Integrative analysis of intraerythrocytic differentially expressed transcripts yields novel insights into the biology of Plasmodium falciparum
Raphael D Isokpehi* and Winston A Hide

Address: South African National Bioinformatics Institute, University of the Western Cape, Bellville 7535, South Africa
Email: Raphael D Isokpehi* - raphael@sanbi.ac.za; Winston A Hide - winhide@sanbi.ac.za
* Corresponding author

Abstract

Background: The intraerythrocytic development of Plasmodium falciparum, the most virulent human malaria parasite involves asexual and gametocyte stages. There has been a significant increase in disparate datasets derived from genomic and post-genomic analysis of the parasite that necessitates delivery of integrated analysis from which biological processes important to the survival of the parasite can be determined.

Methods: In order to resolve genes associated with stage differentially expressed transcripts, we have developed and implemented an integrative approach that combines evidence from P. falciparum expressed sequence tags (ESTs), genomic, microarray, proteomic and gene ontology data.

Results: A total of 143 gametocyte-overexpressed and 51 asexual-overexpressed transcripts were identified. A subset of 74 genes associated with these transcripts showed evidence of stage-correlated protein expression, of which 53 have not been experimentally characterised. Our study has revealed (1) possible regulatory mechanisms in malaria parasites' gametocyte maturation, (2) correlation between EST and microarray data for a P. falciparum gene family to present unique EST-derived information, (3) candidate drug and antigenic targets on which computational and experimental studies can be performed, and (4) the need for more empirical studies on gene and protein expression in malaria parasites.

Conclusion: Applying different domains of data to the same underlying gene set has yielded novel insights into the biology of the parasite and presents an approach to appraise critically the data quality of post-genomic datasets from malaria parasites.
based assessment of disparate datasets similar to gene structure prediction approaches that rely on accumulation of evidence such as similarity to known genes, nucleotide compositional features, intron/exon boundaries and promoter sequences [4].

The high malaria burden in Africa [5,6] necessitates increased efforts to understand the biology of the pathogen with a view to discovering new drugs, candidate vaccines and diagnostics, as well as improving existing ones. The publication of the genomes of the human malaria parasite *P. falciparum* and the rodent malaria parasite *Plasmodium yoelii* as well as ongoing sequencing projects of other *Plasmodium* species presents new opportunities to achieve the above-mentioned goals [7-9]. In addition, there have been efforts to obtain and analyse on a large-scale, gene expression profiles (transcriptome) of *Plasmodium* species using Expressed Sequence Tags (ESTs) [1,10-13], full length cDNAs [14], Serial Analysis of Gene Expression (SAGE) [15,16] and microarrays [17-19]. Protein expression profiles (proteome) on particular stages of the *P. falciparum* life cycle are also available [20,21].

The random single-pass sequencing of a cDNA library to generate short (200–500 bp) nucleotide sequences that tag an expressed gene sequence is an established method of gene discovery [22,23]. EST gene indices are generated by computer-based methods to organise these tags by assigning them into groups to remove redundancies and yield reconstructed transcripts that represent consensus sequences of each group [22,24,25]. These indices are being used to understand the complexity of the human genome, especially in providing information on alternative transcripts, non-translated transcripts, truly unique genes and extremely short genes that will complement the genome data [25]. The availability of the complete genome of *P. falciparum* 3D7 makes it possible to provide similar information for the parasite. In fact, additional EST and full-length cDNA sequences are required to improve the current annotation and verify predicted genes [7]. EST sequencing projects on *Plasmodium* have identified novel genes [1,10,13] but only limited analyses have been performed on ESTs for coordinate and differential gene expression [13].

*Plasmodium* ESTs from a variety of cDNA libraries are available in the GenBank EST database (dbEST). As of February 2003, 11 libraries comprising of nine asexual, one sporozoite and one gametocyte were available in dbEST. ESTs from some of these libraries have been indexed [1,10,13,26]. Microarrays, mRNA differential display and EST-based analysis have been used to study transcriptional differences between asexual and gametocyte stages of *P. falciparum*, revealing stage-specific genes [13,17,27]. These studies were done prior to the publication of the genome sequence of strain 3D7. Furthermore, in the case of Li and colleagues [13], the functional annotation was selective. An EST-based analysis with an improved functional annotation that combines the automated annotation from *P. falciparum* gene indices and the curated annotation in the *Plasmodium* Genome Database (PlasmoDB) [28] is needed. In addition, integration of proteomic data with such analysis has been recognized as an important component in drug target identification and validation in the human genome [29].

The number of ESTs used to generate a consensus sequence in a gene index can provide a rough estimate of the mRNA abundance in the tissue or cell of origin [23]. Furthermore, statistical tests have been developed to identify genes that are differentially expressed (significantly overexpressed) in a particular tissue compared to one or more other tissues [30,31]. The differences in EST counts have been applied to understand gene expression in different metabolic pathways, tissues or stages [32-34]. These differences appear to correlate with biology of the tissue or stage under investigation. Microarray and SAGE methods are more narrow but sensitive for differential gene expression studies and can be used to validate broader EST-based analysis [13].

The life cycle of *P. falciparum* involves stages in the female anopheline mosquito vector and stages in the human host [35]. The parasite goes through pre-erythrocytic and intraerythrocytic stages in the human host. The pre-erythrocytic stage involves invasion and growth within liver cells, whereas the intraerythrocytic cycle is a multi-stage process, which includes differentiation into asexual stages (rings, merozoites, trophozoites and schizonts) as well as sexual stages (male and female gametocytes). The clinical symptoms of malaria are produced primarily as a consequence of the asexual life cycle, while the sexual cycle, which can be divided into early (I-II) and late (III-V) gametocyte stages [36], is necessary for the development of the parasite in the mosquito. The intensive research on gene expression in the asexual stage compared to gametocyte stage can be inferred from the number of cDNA libraries deposited in the dbEST as mentioned above. The late (mature) stage gametocyte cDNA library (ID:10054) should contain transcripts important for gametocyte maturation and also formation of gametes and fertilization [37]. The availability of a cDNA library of 3D7 (ID:9765) asexual mixed stage (rings, trophozoites and schizonts) and genome data from the same strain presents an opportunity to determine differentially expressed transcripts between the two libraries.
chromosomal clusters encoding co-expressed proteins, unspliced mRNA, gene family member-specific expression and translational control [20,38,39]. These features contribute to parasite fitness and ability to undergo a complex life cycle. Understanding the role of these features in the regulation of important intraerythrocytic biological processes can deliver new tools for malaria control. For example, a proportion of genes involved in glycolysis, proteolysis and apicoplast targeting of nuclear encoded genes are thought to be regulated during the transition from asexual to sexual stages [7,40]. The integration of data from EST sequencing with those from genomic, microarray and proteomic technologies could provide insights into molecular mechanisms that contribute to the regulation of these processes.

The significant increase in disparate datasets from genome sequencing and post-genomic analysis of \textit{P. falciparum} necessitates delivery of integrated analysis from which biological processes important to the survival of the parasite can be determined. The integrated approach developed has identified stage-overexpressed genes with computational and experimental evidence to support their functional analysis. Furthermore, the approach is demonstrated as a means to appraise critically the data quality of the increasing number of post-genomic datasets from malaria parasites.

\textbf{Methods}

\textbf{Integrative analysis approach}

The integrative analysis approach that was used to combine genomic, expressed sequence tag, microarray, proteomic and gene ontology data from \textit{P. falciparum} 3D7 is presented in Figure 1. The starting integrative criterion was significant overexpression of a transcript in a stage relative to the other stage. Criteria used and their acceptable ranges are presented in Table 1.

\textbf{Expressed sequence tags and transcript reconstruction}

Expressed Sequence Tags derived from \textit{P. falciparum} 3D7 mixed asexual stage (dbEST ID: 9765) and gametocyte (III-V) stages (dbEST ID: 10054) cDNA libraries were retrieved using Sequence Retrieval System (SRS) version 7.02 from EMBL database (Release 74, March 2003). These sets of ESTs were sequenced by Washington University \textit{Plasmodium} EST Project [13]. A total of 15,126 ESTs consisting of 11,872 asexual and 3,254 gametocyte ESTs were downloaded. Transcript reconstruction of these ESTs was performed using stackPACK clustering system version 2.2 [22,24] as described previously for reconstructing \textit{Plasmodium} transcripts [1]. Briefly, the process starts with removal of artificial sequences such as repeats and vector sequences. The "clean" sequences are grouped using a loose clustering approach into clusters and the clusters assembled into contigs. The alignments of sequences that make up these assembled clusters are analysed to produce consensus sequences of maximal length representing the reconstructed transcripts. stackPACK was chosen for its ability to provide extended consensus sequences [41] (Hide et al. in preparation). Clusters containing only a single sequence are called singletons. A gene index, manufactured by such a method, is therefore a non-redundant representation of a set of reconstructed gene fragments that approximates to the best available representation of genes for that organism. The clustering was unsupervised in that known sequences such as mRNA, full-length cDNA, previously reconstructed ESTs or exon constructs were not used to guide the process. This type of clustering was required to provide valid input data for the software used to calculate the differential expression statistics applied in this study.

\textbf{Differential gene expression analysis}

Audic-Claverie (AC) and the Chi-square ($\chi^2$) 2 × 2 statistical tests for differential gene expression were used to identify stage-overexpressed transcripts. These pairwise tag statistics are based on EST counts of contigs (assembled clusters) with at least five ESTs since for a 95% confidence interval, the first value that is significantly different from 0 is 5 [30,32].

The calculation of these statistics was implemented with the web version of IDEG6 software; \texttt{http://telethon.bio.unipd.it/bioinfo/IDEG6/}, with a significance threshold of 0.05 [31]. A suite of PERL scripts was written to extract EST counts from output of stackPACK 2.2 and present the input dataset in the format required by IDEG6. Data extracted from the output file of IDEG6 were (1) contig description; (2) observed and normalised EST counts from the two libraries; and (3) probability that a transcript is differentially expressed as represented by $P$-values for the two tests. Transcripts for which the $P$-values for both statistics were less than 0.05 were taken as differentially expressed. Since these statistics determined transcripts differentially expressed, the terms asexual-overexpressed and gametocyte-overexpressed were used for transcripts (or genes) with significant overexpression in mixed asexual stage and late stage gametocytes respectively.

\textbf{Protein expression profiles and functional annotation of transcripts}

Annotated protein predictions (release 4.0) of the whole genome sequence of \textit{P. falciparum} 3D7 was obtained from the PlasmoDB website; \texttt{http://www.plasmodb.org}. A total of 5,334 predicted protein sequences were obtained. The overview page for each gene was retrieved using wget and saved as a Hypertext Markup Language (HTML) file on a local computer to allow ease of manipulation without accessing the database over the Internet. A PERL script was
Figure 1
Simplified flowchart of integrative analysis of Plasmodium falciparum data. Flowchart symbols: rounded rectangle, start or end; rectangle, process; diamond, decision.
used to query each page for the words sporozoite, merozoite, trophozoite or gametocyte preceded by an apostrophe (’) followed by a specific text as for the gametocyte; ‘gametocyte stage peptide fragment(s) detected by mass spectrometry’. A match of this text was taken as evidence of expression and protein expression at the stage was assigned 1 or else 0 for no evidence. Thus, a 4-digit binary accession that indicates evidence for expression in sporozoite, merozoite, trophozoite and gametocyte is used to represent the 15 protein expression profiles presented by Florens et al. [20] and an additional accession for lack of evidence in all stages (0000).

Reconstructed transcripts were annotated on the basis of similarity searches using NCBI BLASTX version 2.2.1 against predicted proteins of \textit{P. falciparum} 3D7. Statistical significance cut-off was set at an E-value of $10^{-10}$ following that of Carlton et al. [1]. Since an unsupervised clustering was performed, to support the functional annotation, the annotations obtained were correlated with the TIGR \textit{P. falciparum} Gene Index; \url{http://www.tigr.org/tdb/tgi/pfgi/} (Version 6.0, Release Date – January 11, 2003) and the Apicomplexan EST Database (ApiESTDB); \url{http://www.cbil.upenn.edu/paradbs-servlet/}. Both these indices were generated with supervised clustering. The correlation was done by computational extraction of associated annotation of the TIGR Tentative Consensus (TC) followed by manual checking to determine if the annotation obtained in our analysis was identical to that of the TIGR TCs. This was done for only differentially expressed contigs. If the annotations were not identical, the reconstructed sequence was excluded from further analysis. ApiESTDB was consulted when additional support was required to make a decision.

\textbf{Mining gene ontology annotation associated with transcripts}
Genes classified as being involved in glycolysis (GO:0006096), proteolysis (GO:0006508) or targeted to the plastid (GO:0009536) were retrieved by searching PlasmoDB gene overview page for the respective GO identification (ID) number in a similar way as described for the protein expression profile except the search text was the respective GO ID preceded by the greater than sign (>) for example >GO:0006096. This text limits the search to the Gene Ontology section of the gene overview page. The number of genes retrieved was: 20 for glycolysis, 98 for proteolysis and 553 for plastid component. This corresponds to values obtained from the web-based PlasmoDB query page.

\textbf{Correlation of EST-based abundance with microarray expression levels}
The numbers of ESTs used to generate a reconstructed sequence were retrieved from the FASTA sequence description line of all reconstructed sequences generated by stackPACK 2.2. The levels of expression or average signal intensities obtained from microarray experiments on the serine repeat antigen (SERA) gene family of \textit{P. falciparum} [19,42-44] were used to compare the levels of expression obtained using ESTs. This gene family is characterised by a cysteine proteinase framework [39] and was selected because its members are annotated as being involved in proteolysis. Published microarray studies on this family have been obtained that facilitated comparative analysis with EST data.

\textbf{Results}
\textbf{Transcript reconstruction and functional annotation of transcripts}
Transcript reconstruction using stackPACK 2.2 resulted in 1,760 contigs and 3,391 singletons. A total of 569 transcripts had an EST count of at least five ESTs. Agreement of pairwise differential expression statistics at $P < 0.05$ was 1. Maximum BLASTX E-value of $10^{-10}$ against predicted proteins was 1. Correlation of functional annotation with \textit{Plasmodium falciparum} gene indices was 1. Evidence that protein is expressed in same stage as gene was 1. Gene Ontology classification: proteolysis, glycolysis or localised to plastid was 1. Microarray: Published data on a gene family was 1.

\textbf{Table 1: Threshold values for steps in integrative analysis of \textit{Plasmodium falciparum} data}

| Criterion and acceptable range |
|-------------------------------|
| Reconstructed transcript derived from minimum of 5 ESTs |
| Agreement of pairwise differential expression statistics at $P < 0.05$ |
| Maximum BLASTX E-value of $10^{-10}$ against predicted proteins |
| Correlation of functional annotation with \textit{Plasmodium falciparum} gene indices |
| Evidence that protein is expressed in same stage as gene |
| Gene Ontology classification: proteolysis, glycolysis or localised to plastid |
| Microarray: Published data on a gene family |

\textbf{Differential expression transcripts and protein expression profiling}
The majority of the stage-overexpressed transcripts were from the late gametocyte stage. However, the mixed...
asexual stage had the highest percentage (83%) of genes with evidence of protein expression in the same stage (stage-correlated protein expression) compared to 31% for the late gametocyte stage. The observations are summarised in Tables 2 to 5. The 194 transcripts differentially expressed between the two libraries consisted of 51 from the mixed asexual stage and 143 from the late gametocyte stage. The complete list with transcript identification used in this study, correlated transcripts in the TIGR P. falciparum gene index, gene locus name, gene product description, representative EST or ESTs (for genes with representation from both libraries), observed and normalized EST counts for the two stages, as well as protein expression profile, are presented in the additional files 1 and 2 for mixed asexual stage and late gametocyte stage respectively. A list of stage-overexpressed transcripts that match those of Li et al. [13] is presented in additional file 3.

A total of 128 gametocyte-overexpressed and 48 asexual-overexpressed transcripts had a significant match with the predicted P. falciparum 3D7 proteins. Seventy-four genes (40 asexual-overexpressed, 34 gametocyte-overexpressed) showed evidence of stage-correlated protein expression (Tables 3 and 4). The well-studied S-antigen (PF10_0343) is one of the 8 asexual-overexpressed genes without stage-correlated protein expression. Four gametocyte-overexpressed genes (PFB0730w, PFI1210w, PF10_0115 and PFL0105w) had more than one reconstructed transcript. Multiple transcripts were generated when the reconstructed transcripts associated with a gene are not contiguous, and thus were not assembled into the same contig. Fifty-three of the 74 genes were classified as novel in that either the description of the gene product is labelled hypothetical protein or have the word putative.

In order to identify gametocyte-overexpressed genes that also have stage-correlated protein expression in the proteomics data of Lasonder et al. [21], the spreadsheet file containing 1,289 unique malaria proteins from that study was processed to yield a 3-digit binary accession representing evidence for protein expression of genes in trophozoites/schizonts, gametocytes and gametes. Fifteen of the 34 gametocyte-overexpressed genes were detected by both proteomic analyses (Table 6). Our analysis points to the need to clarify potential confusion in the annotation of the sexual stage specific protein precursor or Pfs16 (PFD0310w), a known marker for the earliest events of sexual differentiation [45]. The locus name (PF11_0318) of another gene, PF16, may be assigned to this gene [21]. PF16 has sequence similarity to a sperm flagella protein localized to the central pair of the axoneme. The gametocyte-overexpressed gene identified in this study was confirmed to be Pfs16 and not PF16 by the identical functional annotation of the associated consensus sequence from this study and that in the TIGR P. falciparum gene index.

The identified asexual-overexpressed genes that have been experimentally characterised have known roles in protein degradation, purine salvage, rhoptry biogenesis and protein trafficking, schizont rupture, merozoite invasion, phospholipid biosynthesis, nuclear metabolism, oxidative stress defense, cell proliferation and membrane biogenesis.

**Mining gene ontology annotation associated with transcripts**

Glyceraldehyde-3-phosphate dehydrogenase (PF14_0598) and ATP-dependent phosphofructokinase (PF11_0294) are two of 20 genes known to be involved in glycolysis. They demonstrate differential expression and show evidence of stage-correlated protein expression.

Microarray average intensities [19] available in PlasmoDB for PF11_0294 support its gametocyte-overexpression
when compared to a closely related gene, PF0755c that also codes for a phosphofructokinase and shows protein expression in intraerythrocytic stages [20,21]. The microarray expression values for PF0755c in trophozoite and schizont stages are 17,223.33 and 7,894 respectively in contrast to ~1,600 in both stages for PF11_0294. Inspection of the predicted protein features of PF11_0294 revealed the presence of two protein domains: gonadotropin-releasing domain, GnRH (Pfam ID: PF00446) and laminin N-terminal (Domain VI) (Pfam ID: PF00055). These domains are found in proteins that are extracellular and have a role in regulation of germ cell development.

**Table 3: Asexual-overexpressed *Plasmodium falciparum* transcripts**

| Transcript | TIGR Tentative Consensus | Gene locus name | Description of gene product | Representative EST(s) |
|------------|--------------------------|----------------|-----------------------------|-----------------------|
| cn672      | TC6879                   | PFI0265c       | rhoptry protein, putative   | BI670632              |
| cn1243     | TC6890 TC6891            | PFL1385c       | 101 kd malaria antigen      | BI670667              |
| cn656      | TC6894                   | PFI1_0098      | endoplasmic reticulum-resident calcium binding protein | BI670528 BM274707 |
| cn346      | TC6883 TC6884 TC6885     | PF14_0598 a    | glyceraldehyde-3-phosphate dehydrogenase | BI670581 BM273393 |
| cn659      | TC6886 TC6887            | PFBO340c f    | cysteine protease, putative | BI670678              |
| cn646      | TC6895                   | PF14_0102      | rhoptry-associated protein 1 | BI670673              |
| cn1292     | TC6896                   | PFI0875w       | Heat shock protein          | BI670644              |
| cn634      | TC6897 TC6898 TC8065     | MAL13P1.214    | phosphoethanolamine N-methyltransferase, putative | BI670572 |
| cn1258     | TC6900                   | PFI1445w       | hypothetical protein        | BI670690              |
| cn1175     | TC6899                   | PFC0120w       | Cytoadherence linked asexual protein, CLAG | BI670808 |
| cn637      | TC6921                   | PFE0165w       | actin depolymerizing factor, putative | BI813965 BM274236 |
| cn1246     | TC6922                   | MALBP1.142 4   | proteasome beta-subunit     | BI670563              |
| cn628      | TC6926                   | PF10_0203      | ADP-ribosylation factor     | BI814382              |
| cn1338     | TC6943                   | PF14_0141      | ribosomal protein L10, putative | BI670722 |
| cn1375     | TC6945                   | MAL7P1.77      | hypothetical protein        | BI814179              |
| cn1569     | TC6954 TC6955            | PFO915c        | proteasome subunit beta type 1 | BI670682 |
| cn1255     | TC6969 TC7520            | PBBO445c       | helicase, putative          | BI670715              |
| cn604      | TC6958                   | PFO210c        | eukaryotic initiation factor 5a, putative | BI670597 |
| cn1249     | TC6970                   | PF07_0054      | histone h2b, putative       | BI670668              |
| cn1465     | TC6959                   | PF14_0368      | 2-Cys peroxiredoxin         | BI670633              |
| cn581      | TC6975                   | PF14_0543 f    | hypothetical protein, conserved | BI814501 |
| cn1219     | TC6956                   | PF10_0345      | merosozite surface protein-3 | BI670568 |
| cn1339     | TC6992                   | PFL1420w       | macrophase migration inhibitor factor homolog, putative | BI815759 |
| cn1396     | TC6971                   | PF10_0121      | hypoxanthine phosphoribosyltransferase | BI814714 |
| cn567      | TC6917                   | PF10_0268      | merosozite capping protein-1 | BI670775 |
| cn1555     | TC7001                   | PFI0155c       | ras family GTP-ase, putative | BI814010 |
| cn561      | TC7038                   | PF10_0016      | acyl CoA binding protein, putative | BI815304 |
| cn1165     | TC7015                   | PFDO240c       | hypothetical protein        | BI816061              |
| cn1379     | TC7007                   | PF07_0087 f    | hypothetical protein        | BI813959              |
| cn1475     | TC6914                   | PF1090w        | s-adenosylmethionine synthetase, putative | BI813864 |
| cn1811     | TC6989 TC6990            | PF14_0323      | calmodulin                  | BI814267              |
| cn564      | TC6993                   | PFE1050w       | adenosylhomocysteinase(S-adenosyl-L-homocysteine hydrolase) | BI814536 |
| cn613      | TC7023 TC8311            | PFB0490c       | hypothetical protein        | BI815328              |
| cn1485     | TC7032                   | PF13_0228      | 405 ribosomal subunit protein S6, putative | BI670560 |
| cn1681     | TC7025                   | PF13_0328      | proliferating cell nuclear antigen | BI813993 |
| cn558      | TC7018                   | PF14_0678      | exported protein 2          | BI670646              |
| cn1605     | TC6904                   | MAL13P1.130    | hypothetical protein        | BI814223              |
| cn1997     | TC7030                   | PFE00660c      | uridine phosphorylase, putative | BI814451 |
| cn557      | TC7036                   | PF13_0092      | cholinephosphate cytidlylytransferase | BI814410 |
| cn1368     | TC7086                   | PF14_0569      | hypothetical protein        | BI814420              |

* Transcript generated by stackPACK 2.2. b TIGR Tentative Consensus correlated with transcript available at [http://www.tigr.org/tdb/tgi/pfgi/](http://www.tigr.org/tdb/tgi/pfgi/). c Gene can be viewed at [http://www.plasmodb.org](http://www.plasmodb.org). EST can be retrieved at [http://www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). e Gene involved in glycolysis. f Apicoplast-targeted gene. g Gene involved in proteolysis.
### Table 4: Gametocyte-overexpressed *Plasmodium falciparum* transcripts

| Transcript | TIGR Tentative Consensus | Gene locus name | Description of gene product | Representative EST(s) |
|------------|--------------------------|-----------------|-----------------------------|-----------------------|
| cn298      | TC6923 TC7297 TC9304     | PFD0310w        | sexual stage-specific protein precursor | BI814617 BM273325 |
| cn156      | TC6995                   | PFL0795c        | hypothetical protein         | BI813971 BM273682 |
| cn144      | TC7077                   | PFI1.0255       | hypothetical protein         | BM273367 |
| cn369      | TC6974                   | PF10.0264       | 4OS ribosomal protein, putative | BI814069 BM273547 |
| cn57       | TC7312 TC7511            | PFL2420w        | hypothetical protein         | BM273440 |
| cn271      | TC6963                   | PFB0730w        | DNA helicase, putative       | BM273418 |
| cn291      | TC6911                   | PF07.0029       | heat shock protein B6        | BM273491 |
| cn43       | TC6936                   | PFL2215w        | actin                        | BM273378 |
| cn105      | TC7084                   | PF1210w         | hypothetical protein         | BM274237 |
| cn168      | TC6963                   | PFB0730w        | DNA helicase, putative       | BM273308 |
| cn178      | TC6987                   | PFI1.0255       | hypothetical protein         | BM273367 |
| cn404      | TC7057                   | PF10.0115       | QF122 antigen                | BM273319 BQ596378 |
| cn46       | TC7235                   | PFL0105w        | hypothetical protein         | BM273988 BQ577236 |
| cn246      | TC7159                   | PFI1.0359       | hypothetical protein, conserved | BI814120 BM273571 |
| cn60       | TC7496                   | PF10.0328       | hypothetical protein         | BM273370 |
| cn155      | TC7437                   | PFI1.0294       | ATP-dependent phosphofructokinase, putative | BM273524 |
| cn269      | TC7203                   | MAL6P1.306      | hypothetical protein         | BI815038 BM273934 |
| cn347      | TC6987                   | PFI1210w        | hypothetical protein         | BM273395 |
| cn19       | TC7561                   | MAL13P1.148     | P. falciparum myosin         | BM274131 |
| cn683      | TC7619                   | PFD0235c        | hypothetical protein         | BM274865 |
| cn833      | TC7170                   | PFL0107c        | endoplasmic homolog precursor, putative | BM274046 |
| cn71       | TC6893                   | PFL0105w        | hypothetical protein         | BM273333 |
| cn93       | TC7763                   | PFI1.0460       | hypothetical protein         | BM273638 |
| cn165      | TC7103                   | PFI1.0316       | hypothetical protein         | BM274252 |
| cn288      | TC7304                   | PF10.0165       | DNA polymerase delta catalytic subunit | BM273631 |
| cn685      | TC7766                   | PFI1.0331       | c-complex protein 1, alpha subunit, putative | BM273631 |
| cn717      | TC7621                   | PF10.0115       | QF122 antigen                | BM273917 |
| cn737      | TC8144                   | PFL1395c        | hypothetical protein         | BM273513 |
| cn832      | TC7423                   | PFI0460w        | hypothetical protein         | BM273947 |
| cn49       | TC7047                   | PF10.0242       | hypothetical protein         | BM274006 BQ597262 |
| cn248      | TC7431                   | PFD0685c        | chromosome associated protein, putative | BM936055 BM274686 |
| cn326      | TC7394                   | PFC0570c        | hypothetical protein         | BM273462 BU496460 |
| cn750      | TC7788                   | PF10.0256       | hypothetical protein         | BM273462 BQ452171 |
| cn945      | TC7533                   | PFA0460c        | tubulin-specific chaperone a, putative | BM273558 BQ451292 |
| cn982      | TC7573                   | MAL6P1.48       | hypothetical protein, expressed | BI814116 BM273303 |
| cn681      | TC7652                   | PFE0845c        | 6OS ribosomal subunit protein L8, putative | BM273443 BU495928 |
| cn805      | TC7301                   | MAL13P1.120     | splicing factor, putative    | BI815872 BM274487 |

*Transcript generated by stackPACK 2.2. b TIGR Tentative Consensus correlated with transcript available at [http://www.tigr.org/tdb/tgi/pfgi/](http://www.tigr.org/tdb/tgi/pfgi/). Gene can be viewed at [http://www.plasmodb.org](http://www.plasmodb.org). d EST can be retrieved at [http://www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). e Gene involved in glycolysis. f Apicoplast-targeted gene.

### Table 5: Distribution of protein expression profiles for *Plasmodium falciparum* stage-overexpressed genes

| Gene category               | Binary accession | Count |
|-----------------------------|------------------|-------|
| **Asexual-overexpressed**   |                  |       |
| With protein expression     | 1111, 0111, 1011, 1110, 0011, 0101, 0110, 1010, 1100, 0010, 0100 | 40 |
| Without protein expression  | 0000, 1001, 0010, 1000 | 8 |
| **Gametocyte-overexpressed**|                  |       |
| With protein expression     | 1111, 0111, 1110, 0011, 0101, 1001, 0001 | 34 |
| Without protein expression  | 0000, 1110, 0110, 1010, 1100, 0010, 0100, 1000 | 87 |

*4-digit binary accession for protein expression evidence in sporozoite, merozoite, trophozoite and gametocyte.
data were available were checked for correlation of functional annotation and their EST count retrieved. As shown in Table 7, the EST counts were variable across the gene family consistent with microarray-based studies [42-44]. There was EST evidence for expression of PFB0345c (SERA4), PFB0340c (SERA5) and PFB0335c (SERA6), the three central genes that were demonstrated to be essential for asexual stage growth [42]. The GenBank accession numbers of a representative EST from these genes are BI936220, BI815392 and BQ633262 respectively. PFB0340c showed the highest EST count and microarray intensity values during asexual development of the parasite. Furthermore, multiple contigs mapped to this gene, which may represent alternative transcripts.

Out of the 17 transcripts (four asexual and 13 gametocyte) associated with genes targeted to the apicoplast, only two genes: MAL13P1.281 and PFE0145w have similarities to known genes (glutamate-tRNA ligase and 50S ribosomal subunit protein L8). There was evidence of protein expression in at least one asexual stage for two (PFB0340c, PFB0345c) of the four asexual-overexpressed genes (Table 3). Six gametocyte-overexpressed genes showed evidence for expression in the sporozoite stage while only PF11_0525 showed evidence in the sporozoite and gametocyte stages. PF11_0525 has predicted protein motifs that indicate its likely function. The domains are IQ (calmodulin-binding motif, Pfam ID: PF00612) and LysM (lysin motif, Pfam ID: PF01476), which is a general peptidoglycan-binding module. A list of apicoplast-targeted genes with stage-overexpressed transcripts is presented in additional file 4.

**Table 6: Gametocyte-overexpressed *Plasmodium falciparum* genes with correlated protein expression in two proteomic studies**

| Gene locus name | Description of gene product | Protein expression binary accession a | Florens et al. [20]b | Lasonder et al. [21]c |
|-----------------|-----------------------------|-----------------------------------|---------------------|----------------------|
| PFA0460c        | tubulin-specific chaperone a, putative | 0001                             | 011                 |
| PFD0310w        | sexual stage-specific protein precursor | 0001                             | 000                 |
| PFD0685c        | chromosome associated protein, putative | 0001                             | 000                 |
| PFE0845c        | 60S ribosomal subunit protein L8, putative | 0001                             | 000                 |
| PF07_0029       | heat shock protein 86          | 1111                             | 111                 |
| PF10_0165       | DNA polymerase delta catalytic subunit | 0111                             | 010                 |
| PF10_0242       | hypothetical protein           | 0111                             | 011                 |
| PF10_0264       | 40S ribosomal protein, putative | 0111                             | 011                 |
| PF11_0294       | ATP-dependent phosphofructokinase, putative | 0111                             | 011                 |
| PF11_0331       | t-complex protein I, alpha subunit, putative | 0111                             | 011                 |
| PF11_0525       | hypothetical protein           | 1001                             | 010                 |
| PFL0795c        | hypothetical protein           | 0001                             | 011                 |
| PFL1070c        | endoplasmic homolog precursor, putative | 0111                             | 011                 |
| PFL2215w        | actin                         | 0111                             | 011                 |
| PF14_0359       | hypothetical protein, conserved | 0111                             | 011                 |

a Evidence of expression: 0, no evidence; 1, with evidence. b 4-digit binary accession for protein expression evidence in sporozoite, merozoite, trophozoite and gametocyte. c 3-digit binary accession for protein evidence in trophozoite/schizont, gametocyte and gametes.

**Discussion**

An integrative approach was used to determine genes associated with transcripts differentially expressed between mixed asexual stage and late stage gametocyte parasites. The publication of the genome sequence of two malaria parasites presents opportunities for post-genomic era malaria research including gene discovery and comprehensive understanding of gene expression [46]. The study has revealed (1) possible regulatory mechanisms in malaria parasites’ gametocyte maturation, (2) correlation between EST and microarray data for a *P. falciparum* gene family to present unique EST-derived information, (3) candidate genes on which computational and experimental studies can be performed, and (4) the need for more empirical studies on gene and protein expression in malaria parasites.

A total of 569 contigs was used to determine stage-overexpression. This presents 366 more contigs than described by Li et al. [13] reflecting inclusion of new mixed asexual stage ESTs deposited after March 2002. Only 21 of the 24 significantly stage-specific transcripts identified by Li et al. [13] were among our stage-overexpressed transcripts after correlation of functional annotation. Both studies demonstrate the asexual-overexpression of the gene for glyceraldehyde-3-phosphate dehydrogenase (GAPDH), an important gene in the glycolytic pathway [47].

Gene and protein expression were observed, as well as protein domain evidence for specialization or adaptation of ATP-dependent phosphofructokinase (PF11_0294) for metabolic coupling of glucose utilization and maturation of gametocytes in malaria parasites. This enzyme is of
major regulatory importance in *Plasmodium* and has been characterised only in *Plasmodium berghei* [48]. In addition, it has been proposed as a potential drug target in protozoan parasites [49]. Two genes (PF11_0294, PFI0755c) annotated as phosphofructokinase are present in the genome [7]. This is consistent with the fact that many key enzymes in the glycolytic pathway occur as isoenzymes [48]. Interestingly, PF11_0294 possesses a gonadotropin-releasing domain GnRH and laminin N-terminal (Domain VI) that are thought to regulate germ cell development. PFI0755c does not contain these domains.

PF11_0525 is the only apicoplast-targeted gene associated with a gametocyte-overexpressed transcript that showed stage-correlated protein expression. The fact that germ cell biology is conserved in evolution enables us to speculate on the possible roles of this protein. The calmodulin (CaM) binding site has been extensively studied in a sperm autoantigen (Sp17), which is a zona binding protein and a member of the family of CaM binding proteins that contain the IQ motif in the CaM binding domain. This domain has a regulatory role and undergoes proteolytic processing at the initiation of an acrosome reaction [50]. Some bacterial proteins such as hydrolytic enzymes contain the general peptidoglycan-binding module (LysM) and have a role in cell-wall penetration [51]. PF11_0525 does not have evidence of a bipartite peptide for apicoplast targeting and thus may be targeted via a different mechanism to the organelle or it may no longer function in the plastid.

The EST counts of the SERA gene family are comparable with the gene expression levels observed in microarray experiments. Both technologies agree that expression levels of members are variable as is expression of central genes during the asexual stage of the parasite. PB0340c (SERA5) is the first described member of the family [39] and is also a malaria vaccine candidate [52]. The EST counts for PB0340c observed is consistent with high gene expression levels in trophozoites and schizonts in published microarray experiments. Specifically, Miller et al. [42] and Aoki et al. [52] observed PB0340c to be substantially more strongly transcribed than other SERA genes.

The increasing amount of published and unpublished data from microarray, SAGE, EST and differential display on malaria parasites shows that pairwise correlation is required. Comparison of such datasets obtained from different gene expression technologies can complement less sensitive technologies, hence adding value to data generation from these methods. For example, this study provides identity of ESTs and also potential alternative transcripts that can be used to further characterize the SERA central genes. Furthermore, PFB0325c (SERA8) did not have EST evidence consistent with low or absent expression observed in the microarray studies. However, there was evidence of its expression in the sporozoite stage, indicating the gene may be functional in other stages of the life cycle as speculated by Miller et al. [42]. Large-scale comparative expression analysis of gene families in multiple malaria parasites is needed to advance the knowledge of their evolution and their role during intraerythrocytic development.

Table 7: Correlation of EST abundance and microarray intensity associated with SERA gene family

| Gene (Locus name) | EST counta | Comments b | Microarray intensity values c | Wu et al. [44] |
|-------------------|------------|------------|-------------------------------|---------------|
|                   | Miller et al. [42] | Le Roch et al. [43] | Bozdech et al. [19] | |
| SERA8 (PFB0325c)  | -           | +/-        | 35.3                         | -             |
| SERA7 (PFB0330c)  | 7           | +/-        | 160.5                        | 104.8         |
| SERA6 (PFB0335c)  | 2           | +          | 200.7                        | 1298          |
| SERA5 (PFB0340c)  | 98          | +          | 1225.4                       | 10265.5       |
| SERA4 (PFB0345c)  | 4           | +          | 200                          | 496.7         |
| SERA3 (PFB0350c)  | -           | +          | 87.3                         | 341           |
| SERA2 (PFB0355c)  | -           | +/-        | 185.4                        | 219.4         |
| SERA1 (PFB0360c)  | 2           | +/-        | 125.9                        | 178.1         |

a, no ESTs observed. b, Comments on gene expression: +/-, low or absent expression; +, expression confirm by RT-PCR and microarray. c, R, Rings; T, Trophozoite; S, Schizont; Asyn, asynchronous culture; -, No expression value reported. d, EST count of TIGR TC7227. e, Central genes in the SERA locus that could not be disrupted in study [42]. f, Gene with multiple transcripts, TC6886 (B1670678) TC6962 (B1814353).
vided that can be intensively studied as drug targets, antigenic targets, epidemiological or clinical markers. Eighty-seven of the 121 gametocyte-overexpressed genes did not show evidence of stage-correlated protein expression while 15 of those with such evidence were corroborated by the two proteomics studies. These corroborated genes represent a set of gametocyte-overexpressed genes with correlated transcription and translation data and thus candidates for studies on gametocyte maturation in malaria parasites. A shortlist of stage-overexpressed genes targeted to the plastid is presented to facilitate studies to understand the regulation of plastid metabolism in malaria parasites.

This study has identified the lack of correlation between gene and protein expression of the asexual-overexpressed S-antigen, consistent with observations from published proteome analysis [20]. This observation and those from the gametocyte-overexpressed transcripts as well as comparing outputs from EST clustering efforts demonstrate that our integrative approach has the utility to compare outputs of different post-genomic analysis. The analysis indicates the need for additional empirical studies on gene and protein expression in malaria parasites. Such studies could improve current understanding on discrepancies between gene and protein expression profiling data as well as the detection of proteins with unique characteristics such as proteolytic processing, post-translational modification and sub-cellular location.

Conclusions

The value of integrating a variety of datasets to unravel undiscovered regulation in biological processes during the gametocyte maturation stages of *P. falciparum* was demonstrated. Furthermore, comparative analysis of EST and microarray data was performed on the SERA gene family to advance the knowledge of their gene regulation and additional functional genomics reagents were presented to facilitate their study. Finally, the integrative approach was shown as a means to appraise critically the data quality of the increasing number of post-genomic datasets from malaria parasites.

Additional material

**Additional File 1**
Plasmodium falciparum asexual-overexpressed transcripts
Click here for file
[http://www.biomedcentral.com/content-supplementary/1475-2875-2-38-S1.xls]

**Additional File 2**
Plasmodium falciparum gametocyte-overexpressed transcripts
Click here for file
[http://www.biomedcentral.com/content-supplementary/1475-2875-2-38-S2.xls]

**Additional File 3**
Correlated stage-overexpressed transcripts in this study and that of Li et al. [13]
Click here for file
[http://www.biomedcentral.com/content/supplementary/1475-2875-2-38-S3.xls]

**Additional File 4**
Plasmodium falciparum candidate genes for studies into plastid metabolism
Click here for file
[http://www.biomedcentral.com/content/supplementary/1475-2875-2-38-S4.xls]

Acknowledgements

The authors thank colleagues at the South African National Bioinformatics Institute for useful suggestions and staff of Electric Genetics for stackPACK support. RDI is a Claude Harris Leon Foundation Fellow and thanks the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) and the Malaria Research and Reference Reagent Resource Center (MR4) for grants to attend workshops on Malaria Bioinformatics and Microarrays.

References

1. Carlton JM, Muller R, Yowell CA, Fluegge MR, Sturrock KA, Pritt JR, Vargus-Serrato E, Galinski MR, Barnwell JW, Mulder N, Kanapin A, Cawley SE, Hide WA, Dame JB. Profiling the malaria genome: a gene survey of three species of malaria parasite with comparison to other apicomplexan species. *Mol Biochem Parasitol* 2001, 118:201-210.
2. Davids W, Gamieldien J, Libeerls DA, Hide W. Positive selection scanning reveals decoupling of enzymatic activities of carbamoyl phosphate synthetase in *Helicobacter pylori*. *J Mol Evol* 2002, 54:458-464.
3. Gamieldien J, Pitsyn A, Hide W: Eukaryotic genes in *Mycobacterium tuberculosis* could have a role in pathogenesis and immunomodulation. *Trends Genet* 2002, 18:5-8.
4. Matce C, Sagot MF, Schiex T, Rouze P: Current methods of gene prediction, their strengths and weaknesses. *Nucleic Acids Res* 2002, 30:1-14.
5. Brennan JG: The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am J Trop Med Hyg* 2001, 64:1-11.
6. WHO/UNICEF: The Africa Malaria Report 2003 Geneva: WHO/UNICEF, 2003.
7. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shalom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AHI, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LH, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrett B. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002, 419:498-511.
8. Carlton JM, Angiuoli SV, Suh BB, Kojoi TW, Pertea M, Silva JC, Ermo laeva MD, Allen JE, Selengut JD, Koo HL, Peterson JD, Pop M, Kosack DS, Shumway MF, Bidwell SL, Shalom SJ, van Aken SE, Riedmuller SB, Feldblyum TV, Cho JK, Quackenbush J, Segedah M, Shoal BA, Cummings LM, Flores L, Yates JR, Raine JD, Sinden RE, Harris MA, Cunningham DA, Preiser PR, Bergman LW, Vaidya AB, van Lin LH, Janse CJ, Waters AP, Smith HO, White OR, Salzberg SL, Venter JC, Fraser CM, Hoffman SL, Gardner MJ, Carucci DJ: Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 2002, 419:519.
9. Carlton J: The *Plasmodium vivax* genome sequencing project. *Trends Parasitol* 2003, 19:227-231.
10. Kappe SH, Gardner MJ, Brown SM, Ross J, Matuschewski K, Ribeiro JM, Adams JH, Quackenbush J, Cho J, Carucci DJ, Hoffman SL, Nussenzeneg K. Exploring the transcriptome of the malaria sporozoite stage. Proc Natl Acad Sci U S A 2001, 98:8995-9000.

11. Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parviz B, Pertea G, Sultan A, White J. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res 2001, 29:1559-161.

12. Kongkasuriyachai D, Kumar N. Functional characterisation of sexual stage specific proteins in Plasmodium falciparum. Int J Parasitol 2002, 32:1559-1566.

13. Li L, Brunk BP, Kissing KC, Pape D, Tang K, Cole RH, Martin J, Wylie T, Lacoste M, Fogarty SJ, Howe DK, Liberator P, Dicz K, Anderson DL, Carlton JM, Dame JB, Chakrabarti D, Martin RK, Chung F, Antonescu V, White J, Holt I, Liang F, Quackenbush J. Co-ordinated programme of clustering EST and full-length cDNA sequences. Nucleic Acids Res 2002, 30:113-14.

14. Watanabe J, Sasaki M, Suzuki Y, Sugano S. A comprehensive approach to transcriptome of human malaria parasite Plasmodium falciparum using full-length enriched library: identification of novel genes and diverse transcription start sites of messenger RNAs. Gene 2002, 291:105-111.

15. Munasinghe A, Patankar S, Cook BP, Madden SL, Martin RK, Kyle DE, Shoabi A, Cummings LM, Wirth DF. Serial analysis of gene expression (SAGE) in Plasmodium falciparum: application of the technique to A-T rich genomes. Mol Biochem Parasitol 2001, 113:233-4.

16. Patankar S, Munasinghe A, Shoabi A, Cummings LM, Wirth DF. Serial analysis of gene expression in Plasmodium falciparum reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. Mol Bio Cell 2001, 12:314-315.

17. Hayward RE, DeRisi JL, Alford SL, Kaslow DC, Brown PO, Rathod PK. Shotgun DNA microarrays and stage-specific gene expression in Plasmodium falciparum malaria. Mol Microbiol 2000, 35:6-14.

18. Ben Mamoun C, Glaunam JY, Hott C, MacMillan SK, Amarakone AS, Anderson DL, Clarkson JM, Dame JB, Chakrabarti D, Martin RK, Brownstein BH, Goldberg DE. Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite Plasmodium falciparum using full-length and partial cDNA sequences. Nucleic Acids Res 2000, 28:3726-3736.

19. Bozdech Z, Zhu J, Joachimiak MP, Cohen FE, Pulliam B, DeRisi JL, Hubbard R. Shotgun DNA microarrays and stage-specific gene expression in Plasmodium falciparum malaria. Mol Microbiol 2000, 35:6-14.

20. Ben Mamoun C, Glaunam JY, Hott C, MacMillan SK, Amarakone AS, Anderson DL, Clarkson JM, Dame JB, Chakrabarti D, Martin RK, Brownstein BH, Goldberg DE. Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite Plasmodium falciparum using full-length and partial cDNA sequences. Nucleic Acids Res 2000, 28:3726-3736.

21. Bozdech Z, Zhu J, Joachimiak MP, Cohen FE, Pulliam B, DeRisi JL. Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray. Genome Biol 2003, 4:89.

22. Flores F, Washburn MP, Raine TD, Anthony RM, Grainger M, Hayes JD, Moch JK, Muster N, Sacc J, Tabb DL, Witney AA, Wolters D, Wu Y, Gardner MJ, Holder AA, Sinen DS, Yates JR, Carucci DJ. A proteomic view of the Plasmodium falciparum life cycle. Nature 2002, 419:520-526.

23. Lech KG, Zhou Y, Batalov S, Winsel EA. Monitoring the chromosome 2 intraerythrocytic transcriptome of Plasmodium falciparum using oligonucleotide arrays. Am J Trop Med Hyg 2002, 67:233-243.

24. Wu Y, Wang X, Liu X, Wang Y. Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite. Genome Res 2003, 13:601-616.

25. Dechering K, Kaan AM, Mbacham W, Wirth DF, Elng W, Konings RN, Stunnenberg HG. Isolation and functional characterization of two distinct sex-related stage-specific promoters of the human malaria parasite Plasmodium falciparum. Mol Cell Biol 1999, 19:967-978.

26. Horrocks P, Bowman S, Kyes S, Waters AP, Craig A. Entering the post-genomic era of malaria research. Bull World Health Organ 2002, 79:142-147.

27. Campanale N, Nickel C, Dauemberger CA, Wehlan DA, Gorman JJ, Klonis L, Becker K, Tilley L. Identification and characterization of heme-interacting proteins in the malaria parasite, Plasmodium falciparum. J Biol Chem 2003, 278:27354-27361.

28. Sherman IW. Carbohydrate metabolism of asexual stages. In Malaria parasite biology, pathogenesis, and protection Edited by: Sherman IW. Washington, DC: ASM Press; 1998:135-145.
49. Chi AS, Deng Z, Albach RA, Kemp RG: The two phosphofructokinase gene products of Entamoeba histolytica. J Biol Chem 2001, 276:19974-19981.

50. Wen Y, Richardson RT, O’rand MG: Processing of the sperm protein Sp17 during the acrosome reaction and characterization as a calmodulin binding protein. Dev Biol 1999, 206:113-122.

51. Bateman A, Bycroft M: The structure of a LysM domain from E. coli membrane-bound lytic murein transglycosylase D (MitD). J Mol Biol 2000, 299:1113-1119.

52. Aoki S, Li J, Itagaki S, Okech BA, Egwang TG, Matsuoka H, Palacpac NM, Mitamura T, Hori T: Serine repeat antigen (SERA5) is predominately expressed among the SERA multigene family of Plasmodium falciparum, and the acquired antibody titers correlate with serum inhibition of the parasite growth. J Biol Chem 2002, 277:47533-47540.