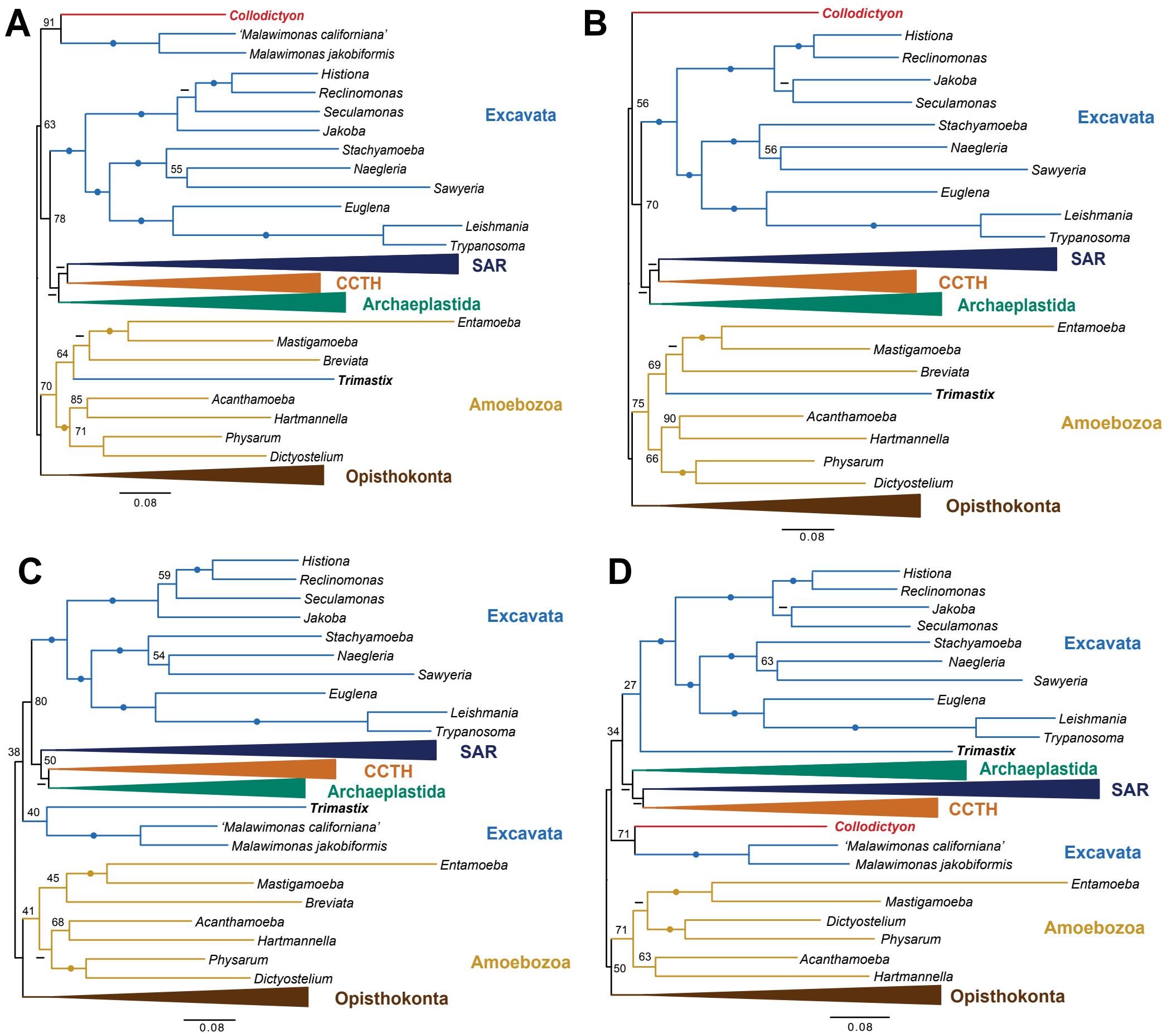
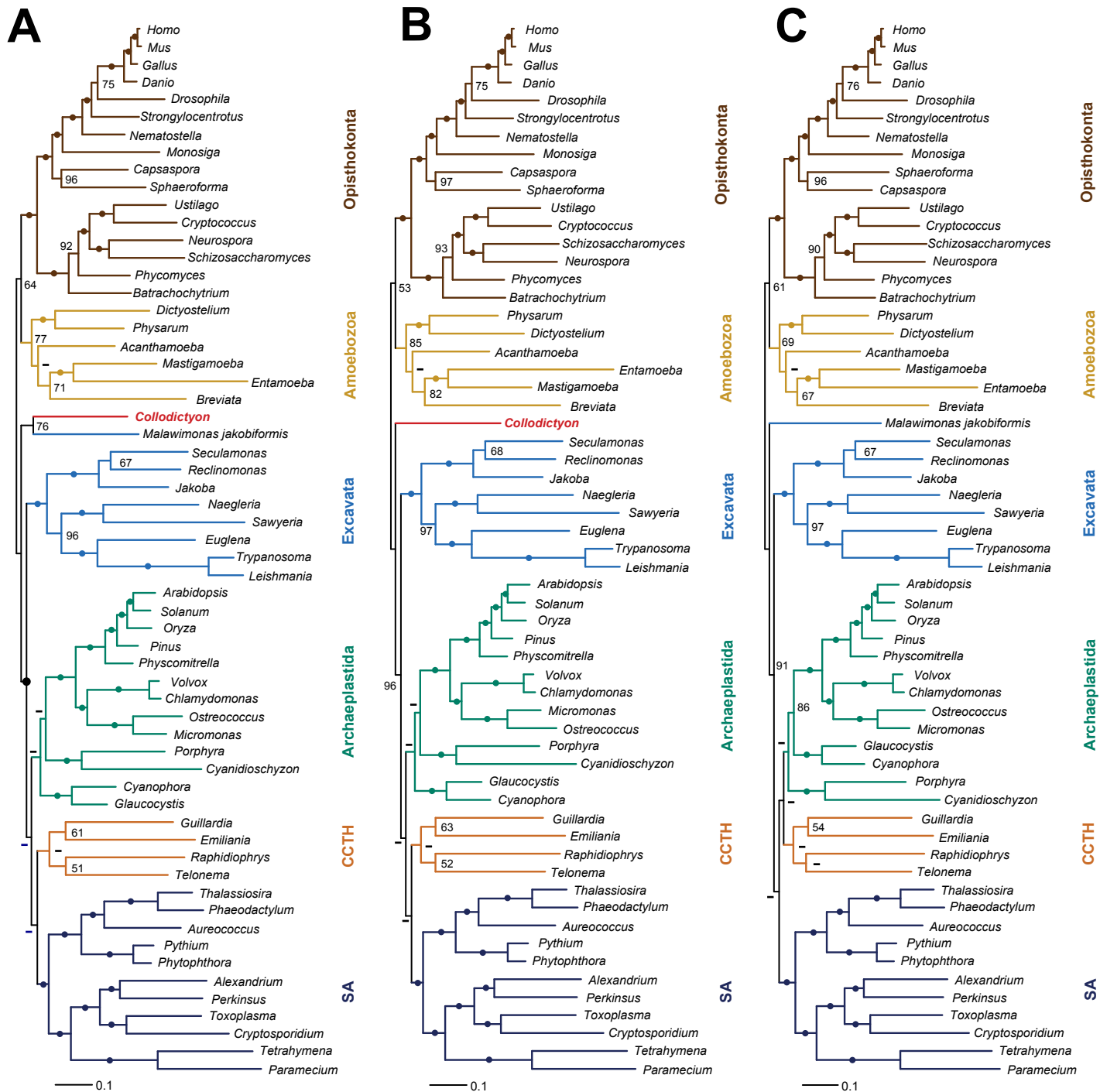


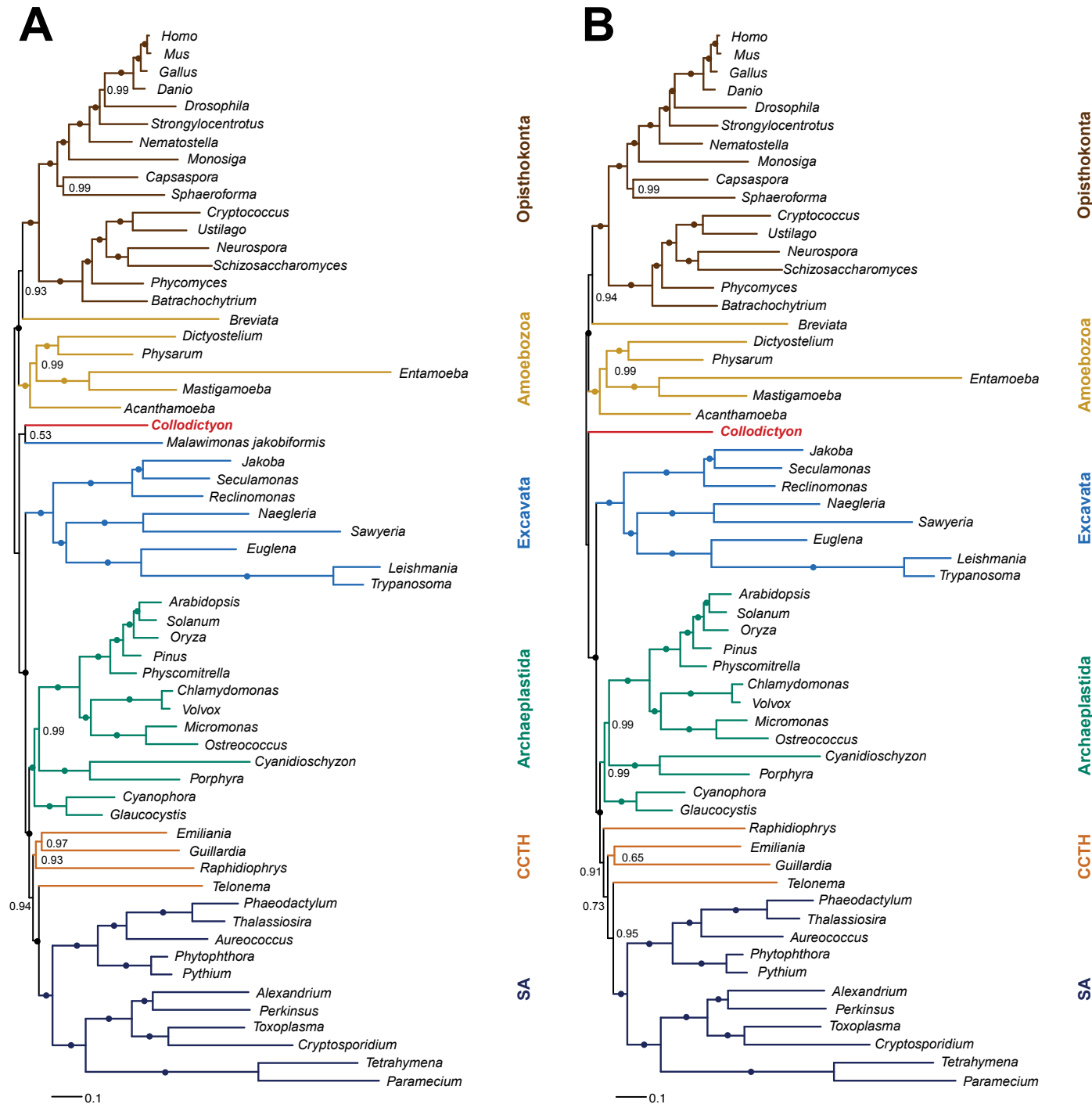
**Supplementary Figure S1.** Maximum likelihood phylogeny of *Collodictyon* and *Malawimonas*. The best topology was inferred using PROTGAMMALGF model in RAxML v7.2.6. CCTH: Cryptophyta, Centrohelida, Telonemia and Haptophyta. Numbers at the nodes indicate bootstrap values calculated under concatenated/separate 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).



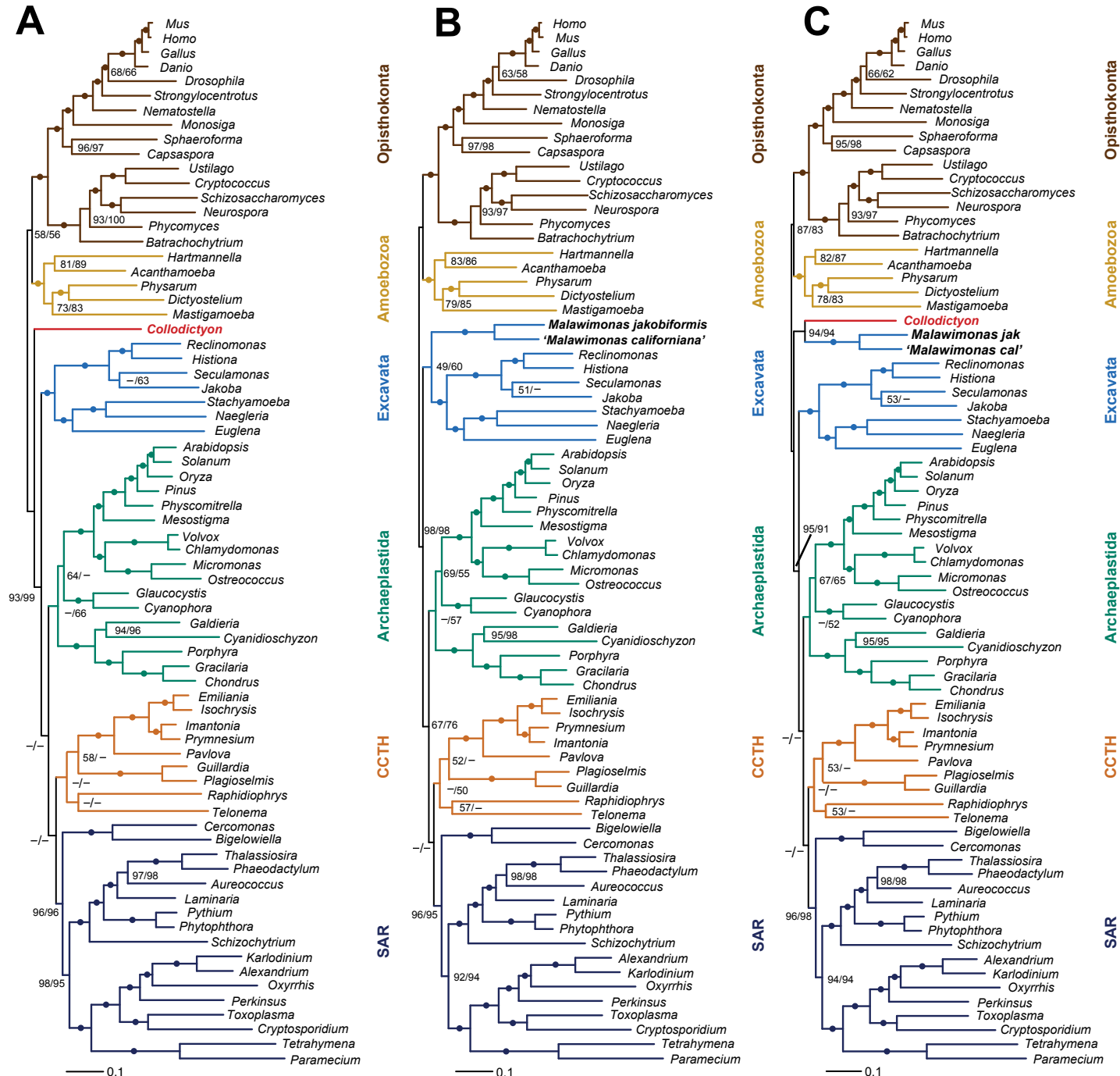
**Supplementary Figure S2.** The maximum likelihood phylogeny of *Trimastix*. The best topology was obtained by RAxML v7.2.6 under the PROTGAMMALGF model. Representative supergroups together with *Collodictyon* were color-coded. Opisthokonta, Archaeplastida, Hacrobia and SAR were schematized in FigTree (<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).



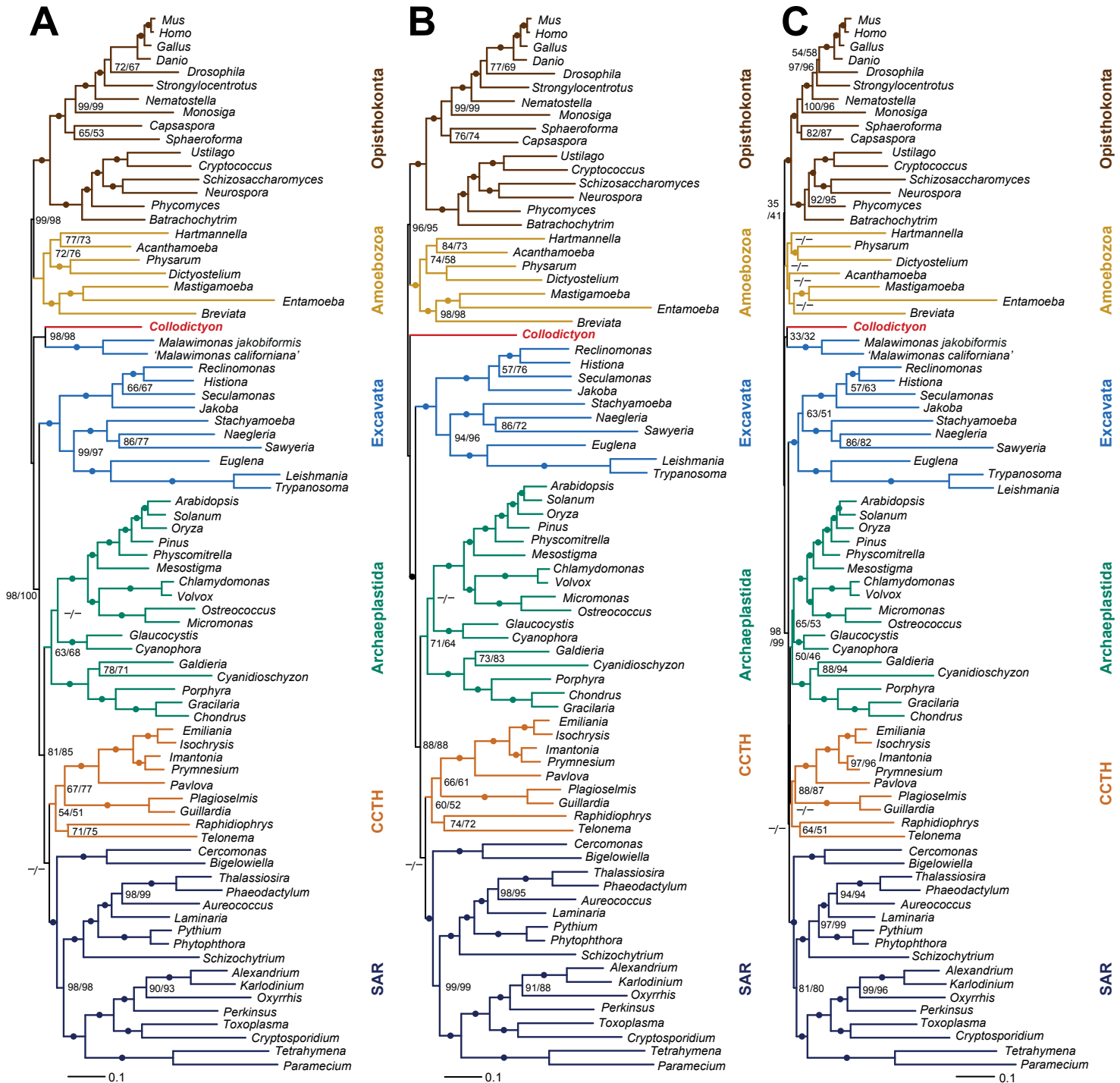
**Supplementary Figure S3.** The maximum likelihood phylogeny of *Collodictyon* after removal of taxa that showed missing percentage of sites over 60% of the alignment (See Figure 3A). The topology was constructed by RAxML v7.2.6 under the PROTGAMMALGF model and representative supergroups were color-coded. 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).



**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.



**Supplementary Figure S6.** Maximum likelihood phylogeny of *Collodictyon* 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.

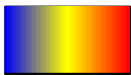






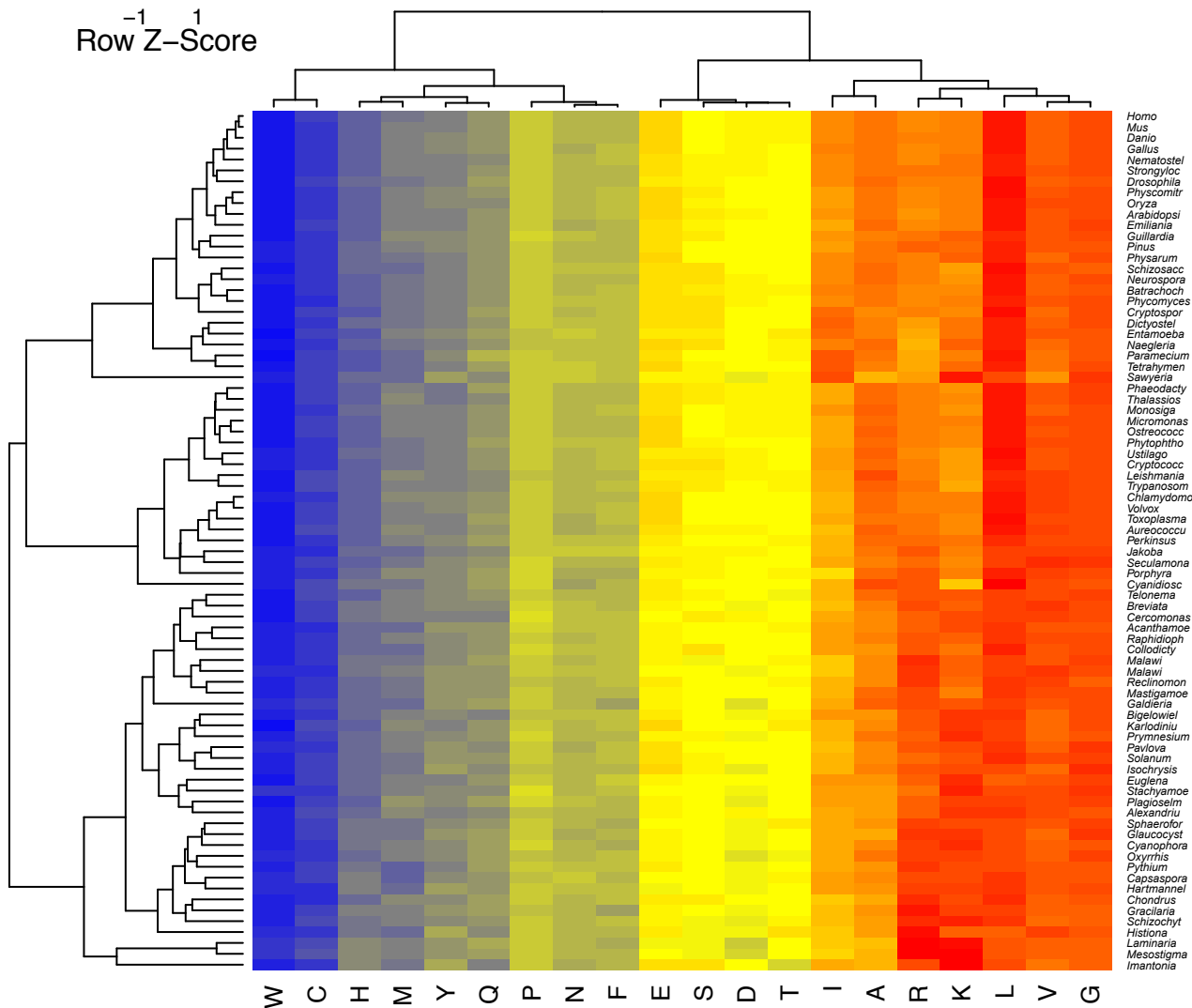


Color Key

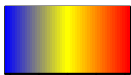


-1 1  
Row Z-Score

B

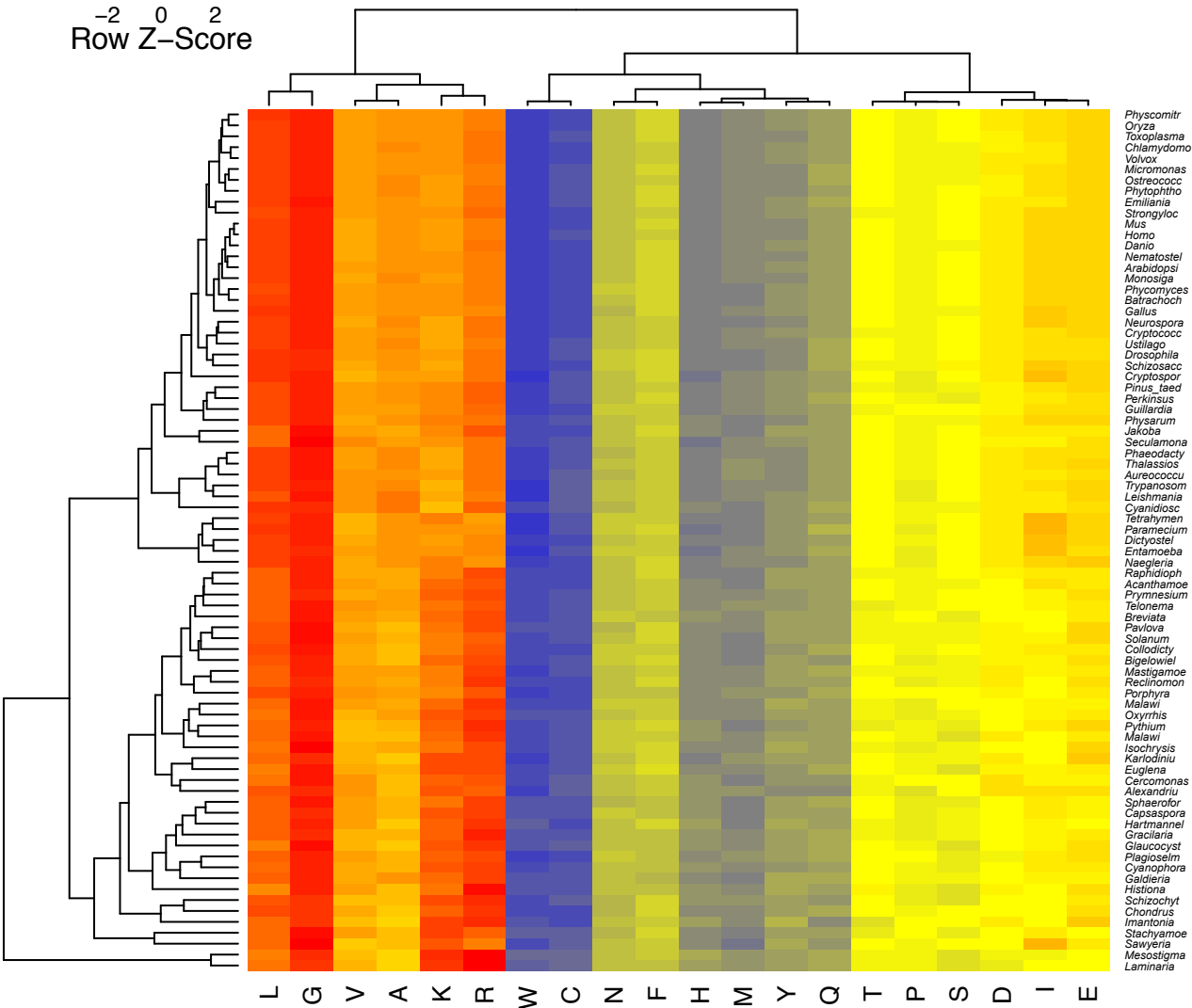


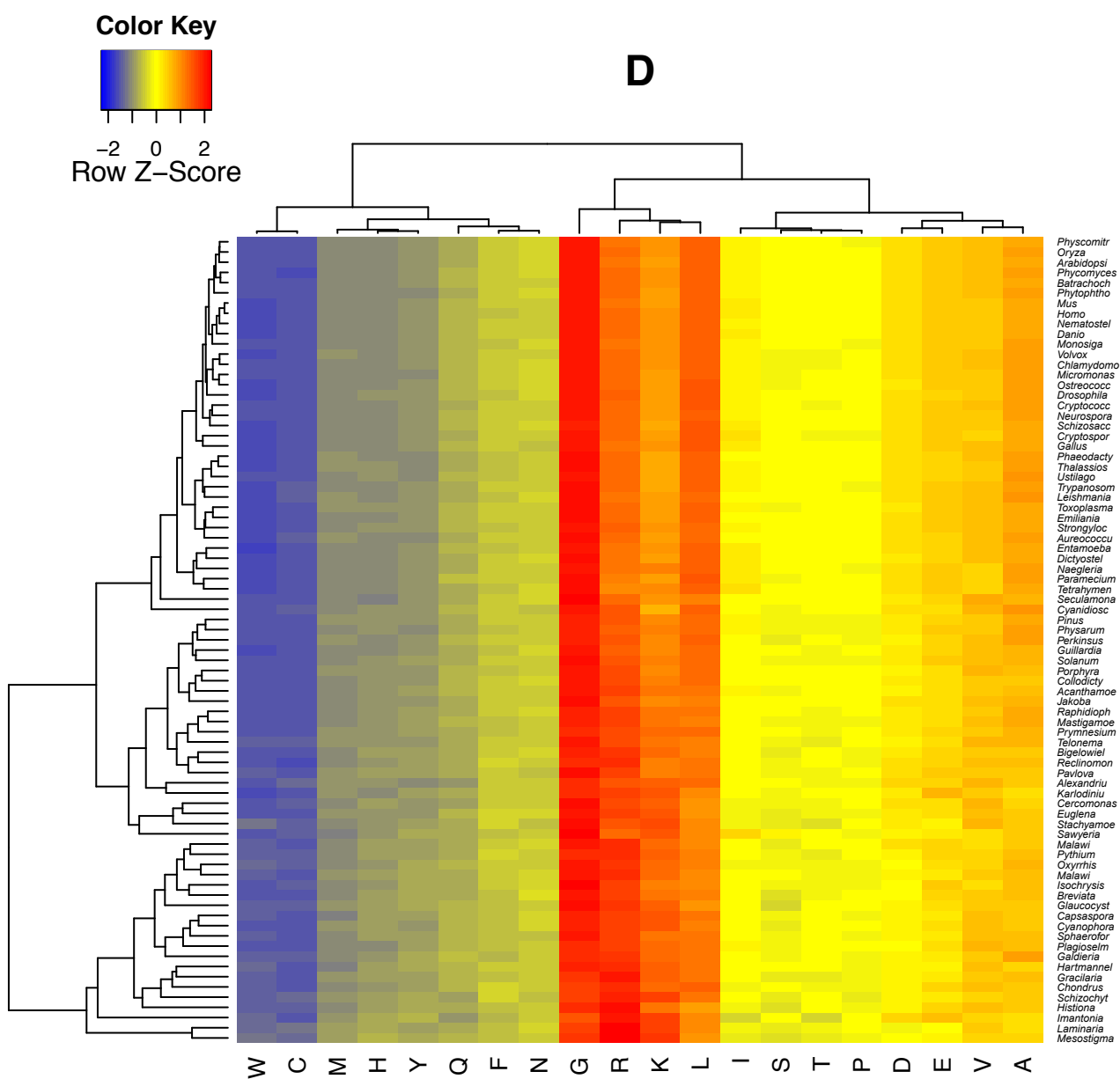
Color Key



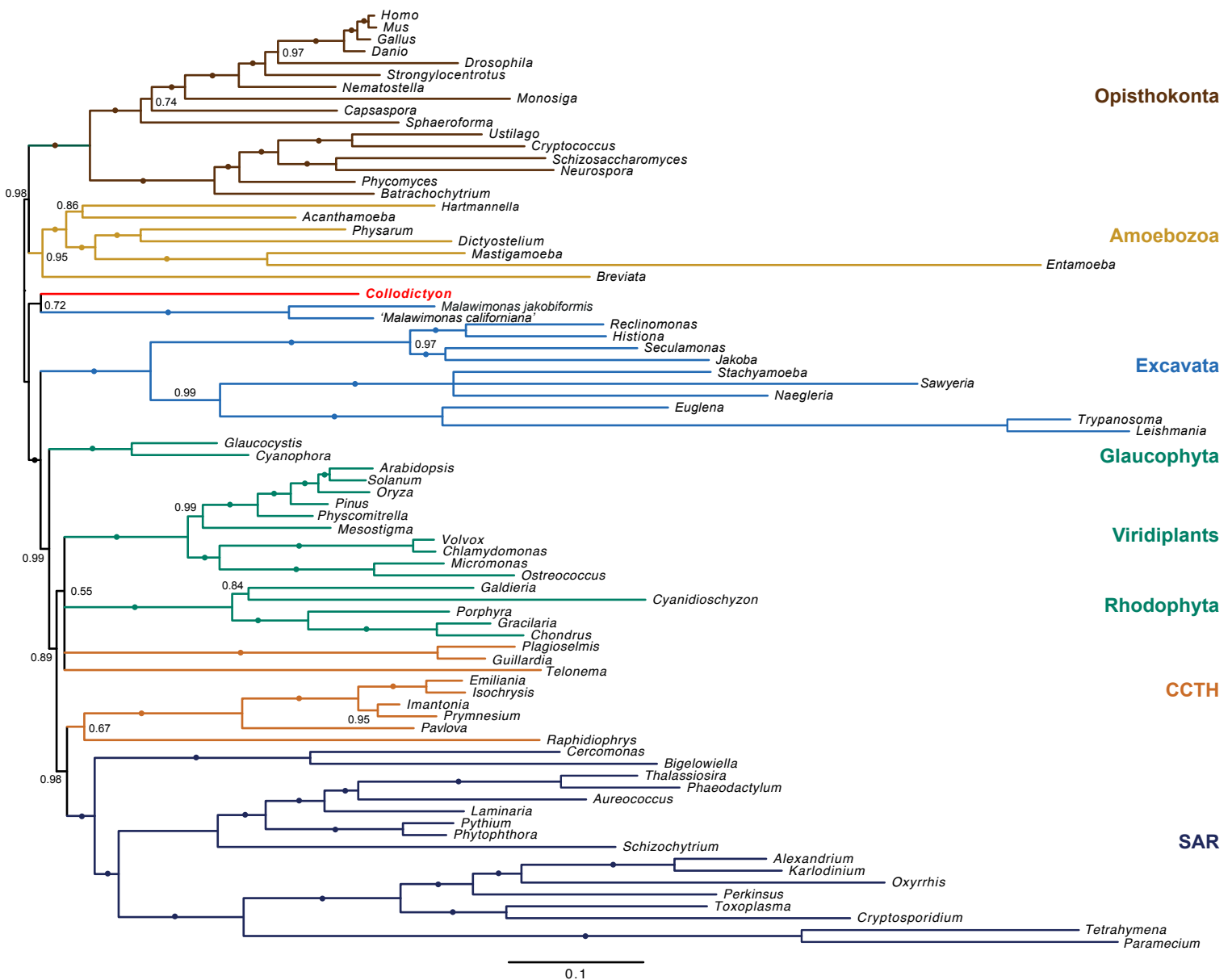
-2 0 2  
Row Z-Score

C

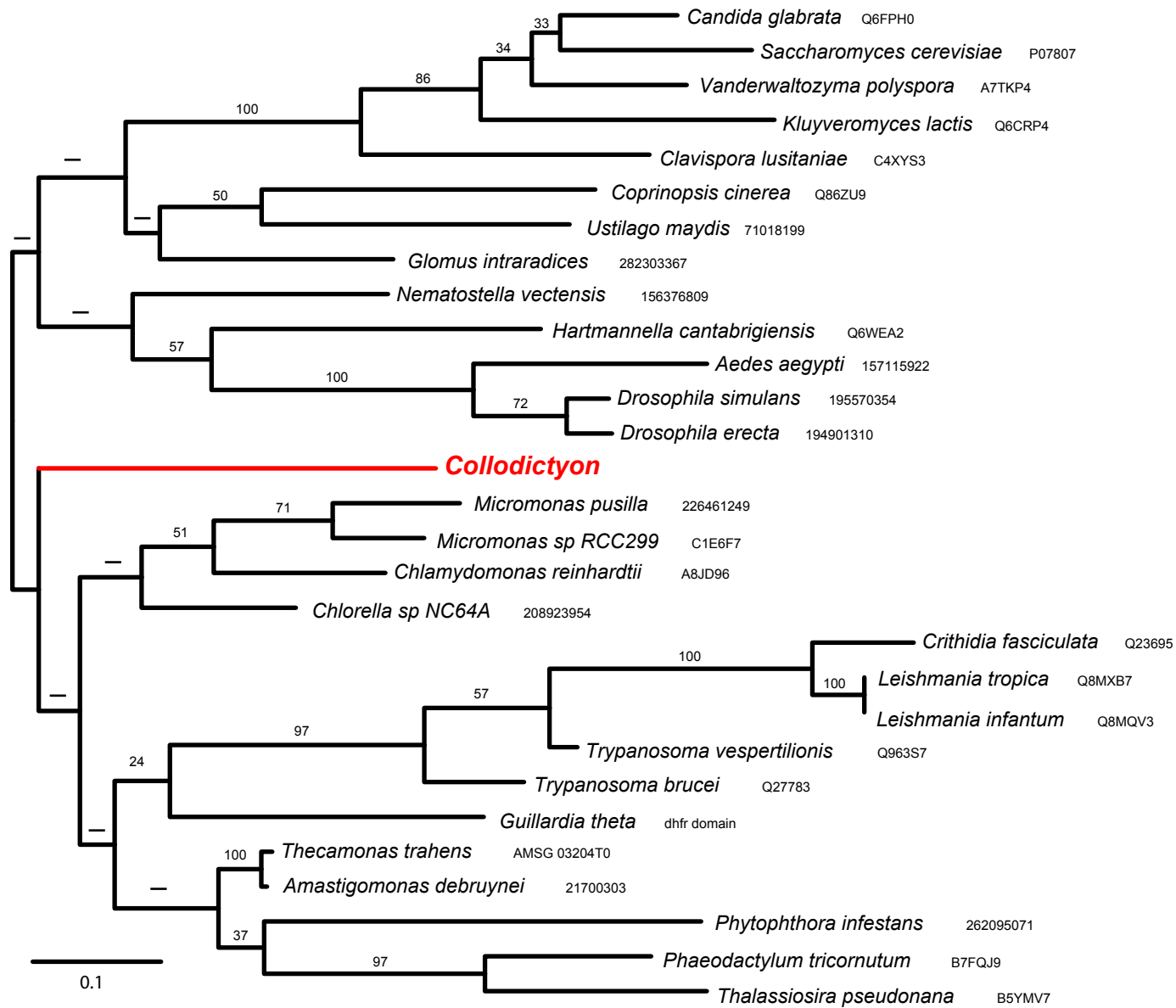




**Supplementary Figure S8.** Two-dimensional 'heatmap' showing the hierarchical clustering of 79 taxa (y-axis) and 20 types of amino acids (x-axis). The clustering was based on the comparison of the amino acid composition for each taxon. (A), (B), (C) and (D) show amino acids composition biases among 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 *Colloidietyon* branch is enlightened by red color.