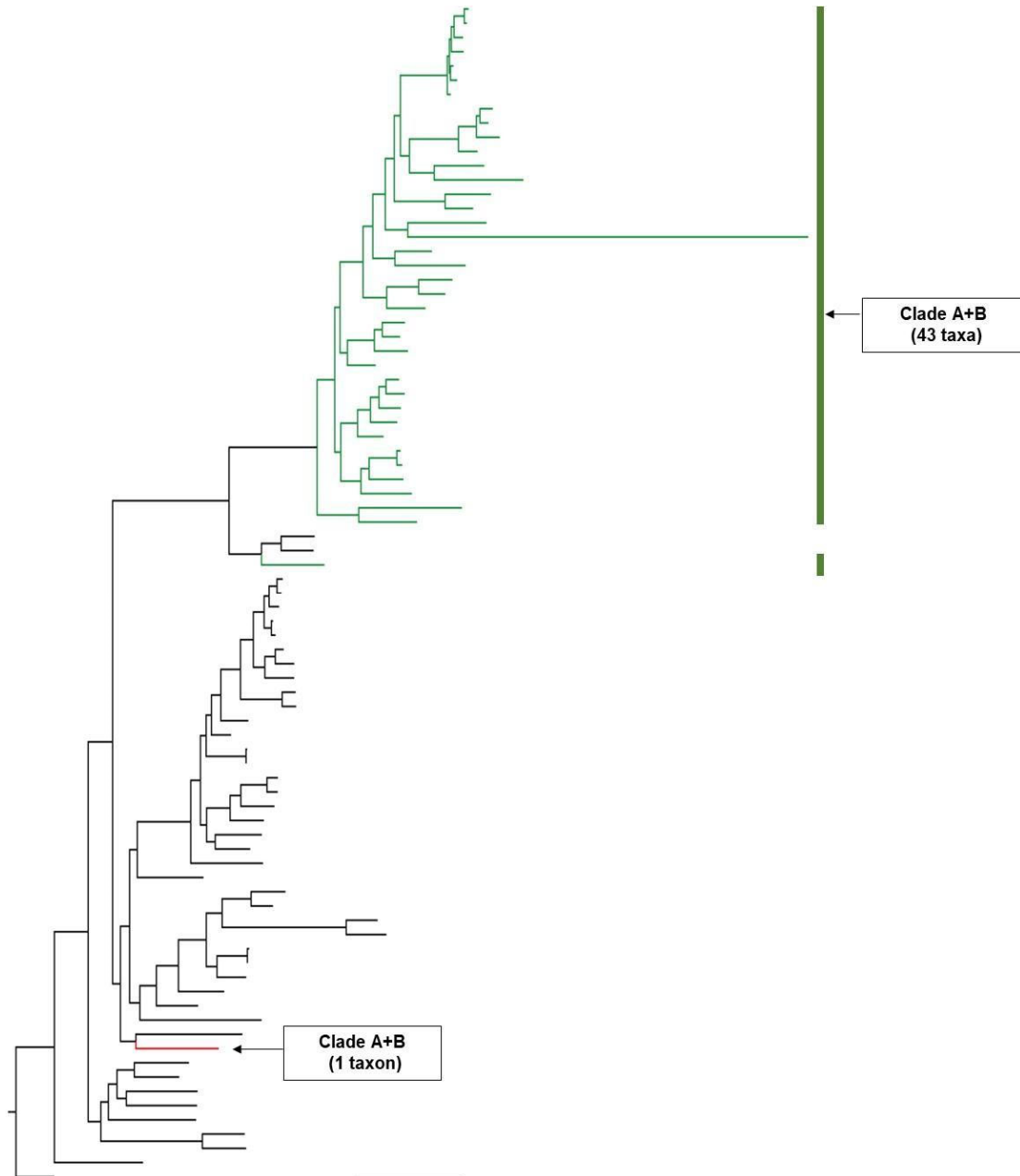


1  
2  
3  
4  
5  
6  
7  
8

**Supplementary Figure 1. Gene tree of BUSCO<sub>f</sub>EOG7W9S51 (7W9S51).**

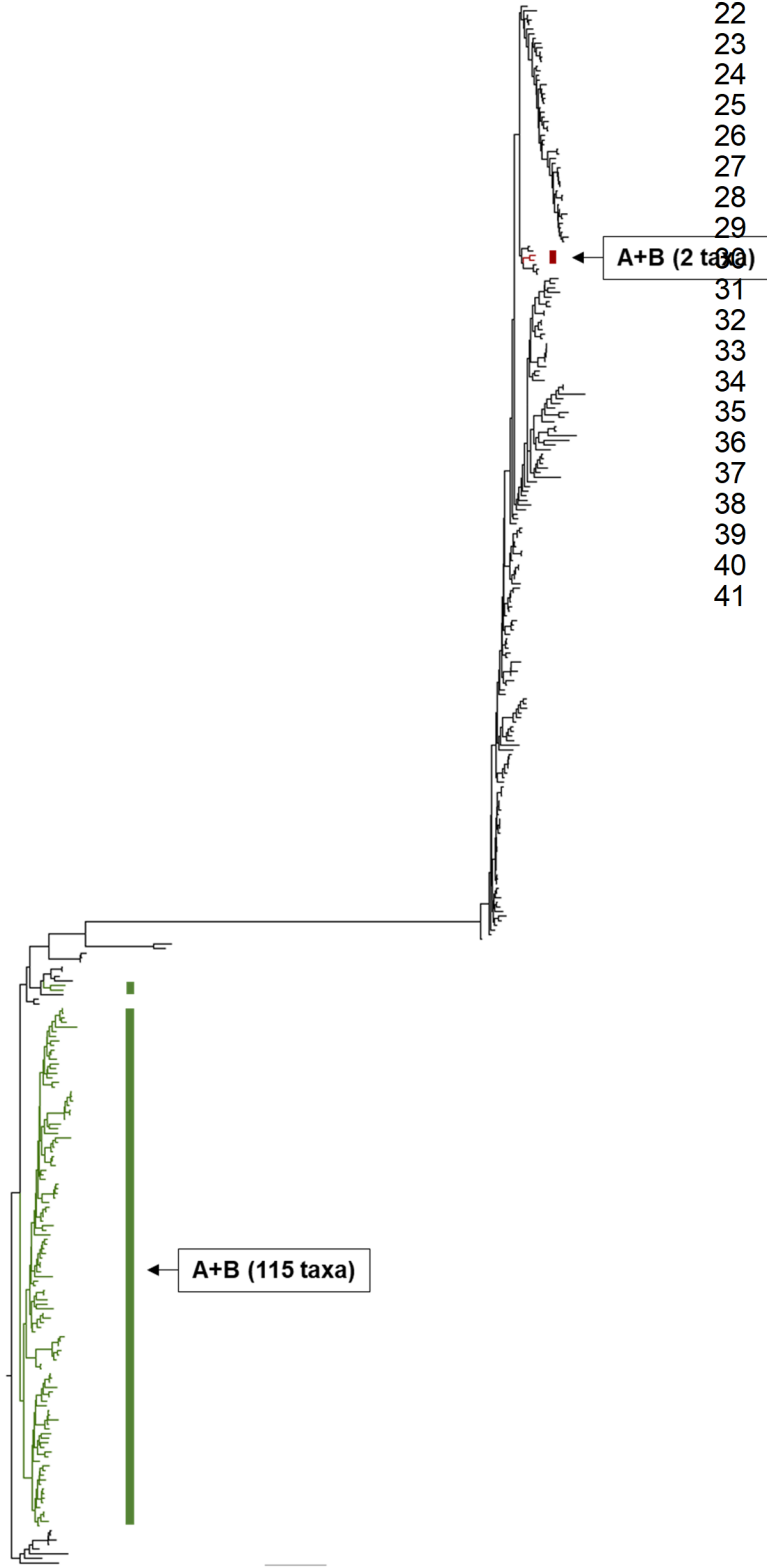
Species indicated by green tip branches belong to clade A+B in the phylogenomic tree (Fig. 2). The species forming clade A (*A. rubescence*) is marked by dotted green branch. The scale bar is in the units of the number of substitutions per site. The gene tree was inferred using the ML method and obtained from Shen et al. (2017).



10

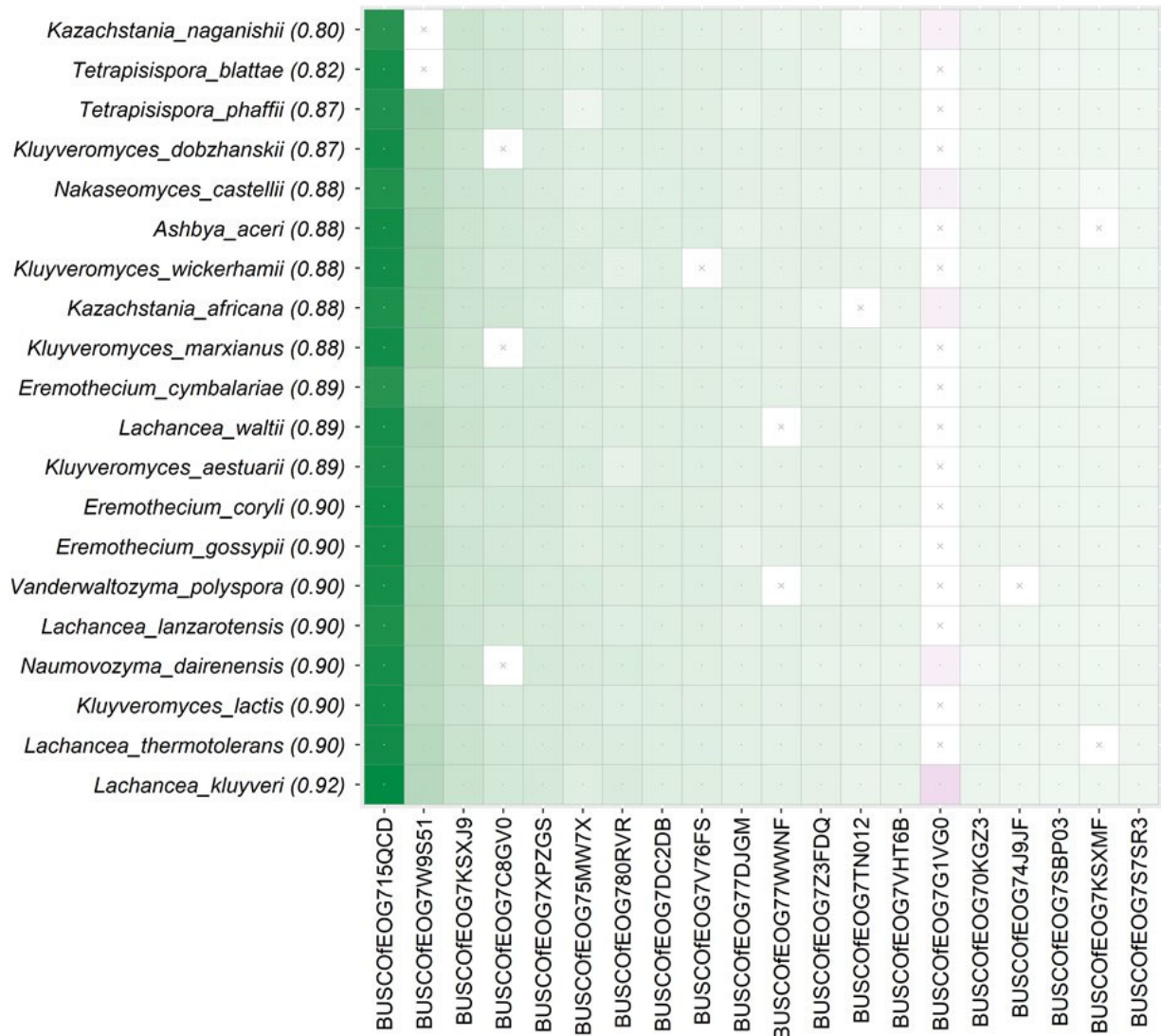
11 **Supplementary Figure 2. Gene tree of *BUSCOfEOG7TN012* (7TN012).**

12 Species indicated by red and green tip branches lead to species belonging to clade A+B in the  
 13 phylogenomic tree (see Fig. 2). Green tip branches led to species that are clustered with other species from  
 14 the clade A+B (Fig. 2) and received a positive GSC score (Fig. 3b). On the other hand, the red tip branch  
 15 is for the species (*A. rubescence*) from clade A+B, and received a negative GSC score for this gene. The  
 16 scale bar is in the units of the number of substitutions per site. The gene tree was inferred by using the ML  
 17 method and obtained from Shen et al. (2017).  
 18



21 **Supplementary Figure 3. Gene tree**  
22 **of *EOG09343FGH*.**

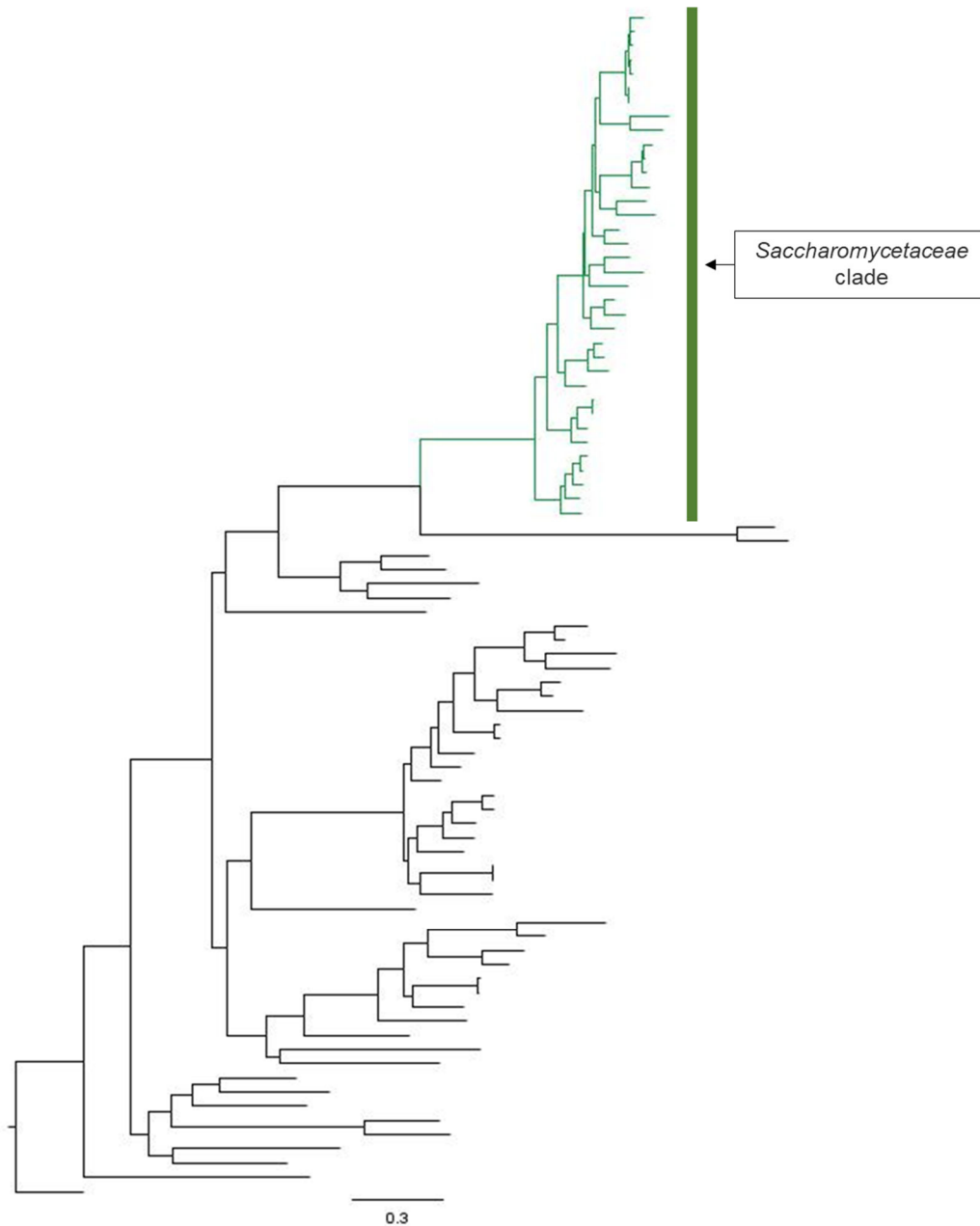
23 Species indicated by red and green  
24 tip branches lead to species  
25 belonging to clade A+B in the  
26 phylogenomic tree (see Fig. 2). Green  
27 tip branches led to species that are  
28 clustered with other species from the  
29 clade A+B and received positive GSC  
30 scores (Fig. 4). On the other hand,  
31 species with tip branch negative GSC  
32 (Fig. 4) for this gene are separated  
33 from the clade A+B by an  
34 unexpectedly long internal branch.  
35 The scale bar is in the units of the  
36 number of substitutions per site. The  
37 gene tree was inferred by using the  
38 ML method and obtained from Shen  
39 et al. (2018).  
40  
41



42  
43  
44  
45  
46  
47

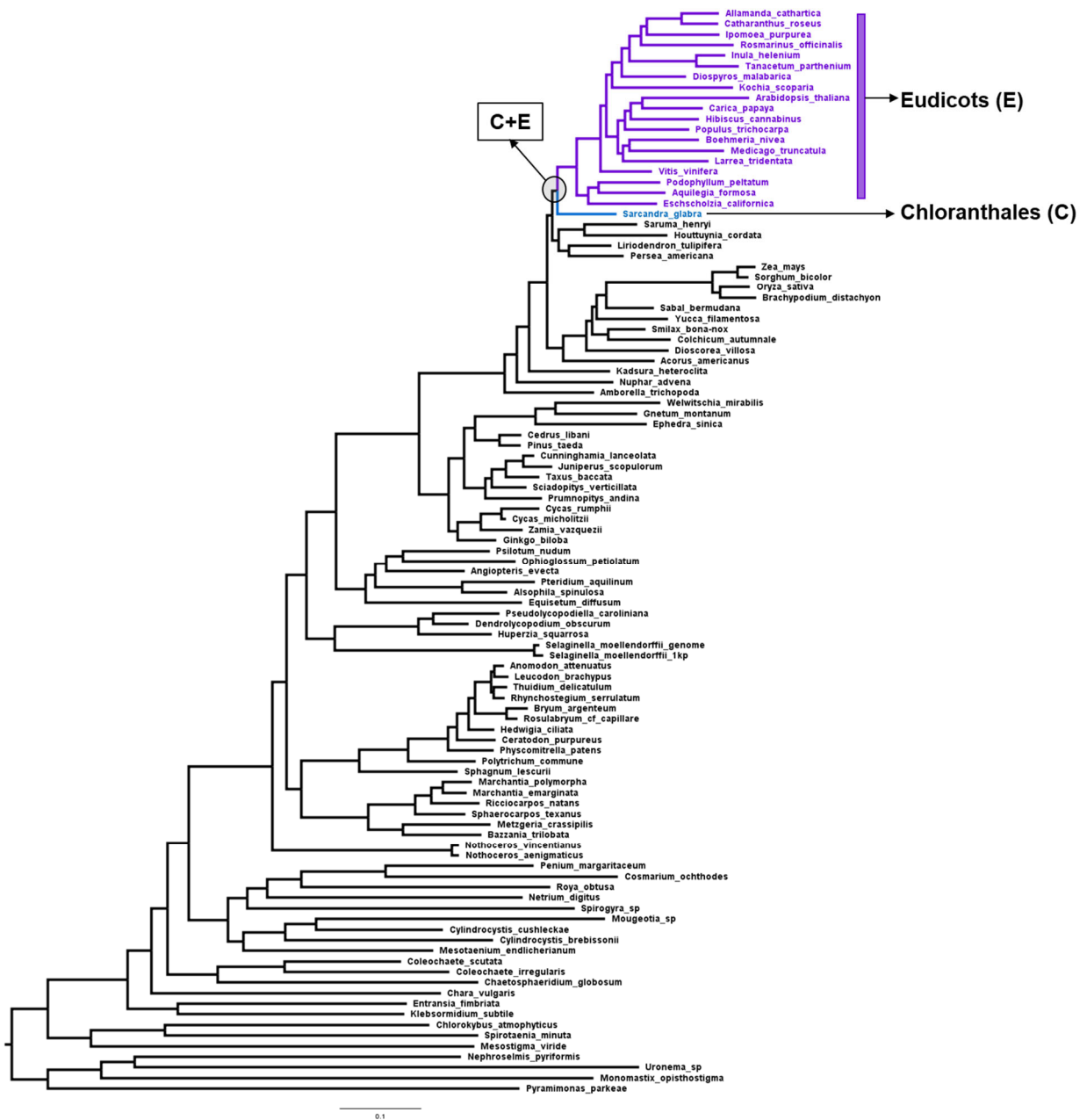
**Supplementary Figure 4. The M-grid for the control clade.**

The model grid for the *Saccharomycetaceae* clade. The M-grid shows that the gene *BUSCOfEOG715QCD* has the highest average GSC scores across all the taxa in the clade, whose CP is 0.80. SCP ranges from 1.00 to 0.80.



48 **Supplementary Figure 5. The ML phylogeny of *BUSCOfEOG715QCD* gene.**

49 Green branches mark tips belonging to species in *Saccharomycetaceae* clade and received positive GSC  
 50 values (Supplementary Fig. 4). No species in the clade received a negative GSC value. The scale bar is in  
 51 the units of the number of substitutions per site. The gene tree was inferred by using the ML method and  
 52 obtained from Shen et al. (2017).

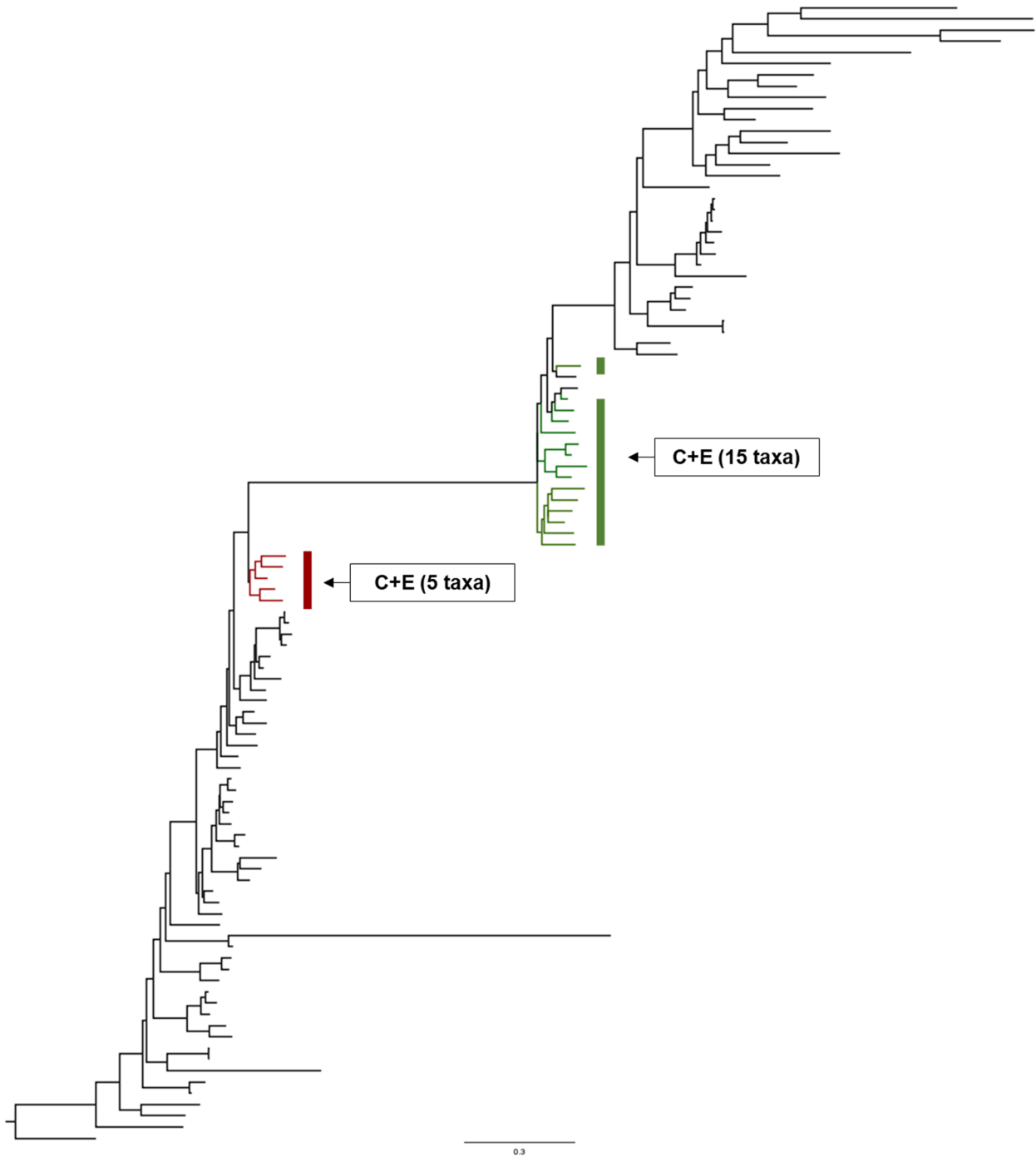


53

54 **Supplementary Figure 6. Plant phylogeny.**

55 The blue circle marks the clade containing 19 species of Eudicots (E; highlighted by purple-colored species)  
 56 and one species from Chloranthales (C; highlighted by blue-colored species). The clade C+E was used to  
 57 build the clade model. The highlighted blue branch leads to *Sarcandra glabra*, the only member of  
 58 Chloranthales. The ML phylogenetic tree of 103 plant species inferred from the concatenated sequence  
 59 alignment of 620 nuclear genes and obtained from Wickett et al. (2014).

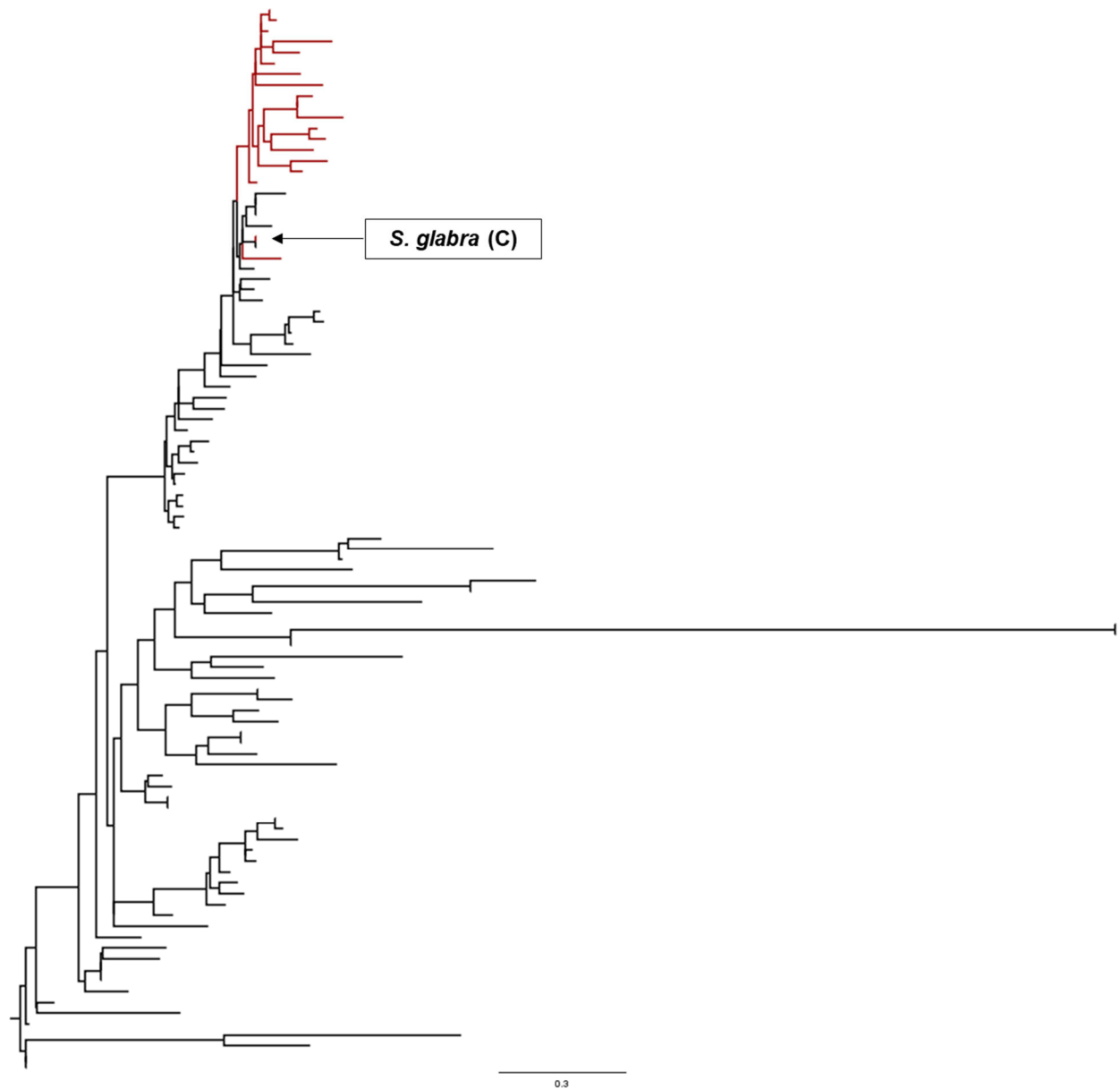
60  
61



62  
63  
64  
65  
66  
67

**Supplementary Figure 7. The ML phylogeny of 6040\_C12 gene.**

Green branches mark tips belonging to species in clade C+E and receive positive GSC (Fig. 6). Red tips are for species that are separated from the majority of species in the C+E clade by an extremely long branch and receive negative GSC (Fig. 6) for this gene. The scale bar is in the units of the number of substitutions per site. The gene tree was inferred by using the ML method and obtained from Shen et al. (2017).

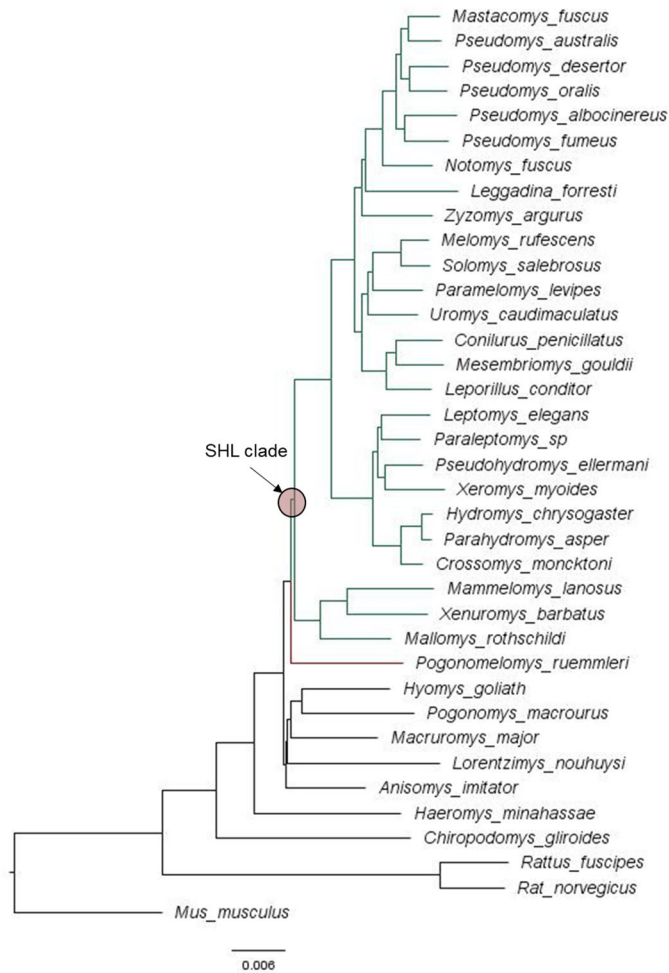


68 **Supplementary Figure 8. ML Gene tree of 5954\_C12.**

69 Red branches mark tips belonging to species in clade C+E and receive negative GSC (Fig. 6). The clade  
70 (Chloranthales; C) has been indicated by an arrow containing one species, *Sarcandra glabra*. The scale  
71 bar is in the units of the number of substitutions per site. The gene tree was inferred by using the ML method  
72 and obtained from Shen et al. (2017).

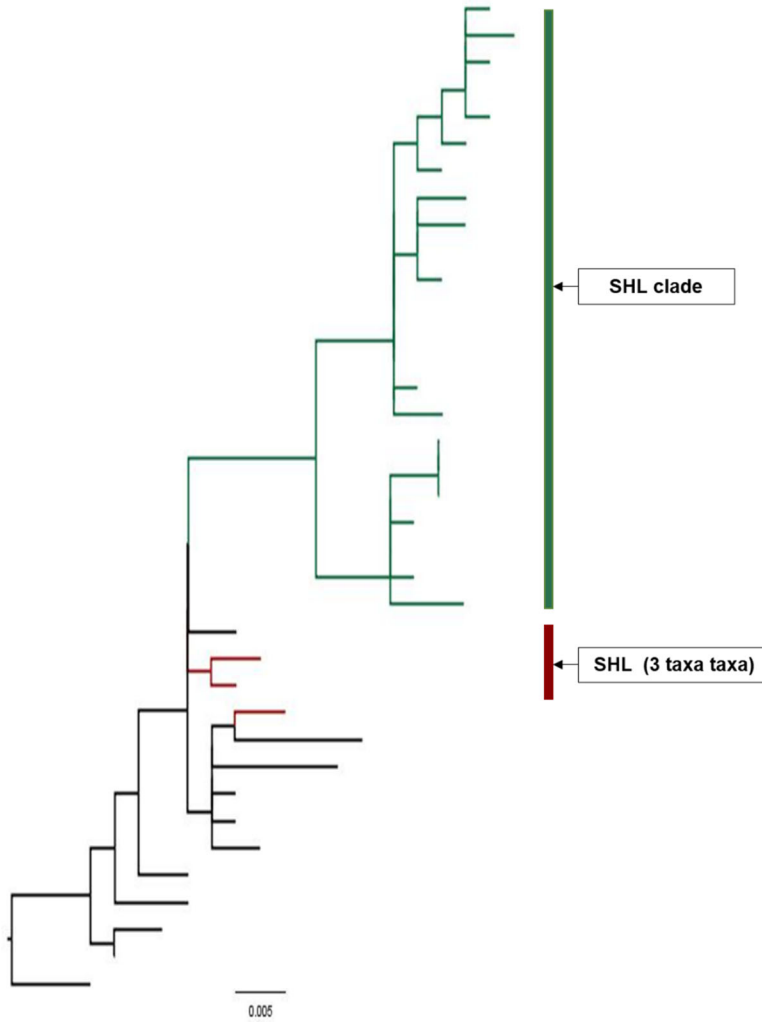


73  
74



75 **Supplementary Figure 9. Animal phylogeny.**

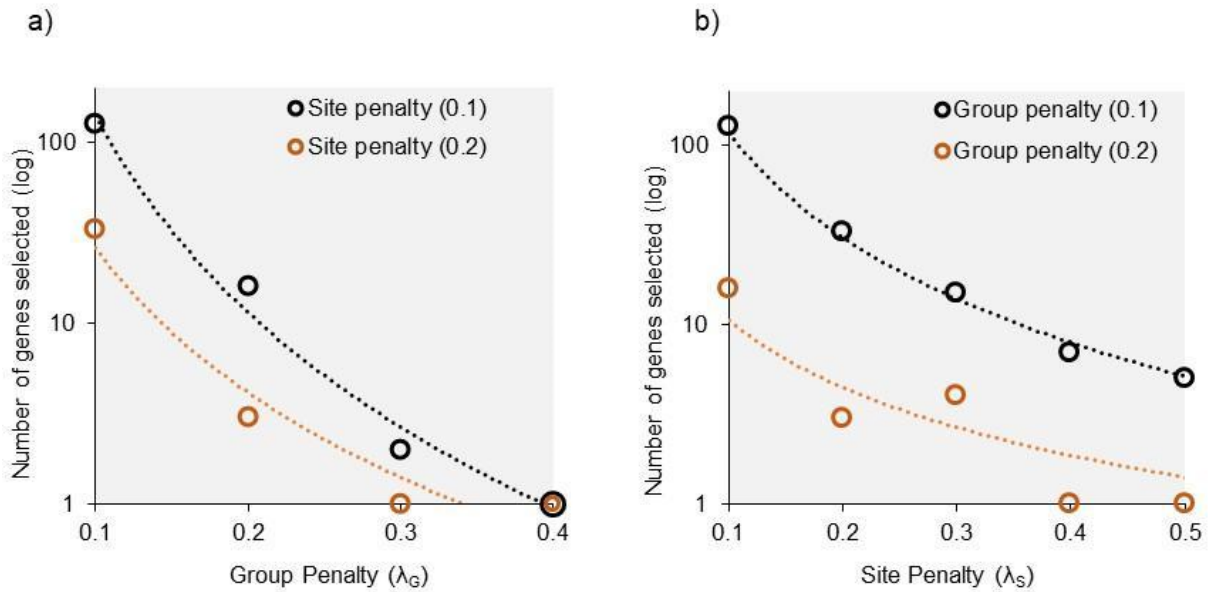
76 The red circle marks the clade containing 26 species from the clade SHL. A total of 26 species were used  
77 to build the clade model. The highlighted red branch leads to *Pogonomelomys ruemmleri*, whose placement  
78 is different in trees inferred from the concatenated supermatrix and MSC approaches. The ML phylogenetic  
79 tree of 37 rodents inferred from the concatenated sequence alignment of 1,245 nuclear genes and obtained  
80 from Roycroft et al. (2020).



81 **Supplementary Figure 10. The ML phylogeny of *KCNH5\_1\_rat*.**

82 Green branches mark tips belonging to species in clade SHL clade and received positive GSC (Fig. 7). Red  
 83 tips are for species that are separated from the majority of species in the SHL clade and receive negative  
 84 GSC (Fig. 7) for this gene. The scale bar is in the units of the number of substitutions per site. The gene  
 85 tree was inferred using ML approach and obtained from Shen et al. (2021).

86



87

88

89 **Supplementary Figure 11. Relationship between the penalty parameters and the number of genes**  
90 **selected in the ESL model.**

91 (a) A scatterplot showing the relationship between the group penalty parameter ( $\lambda_G$ ) and the number of  
92 selected genes (log scale) in the model while keeping the site penalty fixed ( $\lambda_S = 0.1$  and  $0.2$ ). As the group  
93 penalty parameter increases, the number of selected genes decreases exponentially, and only one gene is  
94 included in the clade model beyond a certain point ( $\lambda_G > 0.3$ ). (b) A scatterplot depicting the influence of  
95 the site penalty parameter ( $\lambda_S$ ) on the number of selected genes in the model while holding the group  
96 penalty ( $\lambda_G$ ) fixed. The dotted lines show power curve fits ( $R^2 > 0.8$ ).

97

98 **Supplementary Table 1. AU-test for candidate genes selected for A+B clade in fungi phylogeny.**  
99

Gene ID	<i>P</i> -value (species tree)
BUSCOfEOG7W9S51	$2.94 \times 10^{-55}$
BUSCOfEOG7TN012	0.00103
BUSCOfEOG715QCD	0.01270
BUSCOfEOG7C8GV0	0.00103
BUSCOfEOG77WWNF	0.00840

100  
101 Note: The AU-test was conducted for a pair of tree topologies; inferred gene tree and the species tree,  
102 utilizing 1000 bootstrap replicates in IQTREE. The estimated p-value represents the bootstrap probability  
103 for the species tree topology derived from 1000 bootstrap replicates. A p-value of less than 0.05 suggests  
104 that the species tree is rejected for the given gene sequence alignment.  
105  
106  
107

108 **Supplementary Table 2. Time taken (minutes) in *DrPhylo* analysis.**  
 109

Dataset (Type)	Clade	Number of taxa in the clade	Total number of genes	Genes selected in the model	Count of multi-gene Models built	Total time (minutes)
Fungus (AA)	A+B	44	1,233	124	23	33.0
Fungus (AA)	Control	36	1,233	244	55	58.0
Extended Fungus (AA)	A+B	117	1,292	87	24	52.0
Plant (DNA)	C+E	20	620	286	55	5.5
Animal (DNA)	SHL	10	1,245	260	51	8.7

110  
 111 Note: Dataset names correspond to the species under analysis, while data type refers to the base type  
 112 within each dataset (AA: Amino Acid). The clade ID is the identification code for the clades examined in the  
 113 main text. We have also included the computational time required for comprehensive *DrPhylo* analysis for  
 114 each clade. All the analyses were performed on a windows desktop computer with eight cores and 64 GB  
 115 of RAM.