an effect on the progression of the parasite through its erythrocytic cycle.

Studies in yeast and bacteria have shown that progression through the cell cycle is regulated by a complex process of protein synthesis, degradation and phosphorylation using cyclins, cyclin-dependent kinases (CDKs) and ubiquitin. These proteins ensure that imperative steps are completed before the cell moves into the next stage of the cell cycle. For example, cells must first complete DNA replication before moving into mitosis thereby preventing the formation of new cells which do

not contain the proper number of DNA molecules. Several homologues for cyclins and CDKs have been identified in the *Plasmodium* genome; although their specific roles are poorly understood [13]. No known *Plasmodium* cyclins and CDKs were found in our QTL, indicating that these genes (cyclins and CDKs) are not involved in this phenotype; however, our method allows us to integrate candidates with possible role for these classic cell cycle determinants. This approach is based on the hypothesis that genes underlying QTL effects operate indirectly through physical interactions with key cell cycle regulators, specifically cyclins. The predicted domain-domain based protein interaction network identified interactions between three genes in the candidate list with cyclins, depicting the possibility that the genes could regulate the cell cycle indirectly by relaying regulatory messages. This form of indirect regulation of biological processes such as the cell cycle could enable biological systems to respond to different environmental stimuli to which they have no direct interactions with.

We highlight candidate genes in the chromosome 12 QTL, and one at the interacting loci on chromosome 4, that could play a role in DNA metabolism and have an effect on the length of the parasite's erythrocytic cycle: PFL0185c, nucleosome assembly protein and PFL0465c, Zinc finger transcription factor (Krox1) on chromosome 12 and PFD0960c ribosomal protein L7Ae-related protein on chromosome 4. All of these were identified as candidates by the protein interaction network. Nucleosome assembly proteins (NAPs) are involved in chromatin assembly and remodelling in order to allow for replication, transcription, recombination and repair of DNA. They play a role in the global changes of chromatin structure which control the progression of *P. falciparum* through its lifecycle [36]. A recent study characterized the role of two nucleosome assembly proteins in *P. falciparum*, PfNapS (PFI0930c) and PfNapL (PFL0185c), which bind to histones; phosphorylation of these proteins by casein kinase II results in conformational changes increasing the histone binding affinity [37]. The PfNapS was shown to have a higher affinity for histones than PfNapL and is proposed to strip the histones from PfNapL. It shuttles the histones into the nucleus and deposits them onto DNA for chromatin remodelling. A polymorphism in this gene could affect the function of PfNapL and could affect the rate at which replication occurs resulting in a change of cell cycle length. PfNapL localizes in the cytoplasm of the parasite. Attempts to disrupt the gene were not successful suggesting that its role in *P. falciparum* is essential for parasite survival [38].

The zinc transcription factor (Krox1) is also candidate for regulating differences in cycle duration. Krox-1 was identified in sea urchin embryos and shown to play a

role in the development of the vegetal-plate [39]. It is in the same family as the Blimp-1 transcription factor which is required for the development of plasma cells from B cells [40,41]. Blimp-1 has been shown to be up-regulated in malignant cells compared to B-cells and is involved in the transduction pathway for growth factors in multiple myeloma [42]. A recent study has shown that Krox-1 is unregulated in *P. falciparum* parasites treated with inhibitors of *s*-adenosylmethionine decarboxylase/ornithine decarboxylase which block the synthesis of spermidine and subsequently cell cycle progression [43].

The third candidate identified by the protein interaction network is ribosomal protein L7Ae-related protein, PFD0960c. The gene for this protein is located at the interacting loci on chromosome 4. The network shows a path through which this gene could have an epistatic effect, on the genes located on chromosome 12. A study of larynx carcinoma cells resistant to the drug taxol, which stalls cells in the G2/M phase of the cell cycle, were shown to have elevated levels of ribosomal protein L7a [44]. Expression data between the parents shows the multi-drug resistant Dd2 to have an approximate 3.7-fold increase of PFD0960c compared to HB3, suggesting the malarial ribosomal protein L7Ae may play an important role in regulation of the parasite lifecycle [30].

## Conclusions

Overall, this study provides a new method to elucidate genes and functions involved in the regulation of the *P. falciparum* cell cycle. Regulation of the parasite erythrocytic cycle is complex, involving multiple genes and pathways; consequently, a single genetic cross will capture only a subset of the genetic components existing in the general population. The genes and interactions identified in our analysis point to previously unrecognized mechanisms underlying divergent CT between HB3 and Dd2. We also describe an unbiased method to integrate QTL candidates with the wealth of published data to narrow candidate gene lists. Using a protein interaction network seeded with cell cycle regulators, we found direct contact between 3 positional candidate genes and classical cell cycle genes. The overlap between these methods indicates that systematic, unbiased filtering can reveal new interactions and functions associate with key phenotypes.

## Methods

### Parasite culture

Parasites were grown in complete media (CM) [370  $\mu$ M hypoxanthine (Hx) (Sigma), and 25mM HEPES (Sigma); 0.5% Albumax II (Gibco), 10  $\mu$ g/mL gentamycin (Gibco), and 0.225% sodium bicarbonate (Biosource) in RPMI-1640 medium (Gibco) at approximately 3.5%

hematocrit (hct) in O<sup>+</sup> RBC. Cultures were maintained at a constant pH, 7.0-7.5, temperature, 37°C, and atmosphere, 5% CO<sub>2</sub>/5% O<sub>2</sub>/90% N<sub>2</sub> with daily media changes. Brief descriptions of previously used growth assays are given below [11].

### Determination of cycle time

Cultures were tightly synchronized by two treatments of D-sorbitol on consecutive days and resuspended in 5 mL of CM at 1% parasitemia. Experiments were initiated with parasites in the schizont stage immediately following the second synchronization. Segmenting schizonts provide a visually definitive, narrow morphological window (approximately 4hr) for precise stage identification. Progression of the parasite culture throughout the erythrocytic cycle was monitored using Giemsa-stained smears. The time at which the culture had the highest percentage of segmenting schizonts was defined as the schizont peak. A small volume smear (approximately 50  $\mu$ l) from each culture was made every 2 hr, originating from the same culture for each biological replicate, throughout the 12 hr window spanning the schizont peak for four consecutive life cycles. CT is defined as the duration between schizont peaks. Average CT was calculated from 3-4 consecutive life cycles and used in QTL analysis.

### QTL Analysis

Statistical analysis of the QTL data was carried out by computational approaches previously described [45]. The erythrocyte stages of the parasites used for genotyping and phenotyping are haploid; therefore, only two genetic classes are present for each locus, and the computational approach is equivalent to an analysis of recombinant inbred lines. Regression analysis and QTL mapping was done using R 2.4.1 software available through Jackson Laboratories <http://pga.jax.org/qtl/download.html>. Interval mapping was carried out at the average resolution (5 cM) of the *P. falciparum* genetic linkage map which contains 819 microsatellite markers [29]. One thousand permutations of the trait values determined 10%, 5%, and 1% thresholds [46]. Main QTL and corresponding mean trait values from the primary scans were used to obtain estimates of residual empirical thresholds to search for additional QTL via secondary scans [46]. This secondary scan procedure was equivalent to installing the a priori QTL as covariates, thereby attributing a large portion of the overall variation to these QTL and allowing for detection of additional loci controlling the remaining variation. Pairwise scans were performed at 5 cM spacing, testing all possible pairs of QTL locations for association with the trait. The likelihood calculated for a pair (one main locus and an interacting locus) was compared with that of the null model (no pairwise effects). Significance was

assessed by permutation analysis of every pairwise combination of marker alleles. Where an interaction was identified, a sequential series of statistical tests was performed to distinguish a true epistatic interaction from additive effects (independent gene actions) and hitchhiking (false positives resulting from related genotypes). These subsequent tests used a stringent nominal (comparison-wise) cut-off ( $P = 0.01$ ). For the final model an ANOVA was conducted by incorporating all the main and interacting QTL identified by the main, secondary and pairwise scans into a multiple regression analysis. Contributions of a QTL (or interaction) in combination with all other QTL was determined by calculating  $F$ -statistics from adjusted (type III) sums of squares. For this regression analysis, inheritance of the Dd2 parent allele was coded as +0.5 and inheritance of the HB3 parent allele was coded as -0.5. Accordingly, the main effect estimate for any locus was taken to be the difference between the effect of the Dd2 genotype and the HB3 genotype. Each interaction parameter was computed from the different effects on a main QTL of its paired Dd2- or HB3-type interacting QTL.

#### Constructing a candidate gene list

The candidate gene list was constructed, refined and prioritized using three selection criteria: 1) position, 2) structural and expressional polymorphisms and 3) functional analysis. Any genes falling between markers flanking the 95% confidence range of the QTL peak were included in a preliminary list of candidate genes along with gene IDs, gene names and approximate gene positions (kb) taken from the NCBI database <http://www.ncbi.nlm.nih.gov>.

Structural (SNPs in the ORF) polymorphisms, expressional polymorphisms between Dd2 and HB3 and correlations between transcription and CT phenotype in the progeny were used to refine the list of candidate genes. Genes that satisfied one of these criteria were retained. The top 25% percentile of the genes (or in the case of the transcription data, the bottom 25%) was kept (Additional file 2, Figure s1). The SNP position, type (non-coding, nonsense, synonymous, non-synonymous), and coding sequence SNP density were obtained from PlasmoDB v5.2 <http://plasmodb.org/plasmo/> and recorded for every gene in the preliminary list. The SNP density refers to the number of SNPs in that particular region of the genome. For the initial 165 CT candidate genes, the SNP data was confirmed by comparing Dd2 and HB3 consensus sequences obtained from the Broad Institute (provided by Sarah Volkman). Genes containing multiple SNPs, defined as any gene with a SNP density greater than 1.4 (this was the 25 percentile cutoff), were placed in a refined list of candidate genes. Secondly, expressional polymorphisms were analyzed using

correlation values between the Dd2 and HB3 transcriptional profiles downloaded from the DeRisi Lab Malaria Transcriptome Database <http://malaria.ucsf.edu/comparison/index.php> [47]. If a gene had more than one correlation, the values were averaged and an absolute value was used. All correlation values fell between 0 and 1. Genes with an absolute value correlation less than 0.69 (the 25 percentile cutoff) between the parent lines was maintained in the refined list of genes. Finally, correlation coefficients were calculated between the CT phenotype and gene expression at 18 hr post invasion across the progeny. The expression data from 35 cDNA arrays for 7665 probes representing 5150 ORFs, one for each progeny parasite, at 18 hr post invasion was obtained from Joe Gonzales [30]. A correlation coefficient  $\geq 0.18$  were retained (25 percentile cutoff).

The third criterion, functional analysis, is composed of two steps: 1) defining gene enrichment and 2) assigning a function to genes in the refined list. Gene enrichment is determined as follows: the CT phenotype was compared with 18 hr post-invasion gene expression in the progeny to calculate a correlation coefficient [30]. Of the 5150 ORFs represented, and with a correlation coefficient,  $\geq .030$  was selected and the corresponding gene was used to determine enrichment. The gene ontology (GO) annotation for each selected gene was obtained from PlasmoDB v5.2. The GO annotations were consolidated under broad headings and screened for enrichment within the CT genes. Gene enrichment was calculated using a hypergeometric distribution [32]. Genes within the refined list were assigned a function first based on their GO term taken from PlasmoDB. Hypothetical genes lacking a function were given a putative function using interaction data from yeast two hybrids (provided on PlasmoDB) [33] and PlasmoMAP <http://www.cbil.upenn.edu/plasmoMAP/index-v1.html> [34]. Genes with a term associated with GO in the enriched set were retained. The filtered gene list represents genes falling within the confidence range of the QTL, contain polymorphisms between the parent lines, and bioinformatic evidence of a functional relationship to CT.

#### Protein interaction network

A protein interaction network was constructed between candidate genes and cyclins, using the software Cytoprophet [35], based on physical interactions between known protein domains in the Pfam domain database, database of interacting proteins [48] and protein databank [49]. In order to determine whether the candidate genes may mediate effects on the cell cycle, the network was seeded with *P. falciparum* cyclin related proteins that play a key role in cell cycle regulation and have domains identifiable in the Pfam database. These proteins are cyclin g associated kinase (PFL2280w), cyclin 4

(PF13\_0022), putative cdk 7 (PF10\_0141), putative proliferating cell nuclear antigen, PCNA (PFL1285c), cdc2-related protein kinase 1 (PFD0865c), and cyclin dependent protein kinase (PFF0750w).

## Additional material

**Additional file 1: Table with full list of genes from four Cycle Time QTLs**

**Additional file 2: Scatter plots of criteria used to refine gene list**The x-axis represents the 104 genes in the preliminary gene list. They are arranged in ascending order for each of the three criteria used. A) SNP density between HB3 and Dd2 taken from PlasmoDB. B) Correlations of transcriptome for each gene between HB3 and Dd2 [47] C) Correlation of gene expression at 18 hours post-invasion between HB3 and Dd2 [30].

## Acknowledgements

Expression data used in constructing the candidate list were provided by Joe Gonzales. The consensus sequence data comparing Dd2 and HB3 were provided by the Broad Institute. Funding for this study came from NIH Training Grant T32 AI07030 to HBA, and NIH grant AI071121 to MTF.

## Author details

<sup>1</sup>Life Sciences, Bethel College, 1001 Bethel Circle, Mishawaka, IN 46545, USA.  
<sup>2</sup>Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA.

## Authors' contributions

HBA collected the CT data for all the progeny, performed the QTL analysis, and constructed the candidate gene list. MAW and GS constructed the protein interaction network and performed the corresponding analysis. MTF coordinated the study and helped to draft the manuscript. All authors contributed to writing the manuscript. All authors read and approved the final manuscript.

Received: 2 May 2010 Accepted: 18 October 2010

Published: 18 October 2010

## References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI: **The global distribution of clinical episodes of Plasmodium falciparum malaria.** *Nature* 2005, **434**:214-217.
2. Hawking F: **The clock of the malaria parasite.** *Sci Am* 1970, **222**:123-131.
3. Garcia CRS, Markus RP, Madeira L: **Tertian and Quartan Fevers: Temporal Regulation in Malarial Infection.** *J Biological Rhythms* 2001, **16**:436-443.
4. Chotivanich K, Udomsangpetch R, Simpson JA, Newton P, Pukrittayakamee S, Loareesuwan S, White N: **Parasite multiplication potential and the severity of falciparum malaria.** *J Infect Dis* 2000, **181**:1206-1209.
5. Dondorp AM, Desakorn V, Pongtavornpinyo W, Sahassananda D, Silamut K, Chotivanich K, Newton PN, Pitisuttithum P, Smithyman AM, White NJ: **Estimation of the total parasite biomass in acute falciparum malaria from plasma PfHRP2.** *Plos Med* 2005, **2**:0788-0797.
6. Chulay JD, Haynes JD, Diggs CL: **Plasmodium falciparum: assessment of in vitro growth by [3H]hypoxanthine incorporation.** *Exp Parasitol* 1983, **55**:138-146.
7. Yayon A, Vande Waa JA, Wayaon M, Geary TG, Jensen JB: **Stage-dependent effects of chloroquine on Plasmodium falciparum in vitro.** *J Protozool* 1983, **30**:642-647.
8. Arnot DE, Gull K: **The Plasmodium cell-cycle: facts and questions.** *Ann Trop Med Parasitol* 1998, **92**:361-365.
9. Bannister LH, Mitchell GH: **The role of the cytoskeleton in Plasmodium falciparum merozoite biology: an electron-microscopic view.** *Ann Trop Med Parasitol* 1995, **89**:105-111.
10. Roberts J, Janovy J: *Foundations of Parasitology* McGraw-Hill 2000.
11. Reilly HB, Wang H, Steuter JA, Marx AM, Ferdig MT: **Quantitative dissection of clone-specific growth rates in cultured malaria parasites.** *Int J Parasitol* 2007, **37**:1599-1607.
12. Radke JR, White MW: **A cell cycle model for the tachyzoite of Toxoplasma gondii using the Herpes simplex virus thymidine kinase.** *Mol Biochem Parasitol* 1998, **94**:237-247.
13. Doerig C, Endicott J, Chakrabarti D: **Cyclin-dependent kinase homologues of Plasmodium falciparum.** *Int J Parasitol* 2002, **32**:1575-1585.
14. Doerig C, Doerig C, Horrocks P, Coyle J, Carlton J, Sultan A, Arnot D, Carter R: **Pfcrk-1, a developmentally regulated cdc2-related protein kinase of Plasmodium falciparum.** *Mol Biochem Parasitol* 1995, **70**:167-174.
15. Halbert J, Ayong L, Equinet L, LeRoch K, Hardy M, Goldring D, Reiningger L, Waters N, Chakrabarti D, Doerig C: **A Plasmodium falciparum transcriptional CDK-related kinase with a crucial role in parasite proliferation associates with histone deacetylase activity.** *Eukaryot Cell* 2010.
16. LeRoch 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.** *J Biol Chem* 2000, **275**:8952-8958.
17. Holton S, Merckx A, Burgess D, Doerig C, Noble M, Endicott J: **Structures of P. falciparum PfPK5 test the CDK regulation paradigm and suggest mechanisms of small molecule inhibition.** *Structure* 2003, **11**:1329-1337.
18. Graeser R, Franklin RM, Kappes B: **Mechanisms of activation of the cdc2-related kinase PfPK5 from Plasmodium falciparum.** *Mol Biochem Parasitol* 1996, **79**:125-127.
19. Graeser R, Wernli B, Franklin RM, Kappes B: **Plasmodium falciparum protein kinase 5 and the malarial nuclear division cycles.** *Mol Biochem Parasitol* 1996, **82**:37-49.
20. Rangarajan R, Bei A, Henry N, Madamet M, Parzy D, Nivez MP, Doerig C, Sultan A: **Pbcrk-1, the Plasmodium berghei orthologue of P. falciparum cdc-2 related kinase-1 (Pfcrk-1), is essential for completion of the intraerythrocytic asexual cycle.** *Exp Parasitol* 2006, **112**:202-207.
21. Lipsick JS: **One billion years of Myb.** *Oncogene* 1996, **13**:223-235.
22. Gissot M, Briquet S, Refour P, Boschet C, Vaquero C: **PfMyb1, a Plasmodium falciparum transcription factor, is required for intraerythrocytic growth and controls key genes for cell cycle regulation.** *J Mol Biol* 2005, **346**:29-42.
23. Kelman Z: **PCNA: structure, functions and interactions.** *Oncogene* 1997, **14**:629-640.
24. Janse CJ, Haghparast A, Speranca MA, Ramesar J, Kroeze H, del Portillo HA, Waters AP: **Malaria parasites lacking eef1a have a normal S/M phase yet grow more slowly due to a longer G1 phase.** *Mol Microbiol* 2003, **50**:1539-1551.
25. Falconer DS, Mackay TFC: *Introduction to Quantitative Genetics* Longman 1996.
26. Flint J, Mott R: **Finding the molecular basis of quantitative traits: successes and pitfalls.** *Nat Rev Genet* 2001, **2**:437-445.
27. Welles TE, Walker-Jonah A, Panton LG: **Genetic mapping of the chloroquine-resistance locus on Plasmodium falciparum chromosome 7.** *Proc Natl Acad Sci* 1991, **88**:3382-3386.
28. Ferdig MT, Cooper RA, Mu J, Deng B, Joy DA, Su XZ, Welles TE: **Dissecting the loci of low-level quinine resistance in malaria parasites.** *Mol Microbiol* 2004, **52**:985-997.
29. Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, You J, Wootton JC, Welles TE: **A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum.** *Science* 1999, **286**:1351-1353.
30. Gonzales JM, Patel JJ, Ponmee N, Jiang L, Tan A, Maher SP, Wuchty S, Rathod PK, Ferdig MT: **Regulatory hotspots in the malaria parasite genome dictate transcriptional variation.** *PLoS Biol* 2008, **6**:e238.
31. Mu J, Awadalla P, Duan J, McGee KM, Keebler J, Seyde K, McVean GA, Su XZ: **Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome.** *Nat Genet* 2007, **39**:126-130.
32. Lan H, Chen M, Flowers JB, Yandell BS, Stapleton DS, Mata CM, Mui ET, Flowers MT, Schueler KL, Manly KF, Williams RW, Kendziorski C, Attie AD: **Combined expression trait correlations and expression quantitative trait locus mapping.** *PLoS Gene* 2006, **2**:0051-0061.
33. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, Sahasrabudhe S, Kurschner C, Fields S, Hughes RE: **A protein interaction network of the malaria parasite Plasmodium falciparum.** *Nature* 2005, **438**:103-107.

34. Date SV, Stoeckert CJ: **Computational modelling of the Plasmodium falciparum interactome reveals protein function on a genome-wide scale.** *Genome Res* 2006, **16**:542-549.
35. Morcos F, Lamanna C, Sikora M, Izaguirre J: **Cytosprophet: a Cytoscape plug-in for protein and domain interaction networks inference.** *Bioinformatics* 2008, **24**:2265-2266.
36. Ponts N, Harris EY, Prudhomme J, Wick I, Eckhardt-Ludka C, Hicks GR, Hardiman G, Lonardi S, LeRoch KG: **Nucleosome landscape and control of transcription in the human malaria parasite.** *Genome Res* 2010, **20**:228-238.
37. Navadgi VM, Chandra BR, Mishra PC, Sharma A: **The two Plasmodium falciparum Nucleosome Assembly Proteins play distinct roles in histone transport and chromatin assembly.** *J Biol Chem* 2006, **281**:16978-16984.
38. Gill J, Yogavel M, Kumar A, Belrhali H, Jain SK, Rug M, Brown M, Maier AG, Sharma A: **Crystal structure of malaria parasite nucleosome assembly protein: distinct modes of protein localization and histone recognition.** *J Biol Chem* 2009, **284**:10076-10087.
39. Wang W, Wikramanayake AH, Gonzalez-Rimbau M, Vlahou A, Flytzanis CN, Klein WH: **Very early and transient vegetal-plate expression of SpKrox1, a Kruppel/Krox gene from Strongylocentrotus purpuratus.** *Mech Dev* 1996, **60**:185-195.
40. Lin KI, Tunaplin C, Calame K: **Transcriptional regulatory cascades controlling plasma cell differentiation.** *Immunol Rev* 2003, **194**:19-28.
41. Shapiro-Shelef M, Calame K: **Plasma cell differentiation and multiple myeloma.** *Curr Opin Immunol* 2004, **16**:226-234.
42. Klein B, Tarte K, Jourdan M, Mathouk K, Moreaux J, Jourdan E, Legouffe E, De Vos J, Rossi JF: **Survival and proliferation factors of normal and malignant plasma cells.** *Int J Hematol* 2003, **78**:106-113.
43. van Brummelen AC, Olszewski KL, Wilinski D, Llinas M, Louw AI, Brikholtz LM: **Co-inhibition of Plasmodium falciparum S-adenosylmethionine decarboxylase/ornithine decarboxylase reveals perturbation-specific compensatory mechanisms by transcriptome, proteome, and metabolome analyses.** *J Biol Chem* 2009, **284**:4635-4646.
44. Schmidt M, Schler G, Gruensfelder P, Hoppe F: **Differential gene expression in a paclitaxel-resistant clone of a head and neck cancer cell line.** *Eur Arch Otorhinolaryngol* 2006, **263**:127-134.
45. Sen S, Churchill GA: **A statistical framework for quantitative trait mapping.** *Genetics* 2001, **159**:371-387.
46. Churchill GA, Doerge RW: **Empirical threshold values for quantitative trait mapping.** *Genetics* 1994, **138**:963-971.
47. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL: **Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains.** *Nucleic Acids Res* 2006, **34**:1166-1173.
48. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D: **Dip, the database of interacting proteins: a research tool for studying cellular networks of protein interactions.** *Nucleic Acids Res* 2002, **30**:303-305.
49. Desphande N, Address KJ, Bluhm WF, Merino-Ott JC, Townsend-Merino W, Zhang Q, Knezevich C, Xie L, Chen L, Feng Z, Green RK, Flippen-Anderson JL, Westbrook J, Berman HM, Bourne PE: **The rcsb protein data bank: a redesigned query system and relational database based on the mmcif schema.** *Nucleic Acids Res* 2005, **33**:D232-D237.

doi:10.1186/1471-2164-11-577

**Cite this article as:** Reilly Ayala et al.: Quantitative trait loci mapping reveals candidate pathways regulating cell cycle duration in *Plasmodium falciparum*. *BMC Genomics* 2010 **11**:577.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

