eLIFE elifesciences.org

RESEARCH ARTICLE

Figures and figure supplements

Apicomplexan-like parasites are polyphyletic and widely but selectively dependent on cryptic plastid organelles

Jan Janouškovec *et al*

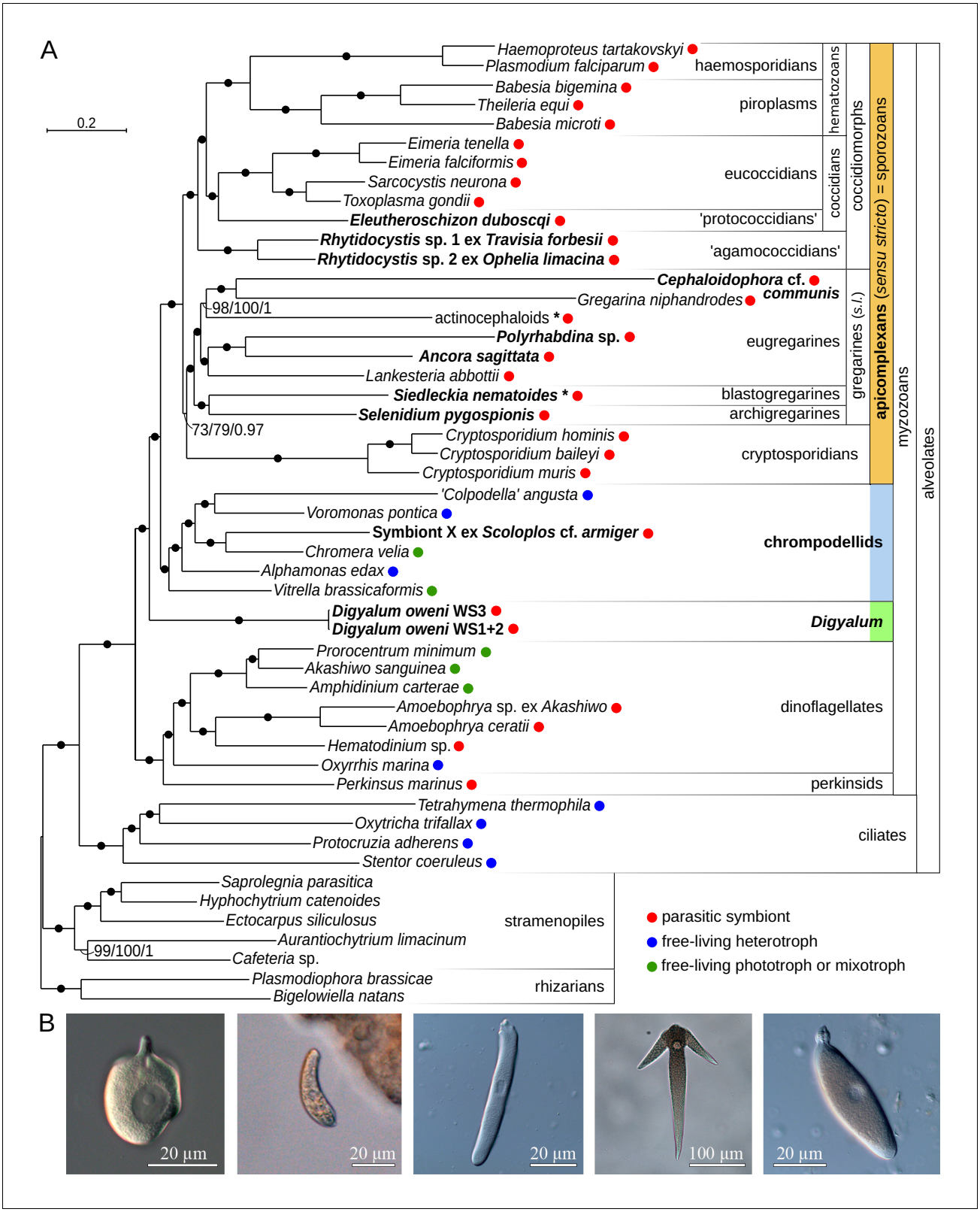


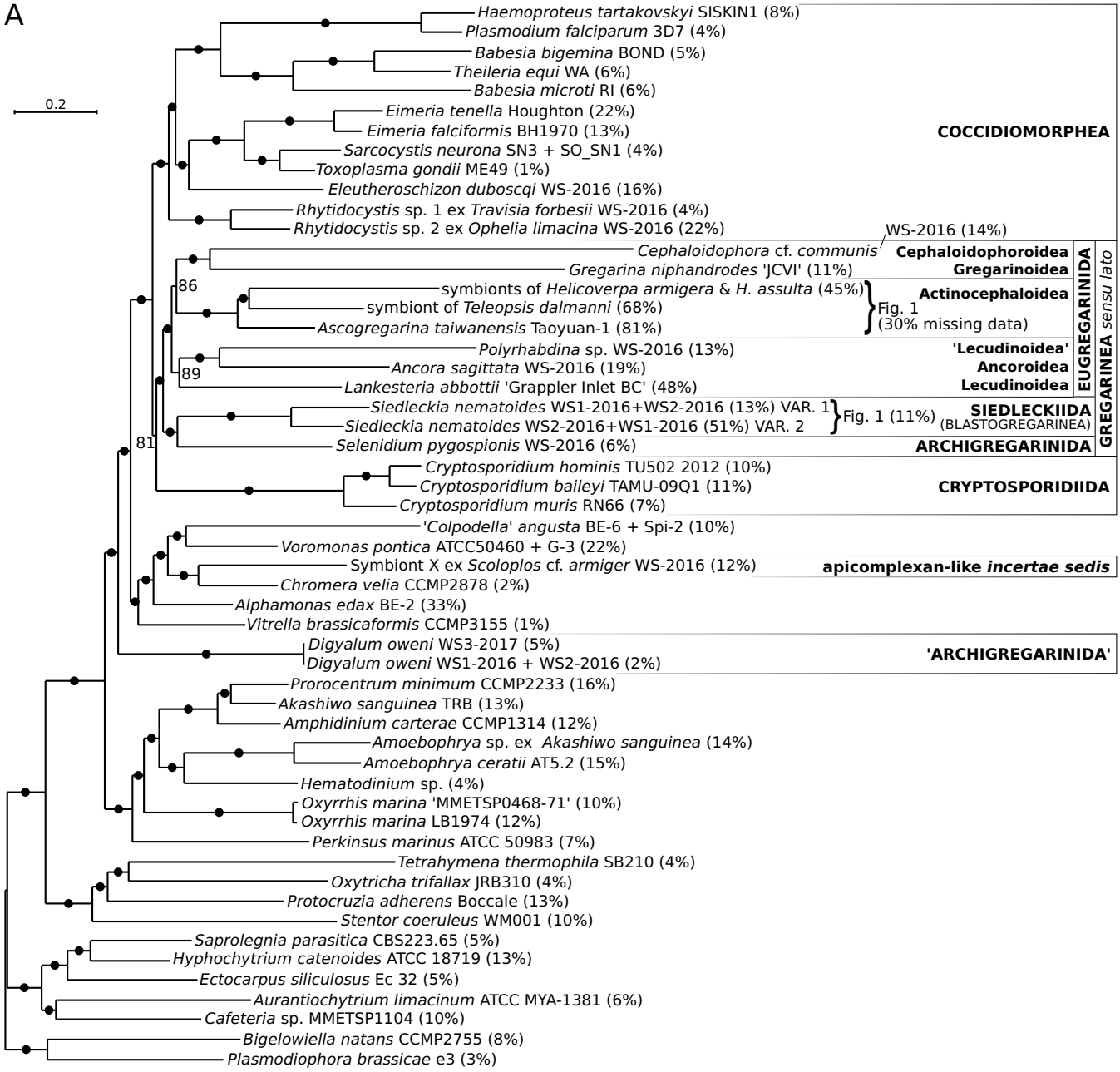
Figure 1. Multiprotein phylogeny of apicomplexans and related taxa. (**A**) Maximum likelihood tree (IQ-TREE) of apicomplexans and their relatives based on 296 concatenated protein markers. Species newly sequenced in this study are in bold. Values at branches correspond to UFBoot2 supports (1000 replicates, LG+G4+F+C60+PMSF model), non-parametric bootstraps (100 replicates, LG+G4+F+C60+PMSF model), and Bayesian posterior probabilities (PhyloBayes, consensus of 10 independent runs, CAT+GTR+G4 model). Black dots indicate 100/100/1 support. Actinocephaloids and

Figure 1 continued on next page

Figure 1 continued

Siedleckia nematoides are hybrid taxa (* symbol) composed from sequences of three parasites and two distant sequence variants, respectively (see ***Figure 1—figure supplement 1***). Values in parentheses behind species names show % of missing data in the phylogenetic matrix. Sequence sources and the phylogenetic matrix are found in ***Supplementary file 2*** and ***Figure 1—source data 1***, respectively. Single quotation marks indicate potentially problematic taxonomic assignments (formal group names are in ***Figure 1—figure supplement 1***). (**B**) Light micrographs of some species studied, left to right: *Digyalum oweni*, Symbiont X, *Selenidium pygospionis*, *Ancora sagittata*, *Polyrhabdina* sp., with the anterior end facing up.

DOI: https://doi.org/10.7554/eLife.49662.003



B

Internal branch \ Dataset (UFBoot2 support)	Fig.1	Fig.1 -*Gre*	Fig.1 -*Cep*	Fig.1 -*Gre*,-*Cep*
Actinocephaloidea + (*Gregarina* + *Cephaloidophora*)	100	98	100	n.a.
Lankesteria + (*Ancora* + *Polyrhabdina*)	100	98	100	100
Cryptosporidiida + Gregarinea *s.l.*	79	77	78	65
All other	100	100	100	100

Figure 1—figure supplement 1. Multiprotein phylogenies of apicomplexans and related taxa. (**A**) Maximum likelihood tree (IQ-TREE) of apicomplexans and their relatives (54 species) based on 296 concatenated protein markers (99948 sites, 14.4% missing data). Values at branches correspond to UFBoot2 supports, black dots indicate 100 support (1000 replicates, LG+G4+F+C60+PMSF model). Sequences of three actinocephaloids and of the two distinct sequence variants in *Siedleckia nematoides* were merged in the ***Figure 1A*** dataset, as shown. Values in parentheses behind species names show % of missing data in the final phylogenetic matrix. Single quotation marks indicate potentially problematic taxonomic assignments.

Figure 1—figure supplement 1 continued on next page

Figure 1—figure supplement 1 continued

Sequence sources are in ***Supplementary file 2***. (**B**) Maximum likelihood phylogenies on datasets that excluded the two fastest evolving taxa, *Gregarina* and *Cephaloidophora*, either individually or together (IQ-TREE). UFBoot2 supports are shown (LG+G4+F+C60+PMSF model); note that all trees were topologically congruent.

DOI: https://doi.org/10.7554/eLife.49662.004

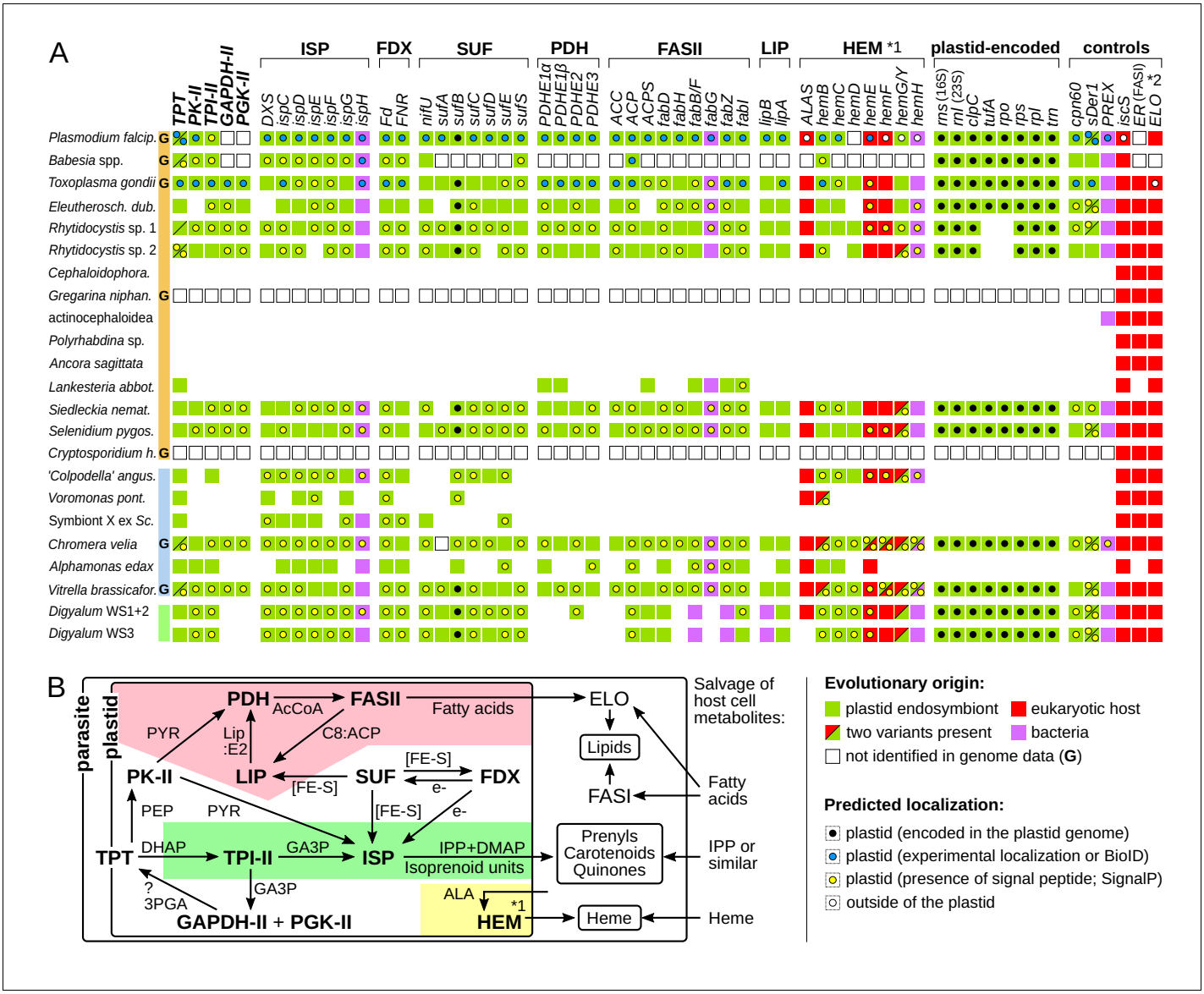


Figure 2. Core plastid metabolism in apicomplexans and their relatives. (**A**) Presence of genes and pathway modules (top; abbreviations in ***Supplementary file 4***) in representative genomes (G) and transcriptomes (left. Each gene (box) is color-coded as to its evolutionary origin, as determined by a maximum likelihood phylogeny (plastid-encoded *rps*, *rpl*, *rpo*, *trn* genes were not analyzed). Empty boxes indicate gene absence in completed genomes and blank spaces indicate absence in transcriptomes. Intracellular localization of corresponding proteins is shown by a circle inside the box and summarizes known experimental data (***Supplementary file 4***) or de novo prediction in silico by SignalP v4.1 (***Supplementary file 5***); it is missing in proteins with incomplete N-termini. Note that only some enzymes of the heme pathway (HEM) are localized in the plastid (*1) and that signal peptides in FAS:ER and ELO were not predicted (*2). (**B**) Dependence network of plastid protein modules for the biosynthesis of key metabolites – isoprenoid precursors IPP and DMAP, fatty acids and heme – which underlie dependency on the plastid organelle in Apicomplexa. Colored regions contain modules specific to one pathway: fatty acid (pink), isoprenoid precursor (pale green) and heme biosynthesis (yellow). Interactions are reconstructed from the literature and substrates are shown near arrows (PYR = pyruvate; AcCoA = acetyl coenzyme A, Lip:E2 = lipoylation on PDHE2; C8:ACP = octanoyl:acyl carrier protein; [FE-S]=iron sulphur cluster; PEP = phosphoenolpyruvate; e-=electron reductive power; GA3p=glyceraldehyde-3-phosphate; 3PGA = 3 phosphoglycerate; ALA = δ-aminolevulinic acid; ?=uncertainty).

DOI: https://doi.org/10.7554/eLife.49662.006

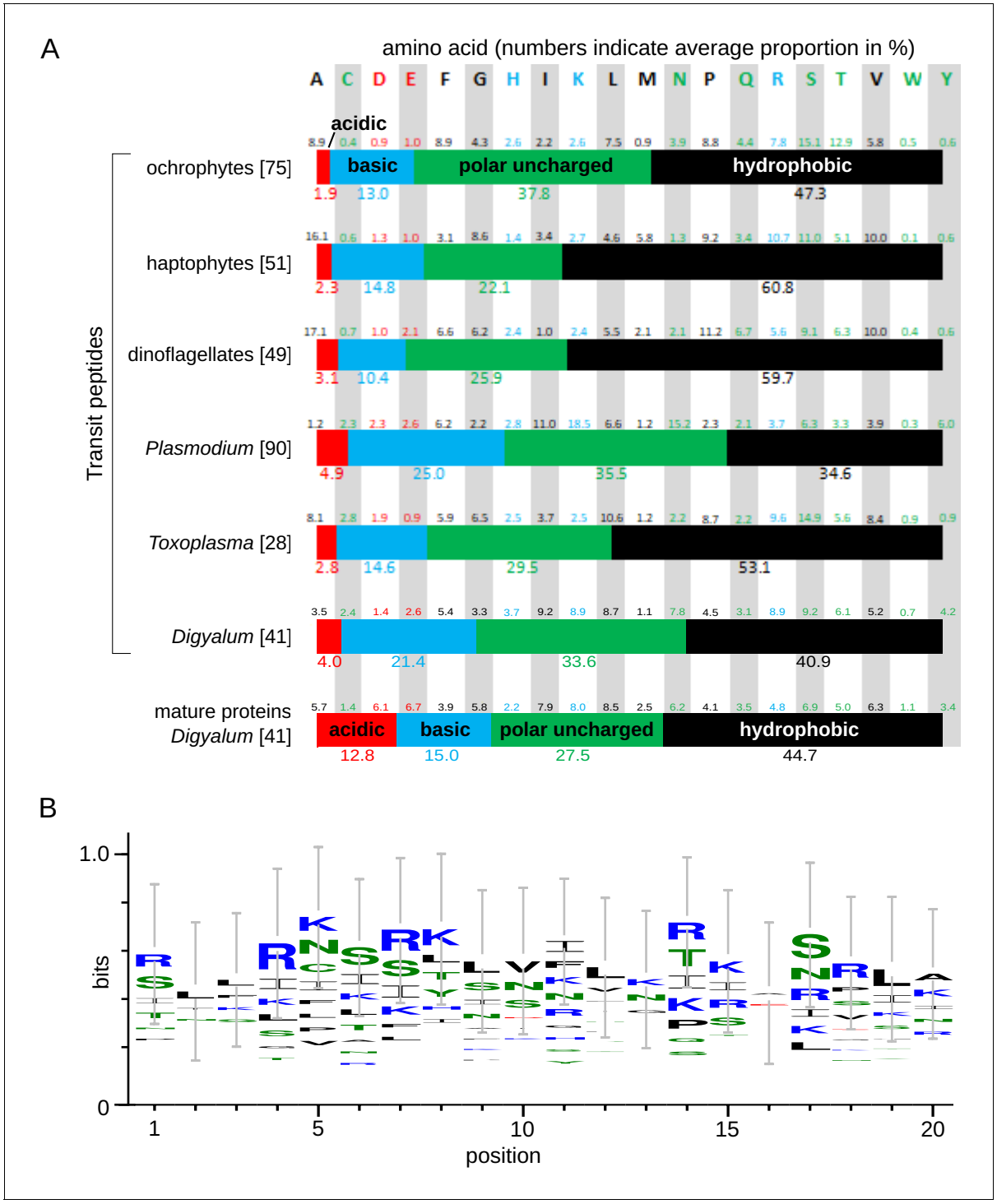


Figure 2—figure supplement 1. Transit peptides in *Digyalum* plastid proteins have positive charge but lack a conserved phenylalanine in the first position. (**A**) Amino acid composition of plastid transit peptides in *Digyalum* in comparison with downstream sequences of mature plastid proteins and transit peptides in other complex plastids as reported by ***Patron and Waller (2007)*** (only the first 14 residues were analyzed to enable consistent comparison) (***Patron and Waller, 2007***). Average proportions of individual amino acids and their classes based on charge (colored bars) are shown: acidic (D + E) in red, basic (H + K + R) in blue, polar uncharged (C + N + Q + S + T + W + Y) in green and hydrophobic (A + F + G + I + L + M + P + V) in black. Number of proteins analyzed are indicated in square brackets behind species/group names. (**B**) WebLogo3 diagram of the relative composition of the first 20 amino acids in 41 target peptides of *Digyalum oweni*. Letter size indicates the probability of an amino acid occurring at a specific position, weighted to account for the relative abundance of each amino acids natural occurrence and unrepresentative results arising from a small sample size.

DOI: https://doi.org/10.7554/eLife.49662.007

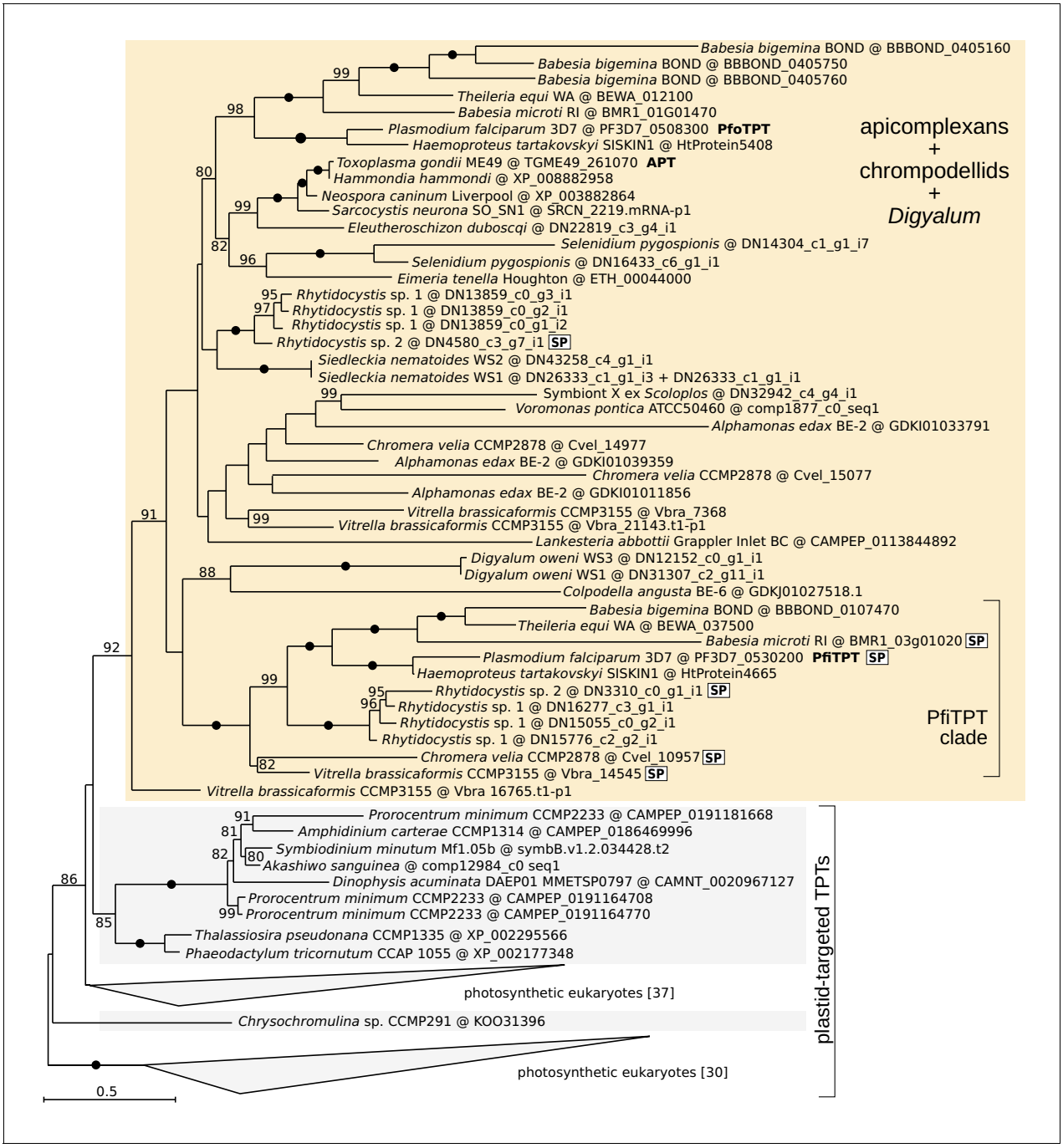


Figure 2—figure supplement 2. Maximum likelihood phylogeny of the triose phosphate translocator, TPT (LG+F+R7 model in IQ-TREE with 10000 UFBoot2 replicates;≥80 are shown; black dots indicate full support). Apicomplexans, chrompodellids and *Digyalum* are shown on orange background and other eukaryotes on gray background. Selected characterized proteins are highlighted in bold.

DOI: https://doi.org/10.7554/eLife.49662.008

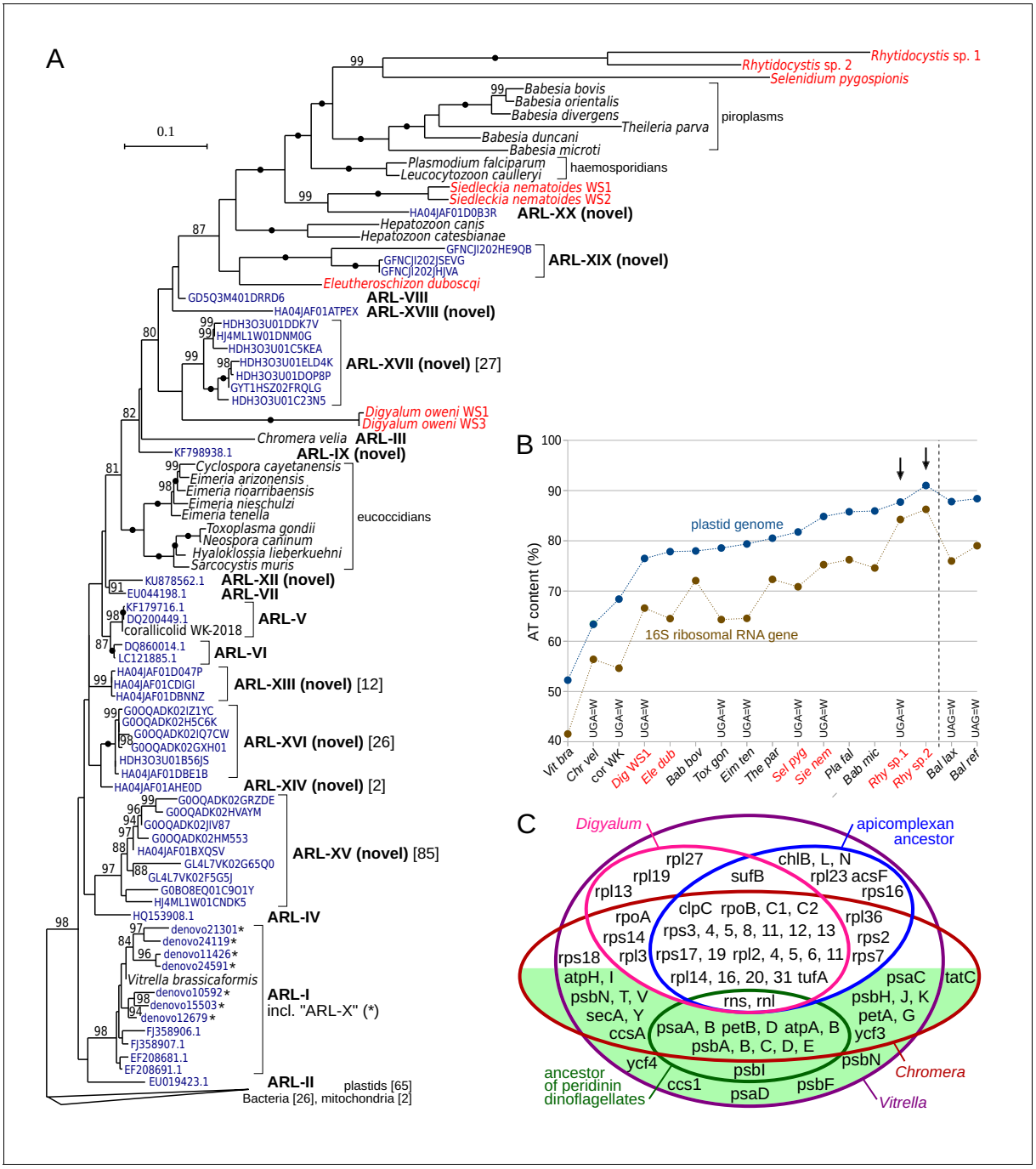


Figure 3. Plastid genomes in apicomplexans and their relatives are widespread and highly divergent. (**A**) Maximum likelihood phylogeny of plastid-encoded 16S ribosomal RNA genes (rDNA) reveals 10 novel apicomplexan-related lineages (ARL-IX and ARL-XII to ARL-XX; ***Supplementary file 6***). Tree was computed with the best-fit TVMe+R5 model in IQ-TREE with UFBoot2 supports at branches (10000 replicates;≥80 are shown; black dots indicate 100 support). Environmental sequences (dark blue) are derived from GenBank or VAMPS (97% identity cluster centroids; numbers of reads are shown in square brackets where > 1). Plastid 16S rDNA transcripts of newly sequenced species are shown in red. Note that sequences in the tree vary greatly in their AT content and substitution rates, which can induce a misleading topology - deep relationships in the tree should therefore be interpreted with caution. The fast-evolving sequences of peridinin dinoflagellates were not included. (**B**) Extremely high AT content in rhytidocystid plastid genomes (arrows). AT content of representative species from part A and *Balanophora laxiflora* and *B. reflexa* parasitic plants (***Su et al., 2019***), all abbreviated to first three letters, is shown for 16S rDNA and plastid genomes. Plastid genomes in the newly sequenced species are only partially reconstructed from transcripts (red color; see ***Supplementary file 7***). Altered genetic codes are indicated. (**C**) Euler diagram of plastid genome contents in apicomplexans, dinoflagellates (ancestral gene sets for each), *Digyalum*, *Chromera*, and *Vitrella*. Genes on the green background are associated solely with photosynthesis. Small RNA genes are not shown.

DOI: https://doi.org/10.7554/eLife.49662.009

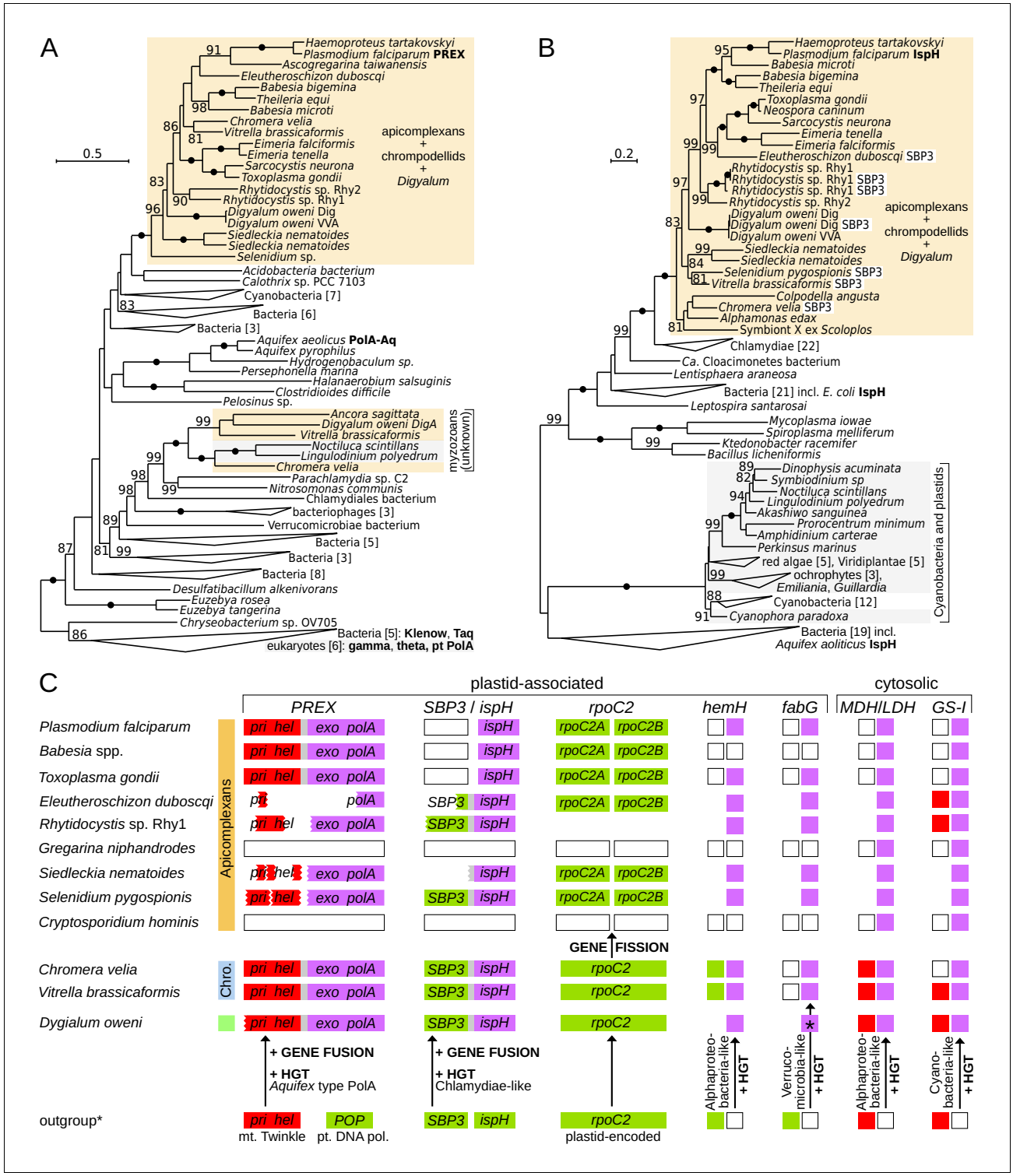


Figure 4. Innovations in plastid and cytosolic genes define early evolution of apicomplexans and their relatives. (**A**) Maximum likelihood phylogeny of the exonuclease/polymerase subunit of the plastid replication and repair complex, PREX (LG+R6 model). (**B**) Maximum likelihood phylogeny of 4-hydroxy-3-methylbut-2-enyl diphosphate reductase, IspH (LG+I+G4 model). Trees were derived from protein sequences in IQ-TREE and have UFBoot2 supports at branches (10000 replicates;≥80 are shown; black dots indicate 100 support). Apicomplexans, chrompodellids and *Digyalum* are highlighted in orange, and other eukaryotes in gray. Characterized enzymes are highlighted in bold. Fusion *ispH* genes with *SBP3* at the N-terminus are shown in white boxes. (**C**) Predicted gain, loss, fusion and fission events in the evolution of five plastid-associated and two cytosolic genes. Genes are shown by boxes (jagged edges indicate truncated genes) and are color-coded by origin as in ***Figure 2A*** (plastid endosymbiont = green, eukaryotic host = red,

Figure 4 continued on next page

Figure 4 continued

bacteria = purple, absent in genome data = white, absence of evidence = blank area). Outgroup (*) shows the state in the closest relevant comparator: dinoflagellates (IspH, HemH, MDH/LDH, GS-I) or other algae (PREX, SBP3, RpoC2, FabG). Note that the ferrochelatase (HemH) is mitochondrial in *Plasmodium* but probably plastidial in *Chromera* and some apicomplexans. Abbreviations: POP = plant organellar DNA polymerase, Twinkle = mitochondrial primase/helicase, SBP3 = sedoheptulose-1,7-bisphosphatase form 3, MDH/LDH = malate/lactate dehydrogenase (other abbreviations in ***Supplementary file 3***).

DOI: https://doi.org/10.7554/eLife.49662.010

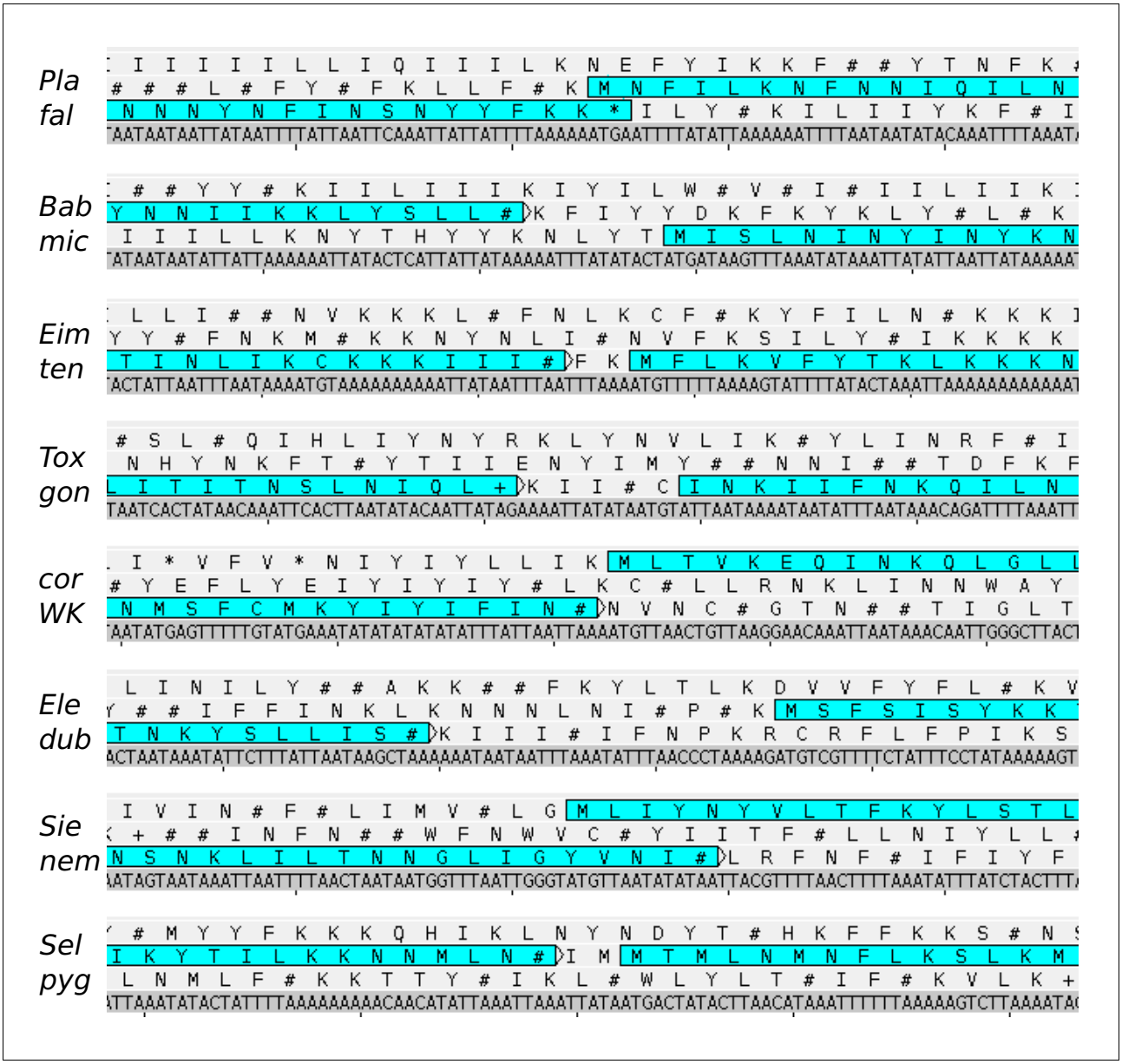


Figure 4—figure supplement 1. The region of the split in the apicomplexan plastid-encoded *rpoC2*. Plastid genome sequences and their three-frame amino acid translations are shown for representative apicomplexan species (snapshots from Artemis 17.0.1). Coding sequences are highlighted by horizontal rectangles. STOP codons are shown by # (=TAA), + (=TAG), and * (=TGA) symbols. In some taxa, TGA encodes for tryptophan (W). Note that the split occurs in a non-conserved part of *rpoC2* so the corresponding region cannot be shown for a non-apicomplexan outgroup - *rpoC2* genes in chrompodellids, *Digyalum* and all other plastids nevertheless lack the split.

DOI: https://doi.org/10.7554/eLife.49662.011

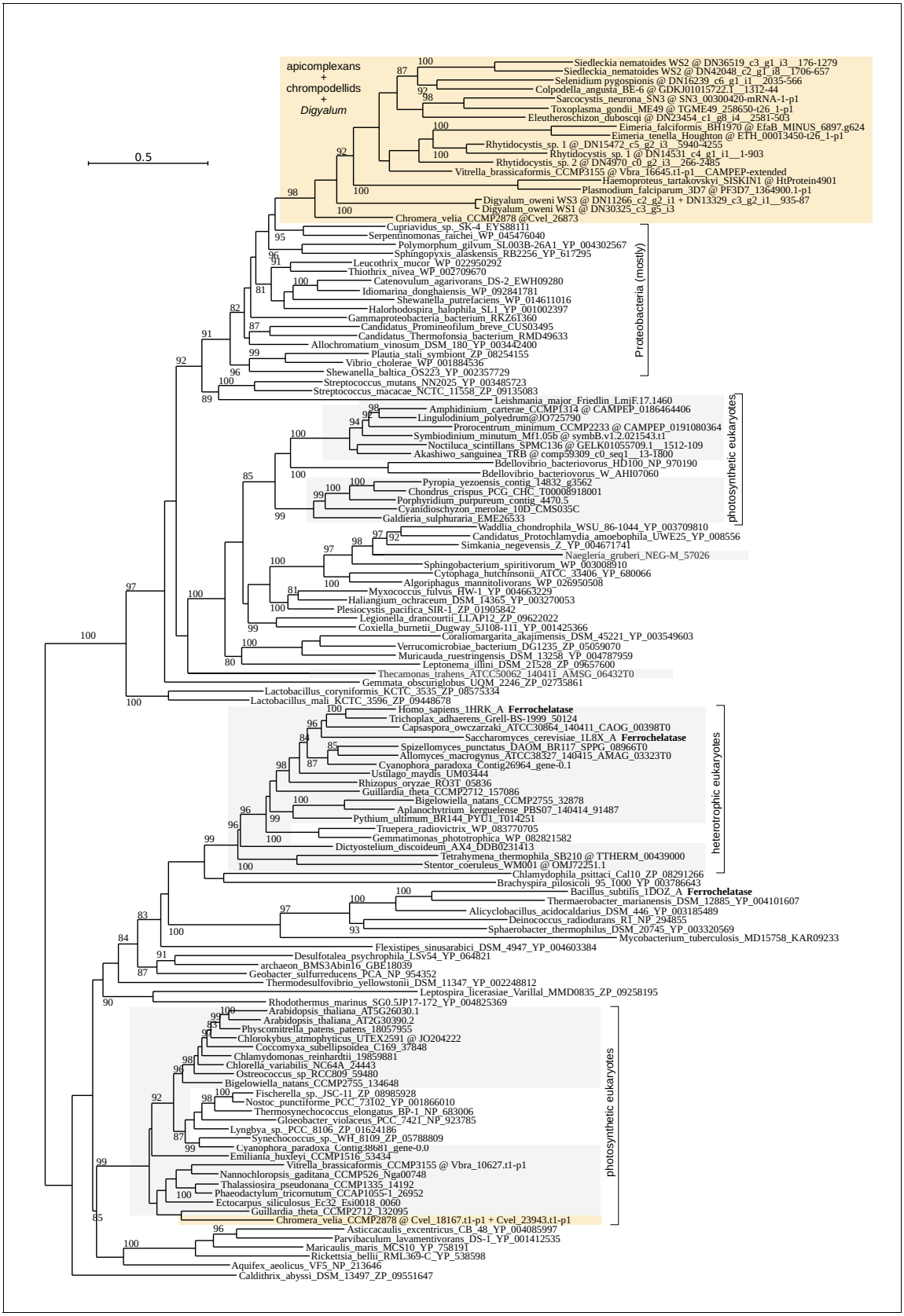

apicomplexans
+
chrompodellids
+
Digyalum
0.5
Siedleckia nematoides WS2 @ DN36519_c3_g1_i3__176-1279
Siedleckia_nematoides WS2 @ DN42048_c2_g1_i8__1706-657
Selenidium pygospionis @ DN16239_c6_g1_i1__2035-566
Colpodella_angusta_BE-6 @ GDKJ01015722.1__1312-44
Sarcocystis_neurona_SN3 @ SN3_00300420-mRNA-1-p1
Toxoplasma_gondi_ME49 @ TGME49_258650-t26_1-p1
Eleutheroschizon_duboscqi @ DN23454_c1_g8_i4__2581-503
Eimeria_falciformis_BH1970 @ EfaB_MINUS_6897.g624
Eimeria_tenella_Houghton @ ETH_00013450-t26_1-p1
Rhytidocystis_sp. 1 @_DN15472_c5_g2_i3__5940-4255
Rhytidocystis_sp. 1 @ DN14531_c4_g1_i1__1-903
Rhytidocystis_sp. 2 @ DN4970_c0_g2_i3__266-2485
Vitrella_brassicaformis_CCMP3155 @ Vbra_16645.t1-p1__CAMPEP-extended
Haemoproteus_tartakovskyi_SISKIN1 @ HtProtein4901
Plasmodium_falciparum_3D7 @ PF3D7_1364900.1-p1
Digyalum_oweni WS3 @ DN11266_c2_g2_i1 + DN13329_c3_g2_i1__935-87
Digyalum_oweni WS1 @ DN30325_c3_g5_i3
Chromera_velia_CCMP2878 @Cvel_26873
Cupriavidus_sp._SK-4_EYS88111
Serpentinomonas_raichei_WP_045476040
Polymorphum_gilvum_SL003B-26A1_YP_004302567
Sphingopyxis_alaskensis_RB2256_YP_617295
Leucothrix_mucor_WP_022950292
Thiothrix_nivea_WP_002709670
Catenovulum_agarivorans_DS-2_EWH09280
Idiomarina_donghaiensis_WP_092841781
Shewanella_putrefaciens_WP_014611016
Halorhodospira_halophila_SL1_YP_001002397
Gammaproteobacteria_bacterium_RKZ61360
Candidatus_Promineofilum_breve_CUS03495
Candidatus_Thermofonsia_bacterium_RMD49633
Allochromatium_vinosum_DSM_180_YP_003442400
Plautia_stali_symbiont_ZP_08254155
Vibrio_cholerae_WP_001884536
Shewanella_baltica_OS223_YP_002357729
Streptococcus_mutans_NN2025_YP_003485723
Streptococcus_macacae_NCTC_11558_ZP_09135083
Proteobacteria (mostly)
Leishmania_major_Friedlin_LmjF.17.1460
Amphidinium_carterae_CCMP1314 @ CAMPEP_0186464406
Lingulodinium_polyedrum@JO725790
Prorocentrum_minimum_CCMP2233 @ CAMPEP_0191080364
Symbiodinium_minutum_Mf1.05b @ symbB.v1.2.021543.t1
Noctiluca_scintillans_SPMC136 @ GELK01055709.1__1512-109
Akashiwo_sanguinea_TRB @ comp59309_c0_seq1__13-1800
Bdellovibrio_bacteriovorus_HD100_NP_970190
Bdellovibrio_bacteriovorus_W_AHI07060
Pyropia_yezoensis_contig_14832_g3562
Chondrus_crispus_PCG_CHC_T000089180001
Porphyridium_purpureum_contig_4470.5
Cyanidioschyzon_merolae_10D_CMS035C
Galdieria_sulphuraria_EME26533
photosynthetic eukaryotes
Waddlia_chondrophila_WSU_86-1044_YP_003709810
Candidatus_Protochlamydia_amoebophila_UWE25_YP_008556
Simkania_negevensis_Z_YP_004671741
Naegleria_gruberi_NEG-M_57026
Sphingobacterium_spiritivorum_WP_003008910
Cytophaga_hutchinsonii_ATCC_33406_YP_680066
Algoriphagus_mannitolivorans_WP_026950508
Myxococcus_fulvus_HW-1_YP_004663229
Haliangium_ochraceum_DSM_14365_YP_003270053
Plesiocystis_pacifica_SIR-1_ZP_01905842
Legionella_drancourtii_LLAP12_ZP_09622022
Coxiella_burnetii_Dugway_5J108-111_YP_001425366
Coraliomargarita_akajimensis_DSM_45221_YP_003549603
Verrucomicrobiae_bacterium_DG1235_ZP_05059070
Muricauda_ruestringensis_DSM_13258_YP_004787959
Leptonema_illini_DSM_21528_ZP_09657600
Thecamonas_trahens_ATCC50062_140411_AMSG_06432T0
Gemmata_obscuriglobus_UQM_2246_ZP_02735861
Lactobacillus_coryniformis_KCTC_3535_ZP_08575334
Lactobacillus_mali_KCTC_3596_ZP_09448678
Homo_sapiens_1HRK_A **Ferrochelatase**
Trichoplax_adhaerens_Grell-BS-1999_50124
Capsaspora_owczarzaki_ATCC30864_140411_CAOG_00398T0
Saccharomyces_cerevisiae_1L8X_A **Ferrochelatase**
Spizellomyces_punctatus_DAOM_BR117_SPPG_08966T0
Allomyces_macrogynus_ATCC38327_140415_AMAG_03323T0
Cyanophora_paradoxa_Contig26964_gene-0.1
Ustilago_maydis_UM03444
Rhizopus_oryzae_RO3T_05836
Guillardia_theta_CCMP2712_157086
Bigelowiella_natans_CCMP2755_32878
Aplanochytrium_kerguelense_PBS07_140414_91487
Pythium_ultimum_BR144_PYU1_T014251
Truepera_radiovictrix_WP_083770705
Gemmatimonas_phototrophica_WP_082821582
Dictyostelium_discoideum_AX4_DDB0231413
Tetrahymena_thermophila_SB210 @ TTHERM_00439000
Stentor_coeruleus_WM001 @ OMJ72251.1
heterotrophic eukaryotes
Chlamydophila_psittaci_Cal10_ZP_08291266
Brachyspira_pilosicoli_95_1000_YP_003786643
Bacillus_subtilis_1DOZ_A **Ferrochelatase**
Thermaerobacter_marianensis_DSM_12885_YP_004101607
Alicyclobacillus_acidocaldarius_DSM_446_YP_003185489
Deinococcus_radiodurans_RT_NP_294855
Sphaerobacter_thermophilus_DSM_20745_YP_003320569
Mycobacterium_tuberculosis_MD15758_KAR09233
Flexistipes_sinusarabici_DSM_4947_YP_004603384
Desulfotalea_psychrophila_LSv54_YP_064821
archaeon_BMS3Abin16_GBE18039
Geobacter_sulfurreducens_PCA_NP_954352
Thermodesulfovibrio_yellowstonii_DSM_11347_YP_002248812
Leptospira_licerasiae_Varillal_MMD0835_ZP_09258195
Rhodothermus_marinus_SG0.5JP17-172_YP_004825369
Arabidopsis_thaliana_AT5G26030.1
Arabidopsis_thaliana_AT2G30390.2
Physcomitrella_patens_patens_18057955
Chlorokybus_atmophyticus_UTEX2591 @ JO204222
Coccomyxa_subellipsoidea_C169_37848
Chlamydomonas_reinhardtii_19859881
Chlorella_variabilis_NC64A_24443
Ostreococcus_sp_RCC809_59480
Bigelowiella_natans_CCMP2755_134648
Fischerella_sp._JSC-11_ZP_08985928
Nostoc_punctiforme_PCC_73102_YP_001866010
Thermosynechococcus_elongatus_BP-1_NP_683006
Gloeobacter_violaceus_PCC_7421_NP_923785
Lyngbya_sp._PCC_8106_ZP_01624186
Synechococcus_sp._WH_8109_ZP_05788809
Cyanophora_paradoxa_Contig38681_gene-0.0
Emiliania_huxleyi_CCMP1516_53434
Vitrella_brassicaformis_CCMP3155 @ Vbra_10627.t1-p1
Nannochloropsis_gaditana_CCMP526_Nga00748
Thalassiosira_pseudonana_CCMP1335_14192
Phaeodactylum_tricornutum_CCAP1055-1_26952
Ectocarpus_siliculosus_Ec32_Esi0018_0060
Guillardia_theta_CCMP2712_132095
Chromera_velia_CCMP2878 @ Cvel_18167.t1-p1 + Cvel_23943.t1-p1
photosynthetic eukaryotes
Asticcacaulis_excentricus_CB_48_YP_004085997
Parvibaculum_lavamentivorans_DS-1_YP_001412535
Maricaulis_maris_MCS10_YP_758191
Rickettsia_bellii_RML369-C_YP_538598
Aquifex_aeolicus_VF5_NP_213646
Caldithrix_abyssi_DSM_13497_ZP_09551647

Figure 4—figure supplement 2. Maximum likelihood phylogeny of the ferrochelatase, HemH (LG+R8 model in IQ-TREE with 10000 UFBoot2 replicates;≥80 are shown). Apicomplexans, chrompodellids and *Digyalum* are shown on orange background and other eukaryotes on gray background. Selected characterized enzymes are highlighted in bold.

DOI: https://doi.org/10.7554/eLife.49662.012

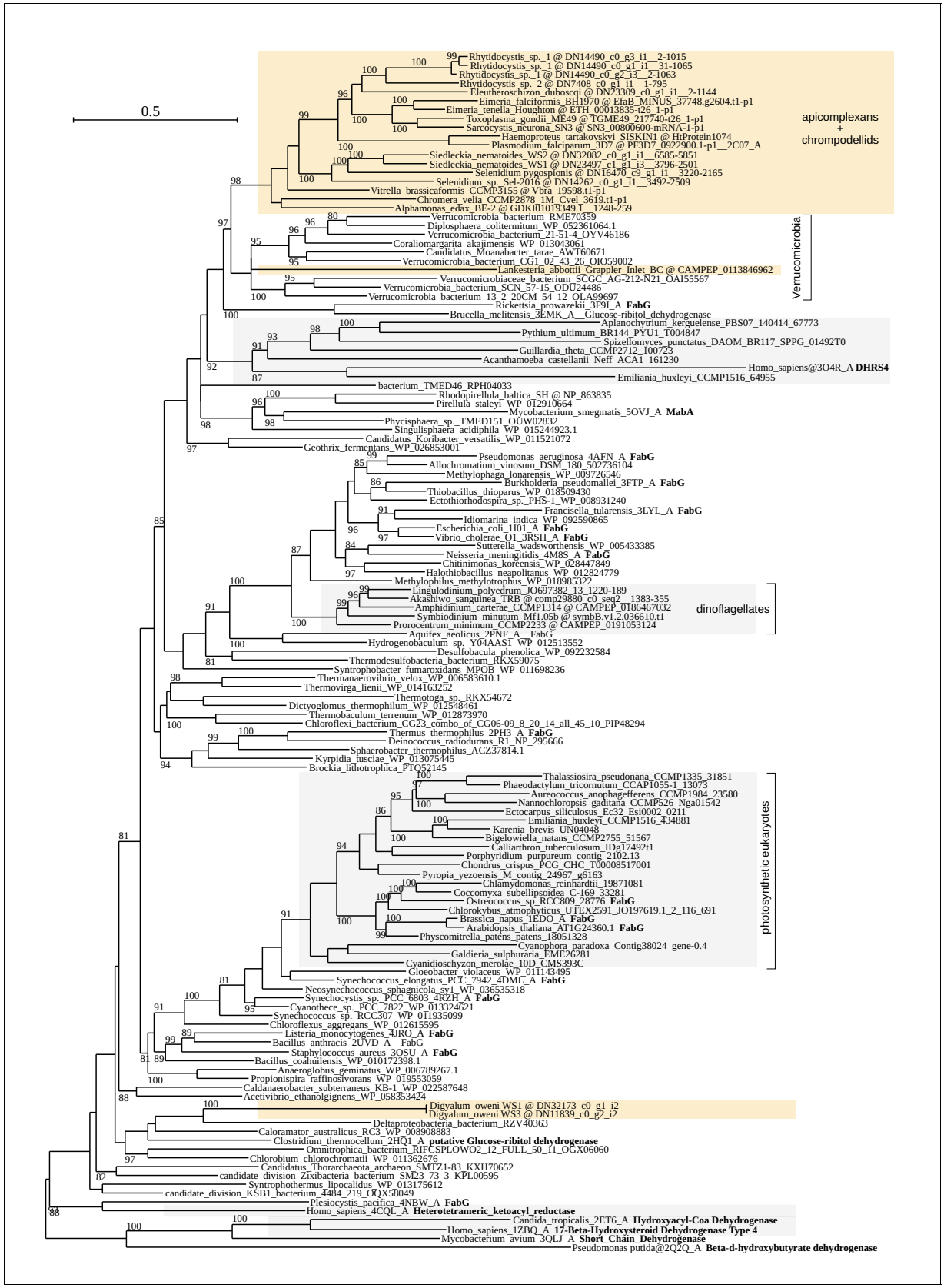

0.5
apicomplexans
+
chrompodellids
Verrucomicrobia
dinoflagellates
photosynthetic eukaryotes
Rhytidocystis_sp._1 @ DN14490_c0_g3_i1__2-1015
Rhytidocystis_sp._1 @ DN14490_c0_g1_i1__31-1065
Rhytidocystis_sp._1 @ DN14490_c0_g2_i3__2-1063
Rhytidocystis_sp._2 @ DN7408_c0_g1_i1__1-795
Eleutheroschizon_duboscqi @ DN23309_c0_g1_i1__2-1144
Eimeria_falciformis_BH1970 @ EfaB_MINUS_37748.g2604.t1-p1
Eimeria_tenella_Houghton @ ETH_00013835-t26_1-p1
Toxoplasma_gondii_ME49 @ TGME49_217740-t26_1-p1
Sarcocystis_neurona_SN3 @ SN3_00800600-mRNA-1-p1
Haemoproteus_tartakovskyi_SISKIN1 @ HtProtein1074
Plasmodium_falciparum_3D7 @ PF3D7_0922900.1-p1__2C07_A
Siedleckia_nematoides_WS2 @ DN32082_c0_g1_i1__6585-5851
Siedleckia_nematoides_WS1 @ DN23497_c1_g1_i3__3796-2501
Selenidium_pygospionis @ DN16470_c9_g1_i1__3220-2165
Selenidium_sp._Sel-2016 @ DN14262_c0_g1_i1__3492-2509
Vitrella_brassicaformis_CCMP3155 @ Vbra_19598.t1-p1
Chromera_velia_CCMP2878_1M_Cvel_3619.t1-p1
Alphamonas_edax_BE-2 @ GDKI01019349.1__1248-259
Verrucomicrobia_bacterium_RME70359
Diplosphaera_colitermitum_WP_052361064.1
Verrucomicrobia_bacterium_21-51-4_OYV46186
Coraliomargarita_akajimensis_WP_013043061
Candidatus_Moanabacter_tarae_AWT60671
Verrucomicrobia_bacterium_CG1_02_43_26_OIO59002
Lankesteria_abbottii_Grappler_Inlet_BC @ CAMPEP_0113846962
Verrucomicrobiaceae_bacterium_SCGC_AG-212-N21_OAI55567
Verrucomicrobia_bacterium_SCN_57-15_ODU24486
Verrucomicrobia_bacterium_13_2_20CM_54_12_OLA99697
Rickettsia_prowazekii_3F9I_A FabG
Brucella_melitensis_3EMK_A__Glucose-ribitol_dehydrogenase
Aplanochytrium_kerguelense_PBS07_140414_67773
Pythium_ultimum_BR144_PYU1_T004847
Spizellomyces_punctatus_DAOM_BR117_SPPG_01492T0
Guillardia_theta_CCMP2712_100723
Acanthamoeba_castellanii_Neff_ACA1_161230
Homo_sapiens@3O4R_A DHRS4
Emiliania_huxleyi_CCMP1516_64955
bacterium_TMED46_RPH04033
Rhodopirellula_baltica_SH @ NP_863835
Pirellula_staleyi_WP_012910664
Mycobacterium_smegmatis_5OVJ_A MabA
Phycisphaera_sp._TMED151_OUW02832
Singulisphaera_acidiphila_WP_015244923.1
Candidatus_Koribacter_versatilis_WP_011521072
Geothrix_fermentans_WP_026853001
Pseudomonas_aeruginosa_4AFN_A FabG
Allochromatium_vinosum_DSM_180_502736104
Methylophaga_lonarensis_WP_009726546
Burkholderia_pseudomallei_3FTP_A FabG
Thiobacillus_thioparus_WP_018509430
Ectothiorhodospira_sp._PHS-1_WP_008931240
Francisella_tularensis_3LYL_A FabG
Idiomarina_indica_WP_092590865
Escherichia_coli_TI01_A FabG
Vibrio_cholerae_O1_3RSH_A FabG
Sutterella_wadsworthensis_WP_005433385
Neisseria_meningitidis_4M8S_A FabG
Chitinimonas_koreensis_WP_028447849
Halothiobacillus_neapolitanus_WP_012824779
Methylophilus_methylotrophus_WP_018985322
Linguodinium_polyedrum_JO697382_13_1220-189
Akashiwo_sanguinea_TRB @ comp29880_c0_seq2__1383-355
Amphidinium_carterae_CCMP1314 @ CAMPEP_0186467032
Symbiodinium_minutum_Mf1.05b @ symbB.v1.2.036610.t1
Prorocentrum_minimum_CCMP2233 @ CAMPEP_0191053124
Aquifex_aeolicus_2PNF_A FabG
Hydrogenobaculum_sp._Y04AAS1_WP_012513552
Desulfobacula_phenolica_WP_092232584
Thermodesulfobacteria_bacterium_RKX59075
Syntrophobacter_fumaroxidans_MPOB_WP_011698236
Thermanaerovibrio_velox_WP_006583610.1
Thermovirga_lienii_WP_014163252
Thermotoga_sp._RKX54672
Dictyoglomus_thermophilum_WP_012548461
Thermobaculum_terrenum_WP_012873970
Chloroflexi_bacterium_CG23_combo_of_CG06-09_8_20_14_all_45_10_PIP48294
Thermus_thermophilus_2PH2_A FabG
Deinococcus_radiodurans_R1_NP_295666
Sphaerobacter_thermophilus_ACZ37814.1
Kyrpidia_tusciae_WP_013075445
Brockia_lithotrophica_PTQ52145
Thalassiosira_pseudonana_CCMP1335_31851
Phaeodactylum_tricornutum_CCAPI055-1_13073
Aureococcus_anophagefferens_CCMP1984_23580
Nannochloropsis_gaditana_CCMP526_Nga01542
Ectocarpus_siliculosus_Ec32_Esi0002_0211
Emiliania_huxleyi_CCMP1516_434881
Karenia_brevis_UN04048
Bigelowiella_natans_CCMP2755_51567
Calliarthron_tuberculosum_IDg17492t1
Porphyridium_purpureum_contig_2102.13
Chondrus_crispus_PCG_CHC_T00008517001
Pyropia_yezoensis_M_contig_24967_g6163
Chlamydomonas_reinhardtii_19871081
Coccomyxa_subellipsoidea_C-169_33281
Ostreococcus_sp_RCC809_28776 FabG
Chlorokybus_atmophyticus_UTEX2591_JO197619.1_2_116_691
Brassica_napus_1EDO_A FabG
Arabidopsis_thaliana_AT1G24360.1 FabG
Physcomitrella_patens_patens_18051328
Cyanophora_paradoxa_Contig38024_gene-0.4
Galdieria_sulphuraria_EME26281
Cyanidioschyzon_merolae_10D_CMS393C
Gloeobacter_violaceus_WP_011143495
Synechococcus_elongatus_PCC_7942_4DML_A FabG
Neosynechococcus_sphagnicola_sy1_WP_026535318
Synechocystis_sp._PCC_6803_4RZH_A FabG
Cyanothece_sp._PCC_7822_WP_013324621
Synechococcus_sp._RCC307_WP_011935099
Chloroflexus_aggregans_WP_012615595
Listeria_monocytogenes_4JRO_A FabG
Bacillus_anthracis_2UVD_A__FabG
Staphylococcus_aureus_3OSU_A FabG
Bacillus_coahuilensis_WP_010172398.1
Anaeroglobus_geminatus_WP_006789267.1
Propionispira_raffinosivorans_WP_019553059
Caldanaerobacter_subterraneus_KB-1_WP_022587648
Acetivibrio_ethanolgignens_WP_058353424
Digyalum_oweni WS1 @ DN32173_c0_g1_i2
Digyalum_oweni WS3 @ DN11839_c0_g2_i2
Deltaproteobacteria_bacterium_RZV40363
Caloramator_australicus_RC3_WP_008908883
Clostridium_thermocellum_2HQ1_A putative Glucose-ribitol dehydrogenase
Omnitrophica_bacterium_RIFCSPLOWO2_12_FULL_50_11_OGX06060
Chlorobium_chlorochromatii_WP_011362676
Candidatus_Thorarchaeota_archaeon_SMTZ1-83_KXH70652
candidate_division_Zixibacteria_bacterium_SM23_73_3_KPL00595
Syntrophothermus_lipocalidus_WP_013175612
candidate_division_KSB1_bacterium_4484_219_OQX58049
Plesiocystis_pacifica_4NBW_A FabG
Homo_sapiens_4CQL_A Heterotetrameric_ketoacyl_reductase
Candida_tropicalis_2ET6_A Hydroxyacyl-Coa Dehydrogenase
Homo_sapiens_1ZBQ_A 17-Beta-Hydroxysteroid Dehydrogenase Type 4
Mycobacterium_avium_3QLJ_A Short_Chain_Dehydrogenase
Pseudomonas putida@2Q2Q_A Beta-d-hydroxybutyrate dehydrogenase

Figure 4—figure supplement 3. Maximum likelihood phylogeny of the beta-ketoacyl-acyl carrier protein reductase, FabG (LG+R6 model in IQ-TREE with 10000 UFBoot2 replicates;≥80 are shown). Apicomplexans, chrompodellids and *Digyalum* are shown on orange background and other eukaryotes on gray background. Selected characterized enzymes are highlighted in bold.

DOI: https://doi.org/10.7554/eLife.49662.013

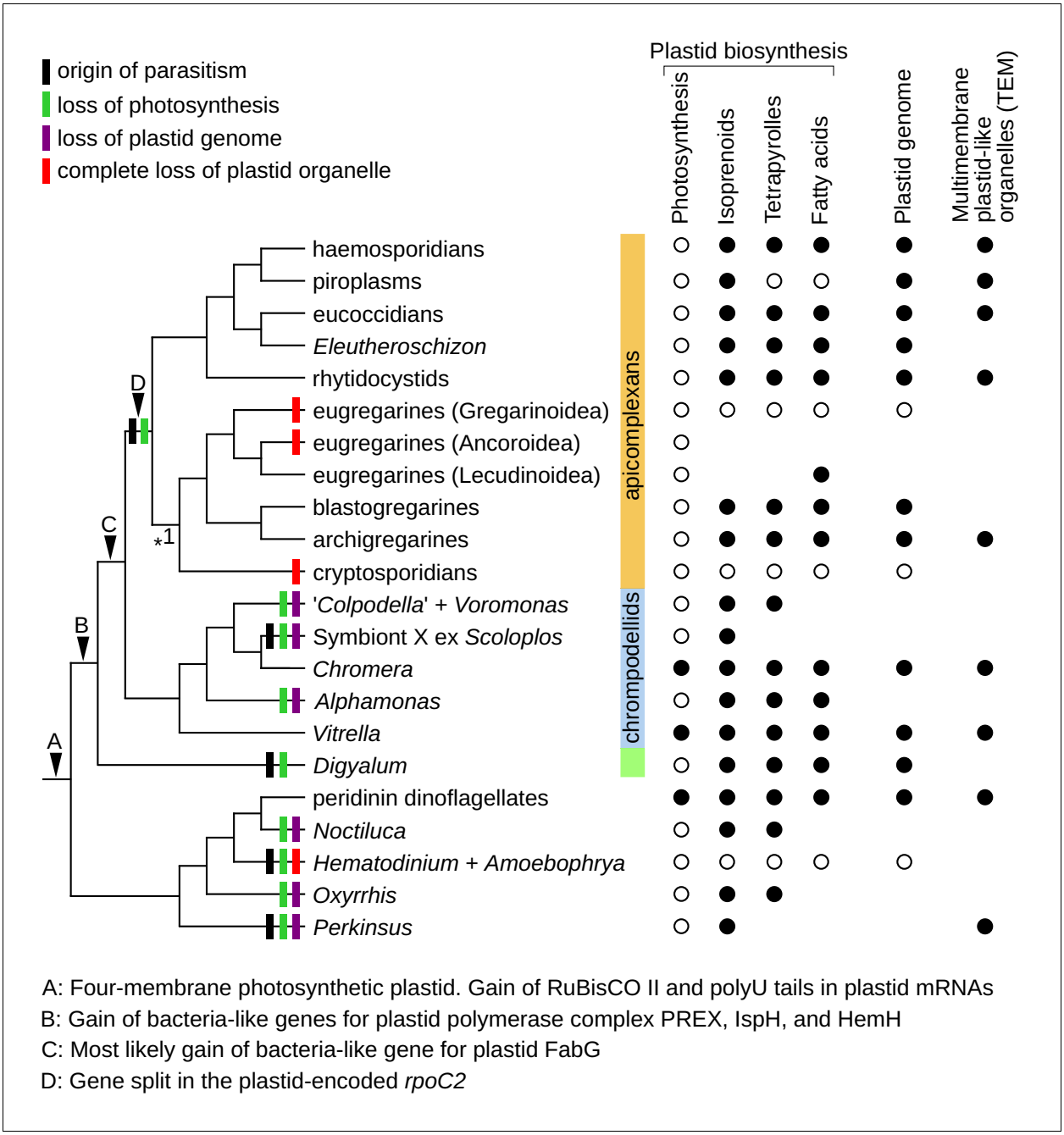


Figure 5. Plastid evolution in apicomplexans and their relatives. Plastid-related characteristics (A–D), origin of parasitism, and predicted losses of photosynthesis, plastid genomes, and plastid organelles were mapped on the updated phylogeny. The phylogeny is fully resolved except one branch where more support is needed (*1). Predicted core plastid anabolic capabilities, presence of plastid genomes, and transmission electron microscopy evidence (TEM) for multimembrane organelles corresponding to plastids by their size, appearance, and position within cells are shown on the right.
DOI: https://doi.org/10.7554/eLife.49662.014