

1 DESCRIPTION OF SUPPLEMENTARY FILES

2 *List of Supplementary datasets*

3 **Supplementary dataset 1:** Supermatrix of 171 gene clusters

4 **Supplementary dataset 2:** Supermatrix of non-missing-gene dataset

5 **Supplementary dataset 3:** Supermatrix of high alignment quality dataset

6 **Supplementary dataset 4:** Supermatrix of dataset with strong phylogenetic signal (average

7 BP  $\geq$  60%)

8 **Supplementary dataset 5:** Supermatrix of dataset with strong phylogenetic signal (average

9 BP  $\geq$  70%)

10

11 *Relationship between genome ID code used in Supplementary datasets 1-5 and species name*

<b>Genome ID Code</b>	<b>Species Name</b>
a_7	<i>Aplosporella prunicola</i>
a_97	<i>Pyrenophora tritici-repentis</i>
a_105	<i>Saitoella complicata</i>
a_130	<i>Agaricus bisporus bisporus H97</i>
a_131	<i>Agaricus bisporus burnettii JB137-s8</i>
a_132	<i>Amanita muscaria</i>
a_133	<i>Amanita thiersii</i>
a_134	<i>Auricularia subglabra</i>
a_135	<i>Bjerkandera adusta</i>
a_136	<i>Boletus edulis</i>
a_137	<i>Botryobasidium botryosum</i>
a_138	<i>Ceriporiopsis subvermispota</i>
a_139	<i>Cerrena unicolor</i>
a_140	<i>Coniophora puteana</i>
a_141	<i>Coprinopsis cinerea</i>
a_142	<i>Cronartium quercuum</i>
a_143	<i>Cryptococcus neoformans var neoformans JEC21</i>
a_144	<i>Cryptococcus neoformans var. grubii H99</i>
a_145	<i>Cryptococcus vishniacii</i>
a_146	<i>Dacryopinax sp. DJM 731 SSP1</i>
a_147	<i>Daedalea quercina</i>
a_148	<i>Dichomitus squalens</i>
a_149	<i>Dioszegia cryoxerica</i>

a_150	<i>Fistulina hepatica</i>
a_151	<i>Fomitiporia mediterranea</i>
a_152	<i>Fomitopsis pinicola</i>
a_153	<i>Galerina marginata</i>
a_154	<i>Ganoderma sp. 10597 SS1</i>
a_155	<i>Gloeophyllum trabeum</i>
a_156	<i>Gymnopus luxurians</i>
a_157	<i>Hebeloma cylindrosporum</i>
a_158	<i>Heterobasidion irregulare</i>
a_159	<i>Hydnomerulius pinastri</i>
a_160	<i>Hypholoma sublateritium</i>
a_161	<i>Jaapia argillacea</i>
a_162	<i>Laccaria amethystina</i>
a_163	<i>Laccaria bicolor</i>
a_164	<i>Laetiporus sulphureus</i>
a_165	<i>Malassezia globosa</i>
a_166	<i>Melampsora laricis-populina</i>
a_167	<i>Mixia osmundae</i>
a_168	<i>Neolentinus lepideus</i>
a_169	<i>Paxillus involutus</i>
a_170	<i>Paxillus rubicundulus</i>
a_171	<i>Phanerochaete carnosae</i>
a_172	<i>Phanerochaete chrysosporium</i>
a_173	<i>Phlebia brevispora</i>
a_174	<i>Phlebiopsis gigantea</i>
a_175	<i>Piloderma croceum</i>
a_176	<i>Piriformospora indica</i>
a_177	<i>Pisolithus microcarpus</i>
a_178	<i>Pisolithus tinctorius</i>
a_179	<i>Pleurotus ostreatus PC15</i>
a_180	<i>Pleurotus ostreatus PC9</i>
a_181	<i>Plicaturopsis crispa</i>
a_182	<i>Polyporus arcularius</i>
a_184	<i>Postia placenta MAD-698-R-SB12</i>
a_185	<i>Puccinia graminis</i>
a_186	<i>Punctularia strigosozonata</i>
a_187	<i>Rhodotorula graminis</i>
a_188	<i>Schizophyllum commune</i>
a_189	<i>Scleroderma citrinum</i>
a_190	<i>Sebacina vermifera</i>
a_191	<i>Serpula lacrymans S7.3</i>
a_192	<i>Serpula lacrymans S7.9</i>
a_193	<i>Serpula lacrymans-shastensis</i>
a_194	<i>Sphaerobolus stellatus</i>
a_195	<i>Sporobolomyces roseus</i>

a_196	<i>Stereum hirsutum</i>
a_197	<i>Suillus luteus</i>
a_198	<i>Trametes versicolor</i>
a_199	<i>Tremella mesenterica</i>
a_200	<i>Tulasnella calospora</i>
a_201	<i>Ustilago maydis</i>
a_203	<i>Wolfiporia cocos</i>
a_204	<i>cryptococcus gattii r265</i>
a_205	<i>cryptococcus gattii wm276</i>
a_206	<i>Exobasidium vaccinii</i>
a_207	<i>Malassezia sympodialis</i>
a_208	<i>Pseudozyma antarctica</i>
a_209	<i>Pseudozyma hubeiensis</i>
a_210	<i>Sporisorium reilianum</i>
a_211	<i>Tilletiaria anomala</i>
a_212	<i>Atractiellales sp.</i>
a_213	<i>Puccinia striiformis</i>
a_214	<i>Agaricostilbum hyphaenes</i>
a_215	<i>Heterogastridium pycnidioideum</i>
a_216	<i>Naiadella fluitans</i>
a_217	<i>Rhodotorula minuta</i>
a_218	<i>Tritirachium sp. CBS 265.96</i>
a_219	<i>Sporobolomyces linderiae</i>

1

2 *List of Supplementary Tables*

3 **Table S1:** Information of 91 fungal genomes investigated in this study

4 **Table S2:** Domain structure and conservation of 171 gene clusters

5 **Table S3:** Enrichment analysis of functional categories of 78 gene clusters that have single-copy  
6 orthologs in yeast genome

7 **Table S4:** Approximately Unbiased test of relationships between the three subphyla

8 **Table S5:** Other tests of relationships between the three subphyla using CONSEL

9 **Table S6:** Gene cluster removal in the evaluation of potential sequence bias

10 **Table S7:** Stability of relationships between the three subphyla in CVTree

11

12 *Supplementary Figure Legends*

1 **Figure S1:** Distribution of peptide length of 171 gene clusters.

2 **Figure S2:** Functional categories of 171 gene clusters. (a-c) Bar chart of GO annotation in the  
3 domain of Biological Process, Cellular Component and Molecular Function. SeqNr denotes the  
4 number of sequences which have annotated hits in the nr database.

5 **Figure S3:** ML phylogeny of 91 fungal genomes using 171 single-copy gene clusters. This tree  
6 is constructed using LG+G+I+F model. Values on branches denote bootstrap support.

7 **Figure S4:** NJ phylogeny of 91 fungal genomes using 171 single-copy gene clusters. This tree is  
8 constructed using JTT+G model. Values on branches denote bootstrap support.

9 **Figure S5:** MP phylogeny of 91 fungal genomes using 171 single-copy gene clusters. Tree  
10 searching method used here is Subtree-Pruning-Regrafting. Values on branches denote bootstrap  
11 support.

12 **Figure S6:** Bayesian phylogeny of 91 fungal genomes using 171 single-copy gene clusters. This  
13 tree is constructed using WAG+I+G model. Values on branches denote posterior probability.

14 **Figure S7:** Full majority rule consensus tree of 91 fungal genomes. This tree is consensus of  
15 phylogeny showed in Figure S3-S6.

16 **Figure S8:** ML phylogeny derived from the non-missing-gene dataset. This tree is constructed  
17 using LG+G model. Values on branches denote bootstrap support.

18 **Figure S9:** ML phylogeny derived from the high alignment quality dataset. This tree is  
19 constructed using LG+G+I+F model. Values on branches denote bootstrap support.

20 **Figure S10:** ML phylogeny derived from the dataset with strong phylogenetic signal (average  
21 BP  $\geq$  60%). This tree is constructed using LG+G+I+F model. Values on branches denote  
22 bootstrap support.

23 **Figure S11:** ML phylogeny derived from the dataset with strong phylogenetic signal (average

1 BP  $\geq 70\%$ ). This tree is constructed using LG+G+I+F model. Values on branches denote  
2 bootstrap support.

3 **Figure S12:** Distribution of indices that indicate sequence bias. (a) Density plot of average upper  
4 quartile of LB score of gene clusters. Here threshold = 60 represent a stringent criterion. (2)  
5 Density plot of average upper quartile of LB score of gene clusters. Here threshold = 70  
6 represent a loose criterion. (c) Density plot of standard deviation of LB score of gene clusters  
7 (threshold = 50). (d) Density plot of  $R^2$  of linear regression between patristic and uncorrected  
8 distances of gene clusters. Here threshold = 0.65 represent a stringent criterion. (e) Density plot  
9 of  $R^2$  of linear regression between patristic and uncorrected distances (threshold = 0.3) of gene  
10 clusters. Here threshold = 0.3 represent a loose criterion. (f) Density plot of slope of the linear  
11 regression between patristic and uncorrected pairwise distances (threshold = 0.1) of gene  
12 clusters.

13 **Figure S13:** ML phylogenies derived from datasets after removal of gene clusters that  
14 potentially introduce sequence bias. All trees were constructed using LG+G+I+F model. Values  
15 on branches denote bootstrap support. (a) Phylogeny derived from the dataset by discarding gene  
16 clusters with high mean value of LB score (threshold = 60). (b) Phylogeny derived from the  
17 dataset by discarding gene clusters with high mean value of LB score (threshold = 70). (c)  
18 Phylogeny derived from the dataset by discarding gene clusters with high standard deviation of  
19 LB score (threshold = 50). (d) Phylogeny derived from the dataset by discarding gene clusters  
20 with high mean value of LB score (threshold = 70) or high standard deviation of LB score  
21 (threshold = 50). (e) Phylogeny derived from the dataset by discarding gene clusters with low  
22 slope value (threshold = 0.1). (f) ML phylogeny derived from the dataset by discarding gene  
23 clusters with low  $R^2$  value (threshold = 0.3). (g) Phylogeny derived from the dataset by

1 discarding gene clusters with low  $R^2$  value (threshold = 0.65). (h) Phylogeny derived from the  
2 dataset by discarding gene clusters with low  $R^2$  value (threshold = 0.3) or low slope value  
3 (threshold = 0.1). (i) Phylogeny derived from the dataset by discarding gene clusters with low  $R^2$   
4 value (threshold = 0.65) or low slope value (threshold = 0.1). (j) Phylogeny derived from the  
5 dataset by discarding gene clusters with low  $R^2$  value (threshold = 0.65) or high mean value of  
6 LB score (threshold = 60) or low slope value (threshold = 0.1).

7 **Figure S14:** Full CVTree of 91 fungal genomes using their whole-gene repertoires.