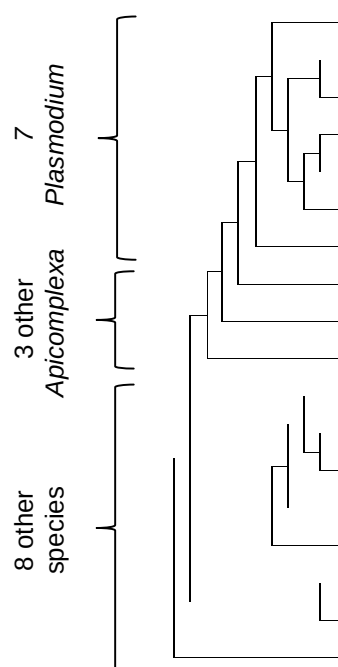




Species	Source	No. proteins	No. longest isoforms
<i>Plasmodium falciparum</i>	PlasmoDB 7.0	5,491	5,458
<i>Plasmodium vivax</i>	PlasmoDB 7.0, genBlastG-improved	5,493	5,493
<i>Plasmodium knowlesi</i>	PlasmoDB 7.0, genBlastG-improved	5,217	5,217
<i>Plasmodium yoelii</i>	PlasmoDB 7.0, genBlastG-improved	7,777	7,777
<i>Plasmodium berghei</i>	PlasmoDB 7.0, genBlastG-improved	4,906	4,906
<i>Plasmodium chabaudi</i>	PlasmoDB 7.0, genBlastG-improved	5,134	5,134
<i>Plasmodium gallinaceum</i>	WTSI, genBlastG-annotated	3,141	3,141
<i>Theileria parva</i>	PiroplasmaDB v1.0	4,082	4,082
<i>Toxoplasma gondii</i>	ToxoDB v6.4	7,940	7,940
<i>Cryptosporidium parvum</i>	CryptoDB 4.4	3,805	3,805
<i>Homo sapiens</i>	Ensembl release 62	88,237	22,111
<i>Drosophila melanogaster</i>	Flybase (FB2011_05)	23,605	14,013
<i>Caenorhabditis elegans</i>	Wormbase WS224	25,010	20,429
<i>Monosiga brevicollis</i>	NCBI genomes	9,203	9,203
<i>Saccharomyces cerevisiae</i>	SGD (June 6, 2011)	6,717	6,717
<i>Arabidopsis thaliana</i>	PlantGDB release 183	35,386	27,416
<i>Chlamydomonas reinhardtii</i>	NCBI genomes	14,489	14,489
<i>Escherichia coli</i>	NCBI Genomes	4,145	4,145
Total		259,778	171,476

Additional File 6: List of the 18 species and number of proteins per species included in the protein sequence data set. The dendrogram to the left shows the phylogenetic relationship of the species and was compiled from the tree of life Web site (<http://tolweb.org>). WTSI... Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/resources/downloads/protozoa/plasmodium-gallinaceum.html>).