eLIFE
elifesciences.org

Figures and figure supplements

Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites

Yong H Woo, et al.

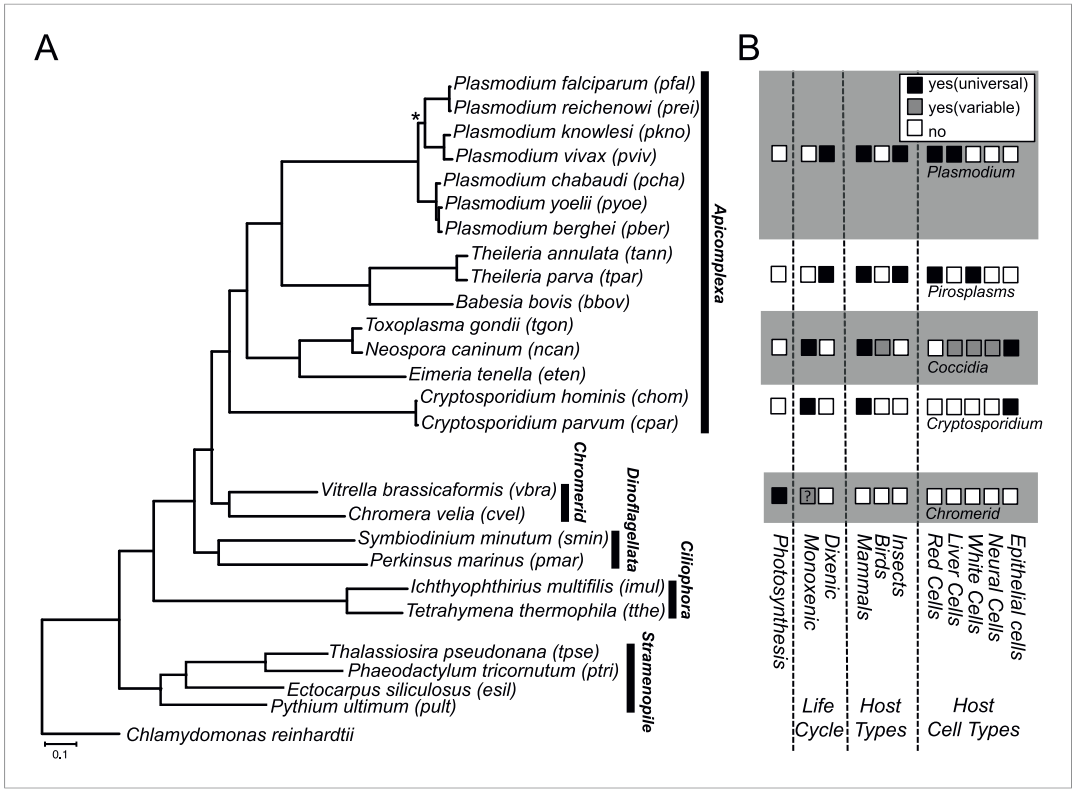

Figure 1. Phylogenetic, parasitological, and genomic context of chromerids. (**A**) Phylogenetic tree of 26 alveolate and outgroup species (see ***Figure 1—source data 1*** for the list of species). Multiple sequence alignments of 101 genes, which have 1:1 orthologs across all species (***Figure 1—source data 2***) were concatenated to a single matrix of 33,997 aligned amino acids. A maximum likelihood tree was inferred using RAxML with 1000 bootstraps, with *Chlamydomonas reinhardtii* as an outgroup. All clades are supported with bootstrap values of 100% except one node (*) with 99%, and also with 1.00 posterior probability from a bayesian phylogenetic tree based on PhyloBayes (***Lartillot and Philippe, 2004***) (CAT-GTR). (**B**) Lifestyles of the apicomplexan and chromerid species under investigation. '?': uncertainty due to lack of relevant data.

DOI: 10.7554/eLife.06974.003

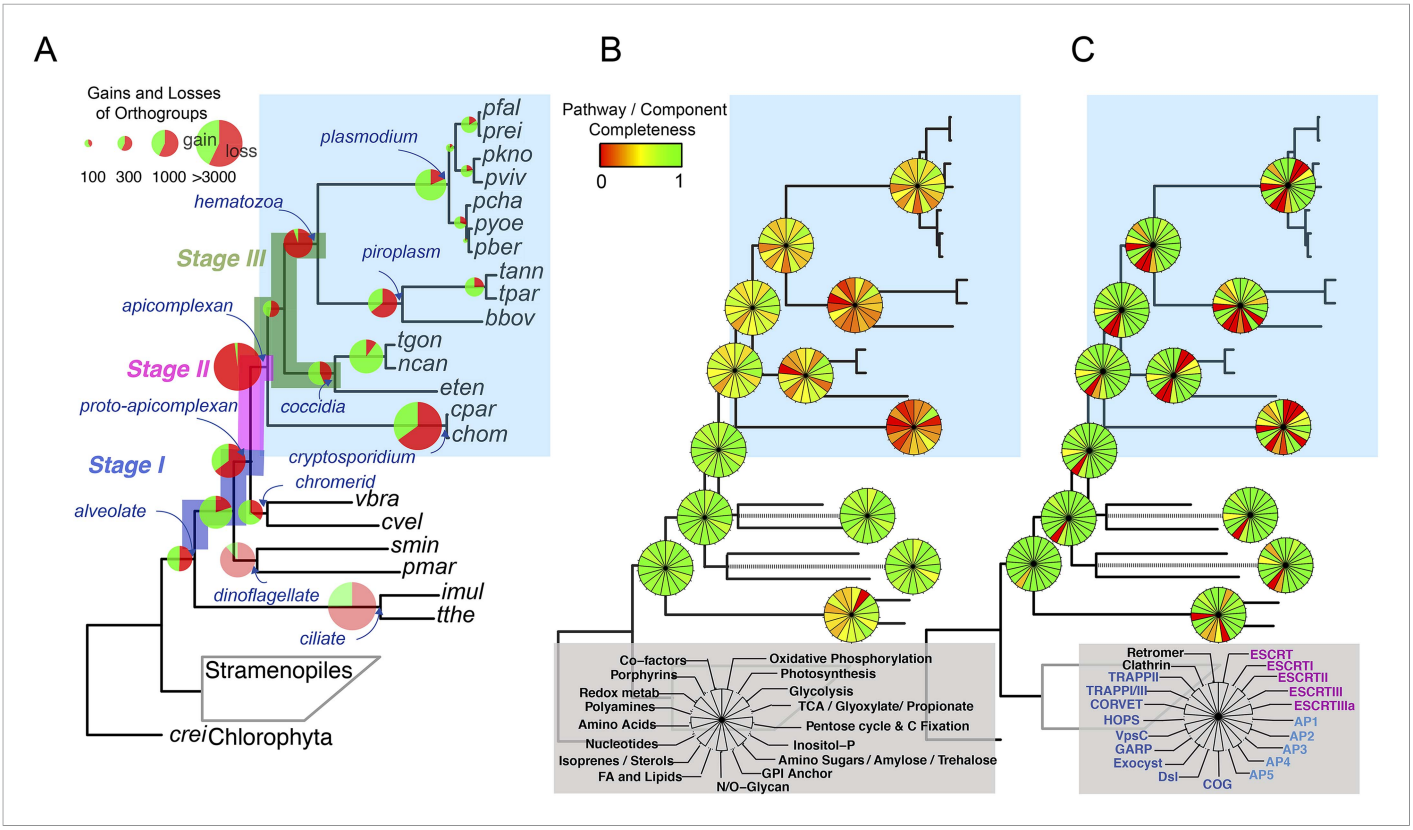

Figure 2. Gene content changes during apicomplexan evolution. (**A**) Gains and losses of orthogroups inferred based on Dollo parsimony (***Csuros, 2010***). Analysis based on a gene birth-and-death model provided similar results (***Figure 2—figure supplement 1A***). Stages I, II, and III (shown in blue, pink and green, respectively) represent groups of branches from the alveolate ancestor to apicomplexan lineage ancestors. Stage III could not be determined for Cryptosporidium lineage because of sparse taxon sampling. The area of a green or red section in a pie is proportional to the number of gained or lost orthogroups, respectively. (**B**, **C**) Overview of metabolic capabilities (**B**) and endomembrane components (**C**) in apicomplexan and chromerid ancestors. Gains and losses of enzymes and components were inferred, based on Dollo parsimony (***Csuros, 2010***). The pie charts are color-coded based on the fraction of enzymes or components present. Additional results from analysis of individual components and enzymes can be found in ***Figure 2—figure supplements 2,3,4,5***, ***Supplementary file 3***. Individual components and enzymes are listed in ***Figure 2—source data 1, 2***. Similar analyses were performed for components encoding flagellar apparatus (***Figure 2—figure supplement 5B***).

DOI: 10.7554/eLife.06974.006

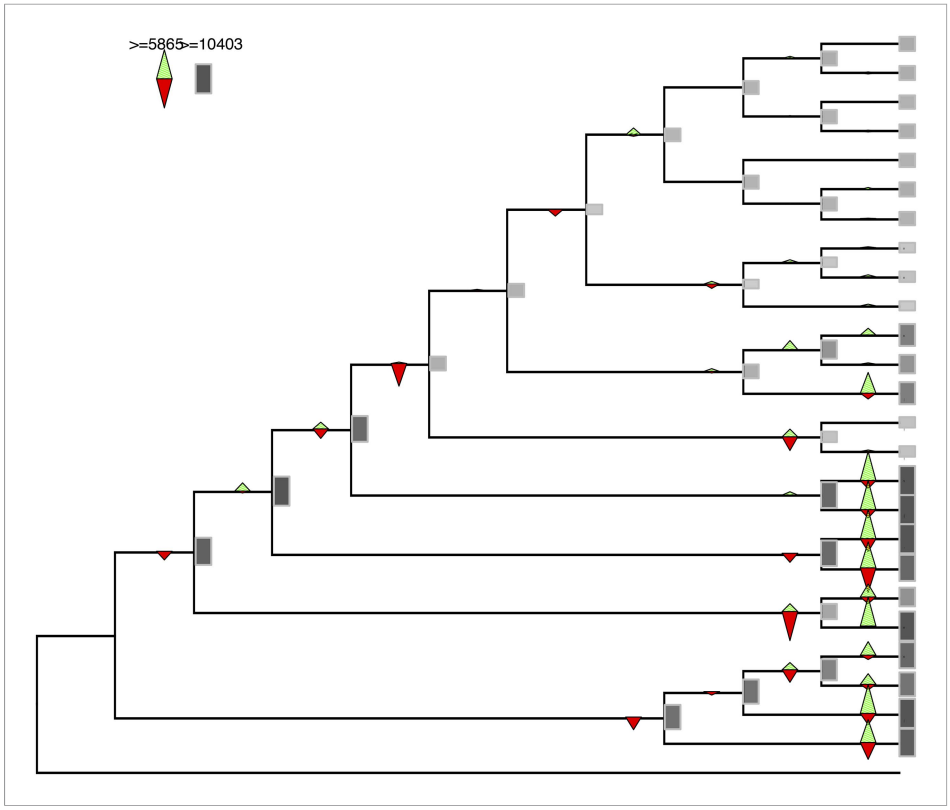

Figure 2—figure supplement 1. Gene gains and losses across the hypothetical ancestors of the 26 species under study.
DOI: 10.7554/eLife.06974.009

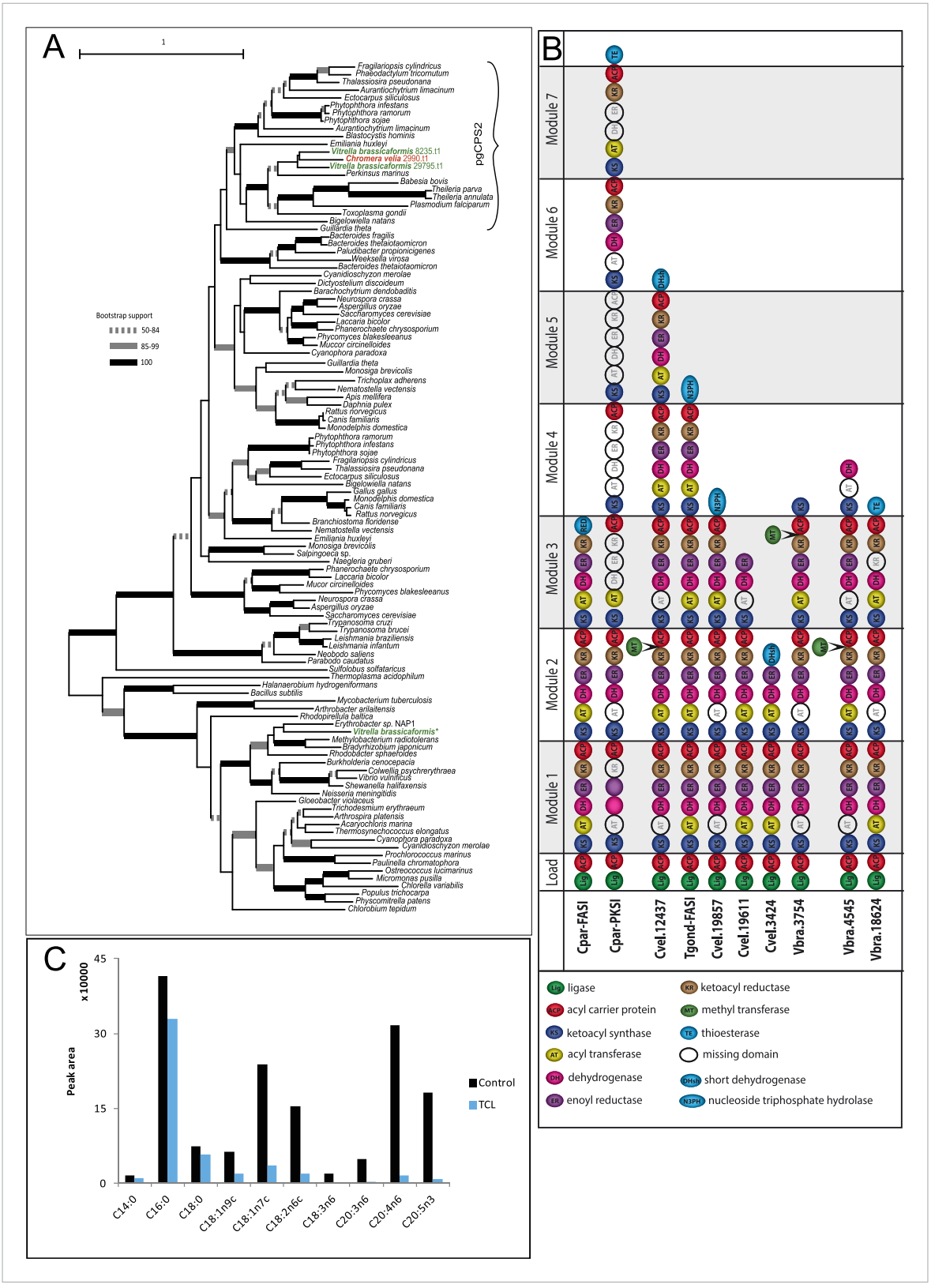

Figure 2—figure supplement 2. Overview of chromerid Carbamoyl Phosphate Synthetase (CPS) and Fatty Acid Synthase I (FAS I).
DOI: 10.7554/eLife.06974.010

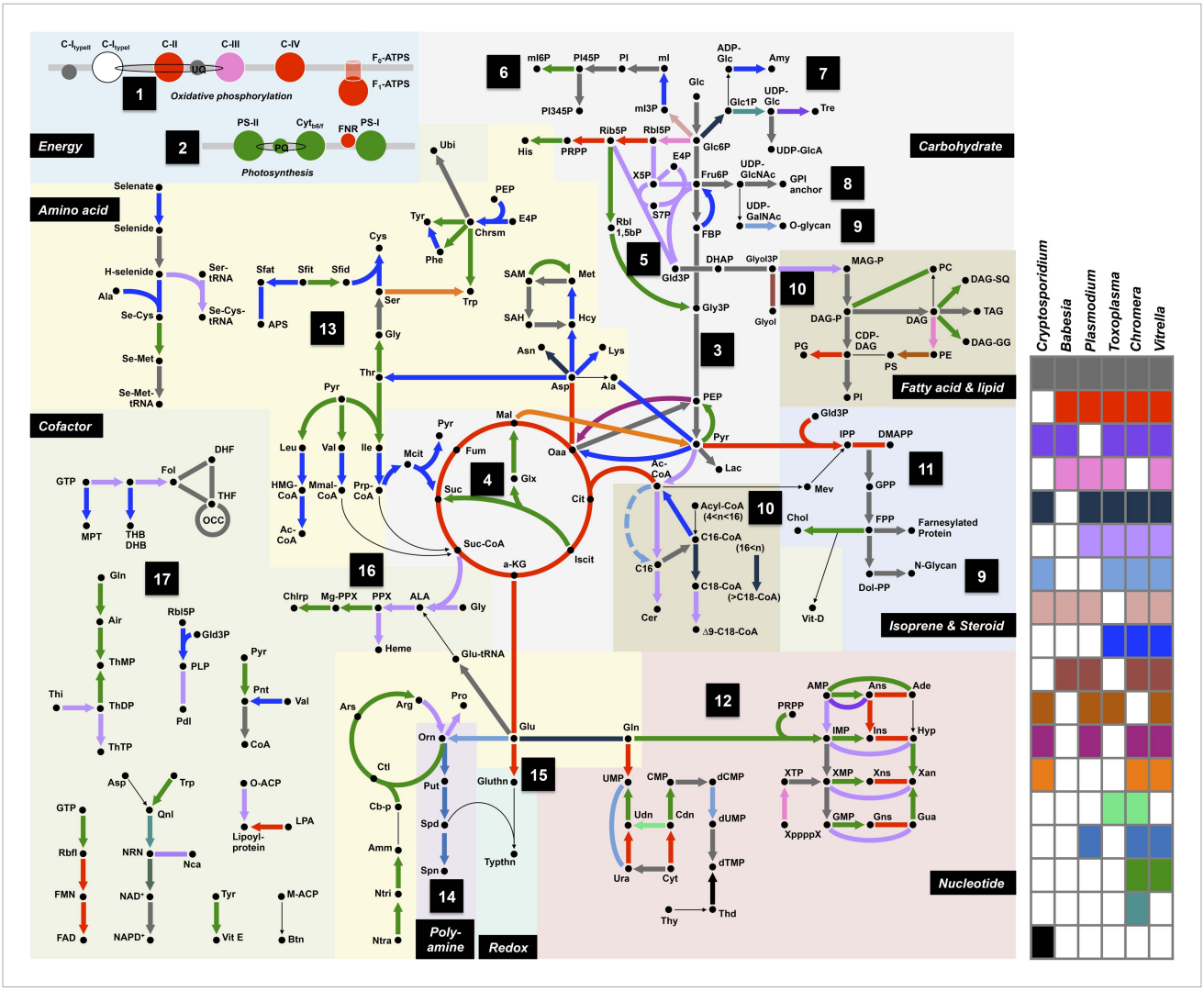

Figure 2—figure supplement 3. Summary of metabolic pathways based on KEGG Assignments.
DOI: 10.7554/eLife.06974.011

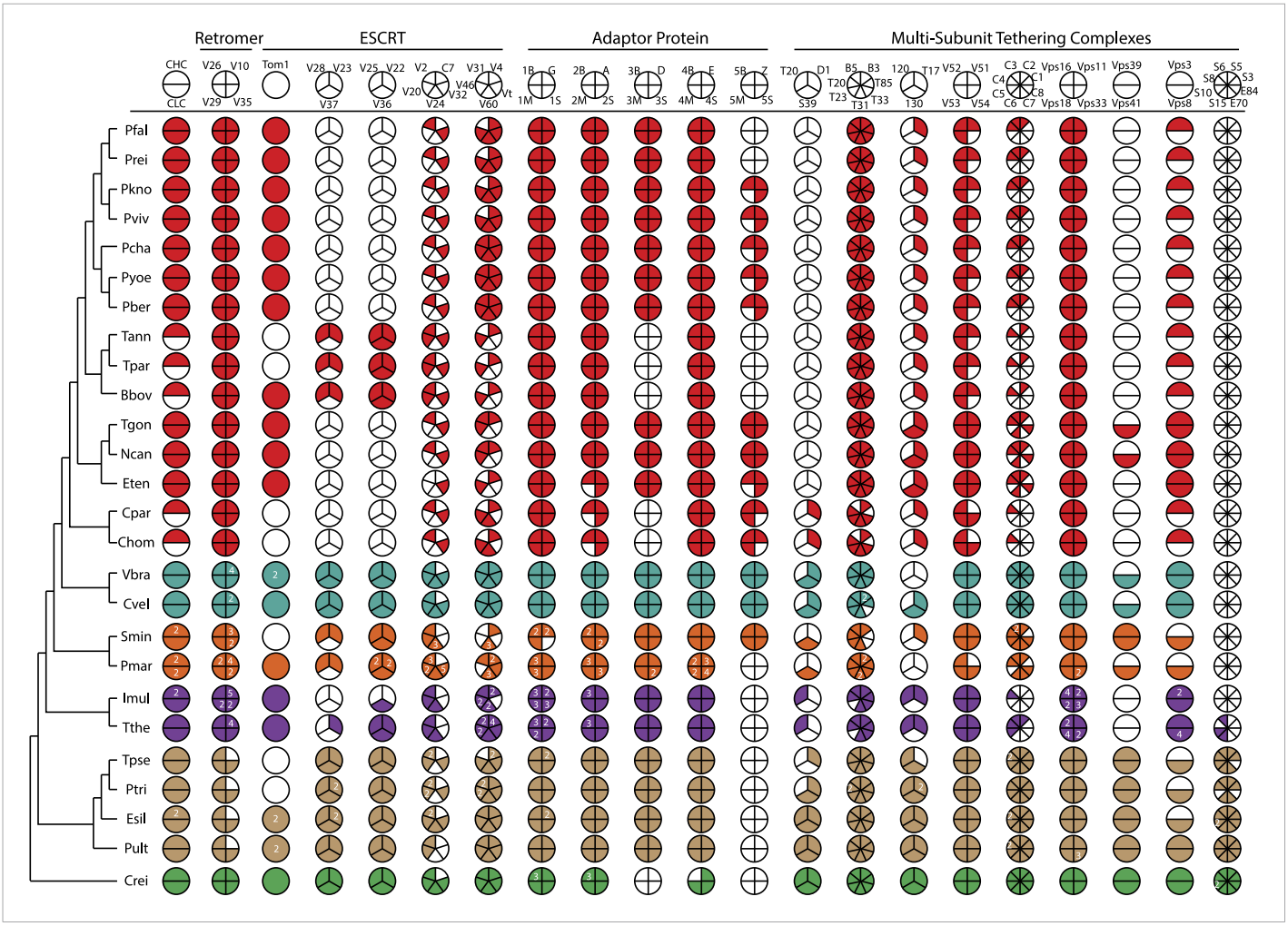

Figure 2—figure supplement 4. An overview of endomembrane trafficking components.
DOI: 10.7554/eLife.06974.012

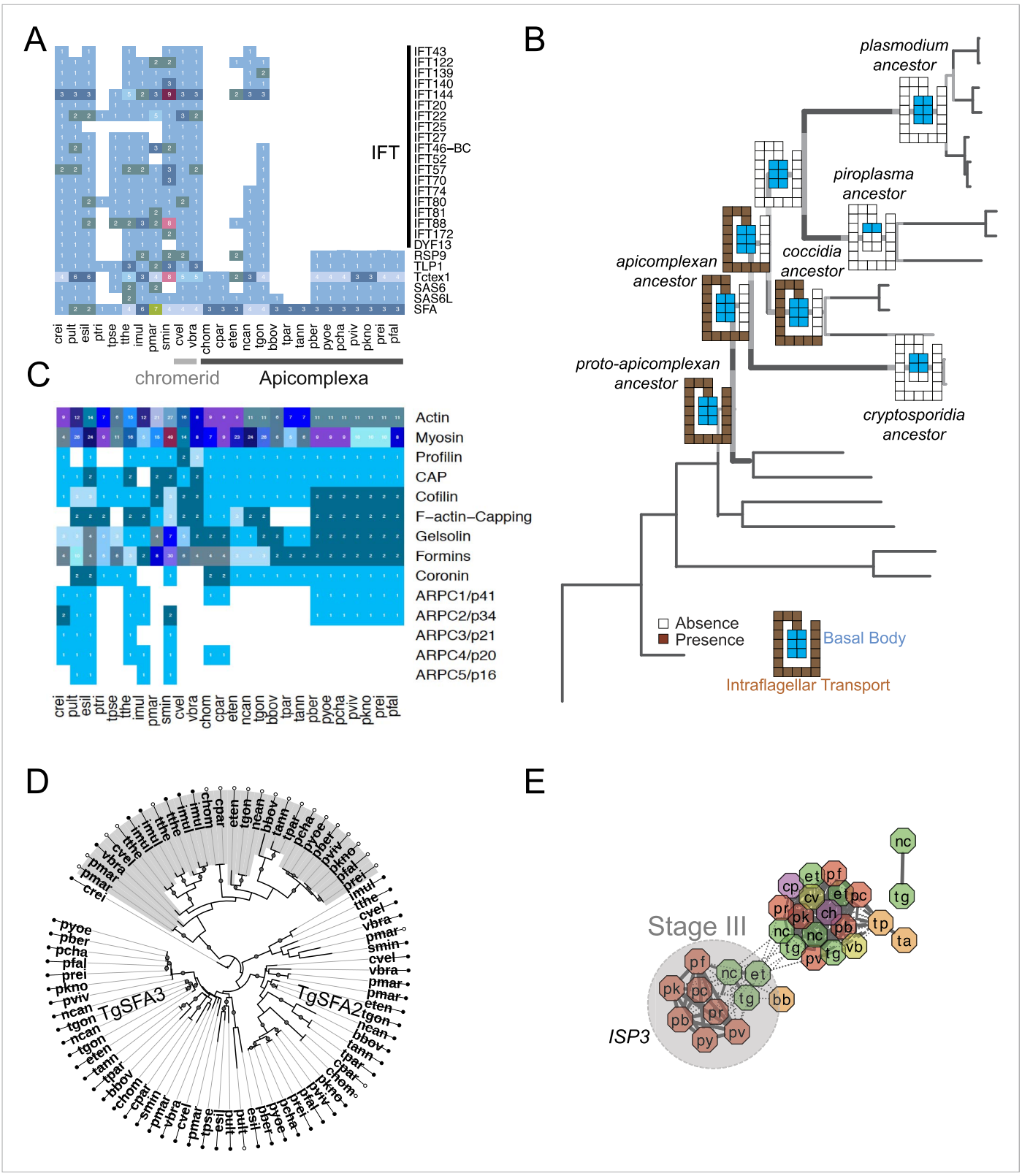

Figure 2—figure supplement 5. Evolutionary history of genes encoding cytoskeleton across 26 species.
DOI: 10.7554/eLife.06974.013

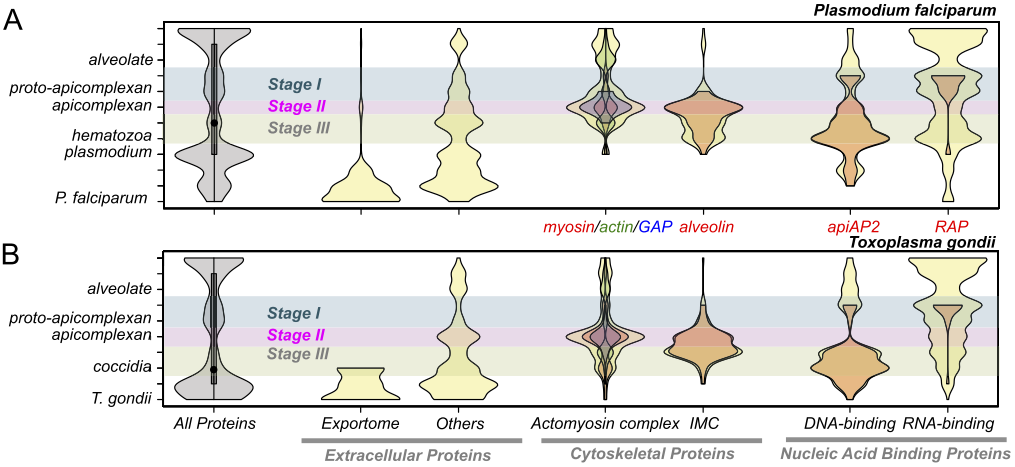

Figure 3. Evolutionary history of *Plasmodium falciparum* and *Toxoplasma gondii* genes. Violin plots showing distribution of evolutionary ages of genes (Y-axis: from species-specific (bottom) to deeply conserved (top)) in *P. falciparum* (**A**) and *T. gondii* (**B**). Evolutionary age of a gene is defined as the earliest node on the evolutionary path of the phylogenetic tree where homolog can be detected ('Materials and methods'). The horizontal thickness of a violin is proportional to the number of genes (gray) or the fraction of genes (yellow) in a functional category (X-axis) out of all with the same evolutionary age. Selected functional sub-categories are overlaid with red, green, or blue violin plots. The maximum width of each violin is scaled to be uniform across categories. Inner boxes in the gray violins indicate inter-quartile ranges and circles indicate medians. Colored shades along the X-axis indicate Stages I–III (***Figure 2***). Extracellular proteins include proteins targeted to host cytoplasm, nucleus, and plasma membrane ('exportome') and all other proteins, which are secreted or localized on the parasite surface ('others'). Cytoskeletal proteins include proteins associated with 'actomyosin motor complex' and 'IMC'. All extracellular and cytoskeletal proteins are listed in ***Figure 3—source data 1, 2***. Nucleic acid-binding proteins are predicted in silico based on presence of DNA-binding domains (DBDs) and RNA-binding domains (RBDs). See 'Materials and methods' for details on how these genes are defined and compiled. Domain architectures of representative extracellular proteins in apicomplexans and chromerids are displayed as schematics in ***Figure 3—figure supplement 4***. Sequence homology networks (***Figure 2—figure supplement 5E*** and ***Figure 3—figure supplements 1B, 2B, 3B***) and gene gains and losses on the phylogenetic tree (***Figure 3—figure supplements 1A, 2A, 3A***) provide complementary views on the evolutionary history of these genes.

DOI: 10.7554/eLife.06974.015

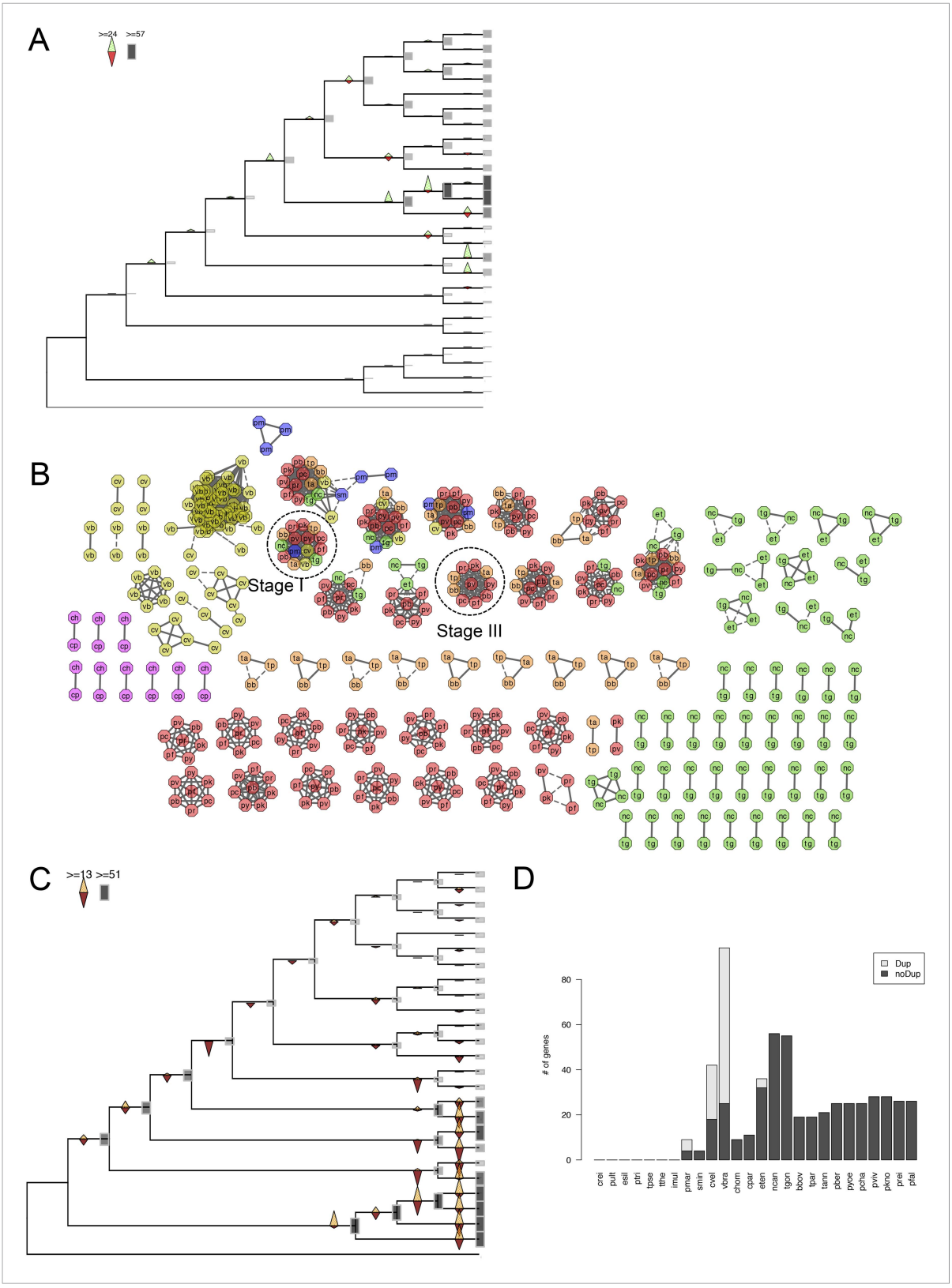

Figure 3—figure supplement 1. Evolutionary history of apiAP2 genes. (**A**) Gains and losses of apiAP2 genes inferred with Dollo parsimony. Triangles pointing upward indicate gains, triangles pointing downward losses, and the total number of orthogroups at that particular node are proportional to the

Figure 3—figure supplement 1. continued on next page

Figure 3—figure supplement 1. Continued

length or the shade thickness of the gray boxes. (**B**) Network view of amino acid sequence homology among all apiAP2. Edges are drawn depending on the strength of the sequence homology: dotted (BLASTP E value $<10^{-20}$) or solid (BLASTP E value $<10^{-30}$). The two letters within the node refer to acronyms of the species name and the node color species group: red (*Plasmodium*); green (coccidians); magenta (Cryptosporidia); orange (piroplasms); yellow (chromerids); and navy (dinoflagellates). For example, nodes from *P. falciparum* are shaded red and lettered with 'pf'. Connected nodes with different or the same species names indicate putative orthologs or paralog, respectively. Nodes without any edges, likely to be species-specific genes without other paralogs, were not displayed. Connections between nodes of different colors indicate deep evolutionary conservation. For example, connections between red and yellow nodes indicate orthologs shared between *Plasmodium* spp. and chromerids, which means that they have been gained by the proto-apicomplexan ancestor after its split from dinoflagellates (Stage I). (**C**) Gains and losses of DBD genes, excluding apiAP2, inferred with Dollo parsimony. (**D**) Bar chart showing putative apiAP2 paralogs (light gray), and singletons (black). We note a paucity of duplicate apiAP2 genes in apicomplexans compared to their abundance in chromerids.

DOI: 10.7554/eLife.06974.018

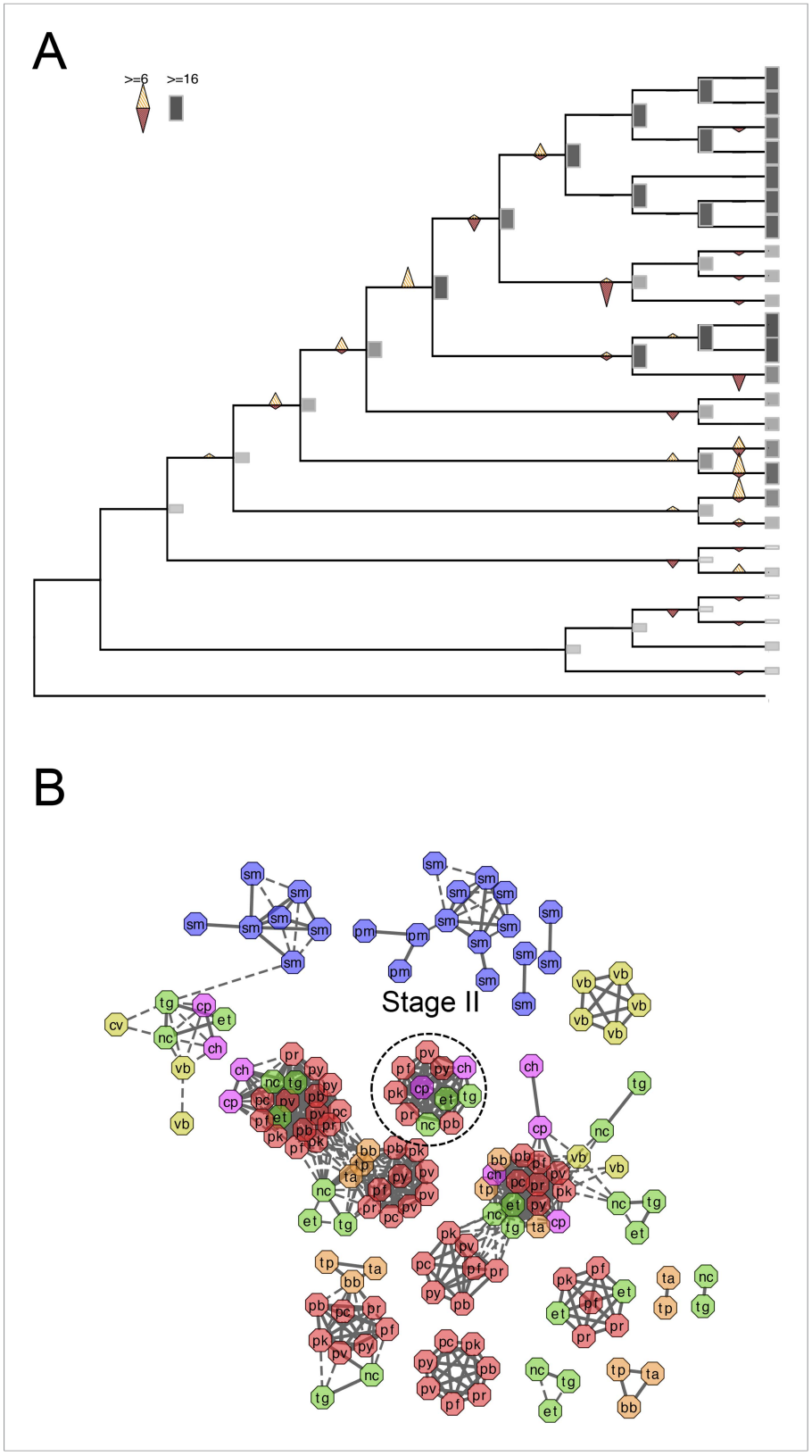

Figure 3—figure supplement 2. Evolutionary history of alveolins. (**A**) Gains and losses of alveolins inferred with Dollo parsimony. Triangles pointing upward indicate gains, triangles pointing downward losses, and the total

Figure 3—figure supplement 2. continued on next page

Figure 3—figure supplement 2. Continued

number of orthogroups at that particular node are proportional to the length or the shade thickness of the gray boxes. (**B**) Network view of amino acid sequence homology among all alveolins. Edges are drawn depending on the strength of the sequence homology: dotted (BLASTP E value $<10^{-20}$) or solid (BLASTP E value $<10^{-30}$). The two letters within the node refer to acronyms of the species name and the node color species group: red (*Plasmodium*); green (coccidians); magenta (Cryptosporidia); orange (piroplasms); yellow (chromerids); and navy (dinoflagellates).
DOI: 10.7554/eLife.06974.019

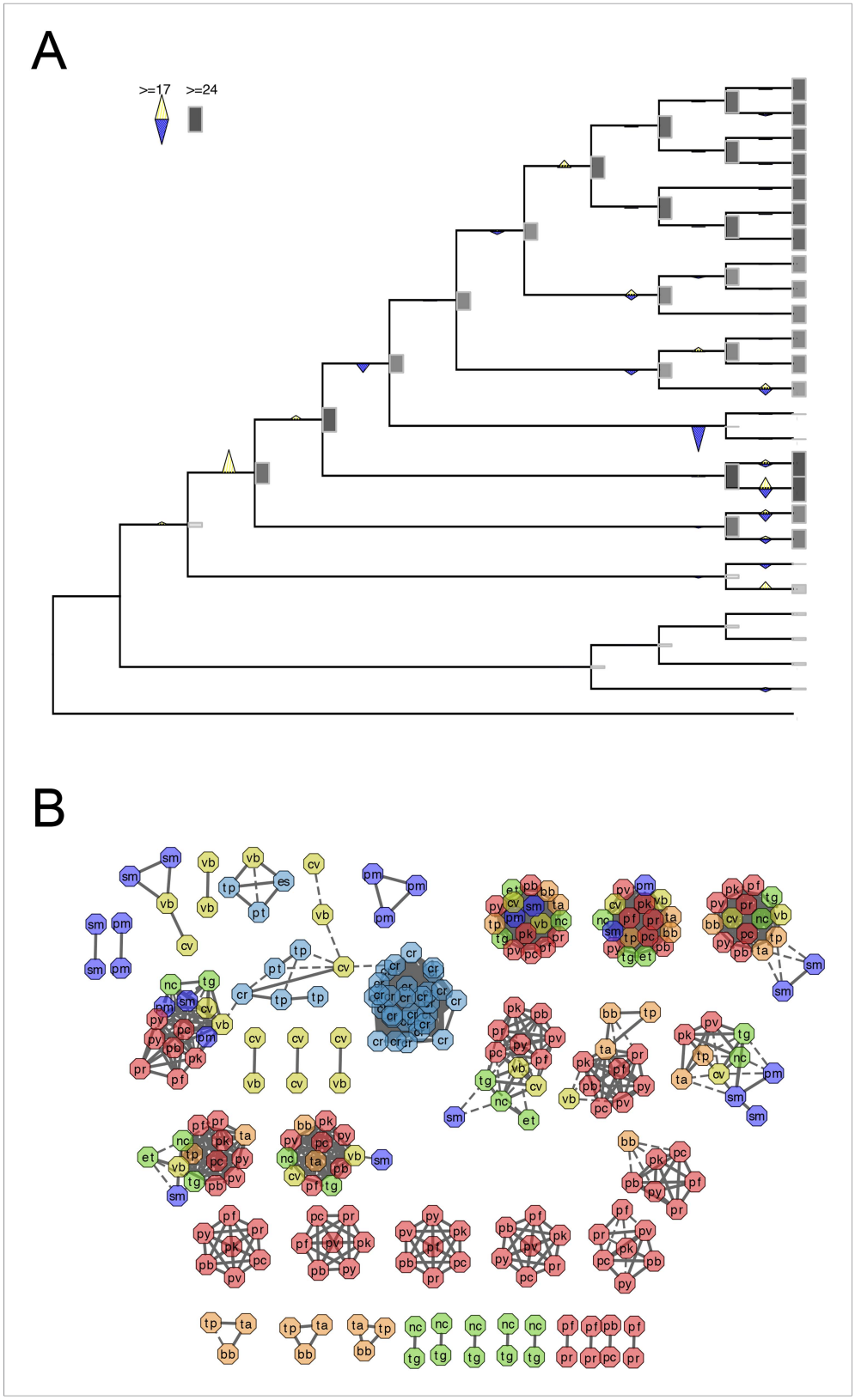

Figure 3—figure supplement 3. Evolutionary history of RAP genes. (**A**) Gains and losses of RAP genes inferred with Dollo parsimony. Triangles pointing upward indicate gains, triangles pointing downward losses, and the total

Figure 3—figure supplement 3. continued on next page

Figure 3—figure supplement 3. Continued

number of orthogroups at that particular node are proportional to the length or the shade thickness of the gray boxes. (**B**) Network view of amino acid sequence homology among all RAP genes. Edges are drawn depending on the strength of the sequence homology: dotted (BLASTP E value $<10^{-20}$) or solid (BLASTP E value $<10^{-30}$). The two letters within the node refer to acronyms of the species name and the node color species group: red (*Plasmodium*); green (coccidians); magenta (Cryptosporidia); orange (piroplasms); yellow (chromerids); and navy (dinoflagellates).
DOI: 10.7554/eLife.06974.020

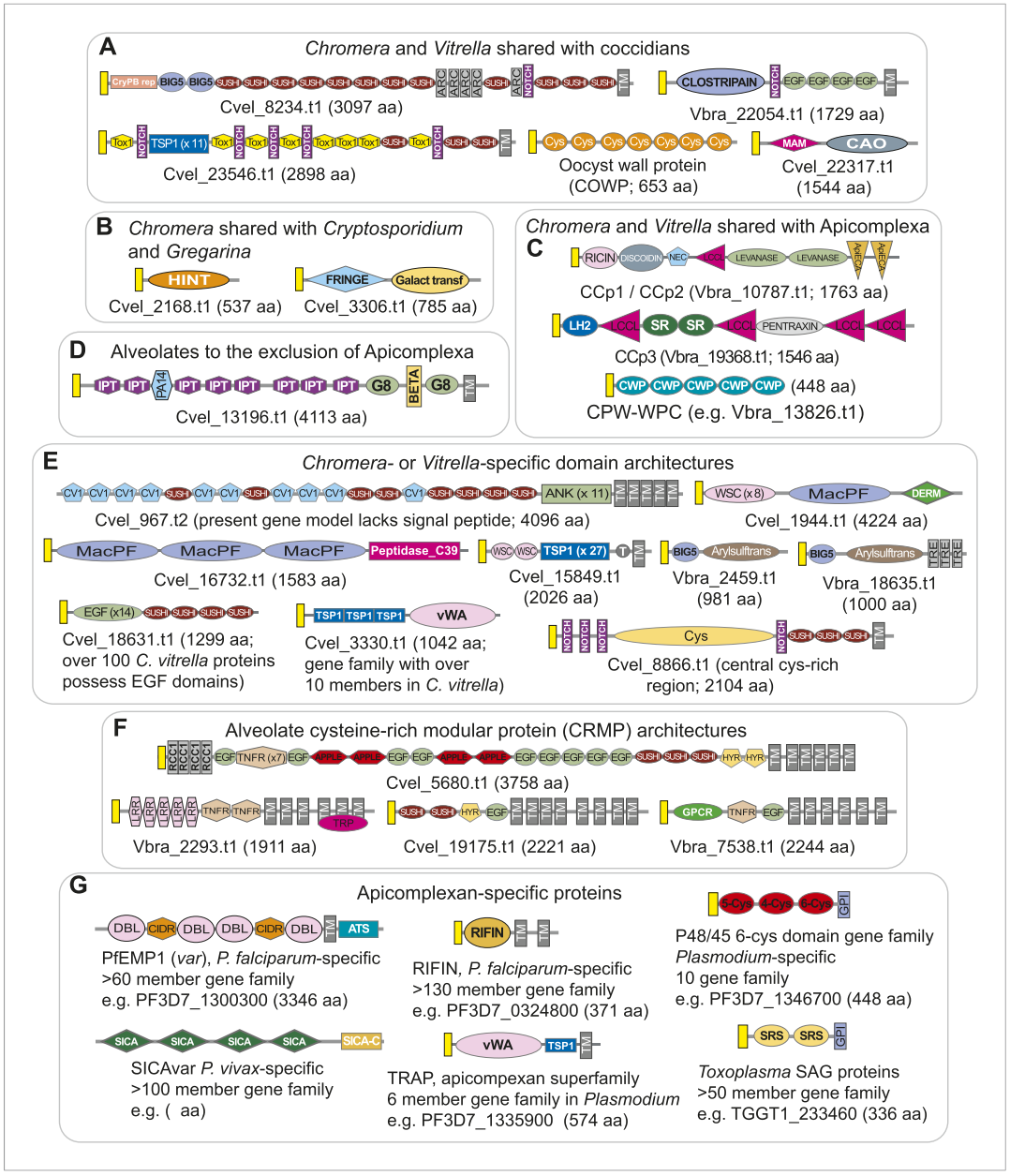

Figure 3—figure supplement 4. Domain architectures of extracellular proteins in chromerids and apicomplexans. Examples of domain architectures of predicted chromerid and apicomplexan extracellular proteins and their phyletic distribution. (**A**) Extracellular proteins with apparent orthologs in coccidians (for example, Toxoplasma) and Cryptosporidium. (**B**) Extracellular proteins shared with Cryptosporidium, but not identified in Toxoplasma and other apicomplexans. (**C**) Extracellular proteins conserved as apparent orthologs throughout the Apicomplexa. (**D**) An example of an apparent alveolate extracellular protein conserved in chromerids, Perkinsus and ciliates, but absent in apicomplexans. (**E**) Lineage-specific extracellular proteins identified only in one or both chromerids. (**F**) Examples of the diversity of domain architectures within chromerids for CRMP family proteins. (**G**) Examples of apicomplexan-specific extracellular proteins. Descriptions of domains and representative genes are provided in Appendix 5. A yellow rectangle indicates predicted signal peptide sequence. TM and GPI denote predicted transmembrane domain and glycosylphosphatidylinositol (GPI anchor), respectively. Protein lengths are not drawn to scale.
DOI: 10.7554/eLife.06974.021

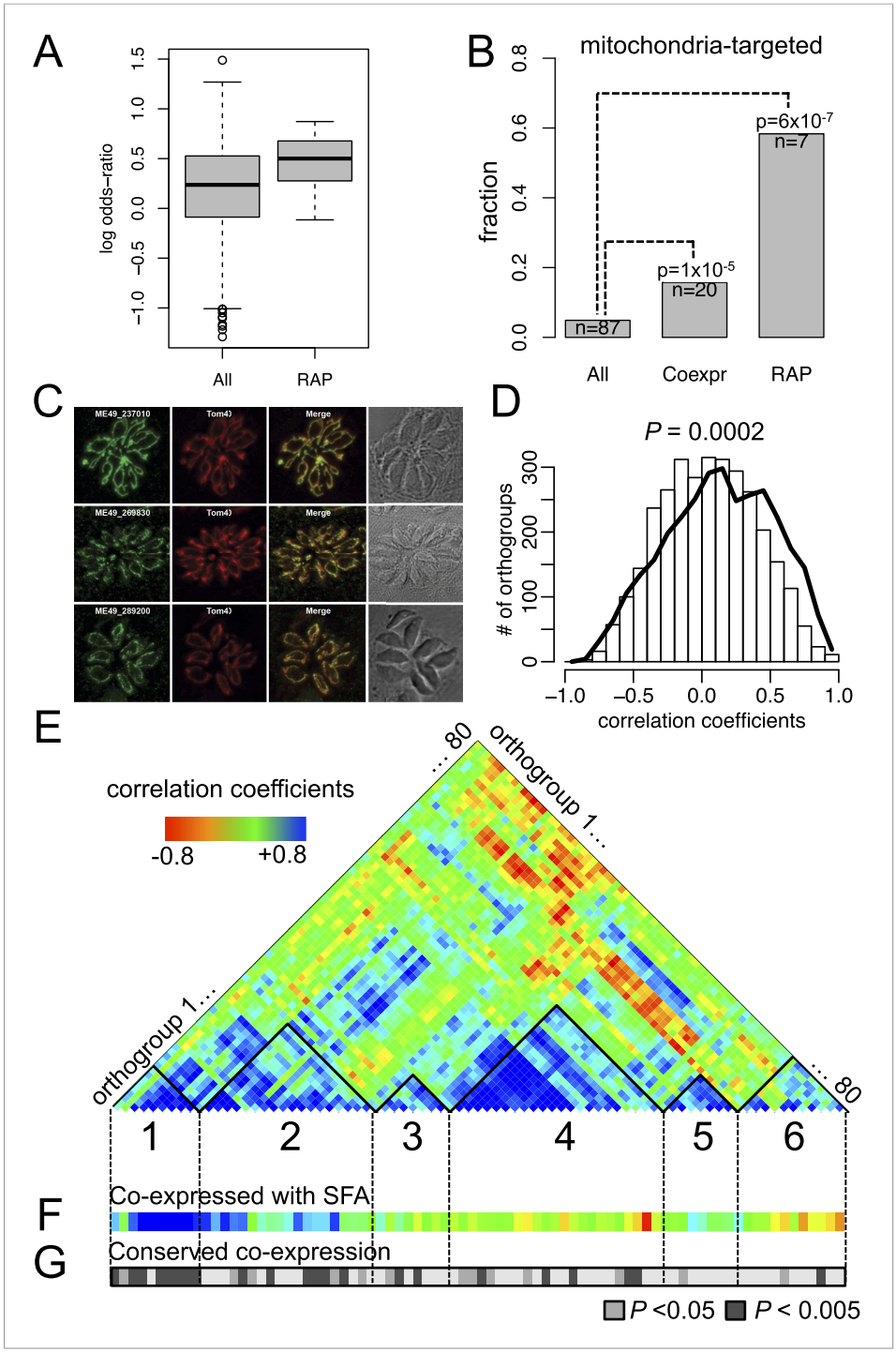

Figure 4. Conserved transcriptional programs in apicomplexans and chromerids. (**A**) Boxplot showing the extent of evolutionary conservation of transcriptional programs for all orthogroups or those with RAP domains. X-axis: 'All' (all orthogroups excluding RAP); 'RAP' (orthogroups with RAP domains). Y-axis: log-transformed odds-ratio, representing, for each orthogroup, the degree of overlap between its co-expressed orthogroups in Chromera and those in *P. falciparum*. (**B**) Bar chart showing the fraction of orthogroups (Y-axis) predicted to be targeted to mitochondria in both species ('Materials and methods'). The number of genes are displayed below each bar. X-axis: 'All' (all orthogroups excluding the other two categories); 'Coexpr' (orthogroups co-expressed with RAP in both species); 'RAP' (orthogroups with RAP domains). The fractions in 'Coexpr' and 'RAP' groups were compared against the fraction in 'All', and p-values based Fisher's exact test are displayed above the bar. Files deposited in European Nucleotide Archive are listed in ***Figure 4—source data 1*** with corresponding conditions. (**C**) Sub-cellular

Figure 4. continued on next page

Figure 4. Continued

localization of RAP proteins encoded by TGME49_237010, TGME49_269830, and TGME49_289200 was tested in *T. gondii* by 3′ tagging of the endogenous genes with the coding sequence for the hemagglutinin epitope, together with a mitochondrial marker Tom40. See ***Supplementary file 6*** for details of the localization predictions.
(**D**) Distributions of Spearman's rank correlation coefficients of gene expression between all possible pairs from the 80 orthogroups implicated in invasion processes in apicomplexans (black outline) were compared against those from 80 randomly selected ones (histogram). The p value indicates statistical significance of the difference based on 10,000 random samplings. The 80 orthogroups and corresponding genes in *Chromera* and *P. falciparum* are listed in ***Figure 4—source data 2***. (**E**) Heatmap showing a matrix of correlation coefficients amongst the 80 orthogroups. Based on a hierarchical clustering, we classified them into six co-expression modules, labeled as numeral 1–6. (**F**) Heatmap showing correlation coefficients with striated fiber assemblin (SFA) (Cvel_872). The color scheme is the same as in (**E**). (**G**) Heatmap indicating statistical significance of conserved transcriptional program, that is, the odds-ratio as defined in (**A**) (Fisher's exact test, $p < 0.05$ (gray); $p < 0.005$ (black)).
DOI: 10.7554/eLife.06974.022

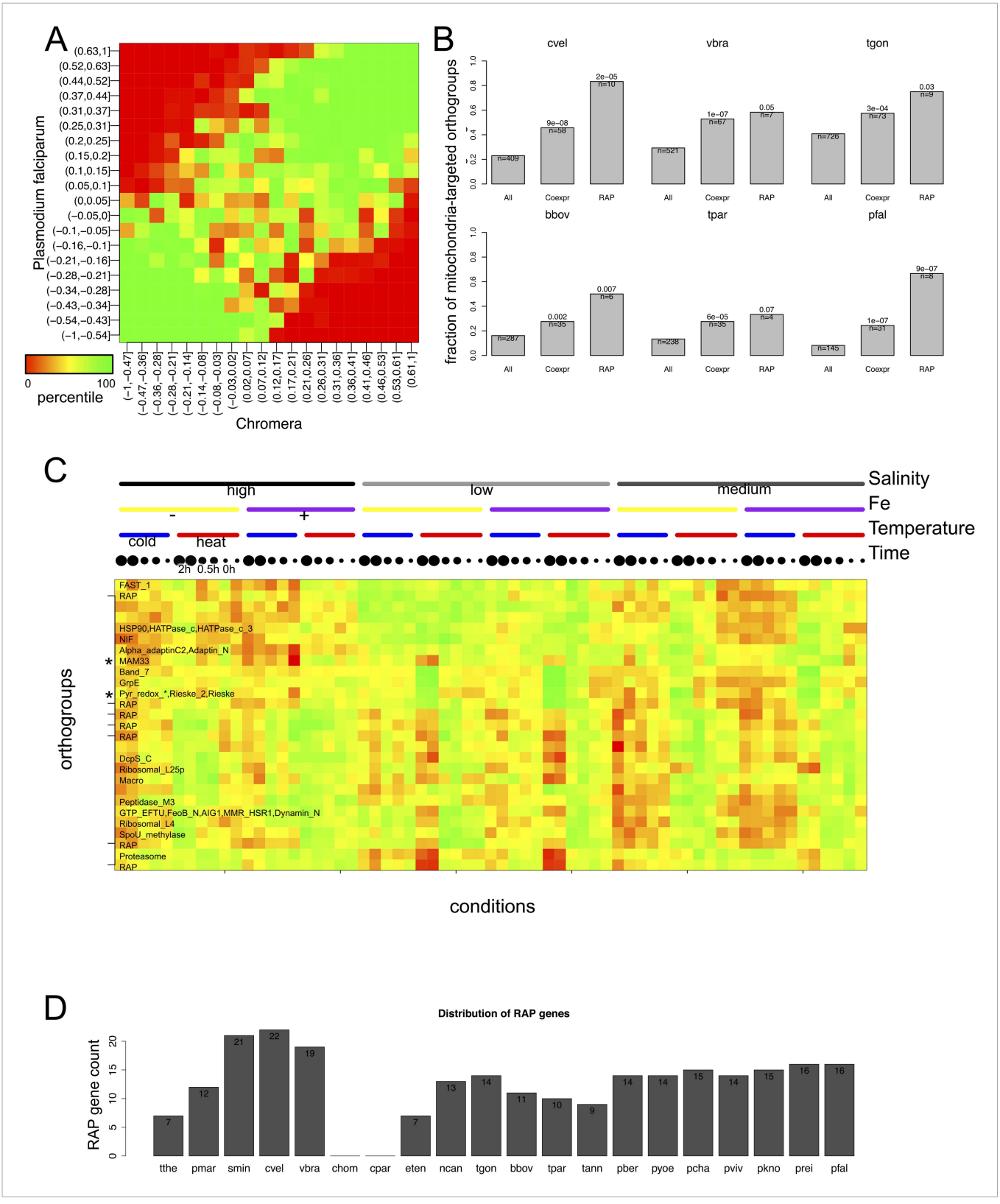

Figure 4—figure supplement 1. Mitochondrial targeting of RAP and its putative role in mitochondrial metabolism. (**A**) Heatmap displaying the extent of association between correlation coefficients of orthogroup-pairs in *P. falciparum* (Y-axis) and those in *Chromera* (X-axis). The color scale represents the

Figure 4—figure supplement 1. continued on next page

Figure 4—figure supplement 1. Continued

percentile of the observed frequency amongst randomly expected frequencies when the orthology were shuffled. We observed high percentiles along the 45˚ diagonal, indicating that the number of orthogroup-pairs that are co-expressed in both species is greater than expected by chance. (**B**) Bar chart showing the fraction of orthogroups (Y-axis) predicted to be targeted to mitochondria in selected species. The orthogroups and the three categories, that is, 'All', 'Coexpr', and 'RAP' (X-asis) are based on those from *Plasmodium* and *Chromera* (***Figure 4B***). (**C**) *Chromera* expression profiles under diverse growth conditions ('Materials and methods') are shown for mitochondria targeted RAPs and co-expressed orthogroups. Expression levels were scaled to have a mean of 0 and a standard deviation of 1. Y-axis: orthogroups ordered based on the hierarchical clustering of their expression patterns. X-axis: combinations of different salt and iron (Fe) concentrations and temperatures in which the *Chromera* cultures were grown. The color scale ranged from red (low expression) to green (high expression). The asterisk (*) denotes genes encoding NADH-dependent oxidoreductase and mitochondrial acidic matrix protein 33, involved in mitochondrial oxidative phosphorylation. (**D**) Abundance of RAP proteins in alveolate species.

DOI: 10.7554/eLife.06974.025