



Published in final edited form as:

*Nature*. 2005 November 3; 438(7064): 108–112. doi:10.1038/nature04135.

## The *Plasmodium* protein network diverges from those of other eukaryotes

Silpa Suthram<sup>1,2,\*</sup>, Taylor Sittler<sup>2,\*</sup>, and Trey Ideker<sup>1,2</sup>

<sup>1</sup>Bioinformatics Program, University of California, San Diego, California 92093, USA

<sup>2</sup>Department of Bioengineering, University of California, San Diego, California 92093, USA

### Abstract

*Plasmodium falciparum* is the pathogen responsible for over 90% of human deaths from malaria<sup>1</sup>. Therefore, it has been the focus of a considerable research initiative, involving the complete DNA sequencing of the genome<sup>2</sup>, large-scale expression analyses<sup>3,4</sup>, and protein characterization of its life-cycle stages<sup>5</sup>. The *Plasmodium* genome sequence is relatively distant from those of most other eukaryotes, with more than 60% of the 5,334 encoded proteins lacking any notable sequence similarity to other organisms<sup>2</sup>. To systematically elucidate functional relationships among these proteins, a large two-hybrid study has recently mapped a network of 2,846 interactions involving 1,312 proteins within *Plasmodium*<sup>6</sup>. This network adds to a growing collection of available interaction maps for a number of different organisms, and raises questions about whether the divergence of *Plasmodium* at the sequence level is reflected in the configuration of its protein network. Here we examine the degree of conservation between the *Plasmodium* protein network and those of model organisms. Although we find 29 highly connected protein complexes specific to the network of the pathogen, we find very little conservation with complexes observed in other organisms (three in yeast, none in the others). Overall, the patterns of protein interaction in *Plasmodium*, like its genome sequence, set it apart from other species.

---

With the recent accumulation of protein interaction maps in public databases, cross-species comparisons are becoming critical for analysing the large networks formed by these interactions to delineate protein function and evolution<sup>7</sup>. At a fundamental level, protein networks can be compared to identify ‘interologues’—that is, interactions that are conserved across species<sup>8</sup>. Beyond the individual comparison of interactions, methods such as PathBLAST (refs 9–10) create a global alignment between protein networks to identify dense clusters of conserved interactions, suggestive of protein complexes. Such comparative approaches are important because they can tease conserved components of cellular machinery out of a highly connected network and increase overall confidence in the underlying interaction measurements.

We compared the protein–protein interaction network of *Plasmodium* reported by LaCount *et al.* (ref. 6) to protein networks for the budding yeast *Saccharomyces cerevisiae*<sup>11</sup>, the nematode

---

© 2005 Nature Publishing Group

Correspondence and requests for materials should be addressed to S.S. (ssuthram@ucsd.edu).

\*These authors contributed equally to this work.

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Author Contributions** S.S. and T.S. contributed equally to this work. All authors discussed the results and wrote the paper.

Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests.

The authors declare no competing financial interests.

worm *Caenorhabditis elegans*<sup>12</sup>, the fruitfly *Drosophila melanogaster*<sup>13</sup> and the bacterial pathogen *Helicobacter pylori*<sup>14</sup>. Surprisingly, the pairwise alignment of these networks using PathBLAST (ref. 9) revealed that *Plasmodium* had only three conserved complexes with yeast (Fig. 1a–c), and no conserved complexes with any of the other organisms examined. However, yeast, fly and worm shared substantial numbers of conserved complexes with each other (Fig. 2a); for instance, yeast and fly had the highest degree of conservation with 61 conserved complexes.

The relatively low similarity between the *Plasmodium* protein network and those of the other eukaryotes suggested that it encodes important functional differences worthy of further investigation. Alternatively, it was possible that differences in the number of complexes were related to network size. Thus, in addition to searching for conserved complexes, we investigated whether the observed similarities and differences were reflected in the probability of conservation of each protein interaction individually (Fig. 2b). For each pair of species, a protein–protein interaction was considered ‘conserved’ if both proteins had homologues that interacted in the opposite species (BLAST *E* value  $\leq 1 \times 10^{-4}$ , normalized for genome size). A global pairwise similarity metric was then defined as the overall fraction of interactions that were conserved, restricted to proteins with at least one homologue in the opposite species.

Figure 2c expresses the pairwise interaction similarities as a phylogenetic tree drawn using the method of Kitchin<sup>15</sup>. This tree was relatively robust to sampling errors as determined by bootstrap analysis: 86.2% of trials placed *Plasmodium* as an outgroup relative to yeast, worm and fly. Among the three model eukaryotes, yeast and worm were closest on the basis of interaction similarity (Fig. 2b), while yeast and fly were closest on the basis of conserved complexes (Fig. 2a). This discrepancy was probably due to network size or coverage. Nonetheless, the particular phylogenetic placement of *Plasmodium* was consistent across both analyses, and also agrees with the accepted taxonomical relationships among these organisms as established by morphological and sequence comparisons<sup>2</sup>.

Another possibility for the low similarity of the *Plasmodium* protein network to the other species was that its interaction network had been measured predominantly among proteins expressed in the asexual stages of the parasite’s life cycle (see ref. 6). There are two ways in which this sampling could have affected network similarity. First, it was possible that a high (or low) level of messenger RNA expression increases (or decreases) the number of interactions identified for the corresponding proteins, and thus alters the topology of the *Plasmodium* network relative to the other species. However, as shown in Supplementary Fig. 1, we found no correlation between expression level in any stage and the number of protein interactions. Second, it was possible that proteins from asexual stages tended to have lower similarity across species than proteins from other stages of the *Plasmodium* life cycle. However, we found that the *Plasmodium* interaction set was enriched for proteins with homologues in the other species, and that the protein interaction networks from all five organisms were enriched for yeast homologues in particular (Table 1; Supplementary Table 1). Such enrichment was observed even in worm, for which baits were explicitly selected to be non-homologous to yeast<sup>12</sup>. This effect requires further study, but might indicate a bias of the yeast two-hybrid system in measuring interactions among yeast homologues, because all two-hybrid constructs must be expressible in the yeast cell.

A final possibility for the low level of conservation between the *Plasmodium* protein network and those of the other organisms was that the *Plasmodium* network might have a substantially higher proportion of false-positive interactions relative to the networks of yeast, fly and worm. Lacking a ‘gold standard’ set of true interactions, we characterized the relative quality of the *Plasmodium* network by examining: (1) its global topological properties, and (2) the signal-to-noise ratio (SNR) of its protein complexes. Several common topological measures<sup>16</sup> were

computed on each network, including the average number of interactions per protein (average degree), the average shortest path length between proteins, and the average clustering coefficient (Table 1). The number of interactions per protein in the *Plasmodium* interaction network followed a scale-free distribution, similar to other networks (Fig. 3a). Moreover, the *Plasmodium* network was never the outlier in any of the various measurements, suggesting that its global organization was consistent with the others.

Next, we applied the PathBLAST procedure to identify dense interaction complexes within each organism independently. A total of 29 single-species complexes were identified for *Plasmodium*, three of which are shown in Fig. 1d–f. This number was the median of the range observed over the five species (Table 1). Single-species complexes were used to assess the overall quality of each network by computing their SNR, a standard measure of assessing data quality in information theory and signal processing<sup>17</sup>. The SNR was computed by comparing the scores of complexes identified in the observed versus random interaction data for each organism (see the Methods). *Plasmodium*, worm and fly had very similar SNR values (Fig. 3b), while the SNR of the yeast network was slightly higher, and that of the *H. pylori* network was slightly lower. The network distances of *Plasmodium* versus yeast, worm or fly do not appear to depend on SNR.

As the observed network differences could not be attributed to bias or error, we next examined the novel functional predictions suggested by the three conserved and 29 *Plasmodium*-specific complexes. The conserved protein complex shown in Fig. 1a predicts that the proteins PF10\_0244 and MAL6P1.286 may have previously uncharacterized roles in endocytosis. The counterpart of PF10\_0244 in the yeast network, Ede1, localizes to the cortical patch<sup>18</sup> of the cell membrane at sites of polarized growth and seems to be involved in endocytosis<sup>19</sup>. Myo5 and Myo3, yeast counterparts of MAL6P1.286, are class I myosins that also localize to actin cortical patches<sup>20</sup>, where the calmodulin protein Cmd1 has been implicated in the uptake step of receptor-mediated endocytosis<sup>21</sup>. Taken together, this evidence suggests a role for this complex in calmodulin-mediated endocytosis. Calmodulin inhibitors have been shown to attenuate growth<sup>22</sup> and chloroquine extrusion (thus effecting drug resistance)<sup>23</sup> in malarial parasites, and endocytosis has recently been linked to the mechanism of anti-malarial drugs including chloroquine and artemisinin<sup>24</sup>. The proximity of calmodulin to the formation of endocytic vacuoles in *Plasmodium* provides for a discrete hypothesis linking endocytosis, drug resistance and drug mechanism-of-action.

Within the 29 *Plasmodium*-specific complexes, chromatin remodelling was a prominent function, as shown in Fig. 1e. This complex involves the chromatin-remodelling protein ISWI (MAL6P1.183) interacting with a nucleosome assembly protein (PFI0930C)<sup>25</sup>. The protein PF11\_0429 has a PHD domain (for plant homeodomain), and PF07\_0029 has an HSP90 domain (for heat shock protein of 90 kDa), both postulated to be involved in the remodelling process<sup>25,26</sup>. Together, these known functions suggest that other proteins in the complex, such as PF08\_0060, PFB0765W and PFL0625C, also participate in chromatin remodelling. For instance, although PFL0625C is annotated as a translation initiation factor, its yeast homologue has been found in complex with histone acetyl-transferases<sup>27</sup>. Further analysis of other complexes shown in Fig. 1 is available in the Supplementary Information.

Several cellular components that we expected to be present, such as the proteasome, were missing from the set of complexes conserved between *Plasmodium* and the other species. To investigate this issue, we plotted the distributions of known functional annotations (according to Gene Ontology Cellular Component Level Three)<sup>28</sup> among *Plasmodium* proteins, protein interactions and conserved interactions (Fig. 4 and Supplementary Fig. 2; note that a protein or an interaction can participate in multiple categories). Considerable proportions of all three data sets were associated with intracellular organelles, membrane-bound organelles or the

cytoplasm (Fig. 4a). Other cellular components, such as the membrane and extra-organismal space, were represented among proteins and interactions but to a lesser extent among conserved interactions (Fig. 4b). Many membrane-associated components were also reported in the 29 *Plasmodium*-specific complexes, and are suggestive of machinery unique to this organism. Finally, components such as the proteasome and cytoskeleton were represented among proteins but were absent from the interaction set, and hence were not found as conserved interactions or complexes (Fig. 4c). Interactions among proteins in these components may have yet to be uncovered. These observations are reinforced by a complementary analysis of the functional distributions of yeast, worm and fly protein networks (Supplementary Fig. 2).

In summary, we have characterized conserved patterns of interaction between the protein network of *Plasmodium falciparum* and those of other species, and reported the specific network regions that are conserved. All of the examined networks contain dense complex-like structures of interactions, some of which are shared by yeast, worm and fly but not *Plasmodium*. These relationships are not clearly related to noise or bias in the *Plasmodium* interaction set. Some of the observed differences are almost certainly due to incomplete coverage in one or more networks: for instance, the present *Plasmodium* interaction set is focused on asexual life-cycle stages. Nevertheless, our comparison reflects the relative degree of similarity between the different networks. These differences are observed even when considering only those genes that are homologous across species.

It is generally expected that conserved genes will retain their functions and interactions. From this comparison, a different principle emerges: conservation of specific groups of related genes does not necessarily imply conservation of interaction among their encoded proteins. Further studies may distinguish the true differences from those related to network coverage and, ultimately, facilitate the discovery of new pharmaceuticals directed at the protein complexes unique to this parasite.

## METHODS

### Identification of conserved and species-specific complexes

Identification of protein complexes was performed using the PathBLAST family of network alignment tools, as previously described<sup>9</sup>. Briefly, these methods integrate protein interaction data from two species with protein sequence homology to generate an 'aligned network', in which each node represents a pair of homologous proteins (one from each organism; BLAST *E* value  $\leq 1 \times 10^{-10}$ ) and each link represents a conserved interaction. The network alignment is searched to identify high-scoring subnetworks, for which the score is based on the density of interactions within the subnetwork as well as confidence estimates for each protein interaction (see below). The search is then repeated over 100 random trials, in which the interactions of both species are arbitrarily reassigned while maintaining the same number of interactions per protein, resulting in a distribution of random subnetwork scores pooled over all trials. Dense subnetworks that score in the top fifth percentile of this random score distribution are considered significant and reported as 'conserved complexes'. The search for 'single-species complexes' is identical to the search for conserved complexes, except that an individual protein network is searched instead of the network alignment. This process identifies dense subnetworks constrained by the interactions of one organism rather than two.

### Interaction confidence scores

We estimated the probability that each measured protein interaction is true using a logistic regression model based on mRNA expression correlation, the network cluster coefficient, and the number of times the interaction had been experimentally observed. Further information on these confidence assignments is provided in the Supplementary Methods.

## Phylogenetic tree construction

The Kitsch algorithm (provided by the PHYLIP package<sup>15</sup>) assumes the presence of an evolutionary clock and is based on pairwise distances between species. For each pair of species, an interaction between proteins  $a$  and  $b$  was considered ‘conserved’ if both proteins had sequence-similar counterparts  $a'$  and  $b'$  (BLAST  $E$  value  $\leq 1 \times 10^{-4}$ ) that interacted in the opposite species. A pairwise similarity between networks was computed as  $s_{1,2} = (c_1 + c_2) / (t_1 + t_2)$ , where  $c$  is the number of conserved interactions and  $t$  is the total number of interactions in species 1 or 2, respectively (with all interactions restricted to the set of proteins with homologues in the opposite species). Pairwise network distance was then defined as  $1 - s_{1,2}$ . The resulting phylogenetic tree shown in Fig. 2c is the consensus over 10,000 bootstrap simulations. Values of  $c$  and  $t$  for each network are listed in Supplementary Table 2.

## Signal-to-noise ratio of protein complexes

SNR was computed for the single-species complexes as follows. The search for dense interaction complexes is initiated from each node (protein) and the highest scoring complex from each is reported (see the ‘Identification of conserved and species-specific complexes’ section of the Methods). This yields a distribution of complex scores over all nodes in the network. A score distribution is also generated for 100 randomized networks, which have an identical degree distribution as the original network. The SNR ratio is computed from these original and random score distributions (representing signal and noise, respectively) according to the standard formula<sup>17</sup> using the root mean square (r.m.s.):

$$\text{SNR} = \log_{10} \frac{\text{r.m.s. (original complex scores)}}{\text{r.m.s. (random complex scores)}},$$

$$\text{with r.m.s. (x)} = \sqrt{\frac{1}{M} \sum_{i=1}^M x_i^2} \quad (1)$$

and where  $x_i$  is the score of a complex and  $M$  is the total number of complexes.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

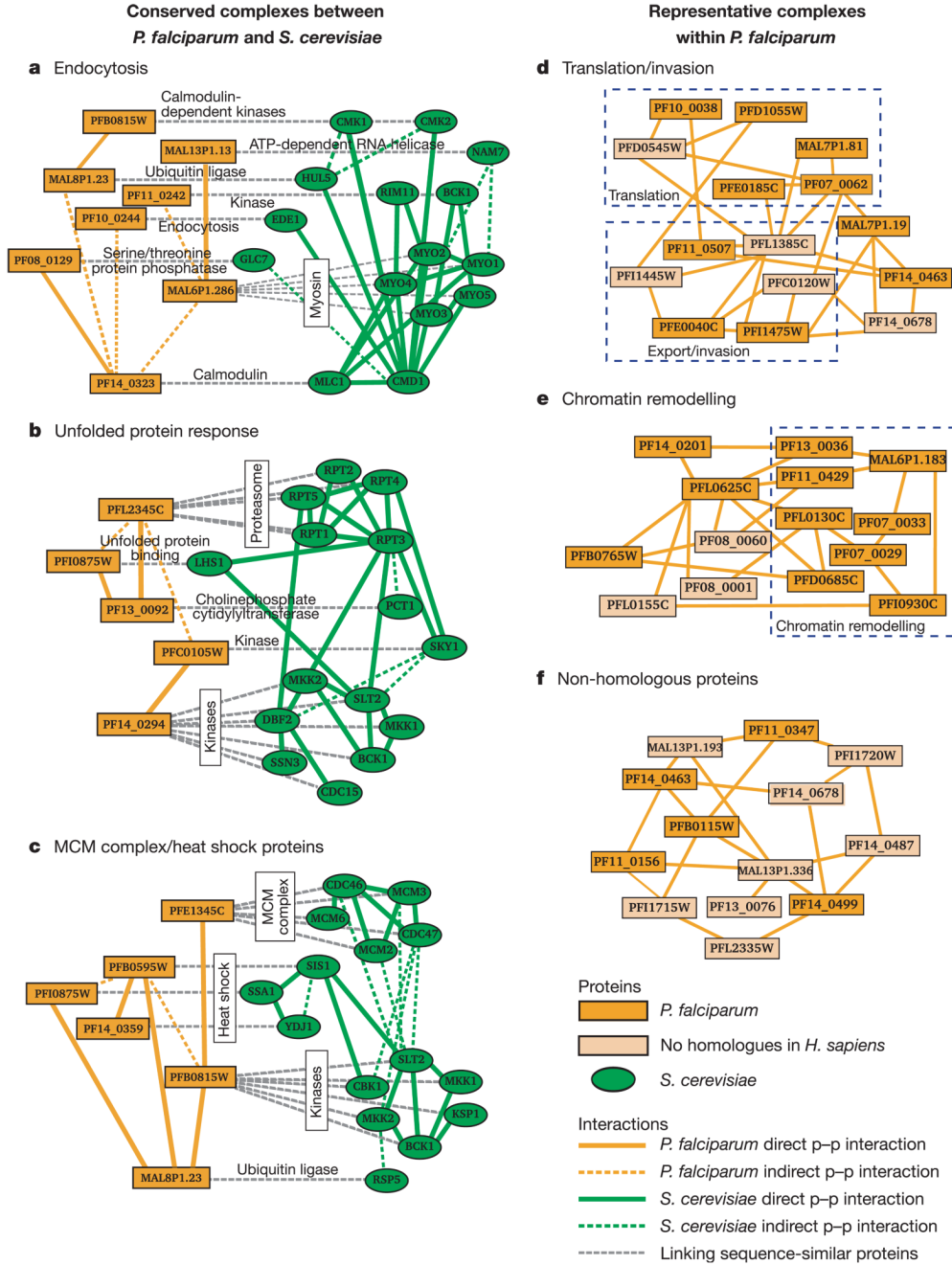
## Acknowledgments

We are indebted to M. Vignali, D. LaCount and S. Fields at the University of Washington, and B. Hughes and S. Sahasrabudhe at Prolexys, for providing us with advance access to the *Plasmodium* interaction data and for suggestions on our manuscript. We also thank E. Winzeler and J. Vinetz for advice on *Plasmodium* protein function, R. Sharan for help with the PathBLAST algorithm, and V. Bafna for assistance with the false-positive analysis. Finally, we acknowledge the following funding support: the National Science Foundation (S.S.); the National Institute of General Medical Sciences (T.I.); a David and Lucille Packard Fellowship award (T.I.); the Howard Hughes Medical Institute (T.S.); and Unilever (T.S.).

## References

1. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. *Nature* 2002;415:673–679. [PubMed: 11832955]
2. Gardner MJ, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002;419:498–511. [PubMed: 12368864]
3. Bozdech Z, et al. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 2003;1:E5. [PubMed: 12929205]
4. Le Roch KG, et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 2003;301:1503–1508. [PubMed: 12893887]

5. Florens L, et al. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 2002;419:520–526. [PubMed: 12368866]
6. LaCount DJ, et al. A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature*. doi:10.1038/nature04104 (this issue).
7. Conant GC, Wagner A. Convergent evolution of gene circuits. *Nature Genet* 2003;34:264–266. [PubMed: 12819781]
8. Yu H, et al. Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. *Genome Res* 2004;14:1107–1118. [PubMed: 15173116]
9. Sharan R, et al. Conserved patterns of protein interaction in multiple species. *Proc. Natl Acad. Sci. USA* 2005;102:1974–1979. [PubMed: 15687504]
10. Kelley BP, et al. Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl Acad. Sci. USA* 2003;100:11394–11399. [PubMed: 14504397]
11. Xenarios I, et al. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 2002;30:303–305. [PubMed: 11752321]
12. Li S, et al. A map of the interactome network of the metazoan *C. elegans*. *Science* 2004;303:540–543. [PubMed: 14704431]
13. Giot L, et al. A protein interaction map of *Drosophila melanogaster*. *Science* 2003;302:1727–1736. [PubMed: 14605208]
14. Rain JC, et al. The protein–protein interaction map of *Helicobacter pylori*. *Nature* 2001;409:211–215. [PubMed: 11196647]
15. Felsenstein J. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 1989;5:164–166.
16. Barabasi AL, Oltvai ZN. Network biology: understanding the cell’s functional organization. *Nature Rev. Genet* 2004;5:101–113. [PubMed: 14735121]
17. Shanmugam, KS. Digital and Analog Communication Systems. New York: Wiley; 1979.
18. Gagny B, et al. A novel EH domain protein of *Saccharomyces cerevisiae*, Ede1p, involved in endocytosis. *J. Cell Sci* 2000;113:3309–3319. [PubMed: 10954428]
19. Engqvist-Goldstein AE, Drubin DG. Actin assembly and endocytosis: from yeast to mammals. *Annu. Rev. Cell Dev. Biol* 2003;19:287–332. [PubMed: 14570572]
20. Goodson HV, Anderson BL, Warrick HM, Pon LA, Spudich JA. Synthetic lethality screen identifies a novel yeast myosin I gene (MYO5): myosin I proteins are required for polarization of the actin cytoskeleton. *J. Cell Biol* 1996;133:1277–1291. [PubMed: 8682864]
21. Salisbury JL, Condeelis JS, Maihle NJ, Satir P. Calmodulin localization during capping and receptor-mediated endocytosis. *Nature* 1981;294:163–166. [PubMed: 7029305]
22. Scheibel LW, et al. Calcium and calmodulin antagonists inhibit human malaria parasites (*Plasmodium falciparum*): implications for drug design. *Proc. Natl Acad. Sci. USA* 1987;84:7310–7314. [PubMed: 3313391]
23. Sanchez CP, McLean JE, Stein W, Lanzer M. Evidence for a substrate specific and inhibitable drug efflux system in chloroquine resistant *Plasmodium falciparum* strains. *Biochemistry* 2004;43:16365–16373. [PubMed: 15610031]
24. Hoppe HC, et al. Antimalarial quinolines and artemisinin inhibit endocytosis in *Plasmodium falciparum*. *Antimicrob. Agents Chemother* 2004;48:2370–2378. [PubMed: 15215083]
25. Langst G, Becker PB. Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J. Cell Sci* 2001;114:2561–2568. [PubMed: 11683384]
26. Sollars V, et al. Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nature Genet* 2003;33:70–74. [PubMed: 12483213]
27. Gavin AC, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002;415:141–147. [PubMed: 11805826]
28. Ashburner M, et al. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nature Genet* 2000;25:25–29. [PubMed: 10802651]
29. Uetz P, et al. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000;403:623–627. [PubMed: 10688190]
30. Mewes HW, et al. MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res* 2004;32:D41–D44. [PubMed: 14681354]



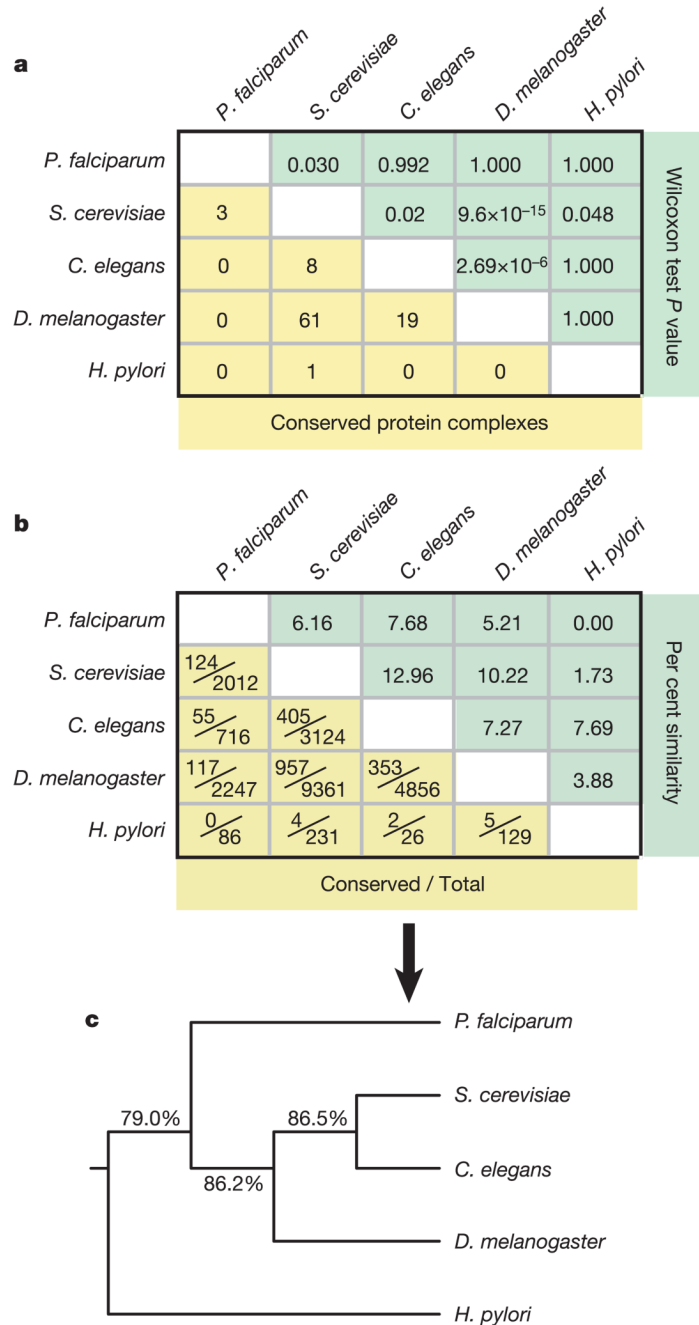
**Figure 1. Conserved and distinct complexes within *P. falciparum***

**a–c**, The three conserved complexes identified between *P. falciparum* and *S. cerevisiae*.

Orange versus green nodes correspond to *P. falciparum* versus *S. cerevisiae* proteins. Solid links represent direct protein–protein (p–p) interactions, while dashed links represent indirect protein–protein interactions mediated by one other protein. Grey dashed lines connect sequence-similar proteins across the two species. MCM, minichromosome maintenance. **d–f**, Three representative complexes found within the *P. falciparum* network only (out of 29 total complexes; see Supplementary Table 7). Cream-coloured nodes denote *Plasmodium* proteins without human homologues (*Homo sapiens*; using a permissive BLAST *E* value threshold  $\leq 1$

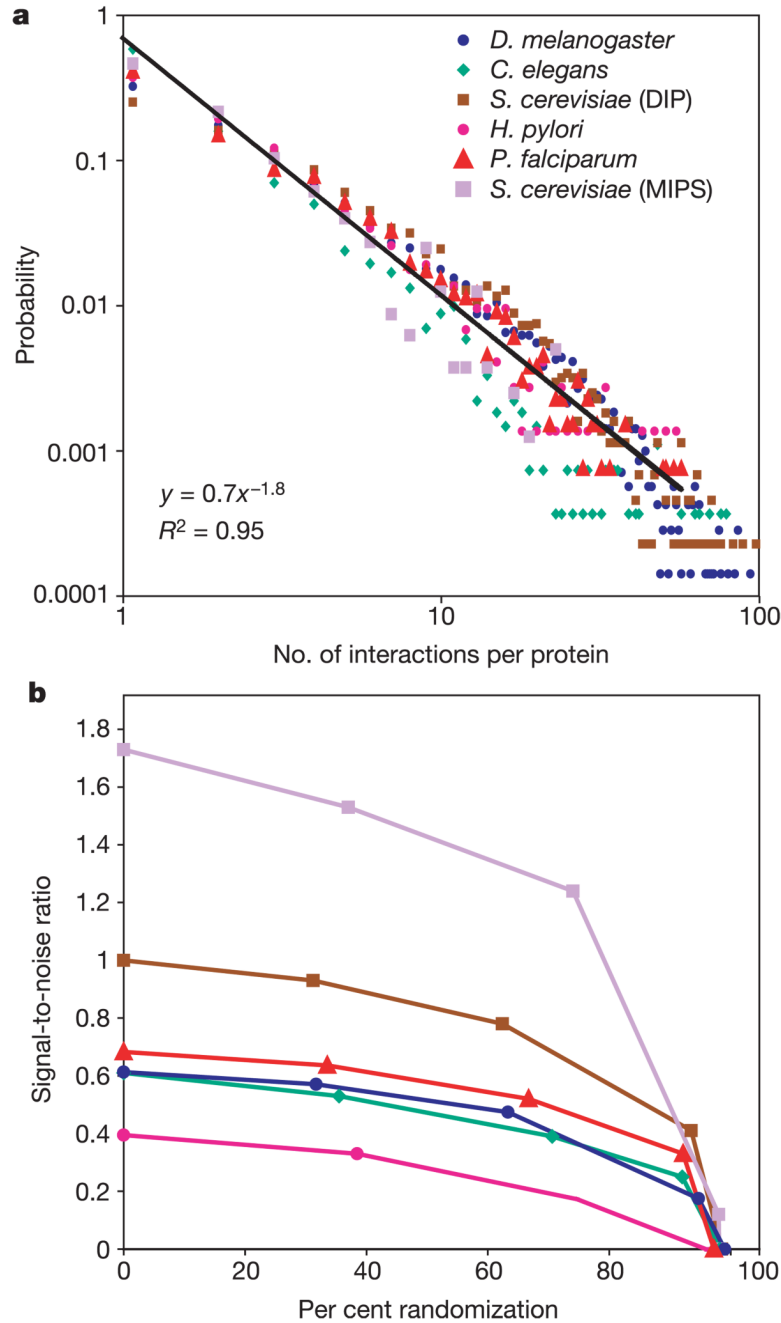
$\times 10^{-2}$  to allow for distant homologues). All figures were drawn using the software Cytoscape (<http://www.cytoscape.org>).





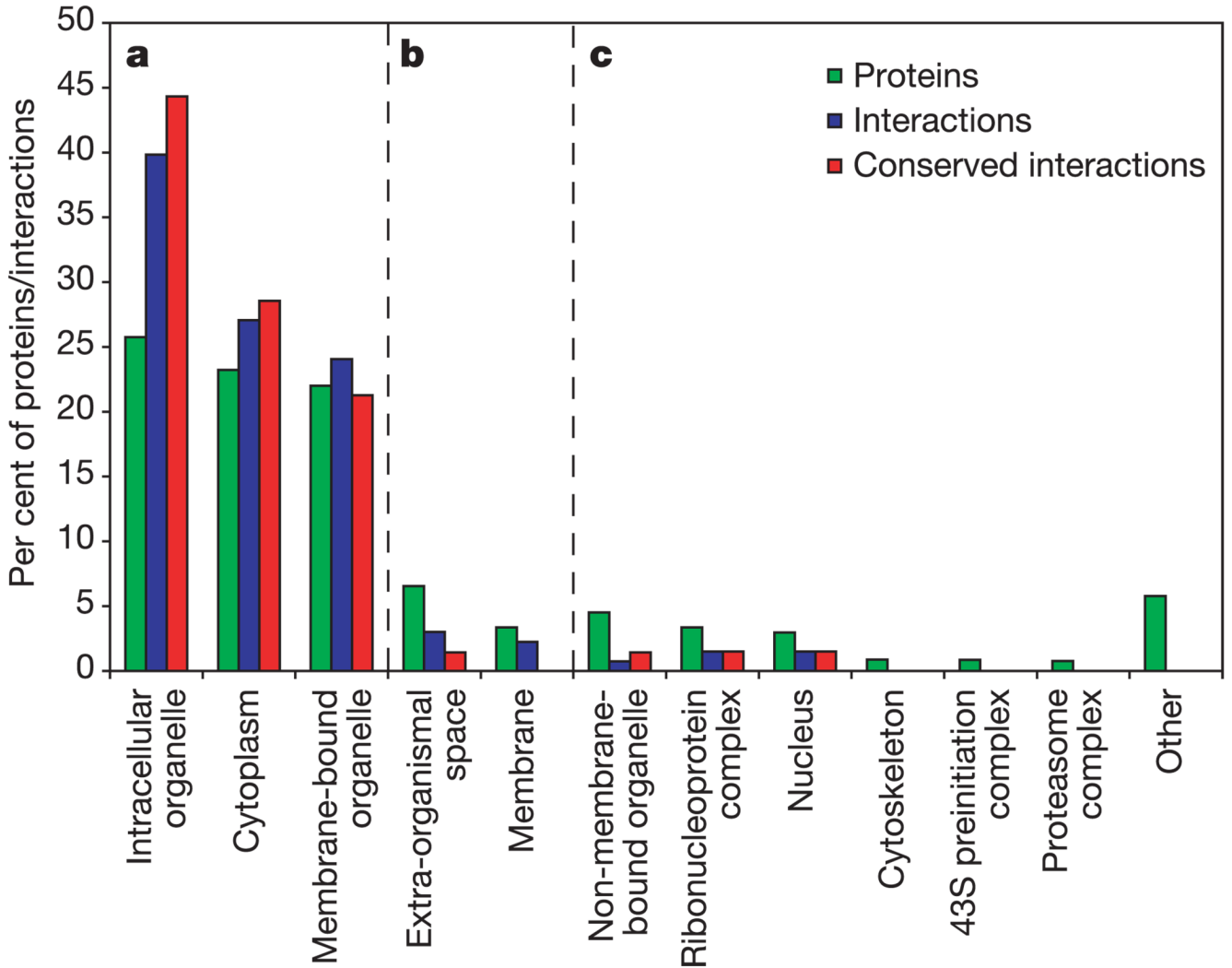
**Figure 2. Network similarity across the five organisms**

**a**, The results of all pairwise PathBLAST comparisons. The number of conserved complexes is shown for each pair of species (yellow). The Wilcoxon rank-sum test *P* value (green) represents the significance of the distribution of all complex scores versus the distribution of complex scores found in equivalent random networks. **b**, The interaction-by-interaction similarity between networks is reported as both fractional values (yellow) and percentages (green). **c**, A phylogenetic tree constructed using these similarities. Percentages indicate the reproducibility of each branch during bootstrap analysis.



**Figure 3. Number of interactions per protein and the signal-to-noise ratio of protein complexes**  
**a**, Scale-free network behaviour is shown in a manner similar to ref. <sup>16</sup>. The linear fit is for *Plasmodium* data only. **b**, Dependence of SNR on error rate. See the key in panel **a** for an explanation of the symbols. Each network was modified by randomly shuffling from 0–100% of its protein interactions to simulate the addition of false positives and negatives. The subsequent decrease in SNR, converging to SNR = 0 at 100% noise, validates that each network contains a substantial fraction of true positive interactions. The high-throughput *S. cerevisiae* data shown is from the network of yeast interactions in the Database of Interacting Proteins (DIP)<sup>11</sup>. In addition to the high-throughput networks in this study, a literature-curated network—*S. cerevisiae* physical interactions according to the Munich Information Center for

Protein Sequences (MIPS)<sup>30</sup>—is provided as a positive control. Similar trends are observed for the average clustering coefficient and the per cent of interactions covered by established protein-domain interactions (Supplementary Table 3).



**Figure 4. Functional roles within the *Plasmodium* protein network**

Functional annotations within the *Plasmodium* protein network associated with intracellular organelles, membrane-bound organelles or the cytoplasm (a), the membrane and extra-organismal space (b), and other components such as the proteasome and cytoskeleton (c). The histograms show the distribution of Gene Ontology Cellular Component assignments over all annotated *Plasmodium* proteins (green bars), protein interactions (blue bars) and conserved protein interactions (red bars). Interactions are considered ‘annotated’ if the interacting proteins share the same Gene Ontology category (these interactions are listed in Supplementary Table 6). For conserved interactions, the percentages in each category are cumulative over the three pairwise comparisons of *Plasmodium* versus the other three eukaryotes (yeast, fly or worm). Note that a protein or interaction can participate in multiple categories.

Table 1

## Topological properties of protein interaction networks

Organism	No. of interactions	Proteins covered	Average degree	Average shortest path	Average clustering coefficient*	No. of yeast homologues <sup>†</sup> ( <i>P</i> value)	No. of single-species complexes
<i>S. cerevisiae</i> (DIP) <sup>‡</sup>	14,319	4,389	6.53	4.12	0.193	–	145
<i>S. cerevisiae</i> (Uetz) <sup>‡</sup>	1,449	1,345	2.16	6.95	0.049	–	66
<i>P. falciparum</i>	2,846	1,312	4.35	4.20	0.032	286 ( $2 \times 10^{-133}$ )	29
<i>D. melanogaster</i>	20,720	7,038	5.89	4.70	0.019	2,429 ( $2 \times 10^{-205}$ )	296
<i>C. elegans</i>	3,926	2,718	2.89	5.10	0.031	673 ( $4 \times 10^{-10}$ )	12
<i>H. pylori</i>	1,465	732	4.00	4.15	0.063	143 ( $1 \times 10^{-2}$ )	21

\*The clustering coefficient measures local density of the network around a protein and is computed as previously described<sup>16</sup>.

<sup>†</sup>Yeast homologues are determined using a conservative BLAST *E* value threshold of  $\leq 1 \times 10^{-10}$ . The *P* values score the significance of enrichment for yeast homologues within the set of proteins covered by each interaction network, using the hypergeometric test. These enrichments are significant over a broad range of *E* value thresholds (data not shown).

<sup>‡</sup>Unlike other networks that are generated from single two-hybrid studies, the network of yeast interactions in the Database of Interacting Proteins (DIP)<sup>11</sup> consists of many experiments and experimental types. A separate analysis is included considering only the data from a single two-hybrid screen by Uetz *et al.* (ref. 29).