January 2006

The *Plasmodium* selenoproteome

Alexey V. Lobanov
University of Nebraska-Lincoln

Cesar Delgado
University of Nebraska-Lincoln

Stefan Rahlfs
Interdisziplinäres Forschungszentrum, Justus-Liebig University Giessen, Giessen, Germany

Sergey V. Novoselov
University of Nebraska-Lincoln, snovoselov2@unl.edu

Gregory V. Kryukov
University of Nebraska-Lincoln

See next page for additional authors

496–505 *Nucleic Acids Research, 2006, Vol. 34, No. 2*
doi:10.1093/nar/gkj450

The *Plasmodium* selenoproteome

Alexey V. Lobanov[1], Cesar Delgado[1,2], Stefan Rahlfs[3], Sergey V. Novoselov[1], Gregory V. Kryukov[1], Stephan Gromer[4], Dolph L. Hatfield[5], Katja Becker[3] and Vadim N. Gladyshev[1,*]

[1]Department of Biochemistry and [2]Department of Computer Science, University of Nebraska, Lincoln, NE 68588, USA, [3]Interdisziplinäres Forschungszentrum, Justus-Liebig University Giessen, Heinrich-Buff-Ring 26-32, D-35392 Giessen, Germany, [4]Biochemie-Zentrum der Universität Heidelberg, Im Neuenheimer Feld 504, D-69120 Heidelberg, Germany and [5]Molecular Biology of Selenium Section, Laboratory of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Received August 22, 2005; Revised October 3, 2005; Accepted January 3, 2006

ABSTRACT

The use of selenocysteine (Sec) as the 21st amino acid in the genetic code has been described in all three major domains of life. However, within eukaryotes, selenoproteins are only known in animals and algae. In this study, we characterized selenoproteomes and Sec insertion systems in protozoan Apicomplexa parasites. We found that among these organisms, *Plasmodium* and *Toxoplasma* utilized Sec, whereas *Cryptosporidium* did not. However, *Plasmodium* had no homologs of known selenoproteins. By searching computationally for evolutionarily conserved selenocysteine insertion sequence (SECIS) elements, which are RNA structures involved in Sec insertion, we identified four unique *Plasmodium falciparum* selenoprotein genes. These selenoproteins were incorrectly annotated in PlasmoDB, were conserved in other Plasmodia and had no detectable homologs in other species. We provide evidence that two *Plasmodium* SECIS elements supported Sec insertion into parasite and endogenous selenoproteins when they were expressed in mammalian cells, demonstrating that the *Plasmodium* SECIS elements are functional and indicating conservation of Sec insertion between Apicomplexa and animals. Dependence of the plasmodial parasites on selenium suggests possible strategies for antimalarial drug development.

INTRODUCTION

Plasmodium falciparum, the organism which causes tropical malaria in man, has a genome of 23 Mb, the entire sequence of which has been determined (1). Significant sequence information is also available for several other plasmodial species (2), as well as for related apicomplexan pathogens, including *Toxoplasma* and *Cryptosporidium* species (3–5). These genomics resources stimulated research on these pathogenic protozoa (6,7). However, much remains unknown about the biology of these parasites.

Selenium is a trace element with significant biomedical potential. It is inserted into protein in response to a UGA codon as selenocysteine (Sec), the 21st amino acid in the genetic code (8–11). However, for most organisms, their dependence on selenium and selenoproteins is not known. Initial sequence and database analyses have shown the presence of a putative Sec-specific elongation factor EFSec/eSelB in *P.falciparum*, suggesting the possibility that *Plasmodium* utilizes Sec insertion and contains Sec-containing proteins. Known selenoproteins have Sec in active sites to carry out redox functions (10). Selenoproteins are highly reactive compared with their cysteine-containing analogs, but are also susceptible to inactivation by a variety of electrophilic agents. In particular, the use of gold compounds and alkylating agents allows specific targeting of Sec residues in proteins (12,13). Thus, the presence of selenoproteins may suggest novel therapeutic strategies.

The identification of selenoproteins in plasmodial species is also of interest from an evolutionary point of view. Selenoproteomes have been characterized in the three major domains of life (14–18). Within prokaryotes, for which complete genomic sequences are available, selenoproteins have been described in several archaea and in diverse bacteria species. However, within eukaryotes, selenoproteins are so far only known in animals and algae (19,20). In fact, a large protozoan community, including Apicomplexa, has been essentially unexplored with regard to the role of the Sec insertion in these organisms.

Recently, $tRNA^{Sec}$ was reported in Plasmodia and a computational search for selenocysteine insertion sequence

*To whom correspondence should be addressed. Tel: +1 402 472 4948; Fax: +1 402 472 7842; Email: vgladyshev1@unl.edu

(SECIS) elements revealed a single SECIS element in these organisms (21). In the current study, we carried out bioinformatics analyses and identified four selenoproteins in Plasmodia. Interestingly, these selenoproteins showed no homology to either known selenoproteins in other species or to non-plasmodial proteins in general.

MATERIALS AND METHODS

Databases and programs

P.falciparum, *P.yoelii*, *P.knowlesi*, *P.berghei*, *P.vivax* and *P.chabaudi* genomes were downloaded from PlasmoDB at www.plasmodb.org (Plasmodium Sequence Release November 04, 2003). The search for SECIS elements was carried out using SECISearch (17,22). *Toxoplasma gondii* genomic sequences were obtained from ToxoDB (www.toxodb.org, Data Release August 13, 2003). To identify Sec tRNA genes, we utilized COVE, tRNAscan-SE (23) and ARAGORN (24) programs. BLAST searches were used to identify other components of the Sec insertion system. To predict the presence of signal peptides, we used SignalP 3.0 Server (25), which utilizes neural networks and hidden Markov models trained on eukaryotes.

Search for SECIS elements

The search for selenoproteins genes was carried out using the following algorithm:

(1) Identification of potential SECIS elements by analyzing primary sequences. This step included the search for NTGA__AA__GA or NTGA__CC__GA motifs in nucleotide sequences with the distances between the Quartet (NTGA) and the unpaired AA in the apical loop of 10–13 nt, and between the AA to the GA of 15–39 nt.
(2) Analyses of secondary structure features of SECIS elements. Searches were performed to identify secondary structures consistent with the eukaryotic SECIS consensus. Additional filters were implemented to filter out candidates with unsuitable secondary structures (i.e. Y-shaped SECIS elements and SECIS elements with more than 2 unpaired nucleotides in a row within the first 7 nt pairs following NTGA).
(3) Calculation of free energy for each candidate structure. Only thermodynamically stable structures were considered. For that purpose, the free energies for the whole structure (with the threshold value of −12.6 kcal/mol) and the upper stem–loop (with the threshold of −3.7 kcal/mol) were calculated.
(4) Identification of candidates for which homologs could be detected in other Plasmodia. The general approach was that, considering evolutionary distance among Plasmodia, candidates should be conserved in at least some species. Thus, the search was restricted to evolutionary conserved selenoproteins (the BLAST expect value below 0.001).
(5) Analyses of SECIS location and identification of upstream open reading frames (ORFs). The purpose of this step was to filter out SECIS candidates, which were located within coding regions of known proteins.
(6) Sequence analyses of predicted selenoprotein ORFs located upstream of candidate SECIS elements. This step included identification of splice sites, ORFs, homologs in other species and searches against the *Plasmodium* EST database. SECIS candidates were filtered out if they were located within annotated ORFs or within introns of annotated genes.

Searches for Sec insertion machinery

To identify components of the Sec insertion machinery, we searched for homologs of known factors involved in Sec insertion. Decoding UGA as Sec requires a unique tRNA, a Sec-specific elongation factor EFSec, SECIS-binding protein SBP2 and selenophosphate synthetase (SelD, SPS). The presence of EFSec, SBP2 and SPS homologs in different *Plasmodium* species was determined using BLAST searches against human counterparts.

Several approaches were attempted for the identification of Sec tRNA in apicomplexan genomes. A common method relies on tRNAscan-SE (23), but this program, when used with the default settings, did not recognize Sec tRNA in *P.falciparum*. However, the 'maximum sensitivity' mode (COVE with a Sec-tRNA profile) of tRNAscan-SE and ARAGORN programs (24) could identify this tRNA in both *Plasmodium* and *Toxoplasma* genomes. In addition, we have developed an algorithm to search for non-canonical tRNA sequences, which is described elsewhere (A.V. Lobanov, G.V. Krykov, D.L. Hatfield and V.N. Gladyshev, manuscript submitted). Use of this program independently identified the same Sec tRNA.

Phylogenetic analyses

Amino acid and nucleotide sequence alignments were generated using the ClustalW multiple sequence alignment program. Distances were calculated between each pair of sequences and these were used to construct the phylogenetic tree using the Neighbor Joining method of Saitou and Nei (26). TreeView was used for visualization of the trees.

Transcription of *P.falciparum* selenoprotein genes

PCR primers were obtained from MWG-Biotech (Ebersberg, Germany), *Taq*-polymerase was from Finnzymes (BioTec, Heidelberg, Germany), and the TA direct cloning vector pDrive was purchased from Qiagen (Hilden, Germany). A *P.falciparum* blood stage cDNA library from the strain 3D7 was used as template for the PCR amplifications. To verify the presence of selenoprotein mRNAs and to test PlasmoDB annotations and our predictions with regard to gene organization of the four *Plasmodium* selenoprotein genes (Figure 2), various primers were developed and employed in various combinations in a PCR procedure. The amplified sections of the four genes were cloned into the cloning vector and verified by sequencing. The following oligonucleotide primers were used (Figure 2): Selenoprotein #1, N: 5′-ATGGATGACAGAAAGGAAAAC-3′; s: 5′-AAGACCCTACTTGGACAATC-3′; C: 5′-TCAGGTACTTCCCAAGCAGG-3′; CU: 5′-ATGTCGGAAAAACATTTAACC-3′. Selenoprotein #2, N: 5′-ATGAAAGCTTATTACGTGGTG-3′; C: 5′-TTATTTATGAATATTTTTCAGATTTG-3′; Ca: 5′-TTAAATTGTAATATAATATAACTC-3′; as: 5′-ATGAATTTCATTTGTATTACATAG-3′. Selenoprotein #3, N: 5′-ATGATTTTAAAAAAAGTGTATATAC-3′; C: 5′-TTATGTCCTTTTAAATT-

TAAAAGG-3′; as: 5′-AAAAGAGTTATTTTGTATATGTT-CC-3′. Selenoprotein #4, N: 5′-ATGGACACAAATGAAAA-TATGAAA-3′; as: 5′-CCTATTTTTTAATGACGGCC-3′; C: 5′-TTATCGTCAGGATTTGCGATTATTG-3′; nC: 5′-TTTT-TTTGTACTTTTAATTCATCC-3′; CD: 5′-CTAATTTTTT-CCTTTTATTATTTTA-3′; gf: 5′-ACTTCACGTGTACATT-CTTC-3′.

Selenocysteine insertion into proteins supported by *Plasmodium* SECIS elements

pEGFP-C-3 Vector (Clontech) was used for preparing green fluorescent protein (GFP) fusion constructs. A GFP-SelH expression construct containing a mouse SelH SECIS element was described previously (22). A GFP-Sel1 construct containing a *P.falciparum* Sel1 SECIS element was obtained by cloning Sel1 coding and 3′-untranslated region (3′-UTR) sequences, which were amplified with primers 5′-GAGAACTCGAGAGATTATGAAATATTTTTAAGAG-3′ and 5′-CTTTACCTGCAGCAAAAGCATATAAATGAA-TATAACATCATTTC-3′ into an XhoI/PstI restriction site immediately downstream of the GFP coding region. A GFP-Sel4 fusion construct was obtained by cloning a PCR fragment, amplified with primers 5′-GTATCCTGCAGCTAA-CATCCCTTTAAAGCTATG-3′ and 5′-CGATGGATCCGT-CATGAAGATATTTTCATTTATATAACATTGTGTAGC-3′, into a PstI/BamHI site. NIH3T3 cells were transfected with these constructs as well as GFP-SelH (positive control) and GFP alone (negative control) constructs using Lipofectamin (Invitrogen) in the presence of a Plus Reagent (Invitrogen). After transfection, the cells were metabolically labeled with ^{75}Se and harvested after 24 h, and proteins were electrophoresed on an SDS–PAGE gel and transferred onto a PVDF membrane. A ^{75}Se radioactivity profile was detected with a PhosphorImager system (Amersham).

RESULTS AND DISCUSSION

Sec insertion in Apicomplexa

To determine if apicomplexan protozoa encode selenoproteins, we initially analyzed *P.falciparum*, *T.gondii* and *Cryptosporidium* genomes for the presence of homologs of known selenoproteins. The full set of human selenoproteins (17), as well as *Chlamydomonas* MsrA (19,20), *Gallus gallus* SelU (27) and protein disulfide isomerase from *Emiliania huxleyi* (28) were used as query sequences. BLAST searches revealed no selenoprotein homologs in *P.falciparum* and *Cryptosporidium*, but detected a distant homolog of SelW, in *T.gondii* (data not shown). We further searched *Plasmodium* and *Cryptosporidium* genomes for the presence of EFSec/eSelB, SBP2 and SPS homologs. These proteins were identified in *P.falciparum*, but were absent in *Cryptosporidium*. Thus, these results suggested that *P.falciparum* and *T.gondii* utilize the Sec insertion, whereas *Cryptosporidium* does not, and verified the recent prediction that *Plasmodium* utilizes the Sec insertion system (21).

Identification of SECIS elements in *Plasmodium*

The presence of the Sec insertion system and the lack of known selenoproteins in *P.falciparum* suggested that this organism might have evolved selenoproteins that are unique to this organism. Initially, we used the default pattern of SECISearch to quickly screen for possible SECIS elements and selenoprotein genes in *P.falciparum*. Two strong candidates were identified (see below), suggesting that indeed Plasmodia evolved a unique selenoproteome.

To increase specificity of the selenoprotein searches in Plasmodia, entire genomes of *P.falciparum* and *P.yoelii* and partial genomic sequences of *P.vivax*, *P.chabaudi*, *P.knowlesi* and *P.berghei* were then analyzed with a loose pattern of SECISearch (Table 1). A total of 38 010 *P.falciparum* hits satisfied primary and secondary structure criteria of the SECIS consensus, and application of thermodynamic criteria reduced the set to 212 SECIS candidates. Other *Plasmodium* genomes produced variable numbers of candidates (Table 1). In the next step, all SECIS candidates were compared with each other using BLASTN and candidates with a homology score below $e = 0.001$ were extracted. This step reduced the number of *P.falciparum* SECIS candidates to 88. Further manual analysis of these structures and analyses of predicted ORFs filtered out all candidates except 4 (Table 1).

The four *P.falciparum* SECIS candidates generated by SECISearch are shown in Figure 1. One of the matches (SECIS #3) was also predicted by Mourier *et al.* (21). This recent study also reported a bioinformatics search for SECIS elements, but identified only a single candidate structure. It was suggested that if additional SECIS elements are present in *P.falciparum*, they may be non-canonical (21).

However, the four SECIS candidates identified in our work are in full compliance with the current consensus SECIS model. The four structures satisfied all criteria of eukaryotic SECIS elements, including the presence of two major helices, unpaired AA in the apical bulge and unpaired A as the nucleotide preceding the Quartet (9). In addition, the Quartet (SECIS core) had a UGAN. . .NGAN sequence, which forms four non-Watson–Crick base pairs (29) and is the most characteristic feature of eukaryotic SECIS elements.

Table 1. Computational search for *Plasmodium* selenoproteins

Species name and strain	Database	Size (bp)	Number of SECIS candidates after		Evolutionary criteria	Selenoproteins found
			Primary/secondary structure analyses	Free energy analysis		
P.falciparum 3D7 strain	Whole genome	23 181 054	38 010	212	88	#1, #2, #3 and #4
P.yoelii 17XNL strain	Contigs	26 029 223	45 141	1255	41	#1, #2, #3 and #4
P.berghei ANKA strain	Whole genome shotgun	29 163 218	49 944	274	28	#1 and #3
P.chabaudi AS strain	Contigs	17 078 234	33 877	1958	21	#1, #2, #3 and #4
P.knowlesi H strain	GSS	20 348 340	28 261	1907	51	#1, #2, #3 and #4
P.vivax (multiple strains)	GSS	6 917 471	6850	767	5	#1, #2 and #3

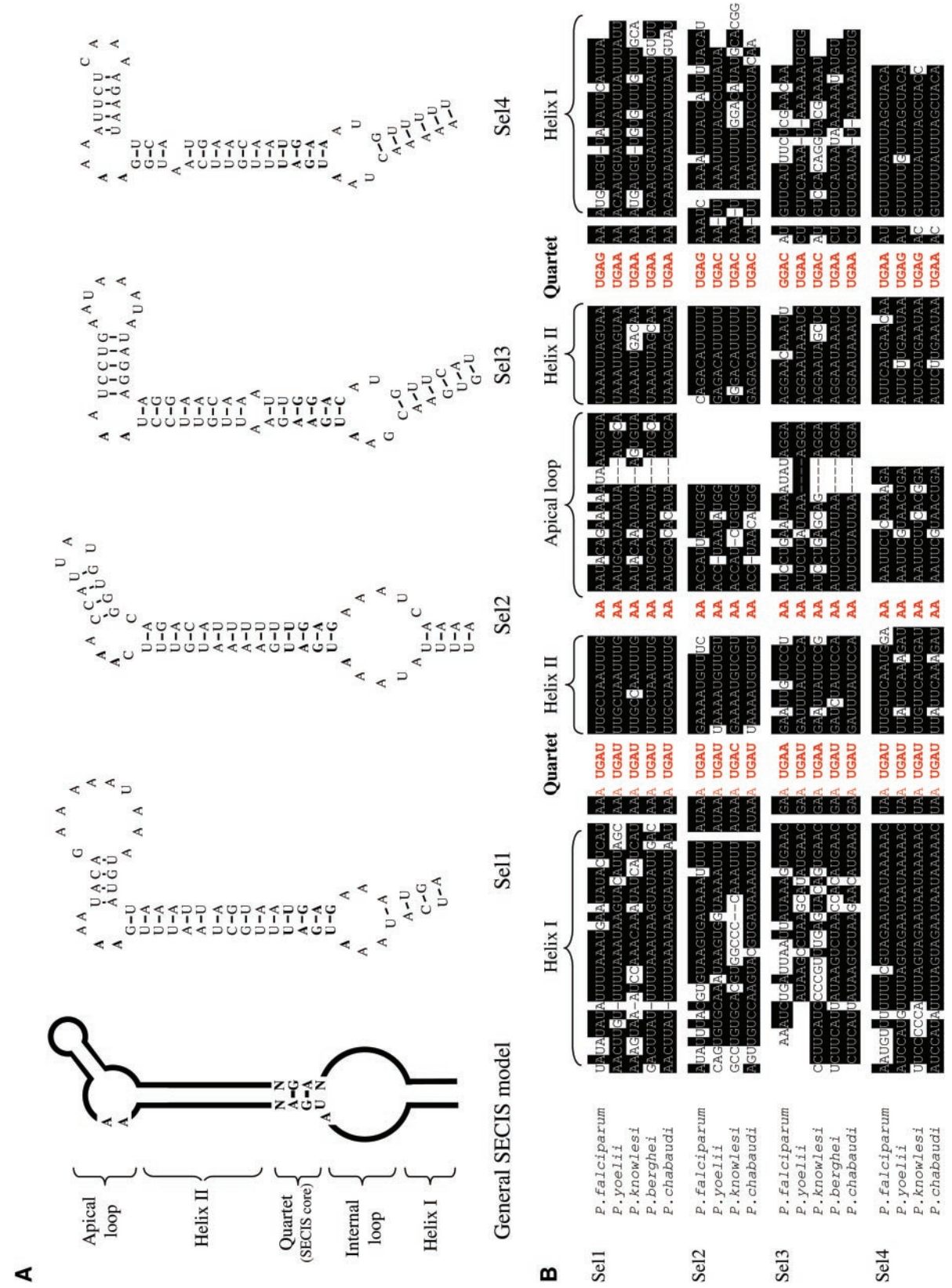


Figure 1. *P.falciparum* SECIS elements. (**A**) Structures of four SECIS elements identified with SECISearch in the *P.falciparum* genome are shown. Conserved nucleotides in the apical loop and the Quartet (SECIS core) are shown in bold. A consensus SECIS structure is shown on the left. It illustrates the overall shape of the structure and indicates the location of various structural elements and conserved nucleotides. (**B**) Nucleotide sequence alignments of SECIS elements in *Plasmodium* selenoprotein genes #1–4. The location of Helices I and II, apical and internal loops, and the Quartet (SECIS core) is indicated. The unpaired A preceding the Quartet, the Quartet itself and the unpaired AA in the apical bulge are shown in red.

Animal SECIS elements were previously described as belonging to form 1 and 2 based on whether the unpaired AA sequence is located in the loop or bulge (30). In this classification, all four *P.falciparum* SECIS elements belonged to form 2 SECIS elements. The four SECIS elements were highly conserved across Plasmodia (Figure 1B) and formed structures very similar to those of *P.falciparum* SECIS elements.

Plasmodium selenoproteome

Genomic sequences upstream of four SECIS elements were analyzed for the presence of ORFs. We found that PlasmoDB had predicted protein sequences upstream of the SECIS elements, but these were misannotated with respect to exon–intron junctions as well as ORFs. Comparative sequence analyses of four selenoprotein genes in different *Plasmodium* species helped us to determine their correct gene structures. The organization of predicted *P.falciparum* selenoprotein genes, which are designated in this study as selenoproteins #1 to #4, is shown in Figure 2. Selenoprotein #1 (Sel1) is organized in three exons, selenoproteins #2 and #3 (Sel2 and Sel3) are encoded by a single exon each and selenoprotein #4 (Sel4) consists of 2 exons. In each *Plasmodium* selenoprotein gene, the Sec-encoding UGAs and the SECIS elements were present in the last exon, although the distance between UGA and SECIS was different (Figure 2).

Exhaustive BLAST searches against a number of databases produced no homologs outside *Plasmodium* for all four selenoprotein genes, whereas these sequences were highly homologous within Plasmodia (Figure 3). In addition, Sec residues were conserved across *Plasmodium* and were flanked by highly conserved regions. In two selenoproteins (Sel1 and Sel4), Sec was present as a C-terminal penultimate residue. This location was reminiscent of Sec in several animal selenoproteins, such as thioredoxin reductase and selenoproteins K, O and S (17). In addition, prediction of secondary structures revealed similarities between *Plasmodium* Sel1 and Sel4 on one side, and mammalian SelK and SelS on the other, suggesting a possible common origin of these proteins.

Sel2 had Sec in the N-terminal region, and it was present in the form of the CxxU motif (cysteine and Sec separated by two other residues). This motif is also found in several animal selenoproteins, including SelM, SelT, SelV and SelW. Moreover, both animal selenoproteins containing the CxxU motif and Sel2 have a predicted α-helix downstream of the motif. These observations suggest that Sel2 might have a redox function, with Sec being a residue that forms a selenenylsulfide bond with the Cys. Sel3 has Sec in the N-terminal region, but it was not present within a known motif.

The possible location of predicted proteins within *Plasmodium* cells was analyzed by searching for the presence of signal peptides. The first 24 amino acids of Sel3 corresponded to a predicted signal peptide (the SignalP probability value of 0.959), whereas other selenoproteins did not have predicted targeting signals. However, using PATS Version 1.2.1, the program that identifies sequences that are targeted to the apicoplast matrix of *P.falciparum* (31), we found that Sel2 (probability score of 0.982) and Sel3 (probability score of 0.936) were candidate apicoplast proteins.

Expression of *P.falciparum* selenoprotein genes

Previous annotations, microarrays with 70 oligomers and Scripps arrays with a higher number of 25 oligomers (all data available in PlasmoDB) suggested that the four predicted selenoprotein genes were transcribed. By carefully analyzing the location of the oligomers used for the arrays, the following conclusions could be drawn: Sel1 was expressed throughout the erythrocytic life cycle with highest levels in schizonts. Sel2 was expressed at constant but relatively low levels throughout the *Plasmodium* life cycle. Sel3 and Sel4 were also expressed in various stages of the *Plasmodium* life cycle, with highest levels in trophozoites and rings/schizonts, respectively. These expression profiles together with the fact that selenoprotein genes were not clustered within *Plasmodium* genomes suggested that selenoproteins may serve distinct functions.

However, these data were not sufficient to determine whether the predicted Sec-encoding UGA codons were indeed present within exons and in the correct reading frame as suggested by our *in silico* analyses, or were parts of introns as annotated in PlasmoDB using Genefinder predictions. The respective PlasmoDB annotation numbers were PF14_0033 for Sel1, PFI1515w for Sel2, MAL8P1.86 for Sel3 and PF14_0251 for Sel4.

We designed primer pairs to examine the gene structure of selenoprotein genes and to verify the presence of predicted mRNAs (Figure 2). For Sel2 and Sel3, in which Sec-encoding UGA codons were well within the ORFs (far from the predicted termination signals), the presence and the correct size of the PCR products clearly indicated the correctness of our predictions as indicated in Figure 2, whereas the PCR products were not consistent with the PlasmoDB predictions. In Sel1 and Sel4, Sec was a C-terminal penultimate residue, thus only small deviations from the PlasmoDB predictions occurred.

Sel1 primer pairs N-CU, N-C and s-C resulted in bands of ~380, 360 and 160 bp, respectively. The 380 bp fragment was cloned and sequenced and found to match our predictions. For Sel2, which we predicted to possess a different C-terminal sequence than that given in PlasmoDB, PCR products could be amplified with primer pairs N-as and N-C (consistent with our predictions), but not with N-Ca, which would be expected if the PlasmoDB prediction was correct. The N-C fragment of ~690 bp was cloned and sequenced and again found to match our prediction. For Sel3, the PCR product of ~834 bp was generated with the N-as primer pair and was verified by sequencing. For Sel4, primer pairs N-as, N-C, N-nC and N-CD gave distinct bands of expected size, which was again consistent with our gene prediction, but with N-gf (PlasmoDB prediction) no product could be detected. These results were further verified by cloning and sequencing an ~444 bp fragment amplified with the N-nC primer pair. Thus, all four selenoprotein genes were in full agreement with our *in silico* predictions, including the exonic, in-frame location of the UGA codons.

Plasmodium SECIS elements support Sec insertion in mammalian cells

Since *Plasmodium* SECIS elements were similar to known eukaryotic SECIS elements, we tested whether these structures could support Sec insertion in mammalian cells. Mammalian cells provide a convenient, commonly used model system to

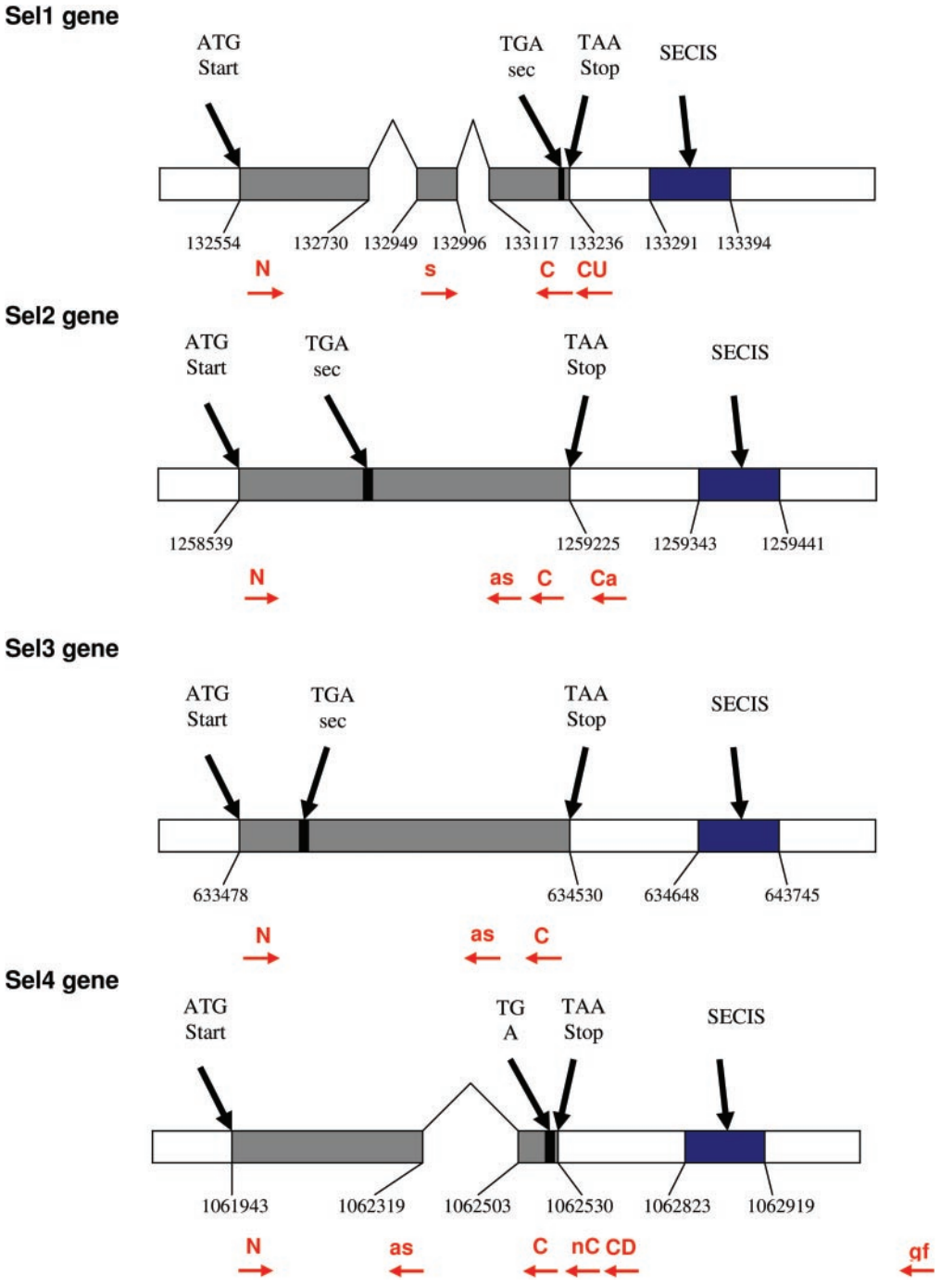
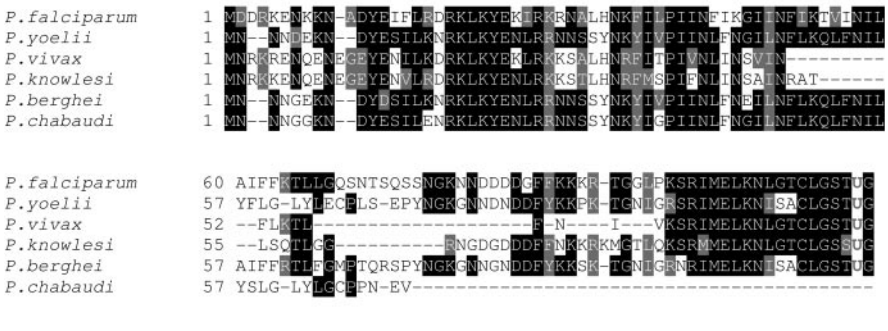


Figure 2. Organization of selenoprotein genes in *P.falciparum*. Exon–intron structure of the four selenoprotein genes is shown with exons indicated by boxes and introns by lines connecting the exons. The location of ORFs and SECIS elements is indicated by filled boxes, and the untranslated regions (other than SECIS elements) by open boxes. The location of initiator ATG codons, Sec-encoding TGA codons, stop codons and SECIS elements is indicated by black arrows. Numbers below the sequences illustrate exon–intron junctions and correspond to nucleotides in *P.falciparum* contigs from PlasmoDB. Red arrows indicate location and orientation of primers, which were used to verify gene expression and examine organization of selenoprotein genes.

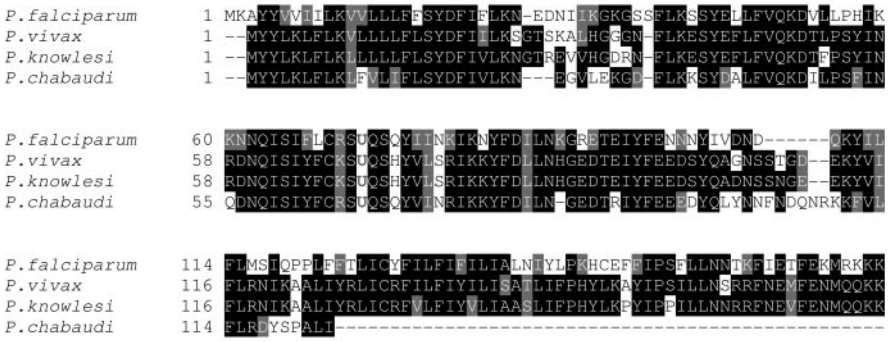
monitor Sec insertion into proteins and to characterize the dependence of Sec insertion on SECIS elements (17,19). We prepared two GFP fusion expression constructs which were designed to test whether *Plasmodium* SECIS element support Sec insertion into *Plasmodium* and mammalian selenoproteins. In the first construct (Figure 4A, #3), a coding region of *P.falciparum* Sel1 was placed downstream of GFP and followed with the entire Sel1 3′-UTR including a SECIS element. The second construct contained a coding region of mouse SelH fused to GFP and also had a Sel4 3′-UTR including a SECIS element (Figure 4A, #2).

These constructs, as well as GFP-SelH with a SelH SECIS element (positive control) and GFP alone (negative control) constructs, were transfected into mouse NIH 3T3 cells and

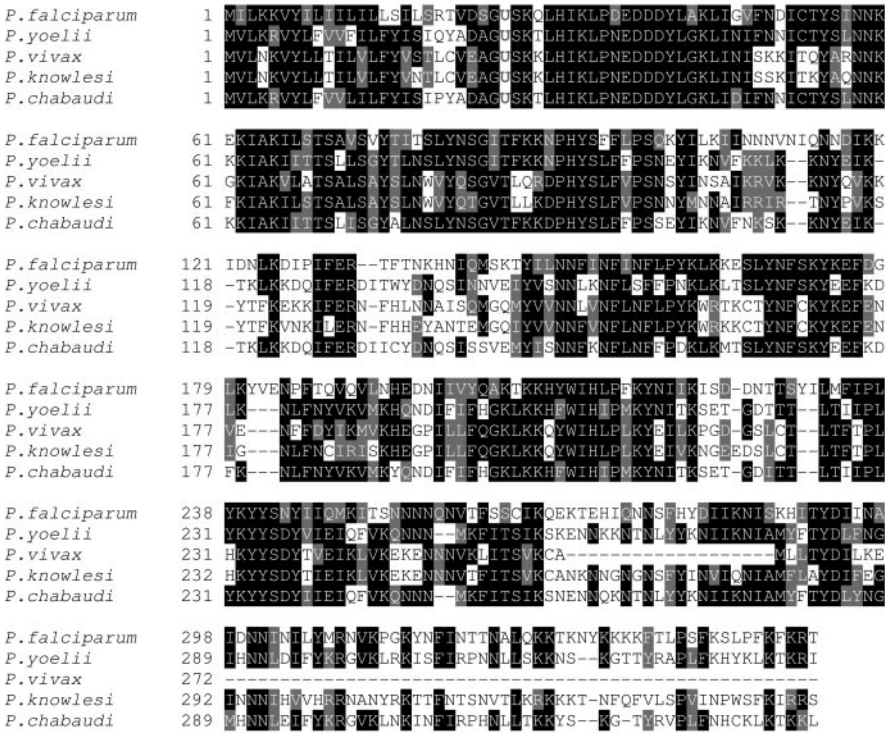
Sel1



Sel2



Sel3



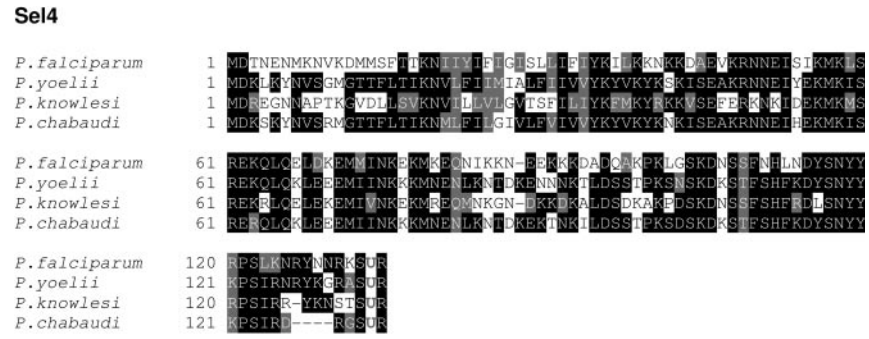


Figure 3. Amino acid sequence alignments of four *Plasmodium* selenoproteins. Conserved residues are highlighted. Sec (indicated by U) is shown in boldface. Only partial sequences are available for *P.chabaudi* selenoprotein 1 and *P.vivax* selenoprotein 3.

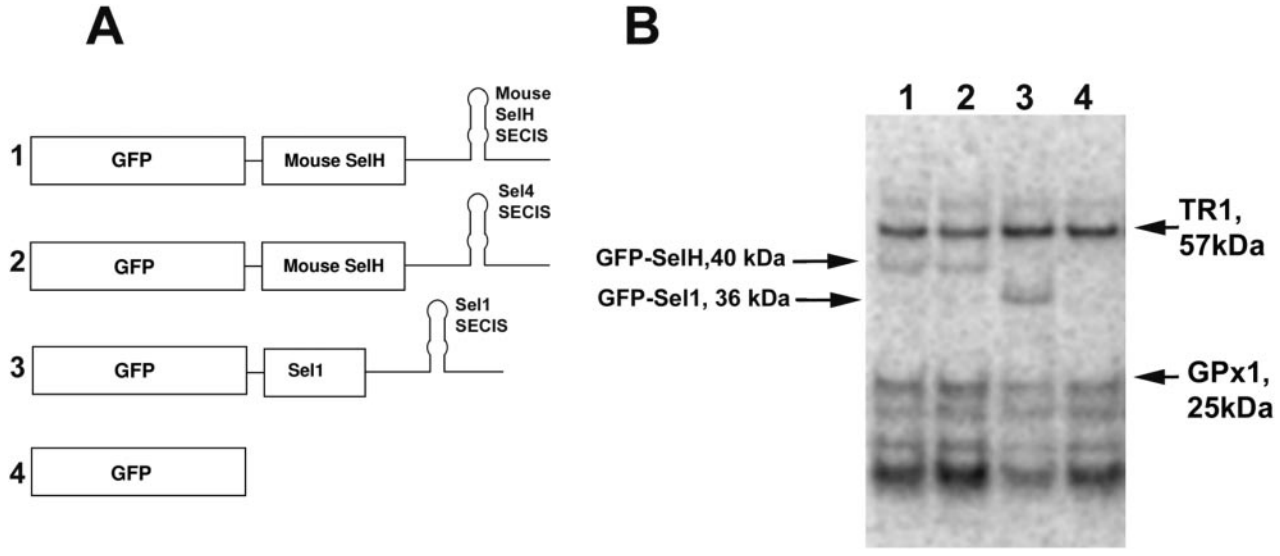


Figure 4. *Plasmodium* SECIS elements support Sec insertion in mammalian cells. (**A**) Schematic illustration of GFP fusion expression constructs. 1, GFP-SelH fusion containing a mouse SelH SECIS element (control); 2, GFP-SelH fusion containing a *P.falciparum* Sel4 SECIS element; 3, GFP-Sel1 fusion containing a *P.falciparum* Sel1 SECIS element; and 4, GFP alone (control). (**B**) Selenoprotein expression in NIH 3T3 cells. Designation of constructs in A corresponds to the designation of lanes in B. The constructs were transfected into NIH3T3 cells, cells were labeled with ^{75}Se, protein resolved by SDS–PAGE gels and analyzed by PhosphorImager. The location of GFP-SelH and GFP-Sel1 fusion selenoproteins is indicated by arrows on the left. Migration of major endogenous selenoproteins thioredoxin reductase 1 (TR1) and glutathione peroxidase 1 (GPx1) is shown on the right.

the cells were labeled with ^{75}Se. PhosphorImager analyses of ^{75}Se-labeled proteins revealed the presence of a 40 kDa GFP-SelH selenoprotein band if either mouse SelH or *Plasmodium* Sel4 SECIS elements were used (Figure 4B, lanes 1 and 2). In addition, a 36 kDa GFP-Sel1 selenoprotein band was observed (Figure 4B, lane 3), whereas transfection with the GFP construct showed only endogenous mouse proteins (Figure 4B, lane 4). These data show that both Sel1 and Sel4 SECIS elements are functional in NIH 3T3 cells and that *Plasmodium* SECIS elements can support Sec insertion into either *Plasmodium* or mouse selenoproteins in mammalian cells. Thus, Sec insertion systems are functionally similar in the Apicomplexa parasites and mammals.

Components of the Sec insertion machinery in *Plasmodium*

All known components of the Sec insertion machinery are present in Plasmodia, including tRNA^{Sec} (Figure 5A), EFSec/eSelB and SPS (21). *Plasmodium* EFSec/eSelB (32,33) was distinct from EF-Tu, which is the elongation factor that inserts all other amino acids. SBP2 was previously described only in animals (11), but hypothetical or putative proteins with accession numbers NP_473214, EAA22418 and Pv_6635.phat_70 for *P.falciparum*, *P.yoelii* and *P.vivax*, respectively, represent good candidates for SBP2 in Plasmodia. Identification of both EFSec/eSelB and SBP2 homologs suggests that the general strategy for Sec insertion (separation of SECIS-binding Sec insertion functions) is conserved across eukaryotes.

Consistent with the recent study (21), the gene sequences of *Plasmodium* tRNA^{Sec} (Figure 5B) revealed that they are 90 nt in length compared with 87 [reviewed in (10)] or 88 nt (34) found in other eukaryotic tRNA^{Sec} genes. The structure of *P.falciparum* tRNA^{Sec} folded into a cloverleaf model including the CCA 3′-terminus that is added post-transcriptionally is 93 nt long as shown in Figure 5A. This tRNA has similar features to those of tRNA^{Sec}s found in other eukaryotes. The acceptor and TψC stems consist of 9/4 base pairs (i.e. 9 bp in the acceptor stem and 4 in the TψC stem) and

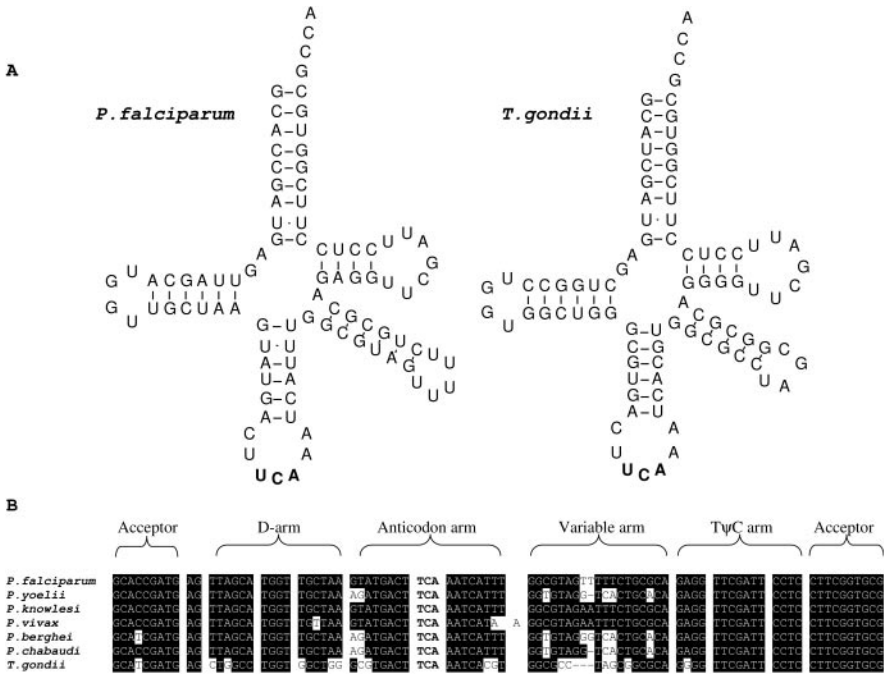


Figure 5. *Plasmodium* and *Toxoplasma* selenocysteine tRNAs. (**A**) Cloverleaf models of *P.falciparum* (left) and *T.gondii* (right) $tRNA^{Sec}$s. The anticodons are shown in bold. (**B**) Nucleotide sequence alignments of $tRNA^{Sec}$s from *P.falciparum*, *P.yoelii*, *P.knowlesi*, *P.vivax*, *P.berghei* and *T.gondii*. The location of tRNA arms is indicated. The anticodon arm features the Sec anticodon (TCA).

the D-stem 6 bp as is commonly observed in other eukaryotic $tRNA^{Sec}$s. The anticodon stem in *Plasmodium* $tRNA^{Sec}$s contains 6 bp as is found in other $tRNA^{Sec}$s with the exception of *P.vivax* that contains a highly improbable G:A base pairing between the two bases most distal to the anticodon loop within this region. Whether this indicates a sequencing error or is indicative that the anticodon stem in *Plasmodium* may only require 5 bp must await further investigation. The three extra bases in *Plasmodium* $tRNA^{Sec}$ compared with those of animal and *Chlamydomonas* $tRNA^{Sec}$s exist in the long extra arm. The *T.gondii* $tRNA^{Sec}$ is shown in Figure 5A, right cloverleaf structure. Its cloverleaf model that includes the CCA terminus is 90 nt long and is very similar to those of other eukaryotic $tRNA^{Sec}$s.

Evolution of Sec insertion in *Plasmodium*

We constructed phylogenic trees for EFSec/eSelB, SPS and $tRNA^{Sec}$ (see Supplementary Figures S1–S3). EFSec/eSelB formed a cluster separate from that of EF-Tu, and in all three trees, plasmodial sequences clustered with other eukaryotes, suggesting a common origin of Sec insertion in eukaryotes. In addition, the $tRNA^{Sec}$ tree suggested that plasmodial $tRNA^{Sec}$s formed a cluster that also included *T.gondii* $tRNA^{Sec}$ and it was separate from the cluster of animal $tRNA^{Sec}$s that also included *Chlamydomonas* $tRNA^{Sec}$. These data suggest that the plasmodial Sec insertion system is perhaps the most divergent among previously characterized eukaryotic Sec insertion systems.

Relevance to malaria

Growing drug resistance to malaria poses a major problem. For example, chloroquine, a safe and cheap drug, is a medication of choice for the treatment of malaria; however, resistance to this compound in Southeast Asia and many areas of Africa has become a problem. In fact, resistance to all major drugs is a general problem in some areas of Southeast Asia. The development of new drugs and drug targets is an important avenue to combat malaria. In this regard, the following facts support the hypothesis that the selenoproteome of *Plasmodium* is a promising drug target: (i) the four selenoproteins of *P.falciparum* are unique to the parasite; (ii) Sec is highly reactive toward electrophilic compounds and potential inhibitors; (iii) two of the selenoproteins are likely to be located in the apicoplast, a cell organelle which does not occur in the human host and bears unique metabolic pathways; and (iv) since malarial parasites are particularly susceptible to oxidative stress, interference with redox metabolism represents a promising strategy for antimalarial drug development [reviewed in (35)]. Similarly to functionally characterized selenoproteins in other species, the four *Plasmodium* selenoproteins are candidate redox proteins.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Rimma Iozef for excellent experimental support and Drs David Sibley and John B. Dame for providing EST clones. The work is supported by NIH GM061603, and in part by DoE DE-FG07-02ID14380 and Deutsche Forschungsgemeinschaft grants BE 1540/6-3 and GR 2028/1-1. Funding to pay the Open Access publication charges for this article was provided by GM061603.

Conflict of interest statement. None declared.

REFERENCES

1. Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S. *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419**, 498–511.
2. Carlton,J.M., Angiuoli,S.V., Suh,B.B., Kooij,T.W., Pertea,M., Silva,J.C., Ermolaeva,M.D., Allen,J.E., Selengut,J.D., Koo,H.L. *et al.* (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature*, **419**, 512–519.
3. Abrahamsen,M.S., Templeton,T.J., Enomoto,S., Abrahante,J.E., Zhu,G., Lancto,C.A., Deng,M., Liu,C., Widmer,G., Tzipori,S. *et al.* (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*, **304**, 441–445.
4. Kim,K. and Weiss,L. (2004) *Toxoplasma gondii*: the model apicomplexan. *Int. J. Parasitol.*, **34**, 423–432.
5. Xu,P., Widmer,G., Wang,Y., Ozaki,L.S., Alves,J.M., Serrano,M.G., Puiu,D., Manque,P., Akiyoshi,D., Mackey,A.J. *et al.* (2004) The genome of *Cryptosporidium hominis*. *Nature*, **431**, 1107–1112.
6. Templeton,T.J., Iyer,L.M., Anantharaman,V., Enomoto,S., Abrahante,J.E., Subramanian,G.M., Hoffman,S.L., Abrahamsen,M.S. and Aravind,L. (2004) Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Res.*, **14**, 1686–1695.
7. Aravind,L., Iyer,L.M., Wellems,T.E. and Miller,L.H. (2003) Plasmodium biology: genomic gleanings. *Cell*, **115**, 771–785.
8. Low,S.C. and Berry,M.J. (1996) Knowing when not to stop: selenocysteine incorporation in eukaryotes. *Trends Biochem. Sci.*, **21**, 203–208.
9. Stadtman,T.C. (1996) Selenocysteine. *Annu. Rev. Biochem.*, **65**, 83–100.
10. Hatfield,D.L. and Gladyshev,V.N. (2002) How selenium has altered our understanding of the genetic code. *Mol. Cell Biol.*, **22**, 3565–3576.
11. Driscoll,D.M. and Copeland,P.R. (2003) Mechanism and regulation of selenoprotein synthesis. *Annu. Rev. Nutr.*, **23**, 17–40.
12. Chaudière,J. and Tappel,A.L. (1984) Interaction of gold(I) with the active site of selenium-glutathione peroxidase. *J. Inorg. Biochem.*, **20**, 313–325.
13. Gromer,S., Urig,S. and Becker,K. (2004) The thioredoxin system—from science to clinic. *Med. Res. Rev.*, **24**, 40–89.
14. Wilting,R., Schorling,S., Persson,B.C. and Böck,A. (1997) Selenoprotein synthesis in archaea: identification of an mRNA element of *Methanococcus jannaschii* probably directing selenocysteine insertion. *J. Mol. Biol.*, **266**, 637–641.
15. Flohé,L., Andreesen,J.R., Brigelius-Flohé,R., Maiorino,M. and Ursini,F. (2000) Selenium, the element of the moon, in life on earth. *IUBMB Life*, **49**, 411–420.
16. Castellano,S., Morozova,N., Morey,M., Berry,M.J., Serras,F., Corominas,M. and Guigo,R. (2001) *In silico* identification of novel selenoproteins in the *Drosophila melanogaster* genome. *EMBO Rep.*, **2**, 697–702.
17. Kryukov,G.V., Castellano,S., Novoselov,S.V., Lobanov,A.V., Zehtab,O., Guigo,R. and Gladyshev,V.N. (2003) Characterization of mammalian selenoproteomes. *Science*, **300**, 1439–1443.
18. Kryukov,G.V. and Gladyshev,V.N. (2004) The prokaryotic selenoproteome. *EMBO Rep.*, **5**, 538–543.
19. Novoselov,S.V., Rao,M., Onoshko,N.V., Zhi,H., Kryukov,G.V., Xiang,Y., Weeks,D.P., Hatfield,D.L. and Gladyshev,V.N. (2002) Selenoproteins and selenocysteine insertion system in the model plant cell system, *Chlamydomonas reinhardtii*. *EMBO J.*, **21**, 3681–3693.
20. Fu,L.H., Wang,X.F., Eyal,Y., She,Y.M., Donald,L.J., Standing,K.G. and Ben-Hayyim,G. (2002) A selenoprotein in the plant kingdom. Mass spectrometry confirms that an opal codon (UGA) encodes selenocysteine in *Chlamydomonas reinhardtii* glutathione peroxidase. *J. Biol. Chem.*, **277**, 25983–25991.
21. Mourier,T., Pain,A., Barrell,B. and Griffiths-Jones,S. (2005) A selenocysteine tRNA and SECIS element in *Plasmodium falciparum*. *RNA*, **11**, 119–122.
22. Kryukov,G.V., Kryukov,V.M. and Gladyshev,V.N. (1999) New mammalian selenocysteine-containing proteins identified with an algorithm that searches for selenocysteine insertion sequence elements. *J. Biol. Chem.*, **274**, 33888–33897.
23. Lowe,T.M. and Eddy,S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
24. Laslett,D. and Canback,B. (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.*, **32**, 11–16.
25. Bendtsen,J.D., Nielsen,H., von Heijne,G. and Brunak,S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.*, **340**, 783–795.
26. Saitou,N. and Nei,M. (1987) The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
27. Castellano,S., Novoselov,S.V., Kryukov,G.V., Lescure,A., Blanco,E., Krol,A., Gladyshev,V.N. and Guigo,R. (2004) Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution. *EMBO Rep.*, **5**, 71–77.
28. Obata,T. and Shiraiwa,Y. (2005) A novel eukaryotic selenoprotein in the haptophyte alga *Emiliania huxleyi*. *J. Biol. Chem.*, **280**, 18462–18468.
29. Walczak,R., Westhof,E., Carbon,P. and Krol,A. (1996) A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA*, **2**, 367–379.
30. Grundner-Culemann,E., Martin,G.W.,III, Harney,W. and Berry,M.J. (1999) Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. *RNA*, **5**, 625–635.
31. Zuegge,J., Ralph,S., Schmuker,M., McFadden,G.I. and Schneider,G. (2001) Deciphering apicoplast targeting signals—feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene*, **280**, 19–26.
32. Fagegaltier,D., Hubert,N., Yamada,K., Mizutani,T., Carbon,P. and Krol,A. (2000) Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J.*, **19**, 4796–4805.
33. Tujebajeva,R.M., Copeland,P.R., Xu,X.M., Carlson,B.A., Harney,J.W., Driscoll,D.M., Hatfield,D.L. and Berry,M.J. (2000) Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Rep.*, **1**, 158–163.
34. Shrimali,R.K., Lobanov,A.V., Xu,X.M., Rao,M., Carlson,B.A., Mahadeo,D.C., Parent,C.A., Gladyshev,V.N. and Hatfield,D.L. (2005) Selenocysteine tRNA identification in the model organisms *Dictyostelium discoideum* and *Tetrahymena thermophila*. *Biochem. Biophys. Res. Commun.*, **329**, 147–151.
35. Becker,K., Tilley,L., Vennerstrom,J.L., Roberts,D., Rogerson,S. and Ginsburg,H. (2004) Oxidative stress in malaria parasite-infected erythrocytes: host-parasite interactions. *Int. J. Parasitol.*, **34**, 163–189.

Supplementary Information

Figure Legends

Figure S1. Phylogenetic analysis of Sec-specific translation factor EFSec. GenBank accession numbers for EFSec are as follows: *G. gallus* (XP_414362), mouse (AAF91471), human (P57772), *D. melanogaster* (AAL49249), *C. elegans* (NP_493276). The PlasmoDB *Plasmodium* sequences for EFSec are follows: *P. falciparum* (chr1.phat_115), *P. yoelii* (PY00420) and *P. vivax* (402918.phat_1). GenBank accession numbers for elongation factor Tu (EF-Tu) are as follows: *C. reinhardtii* (NP_958362), *D. melanogaster* (NP_524752), mouse (NP_766333) and human (S62767). *Plasmodium* sequences for EF-Tu were extracted from PlasmoDB: *P. falciparum* (chr13.phat_330*), P. yoelii* (PY06134) and *P. vivax* (402881.phat_3).

Figure S2. Phylogenetic analysis of Sec tRNA. Accession numbers for sequences are as follows: *S. flexneri* (AE015377), *S. typhimurium* (AE008874), *E. coli* (U00096), *T. tengcongensis* (AE013138), *Y. pestis* (AE014007), *H. hepaticus* (AE017145), *T. denticola* (AE017249), *W. succinogenes* (BX571659), *D. hafniense* (NZ_AAAW00000000), *C. reinhardtii* (AY268554), mouse (L22019), human (K02923), *D. melanogaster* (M34509), *C. elegans* (M34508). *T. gondii* was from ToxoDB (www.toxodb.org) (TGG_995334). *Plasmodium* sequences were extracted from PlasmoDB as described in *Methods*.

Figure S3. Phylogenetic analysis of selenophosphate synthetase (SPS). GenBank accession numbers for sequences are as follows: *Y. pestis* (Q8ZEK1), *E. coli* (P66794), human (NP_036379), mouse (NP_780609), *C. jejuni* (NP_282641), *D. hafniense* (ZP_00100012), *P. straminea* (BAB32693), *D. melanogaster* (CAA04229), *C. elegans* (NP_502604). *Plasmodium* sequences were extracted from PlasmoDB: *P. falciparum* (chr9.phat_117), *P. yoelii* (PY05530) and *P. vivax* (395472.phat_1).

Figure S1

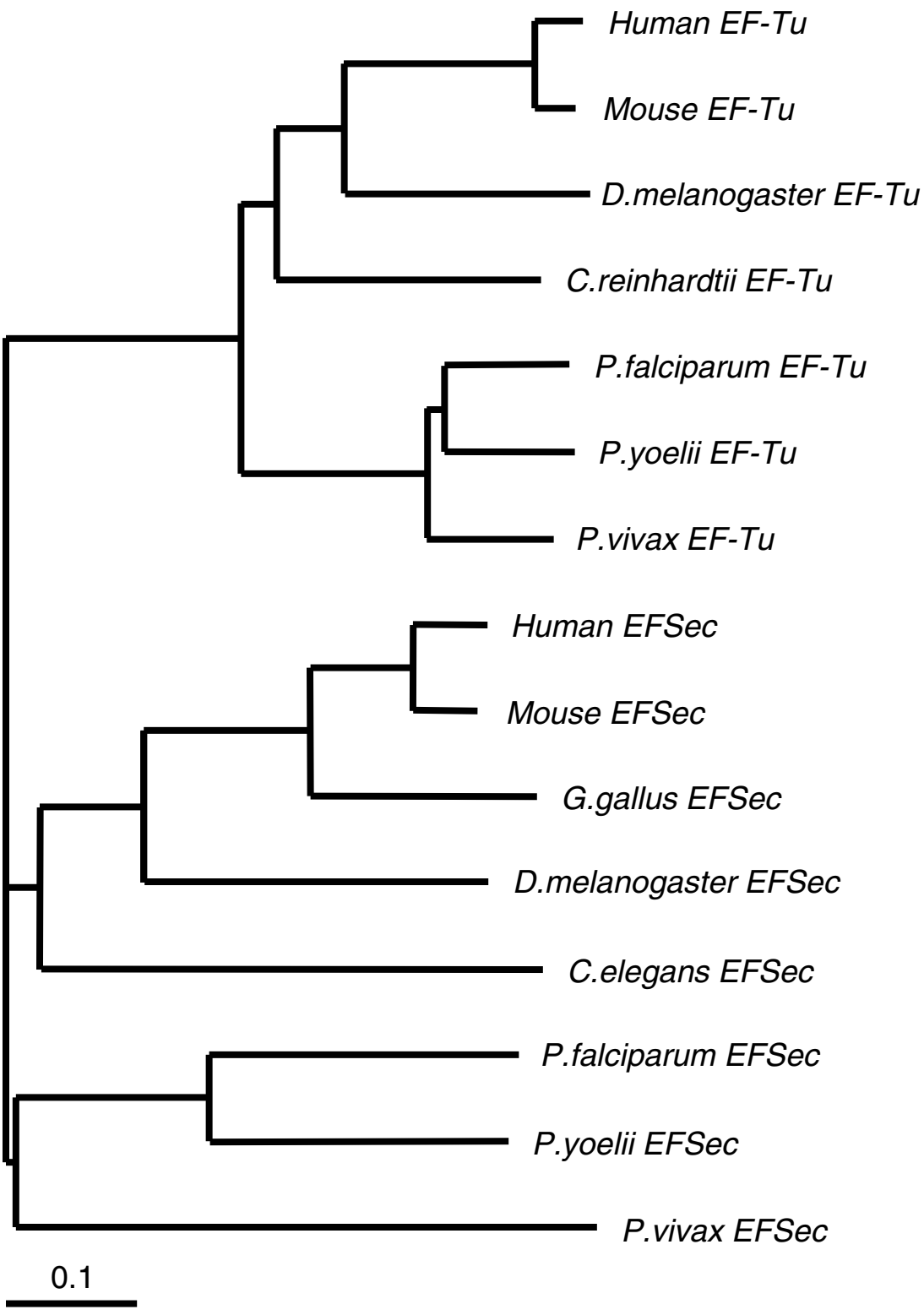

Human EF-Tu
Mouse EF-Tu
D.melanogaster EF-Tu
C.reinhardtii EF-Tu
P.falciparum EF-Tu
P.yoelii EF-Tu
P.vivax EF-Tu
Human EFSec
Mouse EFSec
G.gallus EFSec
D.melanogaster EFSec
C.elegans EFSec
P.falciparum EFSec
P.yoelii EFSec
P.vivax EFSec
0.1

Figure S2

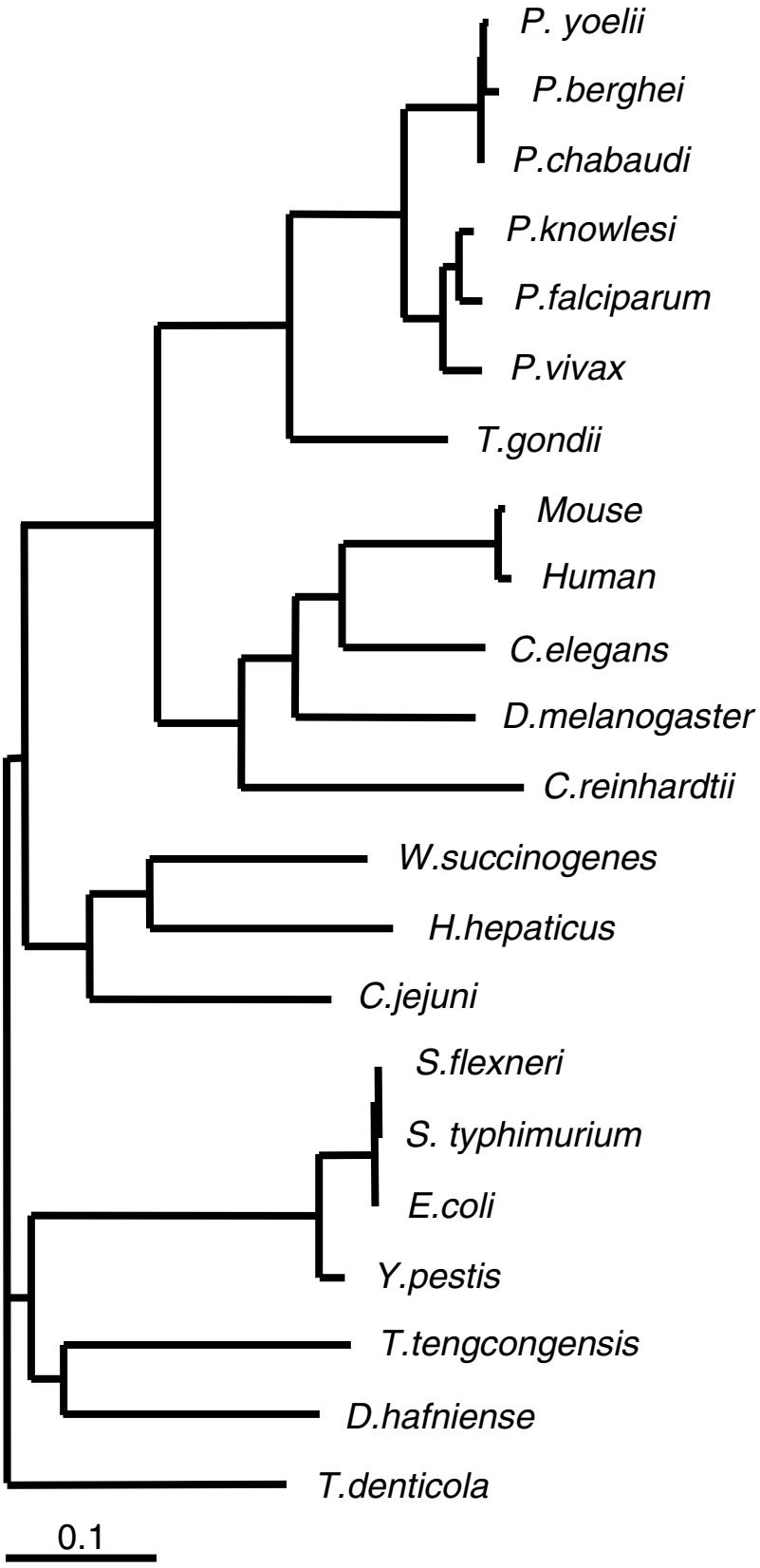
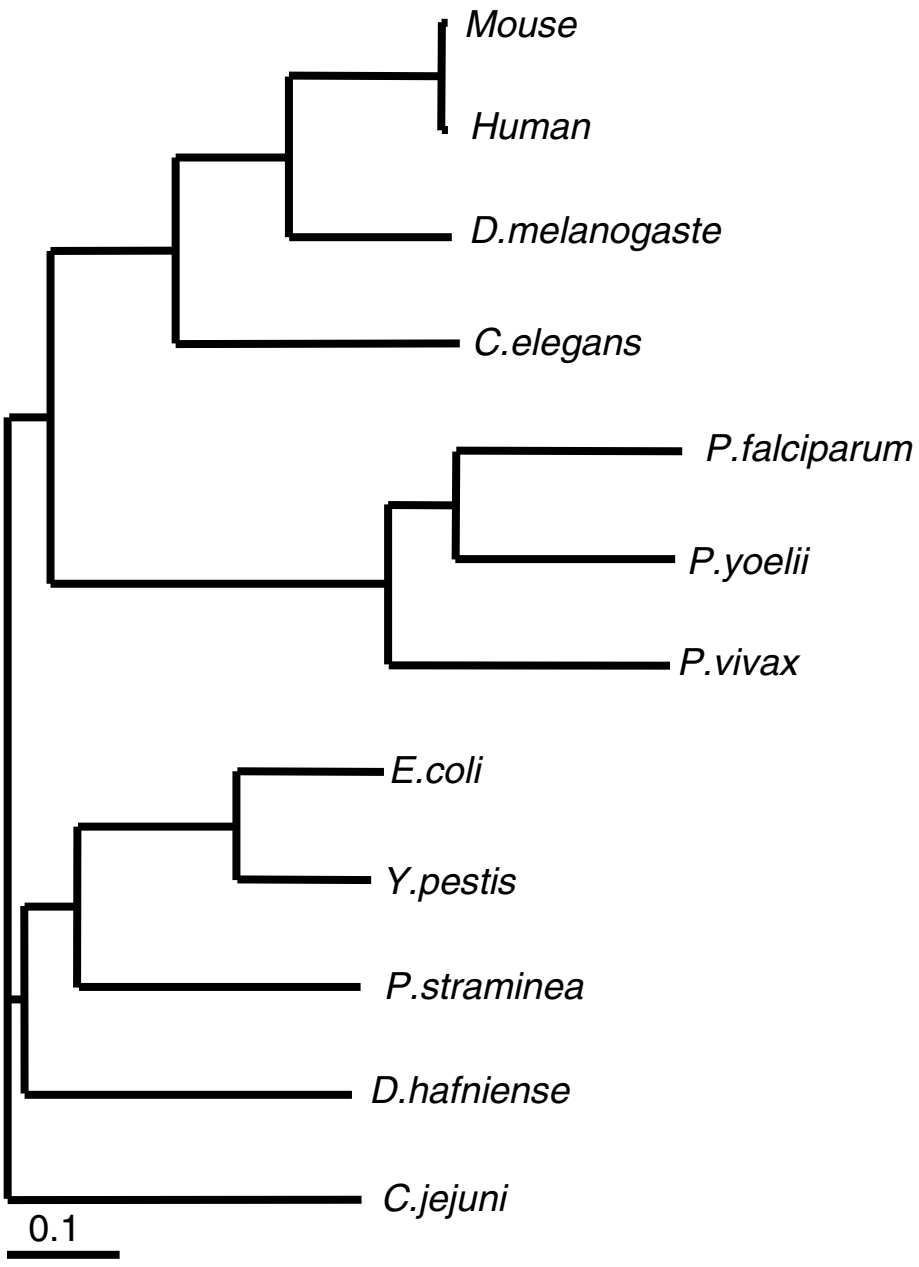

P. yoelii
P.berghei
P.chabaudi
P.knowlesi
P.falciparum
P.vivax
T.gondii
Mouse
Human
C.elegans
D.melanogaster
C.reinhardtii
W.succinogenes
H.hepaticus
C.jejuni
S.flexneri
S. typhimurium
E.coli
Y.pestis
T.tengcongensis
D.hafniense
T.denticola
0.1

Figure S3


Mouse
Human
D.melanogaste
C.elegans
P.falciparum
P.yoelii
P.vivax
E.coli
Y.pestis
P.straminea
D.hafniense
C.jejuni
0.1