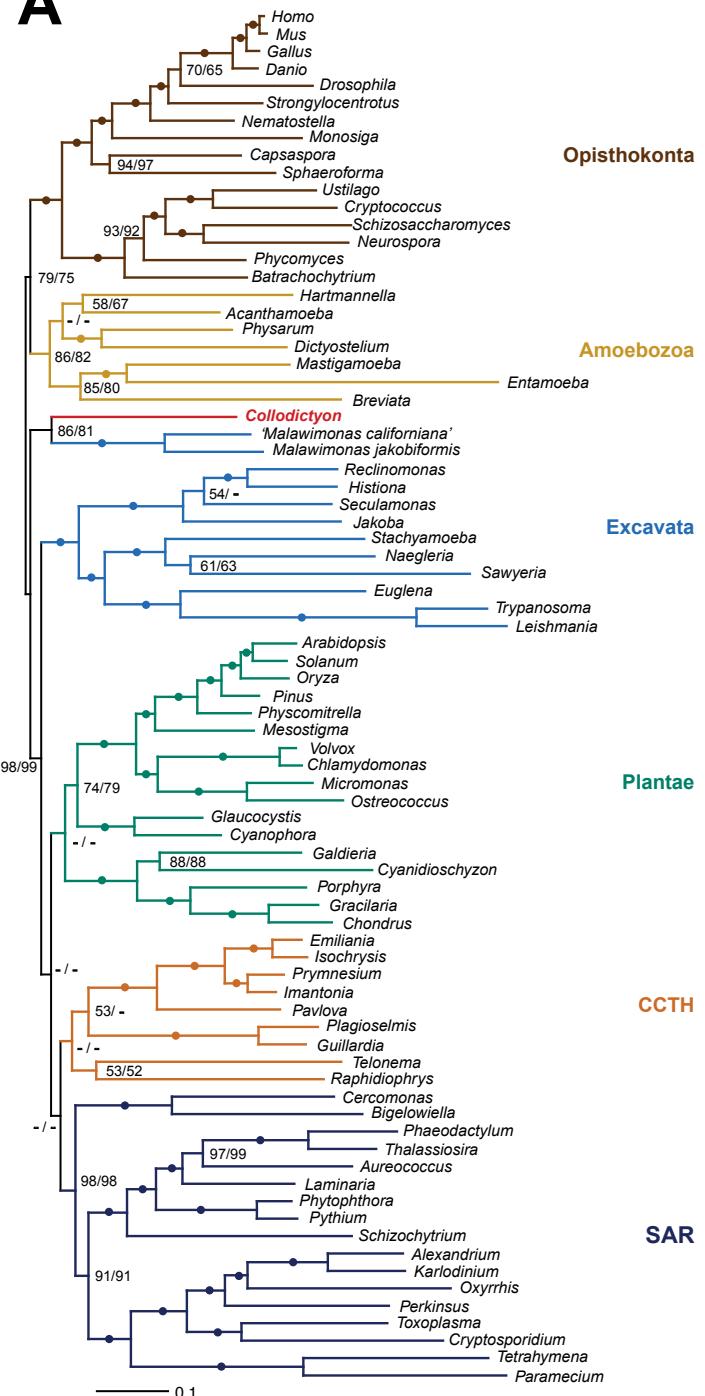
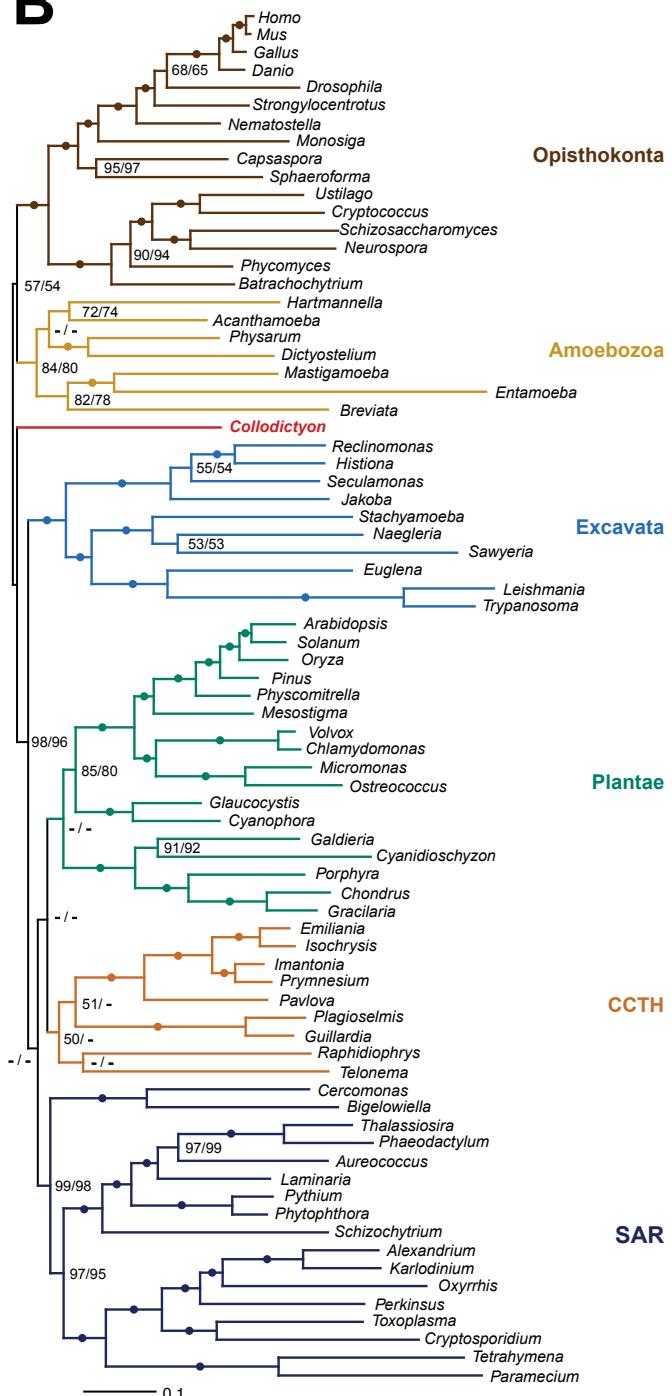
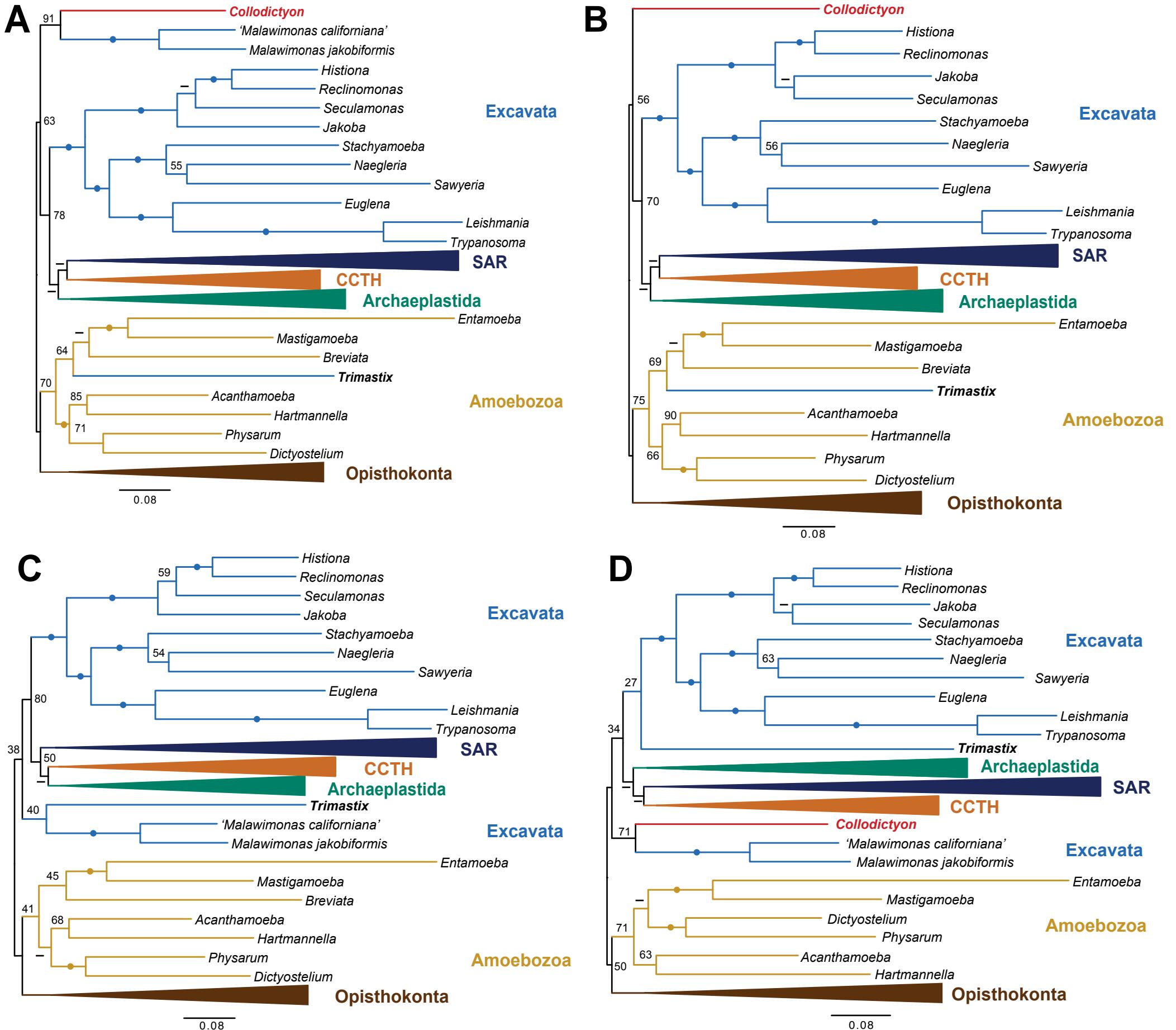
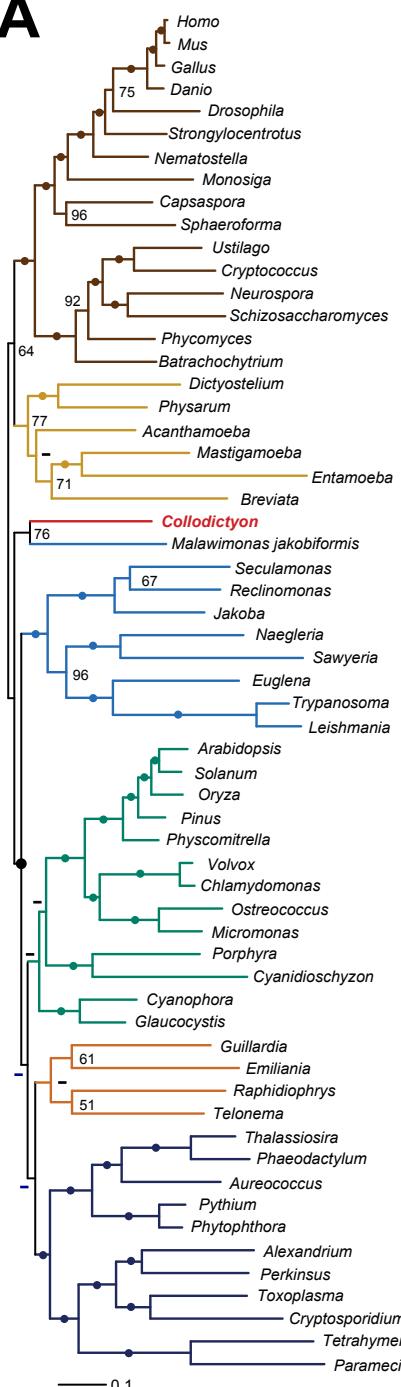
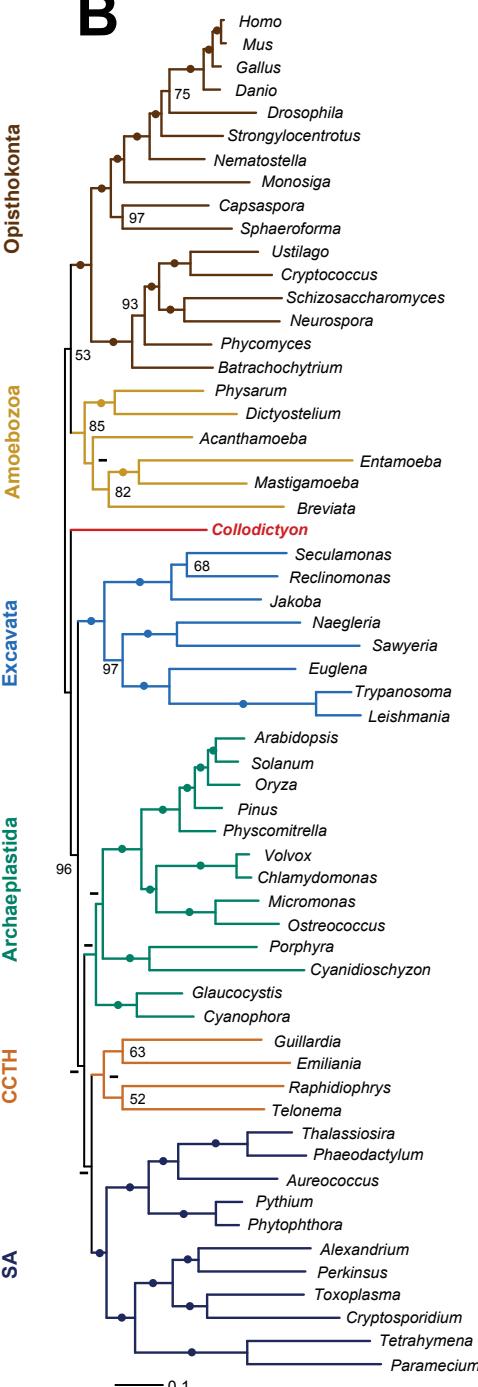
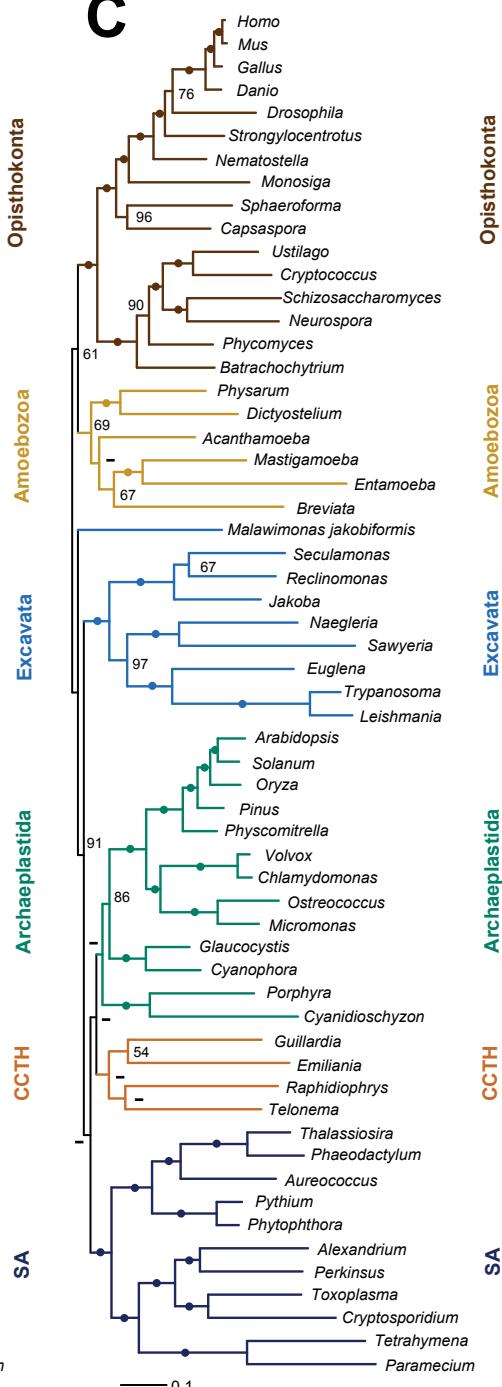


**A****B**

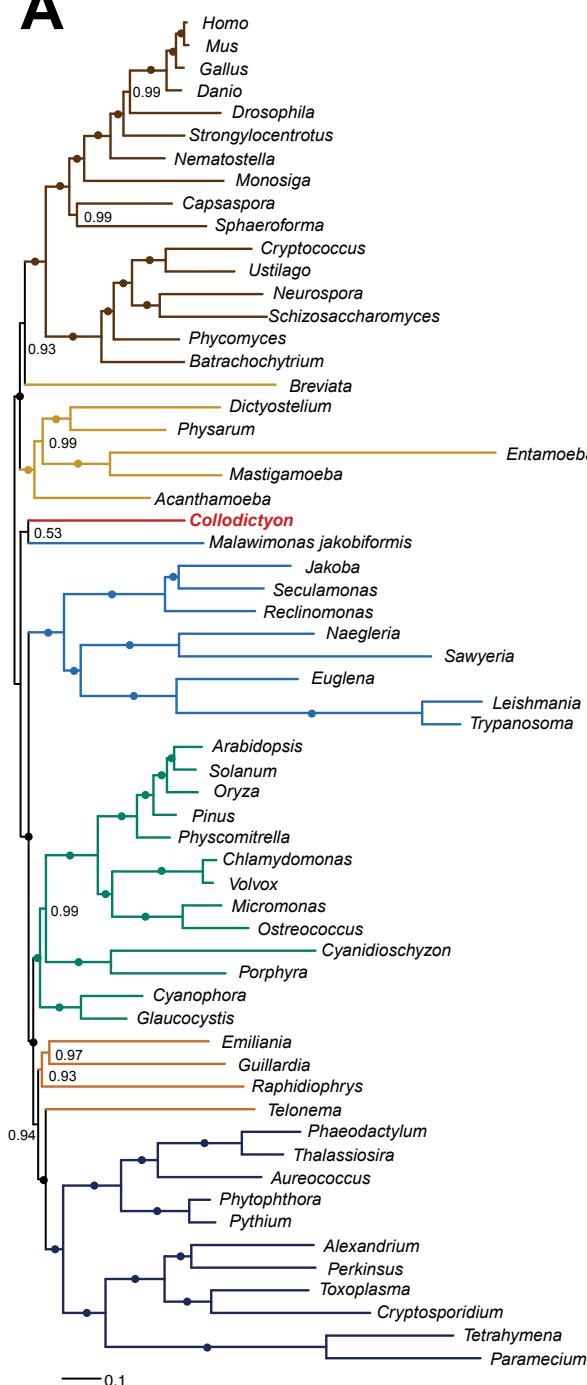
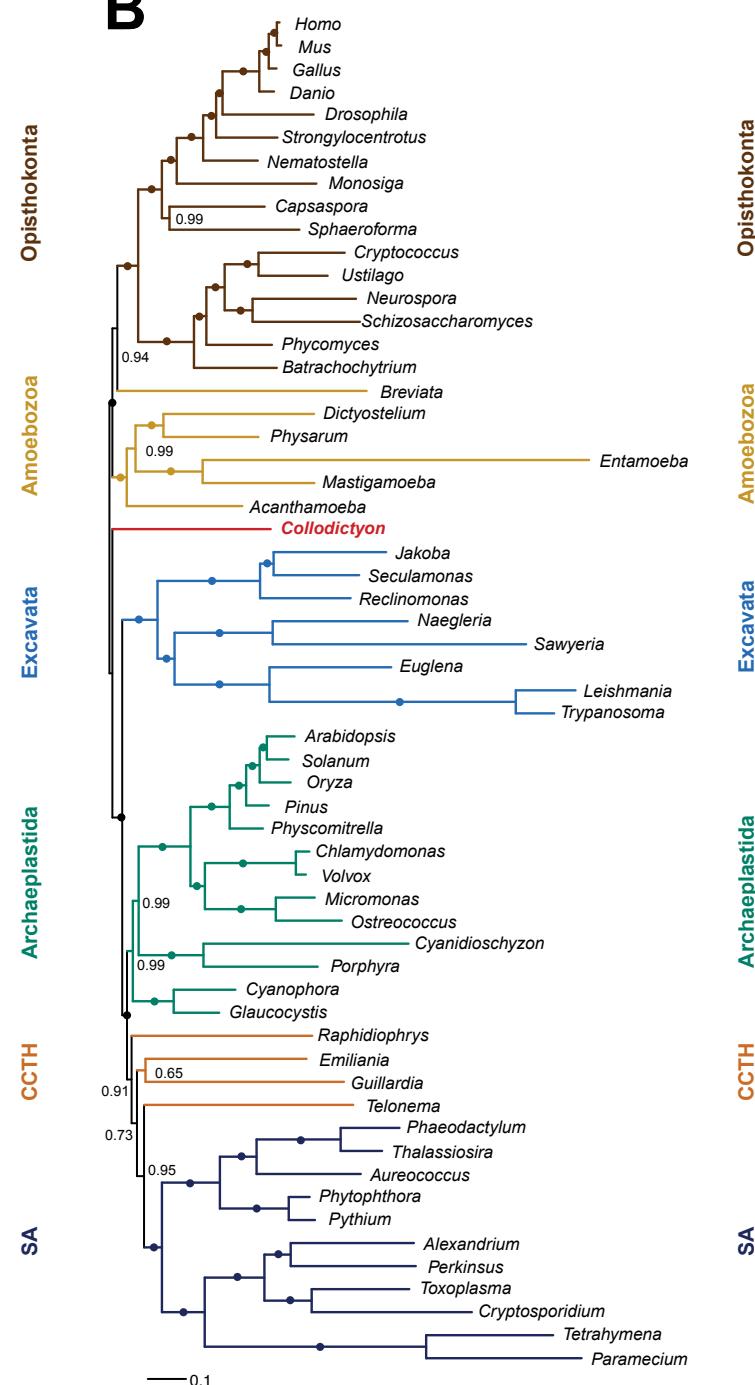
**Supplementary Figure S1.** Maximum likelihood phylogeny of *Collodictyon* and *Malawimonas*. The best topology was inferred using PROTGAMMALGF model in RAxML v7.2.6. CCTH: Cryptophyta, Centrohelida, Telonemia and Haptophyta. Numbers at the nodes indicate bootstrap values calculated under concatenated/separate model. Branches that received 100% bootstrap support (BP) are marked by full circles. Dashes '-' indicate BP < 50%. (A) The topology constructed with 79 taxa. (B) The topology constructed with 77 taxa (i.e. two *Malawimonas* excluded).



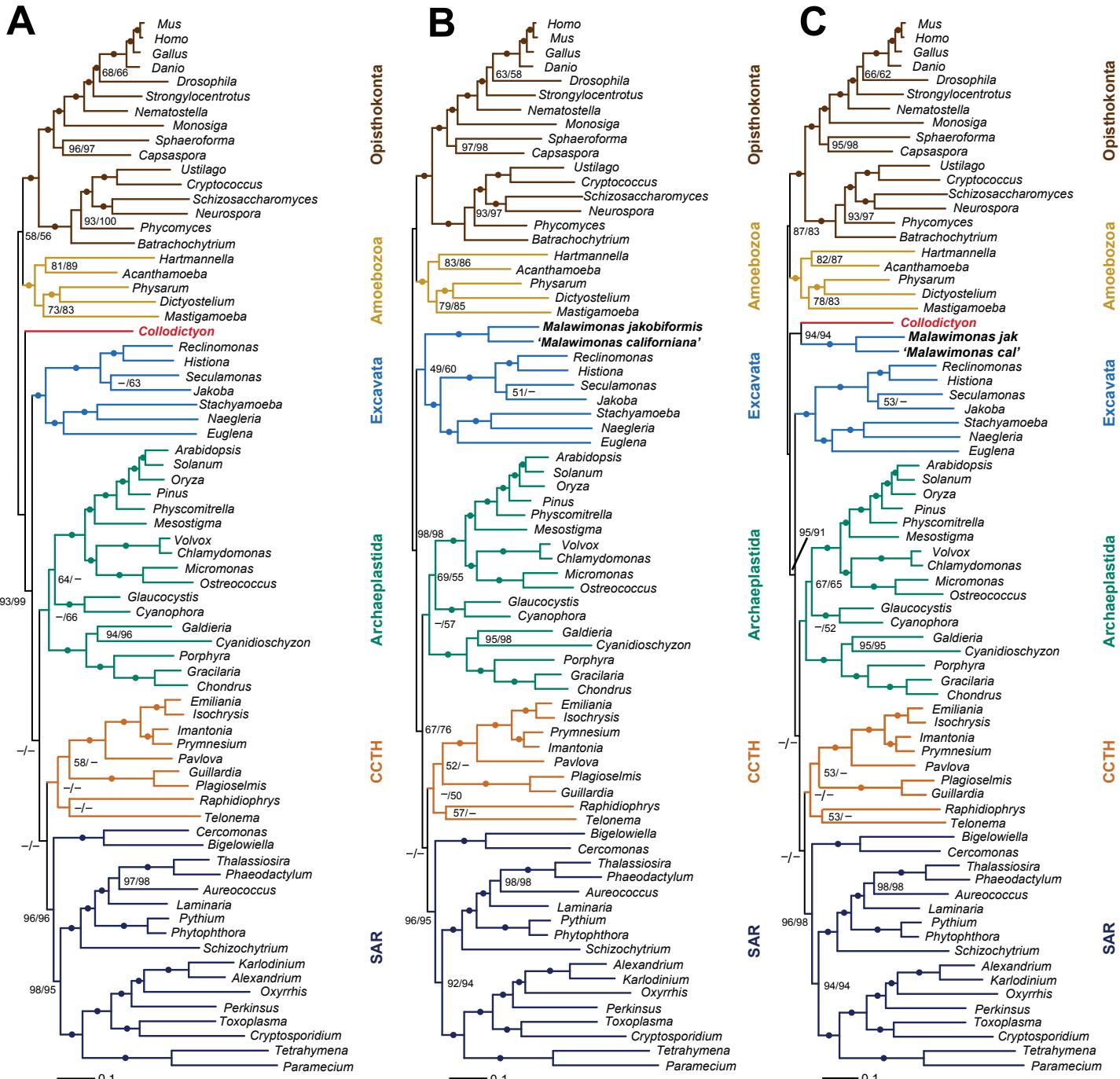
**Supplementary Figure S2.** The maximum likelihood phylogeny of *Trimastix*. The best topology was obtained by RAxML v7.2.6 under the PROTGAMMALGF model. Representative supergroups together with *Collodictyon* were color-coded. Opisthokonta, Archaeplastida, Hacrobia and SAR were schematized in FigTree (<http://tree.bio.ed.ac.uk/software/figtree>). Full circles correspond to 100% bootstrap supports (BP); dashes '-' indicate BP < 50%. CCTH: Cryptophyta, Centrohelida, Telonemia and Haptophyta. (A) The tree was constructed with 80 taxa. (B) The tree was constructed with 78 taxa (i.e. two *Malawimonas* excluded). (C) The tree was constructed with 79 taxa (i.e. *Collodictyon* excluded). (D) The tree was constructed with 79 taxa (i.e. *Breviata* excluded).

**A****B****C**

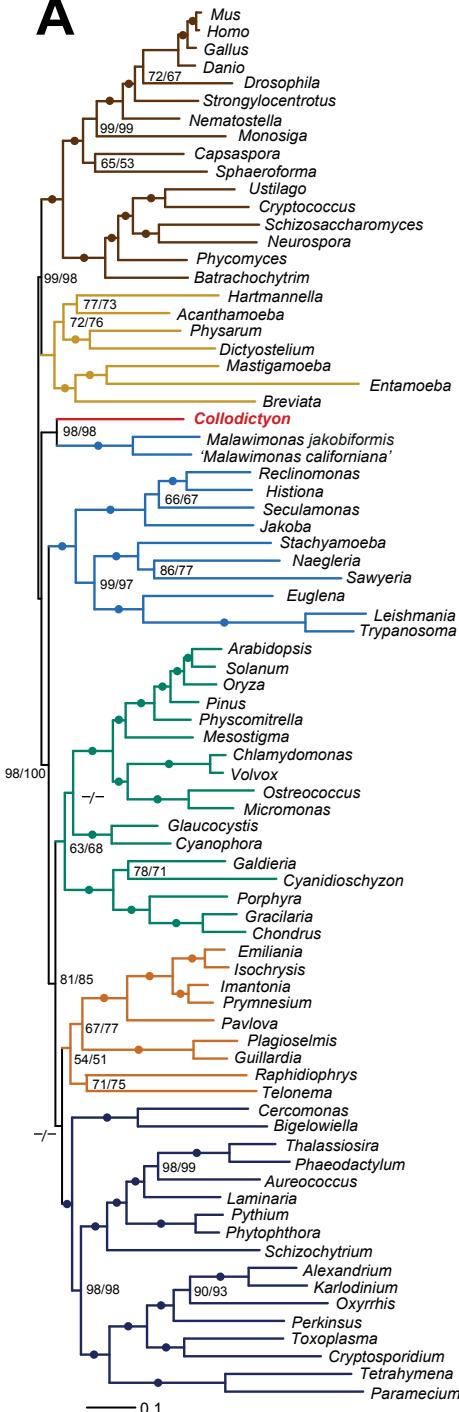
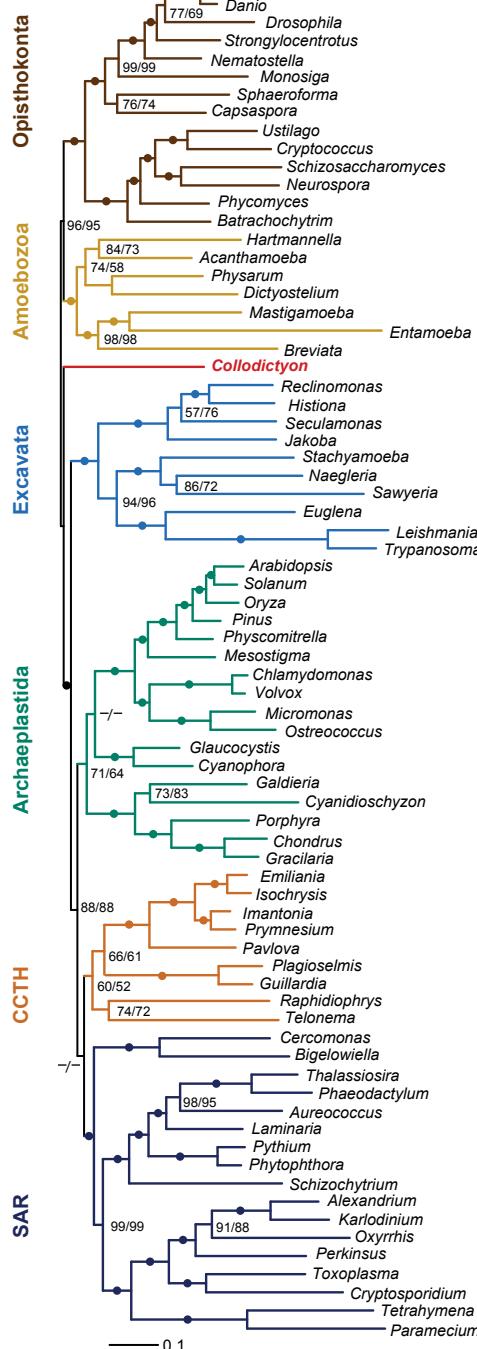
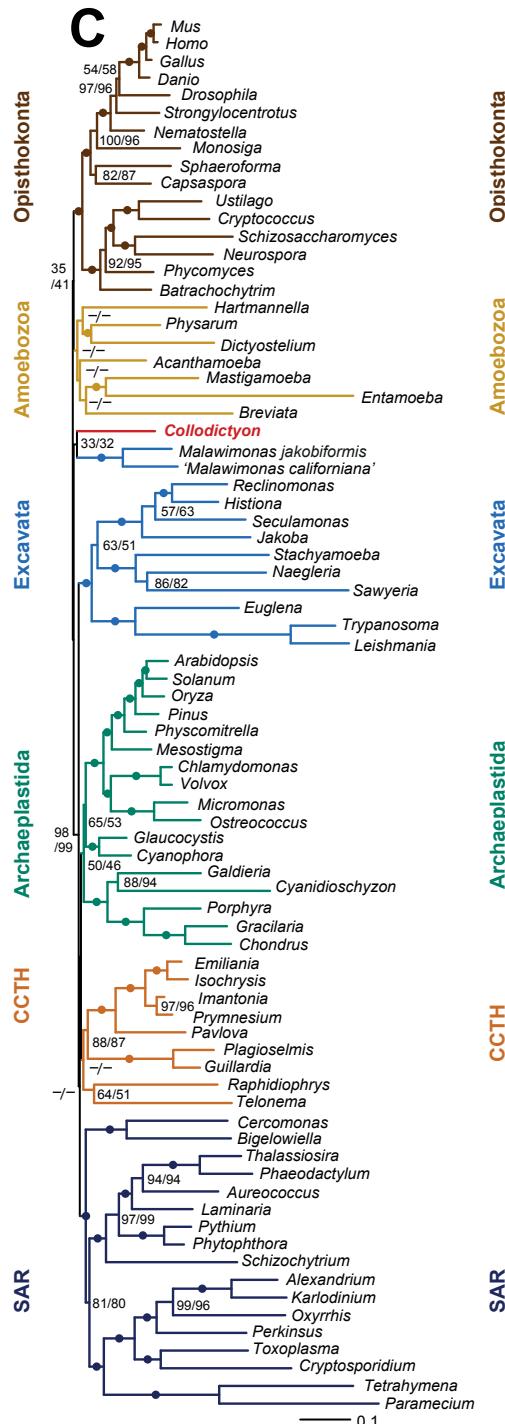
**Supplementary Figure S3.** The maximum likelihood phylogeny of *Collodictyon* after removal of taxa that showed missing percentage of sites over 60% of the alignment (See Figure 3A). The topology was constructed by RAxML v7.2.6 under the PROTGAMMALGF model and representative supergroups were color-coded. Branches that received 100% bootstrap values (BP) are marked by full circles; dashes ‘-’ indicate BP < 50%. SA is the abbreviation of Stramenopila and Alveolata. CCTH: Cryptophyta, Centrohelida, Telonemia and Haptophyta. (A) The tree was constructed with 60 taxa. (B) The tree was constructed with 59 taxa (i.e. *Malawimonas* excluded). (C) The tree was constructed with 59 taxa (i.e. *Collodictyon* excluded).

**A****B**

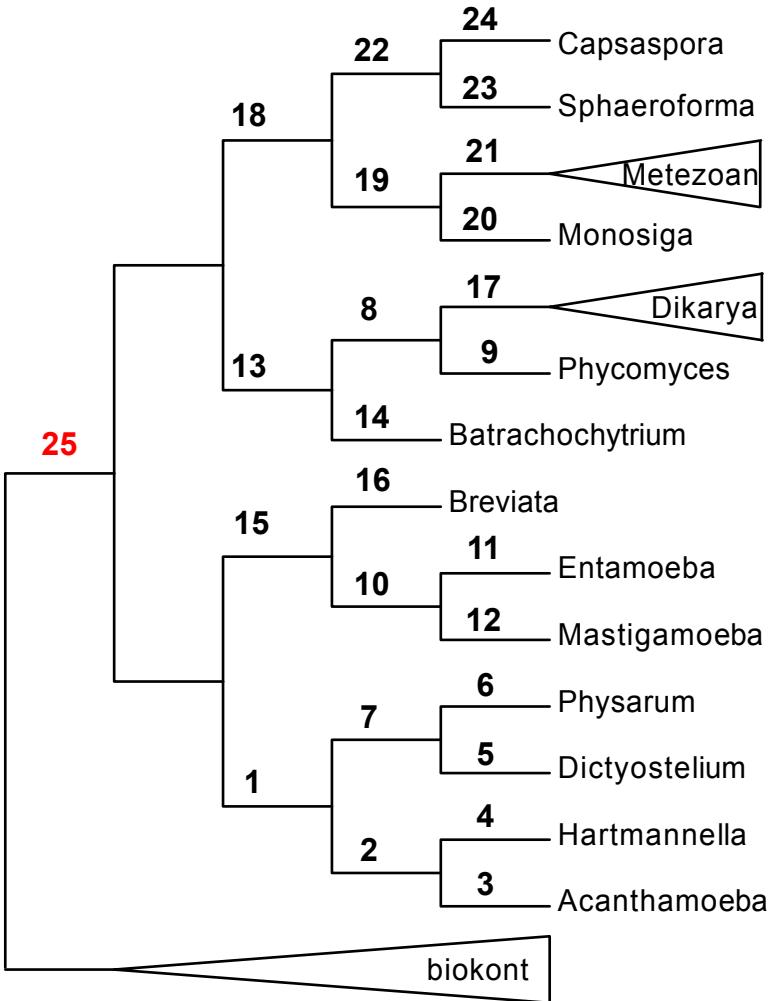
**Supplementary Figure S4.** The bayesian phylogeny of *Collodictyon* after removal of taxa that showed missing percentage of sites over 60% of the alignment (See Figure 3A). The consensus topology was inferred under CAT model from 12,000 saved trees after discarding the first 4,000 cycles as burn-in. Branches that received 1.00 posterior probabilities (PP) are marked by full circles. CCTH: Cryptophyta, Centrohelida, Telonemia and Haptophyta. SA is the abbreviation of Stramenopila and Alveolata. (A) The tree was constructed with 60 taxa and the convergence was ascertained by maxdiff = 0.051. (B) The tree was constructed with 59 taxa (i.e. *Malawimonas* excluded) and the convergence was ascertained by maxdiff = 0.127.



**Supplementary Figure S5.** Maximum likelihood phylogeny of *Collodictyon* and *Malawimonas* after removal of five taxa (i.e. *Leishmania*, *Trypanosoma*, *Sawyeria*, *Entamoeba* and *Breviata*). The ML topology was inferred by RAxML v7.2.6 under PROTGAMMALGF model. The numbers at the nodes indicate bootstrap values calculated under concatenated/separate model. Full circles correspond to 100% bootstrap values (BP); dashes '-' indicate BP < 50%. CCTH: Cryptophyta, Centrohelida, Telonemia and Haptophyta. (A) The tree constructed with 72 taxa (i.e. two *Malawimonas* excluded). (B) The tree constructed with 73 taxa (i.e. *Collodictyon* excluded). (C) The tree constructed with 74 taxa.

**A****B****C**

**Supplementary Figure S6.** Maximum likelihood phylogeny of *Collopyctyon* and *Malawimonas* after removal of the fastest evolving sites. The ML topology was inferred by RAxML v7.2.6 under PROTGAMMALGF model. The numbers at the nodes indicate bootstrap values calculated under concatenated/separate model. Full circles correspond to 100% bootstrap values (BP); dashes '-' indicate BP < 50%. CCTH: Cryptophyta, Centrohelida, Telonemia and Haptophyta. (A) The tree inferred from the trimmed alignment with the 20% fastest evolving sites removed (marked by gray rectangles in Fig. 4A). (B) The tree inferred from the trimmed alignment (i.e. two *Malawimonas* excluded) with the 20% fastest evolving sites removed (marked by gray rectangles in Fig. 4B). (C) The tree inferred from the trimmed alignment with the 45% fastest evolving sites removed in Fig. 4A.



(A)

ID	$\Delta \ln L$	P value
25	<b>-74.8</b>	<b>0.987</b>
15	74.8	0.023
16	94.6	0.002
13	128.4	0.002
4	135.0	8e-07
11	161.6	9e-39
9	186.9	1e-08
10	221.5	2e-62
14	221.5	2e-62
8	389.2	1e-04
7	404.1	2e-34
22	404.5	4e-06
5	411.6	1e-48
6	422.1	2e-39
17	437.7	2e-76
18	842.8	3e-62
19	843.6	2e-87
1	921.8	5e-09
12	930.4	3e-10
20	1171.5	4e-05
21	1222.9	2e-11
23	1393.2	9e-08
24	1398.7	4e-13
2	1725.7	6e-06
3	1738.3	2e-06

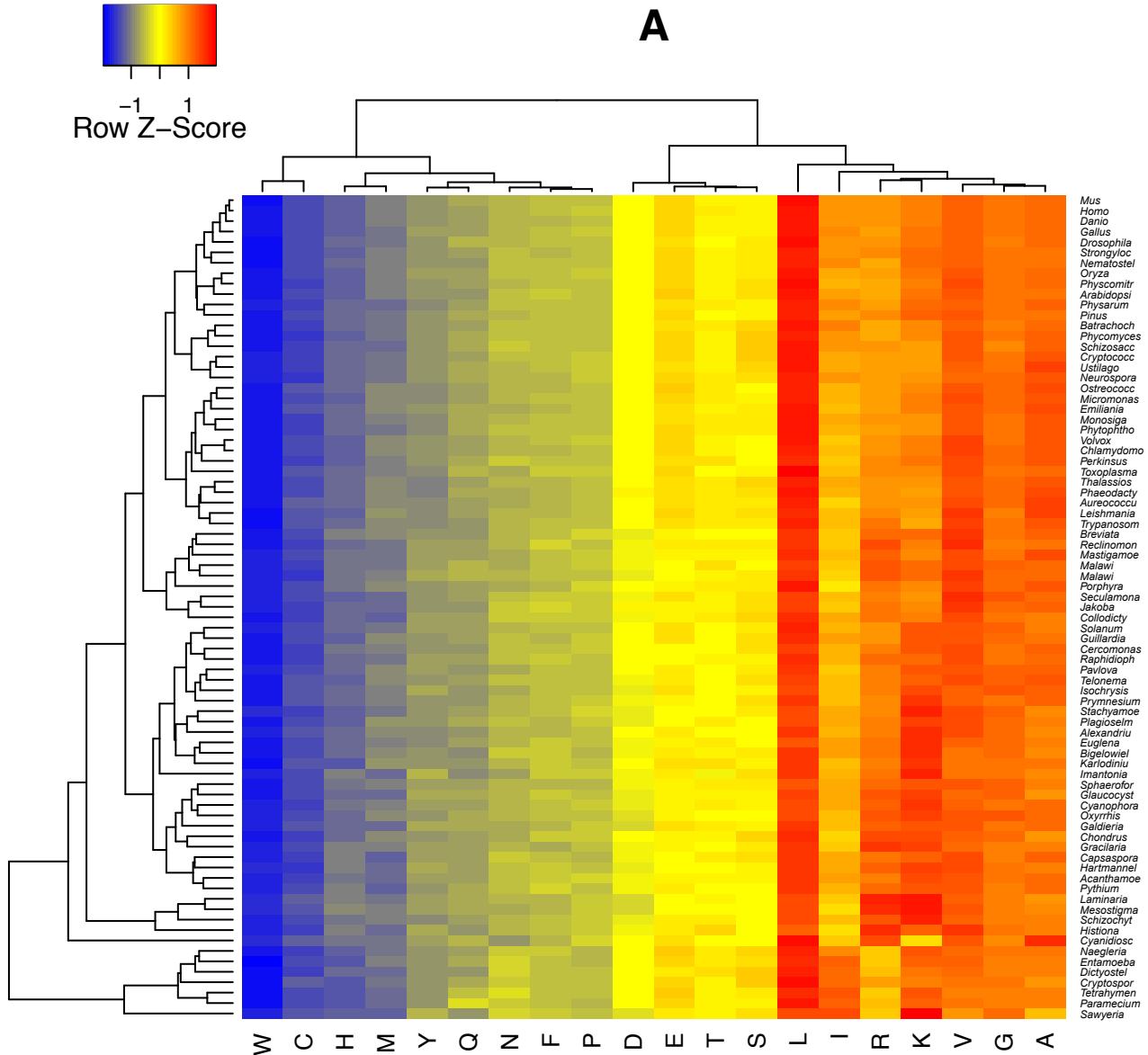
(B)

ID	$\Delta \ln L$	P value
25	<b>-82.4</b>	<b>1.000</b>
15	82.4	0.001
16	82.8	4e-04
13	112.4	0.001
4	141.2	1e-13
9	146.7	1e-56
11	158.6	9e-07
10	167.9	3e-104
14	167.9	3e-104
22	281.9	2e-06
8	297.2	1e-45
17	300.7	2e-09
7	305.1	2e-48
5	336.3	7e-44
6	345.3	3e-84
18	593.6	7e-40
19	593.9	6e-45
1	626.2	4e-04
12	628.4	3e-04
20	840.2	1e-04
21	874.7	2e-08
23	920.5	3e-04
24	921.8	1e-09
2	1123.6	2e-05
3	1136.2	5e-05

**Supplementary Figure S7.** Summary of AU tests on the alternative topologies showing *Collopyctyon* within unikonts (numbers on branches). The values enlightened by red correspond to positions that were not rejected by AU tests at the 5% level of significance. (A) AU tests were run using the original alignment of 77 taxa (i.e. no sites removal). (B) AU tests were run using the trimmed alignment of 77 taxa with the 20% fastest evolving sites removed (marked by gray rectangles in Fig. 4B).

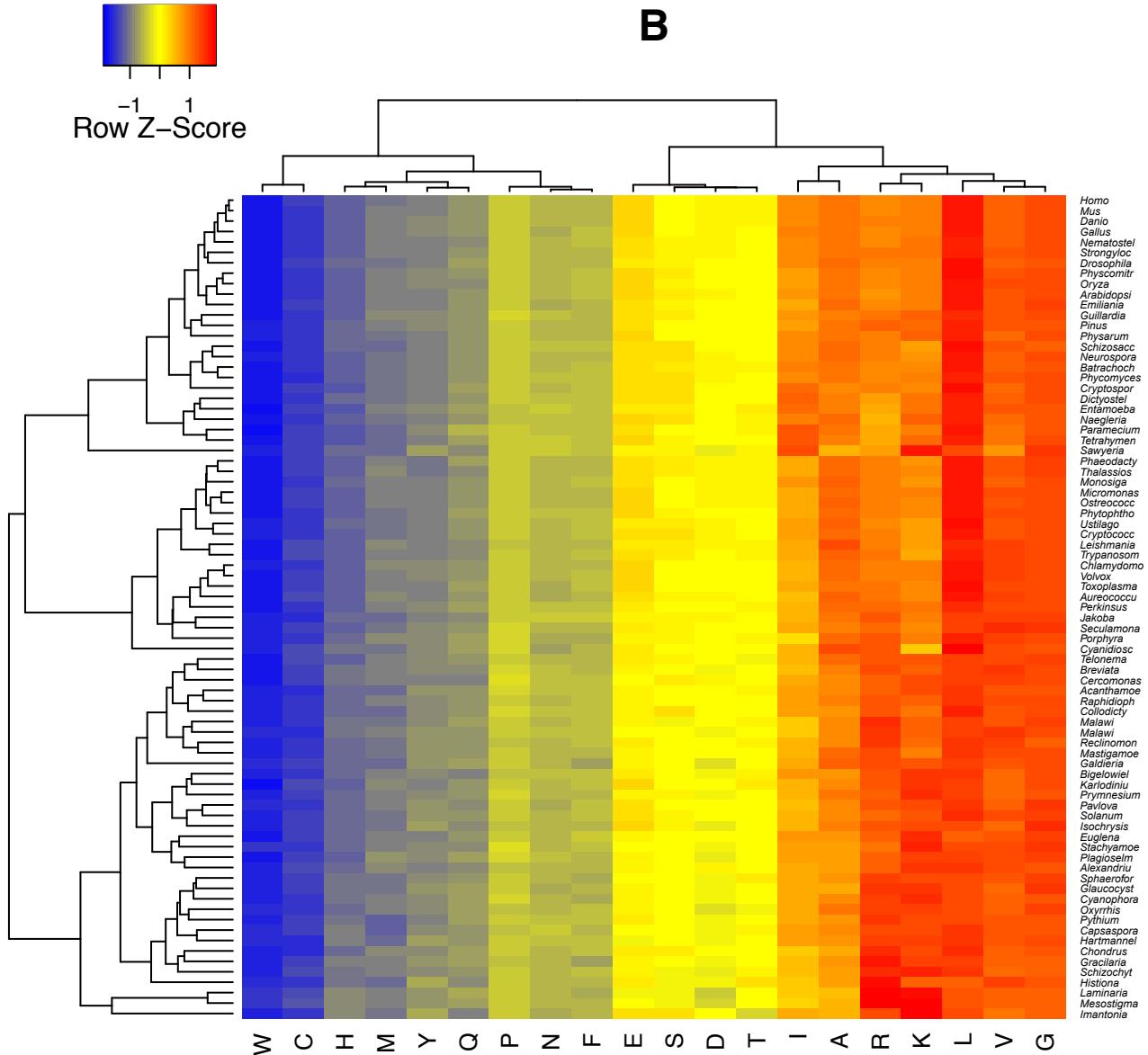
# Color Key

A



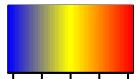
# Color Key

B



# Color Key

C



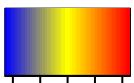
Row Z-Score

Physcomit  
Oryza  
Toxoplasma  
Oikopleuridomo  
Volvox  
Micromonas  
Ostreococc  
Phytophyto  
Emiliania  
Strongyloc  
Mus  
Homo  
Danio  
Nematostel  
Arabidopsi  
Musca  
Phycodrys  
Batrachoch  
Gallus  
Neurospora  
Cryptococc  
Ustilago  
Drosophil  
Schizosacc  
Cryptospor  
Pinus, taed  
Perkinsus  
Gutierrezia  
Physarum  
Jakoba  
Secularmona  
Phaeodactyl  
Thalassios  
Aureococcu  
Trypanosom  
Leishmania  
Cyanidisc  
Tetrahymen  
Paramecium  
Dictyostel  
Entamoeba  
Naegleria  
Raphidioph  
Acanthamo  
Pyrenomon  
Telonema  
Breviate  
Pavlova  
Solanum  
Colocidio  
Diplospor  
Mastigamo  
Reclinomon  
Porphyra  
Malava  
Oxymirs  
Pyram  
Malawi  
Isochrysis  
Karliodinu  
Euglena  
Cercononas  
Aleuridru  
Sphaerotil  
Capsaspora  
Hartmannel  
Gracilaria  
Glaucococc  
Raphidiom  
Cyanophora  
Galdieria  
Histiona  
Schizochyt  
Gigas  
Imantonia  
Stachyamo  
Sawyeria  
Mesotistigma  
Laminaria

L G V A K R W C N F H M Y Q T P S D I E

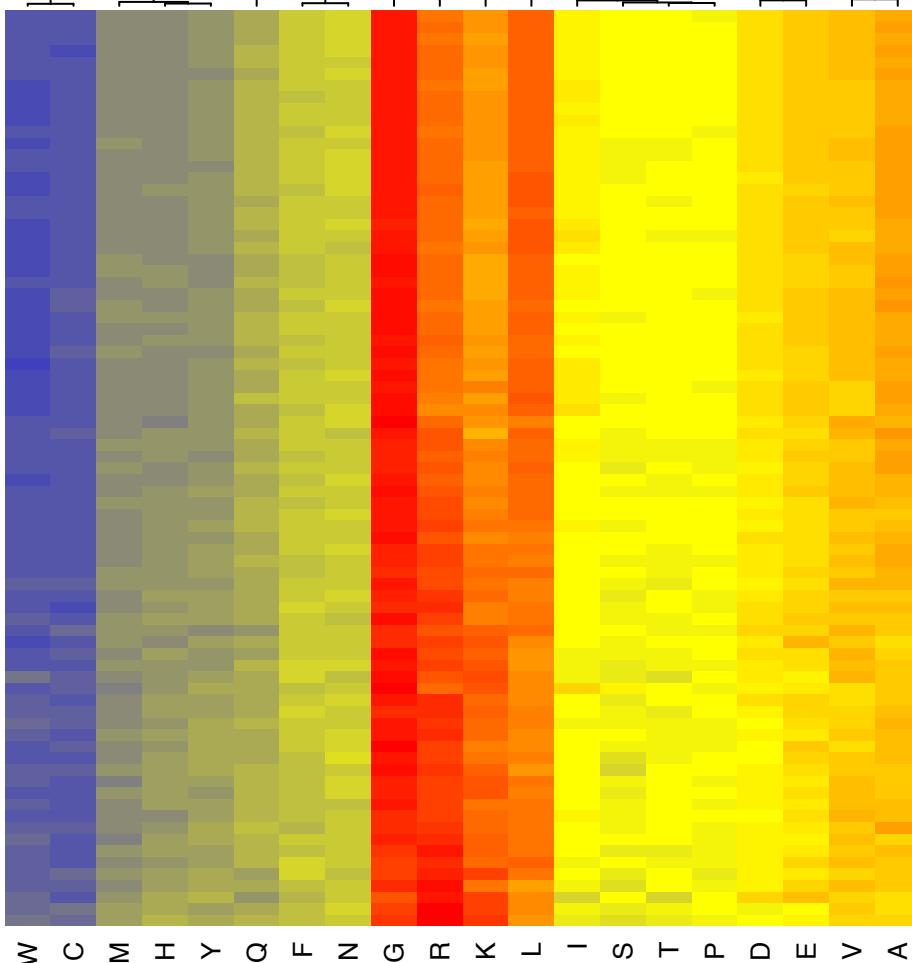
## Color Key

D

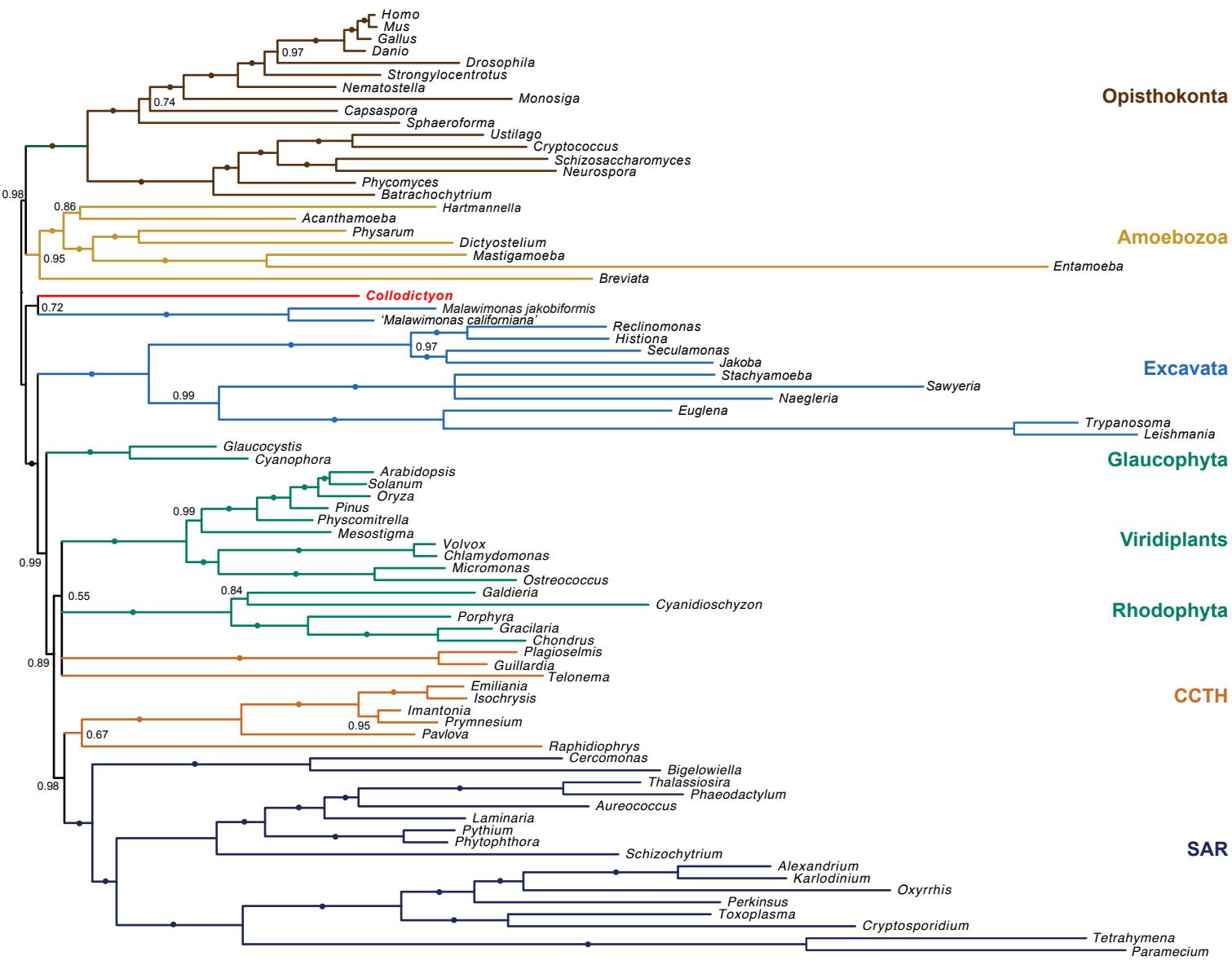


Row Z-Score

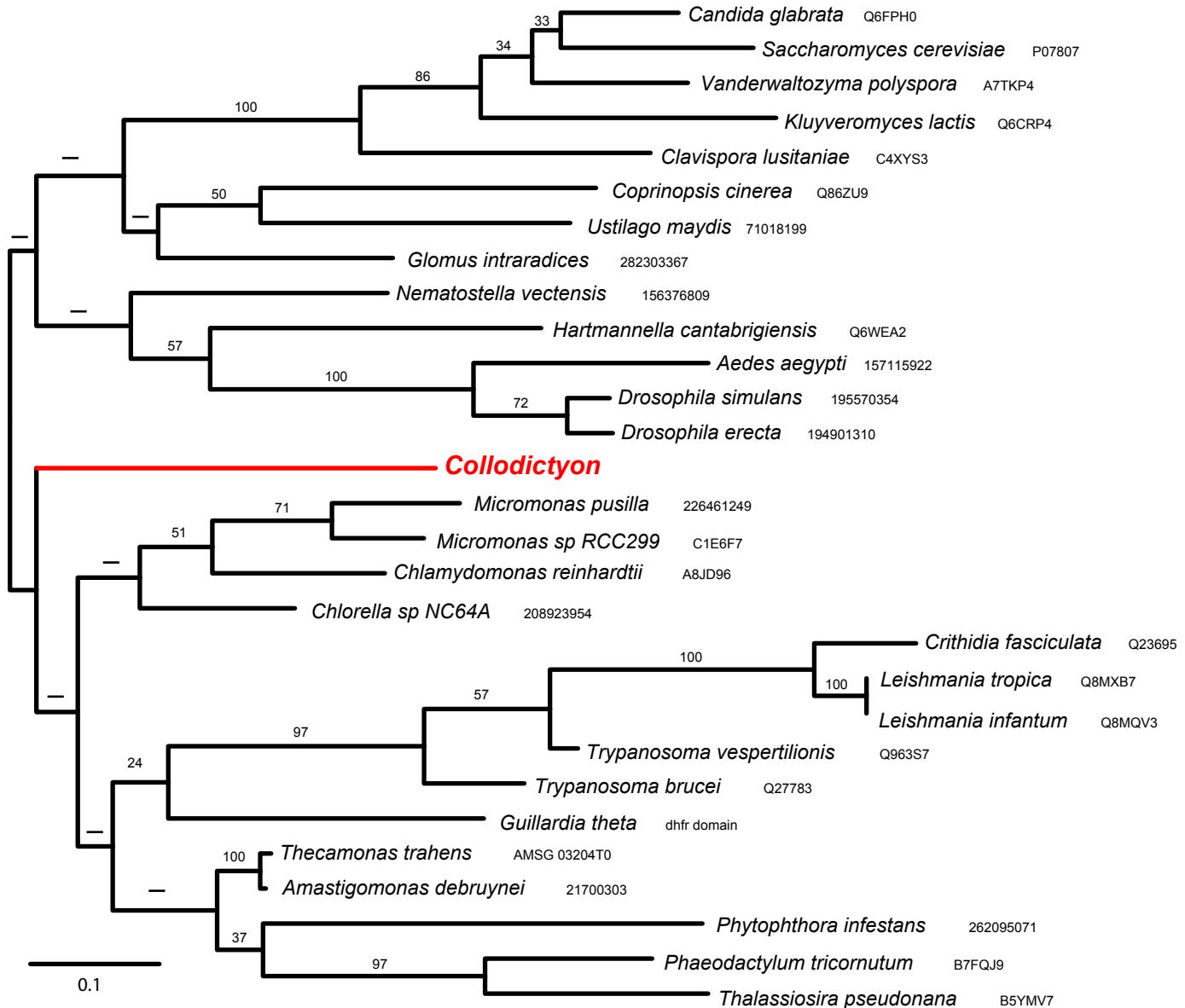
*Physcomitrium*  
*Oryza*  
*Arabidopsis*  
*Pyrenopezia*  
*Bathochochlaena*  
*Phytophthora*  
*Mus*  
*Homo*  
*Nematostel*  
*Danio*  
*Monosiga*  
*Volvox*  
*Chlamydomonas*  
*Micromonas*  
*Ostreococcus*  
*Drosophila*  
*Cryptococcobionatum*  
*Neurospora*  
*Schizosaccharomyces*  
*Cryptosporangium*  
*Gallus*  
*Primates*  
*Thalassiosira*  
*Ustilago*  
*Trypanosoma*  
*Toxoplasma*  
*Entamoeba*  
*Strongyloides*  
*Aureococcus*  
*Entamoeba*  
*Dictyostelium*  
*Naegleria*  
*Paramecium*  
*Tetrahymena*  
*Secularia*  
*Cyanidioscincus*  
*Pinus*  
*Physarum*  
*Perennipous*  
*Guillardia*  
*Solanum*  
*Porphyra*  
*Colloidity*  
*Acanthamoeba*  
*Jekelius*  
*Raphidiophorus*  
*Mastigamoeba*  
*Prymnesium*  
*Telomera*  
*Bigelovielvii*  
*Reduvionon*  
*Pavlovia*  
*Alexandriella*  
*Karliodinu*  
*Cercomonas*  
*Euglypha*  
*Strobilomonas*  
*Savoryna*  
*Malawi*  
*Pythium*  
*Oxymiris*  
*Marine*  
*Leucophysa*  
*Brevitata*  
*Glaucocystis*  
*Capsaspora*  
*Cyanophora*  
*Sphaerotilus*  
*Planctomycetes*  
*Gardieria*  
*Hartmannellula*  
*Gracilaria*  
*Chondrus*  
*Schizothrix*  
*Histiona*  
*Imantonia*  
*Laminaria*  
*Mesostigma*



**Supplementary Figure S8.** Two-dimensional ‘heatmap’ showing the hierarchical clustering of 79 taxa (y-axis) and 20 types of amino acids (x-axis). The clustering was based on the comparison of the amino acid composition for each taxon. (A), (B), (C) and (D) show amino acids composition biases among 79 taxa when 0 (i.e. no sites), 20%, 40% and 50% of the fastest evolving sites were removed from the alignment. Red, yellow and blue represent high, intermediate and low composition values (%).



**Supplementary Figure S9.** The phylogenomic tree constructed with amino acids recoded into six categories. The topology was inferred by Phylobayes v3.2 under mixture model CAT. Representative supergroups together with *Collodictyon* were color-coded. Branches that received 1.00 posterior probabilities (PP) are marked by full circles. Numbers at nodes show posterior probabilities that are not maximal. CCTH: Cryptophyta, Centrohelida, Telonemia and Haptophyta.



**Supplementary Figure S10.** Maximum likelihood phylogeny of the DHFR domain. Numbers at nodes indicate bootstrap supports (BP) calculated by RAxML v7.2.6 under PROTGAMMALGF model; dashes '-' show BP < 50%. The accession numbers for Refseq or Uniprot are added to the species names. The *Collodictyon* branch is enlightened by red color.