The large-scale *Agaricales* genomes shed light on the unique origin, evolution, development characteristics of subterranean mushroom *Agaricus sinodeliciosus*

Hongyun Lu[1,2*], Shingo Miyauchi[3*], Yingqi Jiang[1], Ying Shi[1], Elodie Drula[4,5], Brian Looney[6], Miroslav Caboň[7], Laure Fauchery[2], Nattapol Kraisitudomsook[8], Steven Ahrendt[9], Alan Kuo[9], Sajeet Haridas[9], Robert Riley[9], Stephen Mondo[9], Byoungnam Min[9], Kerrie Barry[9], Kurt LaButti[9], Bill Andreopoulos[9], Jasmyn Pangiinan[9], Andrew Tritt[9], Joe Wang[9], Anna Lipzen[9], Navneet Kaur[9], Mei Wang[9], Annegret Kohler[2], Otto Miettinen[10], Sundy Maurice[11], Slavomír Adamčík[7], Laszlo G. Nagy[12,13], Kabir Peay[14], Matthew E. Smith[15], Rytas Vilgalys[6], Joseph W. Spatafora[16], Igor V. Grigoriev[9,17], Qihe Chen[1,18**], Francis M. Martin[2, 19**]

List of Supplementary Figures

ID	Title
Fig S1	Habitat and fruiting body of *A. sinodeliciosus* in Qinghai-Tibet Plateau
Fig S2	The coverage of transposable elements identified in Agaricales
Fig S3	Overview of major transposable elements covering the genomes
Fig S4	Divergence of transposable elements
Fig S5	Genomic features explained by the fungal lifestyle
Fig S6	Trends of fungal lifestyle groups by total CAZymes.
Fig S7	Trends of fungal lifestyle groups by secreted protease
Fig S8	Trends of fungal lifestyle groups by total protease
Fig S9	Trends of fungal lifestyle groups by small secreted proteins
Fig S10	Trends of fungal lifestyle groups by secreted lipases
Fig S11	Trends of fungal lifestyle groups by total lipases
Fig S12	Top 10 key features identified in the common orthologous genes among the 95 species
Fig S13	Key secreted proteases between wood decayers and litter/soil/grass decomposers
Fig S14	Lifestyle-specific orthologous genes in the 95 fungi
Fig S15	Identified genes distinguishing wood and grass saprotrophs
Fig S16	Overview of orthologous gene groups
Fig S17	Orthologous gene groups identified among 122 Agaricales fungi
Fig S18	Fungal lifestyle-wise associated common orthologous genes
Fig S19	Heatmap and principal coordinate map of lignin-, hemicellulose-, and cellulose- degrading secretome gene in grass/litter/soil saprotroph
Fig S20	Overview of CAZymes specific to the substrates among the three Agaricus species
Fig S21	GO and KEGG enrichment of *A.sinodeliciosus* core gene
Fig S22	Average Nucleotide Identity (ANI) among *A. sinodeliciosus*, *A. sinodeliciosus* ZRL, *A. bisporus* JB 137-S8, *A bisporus* ARP23, *A bisporus* and *A. bisporus* H97 based on MUMmer (ANIm)
Fig S23	Collinearity analysis of reported *A. sinodeliciosus*, *A. sinodeliciosus* ZRL, *A. bisporus* JB 137-S8, *A bisporus* ARP23, *A bisporus* and *A. bisporus* H97
Fig S24	The functional annotation of the 610 species-specific genes in *A. sinodeliciosus* with various databases
Fig S25	Overview of tyrosine metabolic pathways in *A. sinodeliciosus* under different development stage
Fig S26	Whole genome sequencing of *A. sinodeliciosus* from Qinghai-Tibet Plateau
Fig S27	Soil physicochemical index detection of mycelia soil, topsoil of *A. sinodeliciosus*. and mushroom free soil in Qinghai-Tibet Plateau

Supplementary Methods

Cultivation of *A. sinodeliciosus*

The culture material of *Agaricus sinodeliciosus* was composed of 65% fermented wheat straw, 30% dry cow dung, 2% bran, 1% gypsum and 1% lime, with a moisture content of 60%-65%. A 35 cm × 9 cm × 9 cm polypropylene cultivation bag was used, with each bag containing about 300 g. The temperature of the mycelium germination stage was controlled at 20-25 ºC, protected from light, and the relative humidity of the air was maintained at about 65%. The temperature of the primary stage was 17-22 ºC and the relative humidity was 90%, and the substrates were harvested after 7-10 d of incubation after the formation of the primary stage.

Genome sequencing and assembly for *A. sinodeliciosus*

Pacbio Sequel II and Illumina Hiseq 4000 platforms were used to perform whole-genome CCS sequencing of monokaryotic mycelium from *Agaricus sinodeliciosus at* the Beijing Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). The subreads sequenced in PacBio CCS mode based on single-molecule real-time (SMRT) sequencing were processed using the CCS program of smrtlink to obtain higher accuracy CCS reads, and then assembled using the Hifiasm software [1]. finally, the assembled genomes were further error corrected using second-generation data using the Pilon software [2]. To obtain the final more accurate genome, and BUSCO v2.0 software was used to assess the integrity of the genome assembly [3]. Using HiC technology, combined with Illumina Hiseq high-throughput sequencing platform, sequencing data are evaluated and filtered to obtain high-quality data, which are used for genome assembly, assessment of assembly results, repetitive sequence prediction, gene prediction analysis, functional annotation, and other analyses to realize genome assembly at the chromosome level.

Gene prediction and annotation for *A. sinodeliciosus*

Protein-coding gene models were de novo predicted using TransDecoder v2.0 [4]. RepeatMasker (http://www.repeatmasker.org/) was adopted to predict repetitive sequences [5]. tRNA genes were predicted using tRNAscan-SE version 3.0 [6]. Gene functions were predicted with GO (http://geneontology.org/), KEGG (https://www.kegg.jp/), EOG (https://www.creative-proteomics.com/services/kogannotation-analysis-service.htm), NR (https://www.ncbi.nlm.nih.gov/protein/), Transporter Classification Database (http://www.tcdb.org), Fungal Cytochrome P450 Database (http://p450.riceblast.snu.ac.kr/cyp.php), Swiss-Prot (https://www.uniprot.org/), Pfam (http://pfam.xfam.org/) and CAZymes Database (http://www.cazy.org/). All predicted coding genes were aligned using DIAMOND version 2.0.2 with the cut-off values of e-value $\leq 1 \times 10^{-5}$ [7]. Secretory proteins were detected by SignalP and filterd by TMHMM [8]. Further analysis of secreted proteins by EffectorP[9] to predict effector proteins. antiSMASH program version 2.0.2 was employed to predict the gene clusters of secondary metabolites with default parameters[10].

Transcriptome sequencing analysis for *A. sinodeliciosus*

A. sinodeliciosus (Qinghai isolate) was cultivated on sterilized soil-compost mixture (2:1) in controlled environment chambers at 18°C with 70% relative humidity and 12:12 h light:dark cycle (100 μmol m^{-2} s^{-1} photosynthetically active radiation). Three developmental stages were harvested: vegetative mycelium (M, 21 days post-inoculation), primordia (P, 35 days, 3–5 mm in diameter), and mature fruiting bodies (CFB, 45 days, fully expanded caps). Three biological replicates per stage (each from independent cultures) were collected, flash-frozen in liquid nitrogen, and stored at −80°C.

Total RNA was extracted using TRIzol reagent (Invitrogen) and further purified using the RNeasy Plant Mini Kit (Qiagen) to remove polysaccharide contaminants. RNA integrity was assessed using an Agilent 2100 Bioanalyzer (RIN >8.0 for all samples). Poly(A)-enriched RNA-seq libraries were prepared using NEBNext Ultra II RNA Library Prep Kit and sequenced on Illumina NovaSeq 6000 (2×150 bp paired-end, ~30 million read pairs per library).

The raw sequencing reads were filtered using SeqPrep and Sickle with default parameters for the removal of Illumina adaptors and quality trimming [11]. The clean reads were mapped with TopHat2 [12]. Bioinformatics analysis was performed using a cloud platform (www.majorbio.com). After obtaining the read counts of genes/transcripts for each sample using the results of comparison to genome and genome annotation files, TPM conversion was

performed to obtain normalized gene/transcript expression levels using RSEM software [13].

Differential expression analysis was performed using DESeq2 v1.36.0 [14] with the default parameters. Genes with $|\log_2(\text{fold-change})| \geq 1$ and FDR-adjusted $P < 0.05$ were considered differentially expressed. Gene Ontology and KEGG pathway enrichment for differentially expressed genes were performed using clusterProfiler v4.4.4 with Fisher's exact tests (FDR < 0.05 threshold). Transcription factors were identified by Pfam domain annotation (PF00096, PF00172, PF00249, PF04082, and PF00010). Heatmaps were generated using pheatmap v1.0.12 with Z-score normalization.

Gene family clustering analysis

Gene families were identified and analyzed using OthoMCL (http://orthomcl.org/orthomcl/) [15]. Gene sets from each species were first filtered. When multiple transcripts (variable splicing) existed for a gene, only the transcript with the longest coding region was retained. Next, genes coding for proteins smaller than 50 amino acids and with internal stop codons were filtered out. The similarity relationship between protein sequences of all species was obtained by all-vs-all blastp, with the e value defaulting to e^{-5}. Finally, the above results were clustered using the OrthoMCL software comparison, with an expansion factor used of 1.5. The analysis resulted in conserved single-copy and multi-copy gene families between species.

Gene family expansion and contraction analysis

Based on the statistical results of gene family clustering, gene families whose gene numbers were anomalous in individual species were filtered (one species had a gene number greater than or equal to 200, and all other species had a gene number less than or equal to 2). Gene family expansion and contraction analyses were performed using CAFE software (http://sourceforge.net/projects/cafehahnlab/) based on divergence times and family clustering statistics [16]. CAFE results were filtered and significantly enriched results were screened for Family-wide P-value<0.05 and Viterbi P-values<0.05.

Positive selection analysis

The protein sequences of single-copy gene families of the species analyzed for positive selection were compared with multiple sequences by MUSCLE. The protein sequence alignment results were then used as a template to generate multiple sequence alignment results for the corresponding CDS. For each gene family, the branch-site model in the codeml tool of PAML was used to detect whether the genes of the foreground species of the gene family were under positive selection [17].

Soil physicochemical index detection

To investigate the physicochemical properties of soils in the habitat of *A. sinodelicuous*, a unique subterranean edible fungus of the Qinghai-Tibet Plateau, analyses were conducted on the mycelial soil, topsoil, and non-fungal soil from the fungal growth area, thus clarifying the influence of soil organic nutrients and salinity-alkalinity levels on the growth of this fungus. As depicted in Fig.S27, pH in mycelia soil is neutral, lower than topsoil, and mushroom free soil, suggesting that the soil where *A. sinodelicuous* produce fruiting body is neutral soil with low alkalinity. The TK in mycelia soil is relatively higher than topsoil and mushroom free soil, whereas the AK is significantly lower than topsoil and soil without mushroom. The SOM, TN, HN, TP, AP, OC and CEC were higher in mycelia soil, suggesting that soil where *A. sinodelicuous* growth possess the direct supply capacity of C, N, P and K nutrients in soil, and have good ability to retain positive ions in soil. Whereas, the topsoil and mushroom-free soil is low in nutrients and strong in salinity, which is not suitable for *A. sinodelicuous* growth.

Descriptions of *A. sinodeliciosus* genome

The genome of monokaryotic *A. sinodeliciosus* was 36.37 Mb, with a GC content of 45.6%, composed of 106 scaffolds and an N50 of 2.87 Mb, exhibiting 0.16% heterozygosity. BUSCO analysis showed 99.87% genome coverage, with 265 complete BUSCO genes, reflecting 91.38% gene completeness. NR database comparison revealed 95.23% similarity to *A. bisporus* (Fig. S26A-B). Statistical analysis indicated that 29.12% of the genome consists of repetitive sequences (Table S1), a proportion larger than in *A. bisporus*. Hi-C assembly (Fig. S26C) showed a peak at 350 bp in the insertion fragment length distribution, indicating minimal dispersion and

confirming the suitability of the Hi-C library for further analyses. The high-quality whole-genome mapping was depicted in Fig. S26D. The Hi-C assembly anchored 32.15 Mb to 12 pseudo-chromosomes, achieving an 88.39% mounting rate and resulting in a chromosome-level assembly (Fig. S26E; Table S2). A total of 448 genes were annotated in the CAZy database at University of Marseille, with GHs being the most abundant. These genes likely support the catabolism of lignocellulose for nutrient acquisition, including GH families involved in cellulose (GH1, GH3, GH5), hemicellulose (GH10, GH11, GH27, GH43, GH115), lignin (GH12, GH74), pectin (GH28, GH78, GH43, GH105), chitin (GH18), and starch (GH13, GH31) degradation [18].The GH18 chitinase gene family has been shown to be involved in fungal morphogenesis, including spore germination, mycelial elongation and branching [19, 20]. Therefore, these findings indicated that the *A. sinodeliciosus* can utilize lignocellulose for growth and development.

References

1. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722-736 (2017).
2. Walker BJ*, et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *Plos One* **9**, (2014).
3. Huang N, Li H. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39**, (2023).
4. Wang ZJ*, et al.* Transcriptomic analysis of codon usage patterns and gene expression characteristics in leafy spurge. *BMC Plant Biol.* **24**, (2024).
5. Tang DTP, Glazov EA, McWilliam SM, Barris WC, Dalrymple BP. Analysis of the complement and molecular evolution of tRNA genes in cow. *BMC Genomics* **10**, (2009).
6. Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic. Acids Res.* **25**, 955-964 (1997).
7. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59-60 (2015).
8. Teufel F*, et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023-+ (2022).
9. Sperschneider JJNP. EffectorP: predicting fungal effector proteins from secretomes using machine learning. (2016).
10. Medema MH*, et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic. Acids Res.* **39**, W339-W346 (2011).
11. St. John J. SeqPrep: Tool for stripping adaptors and/or merging paired reads with overlap into single reads.). https://github.com/jstjohn/SeqPrep (2023).
12. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
13. Tian J*, et al.* Discovery and remodeling of Vibrio natriegens as a microbial platform for efficient formic acid biorefinery. *Nat. Commun.* **14**, 7758 (2023).
14. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).
15. Chen F, Mackey AJ, Stoeckert CJ, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic. Acids Res.* **34**, D363-D368 (2006).
16. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269-1271 (2006).
17. Marra NJ*, et al.* White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. *Proc. Natl. Acad. Sci*. **116**, 4446-4455 (2019).
18. Dong W-g*, et al.* Chromosome-level genome sequences, comparative genomic analyses, and secondary-metabolite biosynthesis evaluation of the medicinal edible mushroom *Laetiporus sulphureus*. *Microbiol. Spectr.* **0**, e02439-02422.
19. Kuranda MJ, Robbins PW. Chitinase is required for cell separation during growth of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **266**, 19758-19767 (1991).
20. Guo Y, Liu Z, Fu Y, Li Y, Dai Y, Xiao S. Pan-Genomes Provide Insights into the Genetic Basis of Auricularia heimuer Domestication. **8**, 581 (2022).

Supplementary Figures

Fig S1. Habitat and fruiting body of *A. sinodeliciosus* in Qinghai-Tibet Plateau.

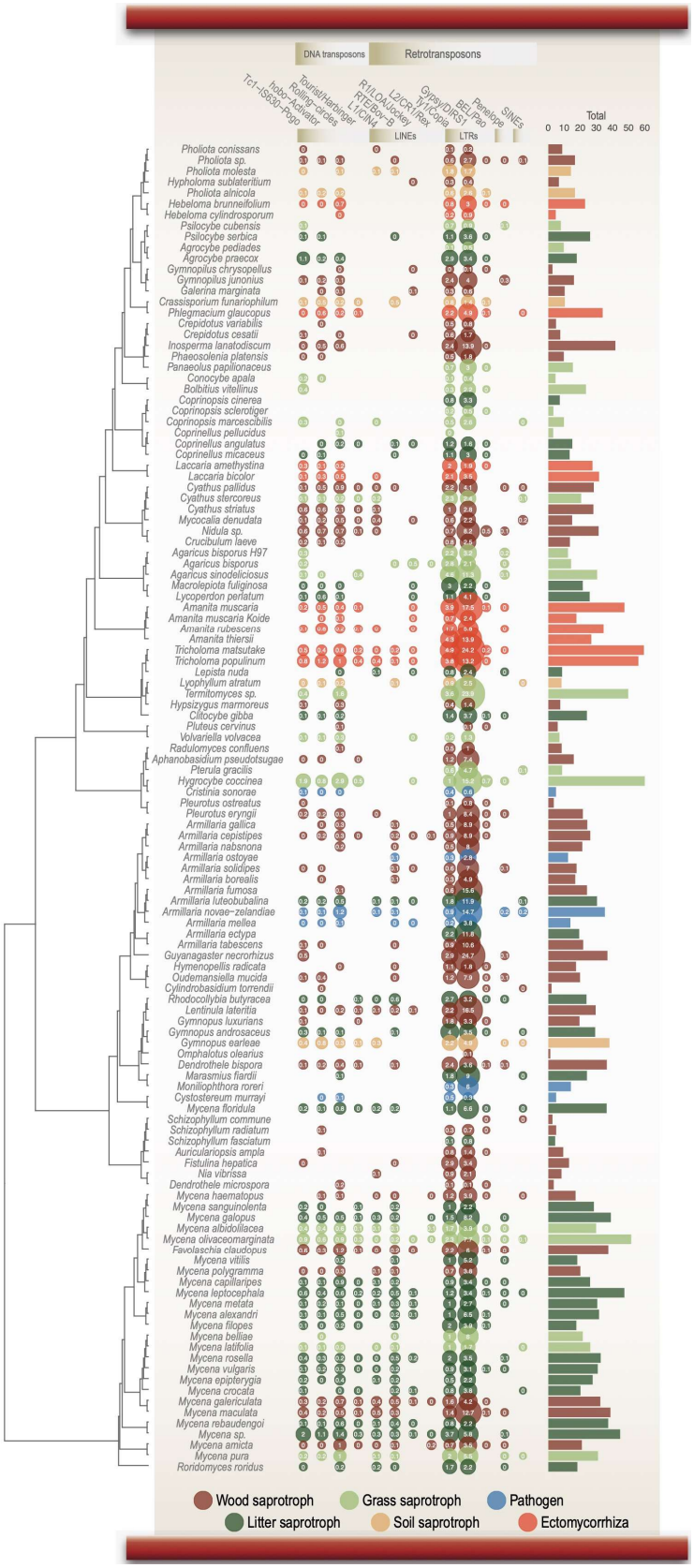

Fig.S2. The coverage of transposable elements (TEs) identified in *Agaricales*. LTR, long terminal repeat retrotransposons; non-LTR, non-long terminal repeat retrotransposons; DNA: DNA transposons; repetitive/SAT: simple repeats; unknown: unclassified repeated sequences. The bubble size is proportional to the coverage / copy number of each of TE (showed inside the bubbles). The bars show the total coverage per genome.

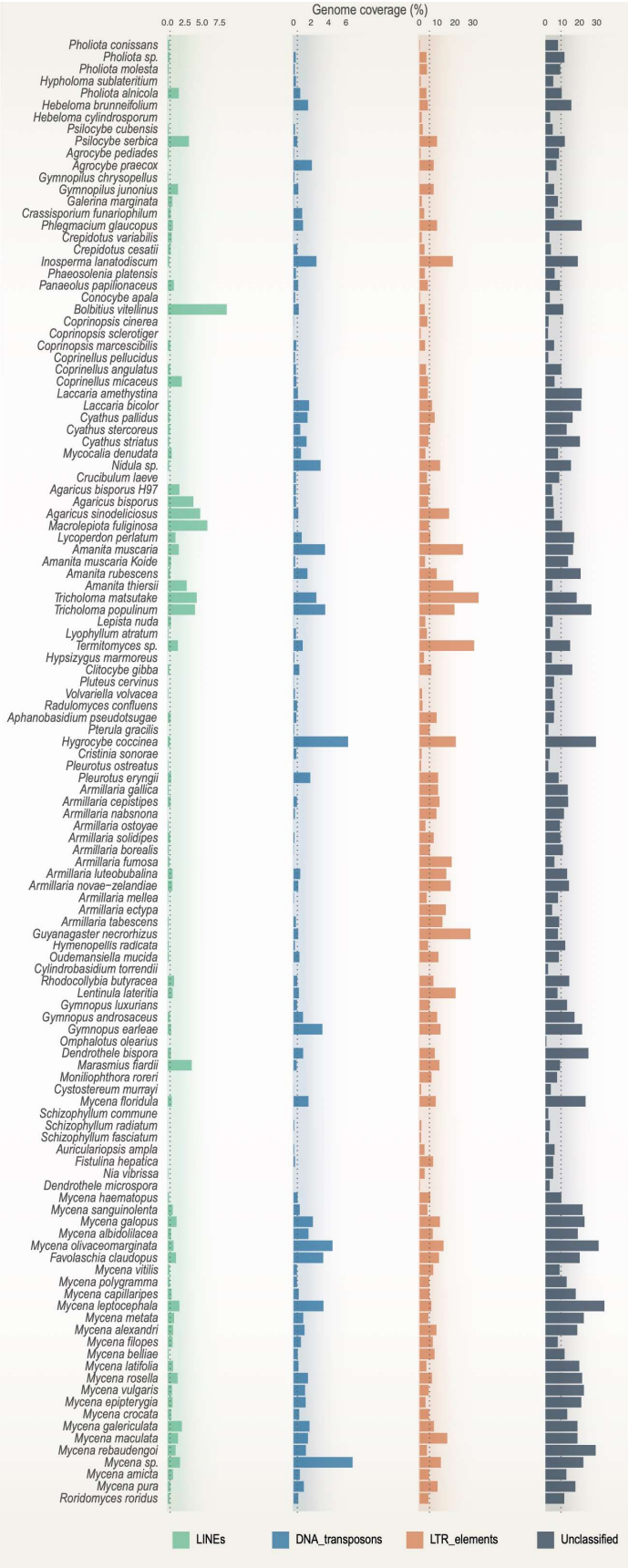

Fig S3. Overview of major transposable elements covering the genomes**.** The major TE categories are shown in color. LINEs: Long interspersed nuclear elements (non-LTR/ class I retrotransposons), DNA transposons: Class II transposons, LTR-elements: Long terminal repeat elements, class I retrotransposons. Unclassified repeats. The fungal species are in the evolutionary order.

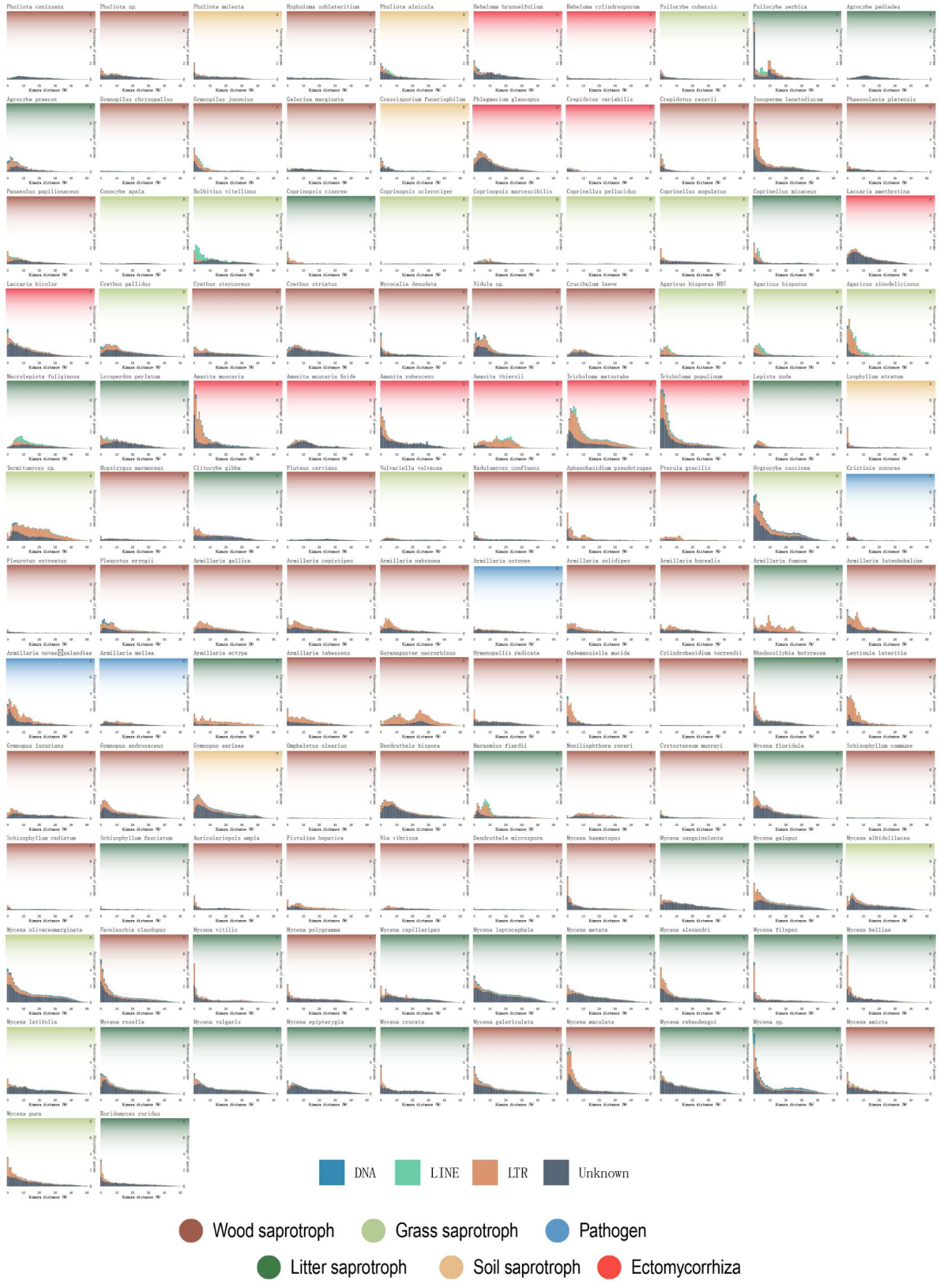

Fig.S4. Divergence of transposable elements. The sequence divergence of transposable elements was estimated based on Kimura distance (%) on the X-axis. The TE abundant peaks towards left indicates recently inserted TEs with fewer nucleotide substitution. The peaks towards right indicates old inserted TEs containing a higher amount of substitution. The Y-axis is the percentage of genomes covered with TEs.

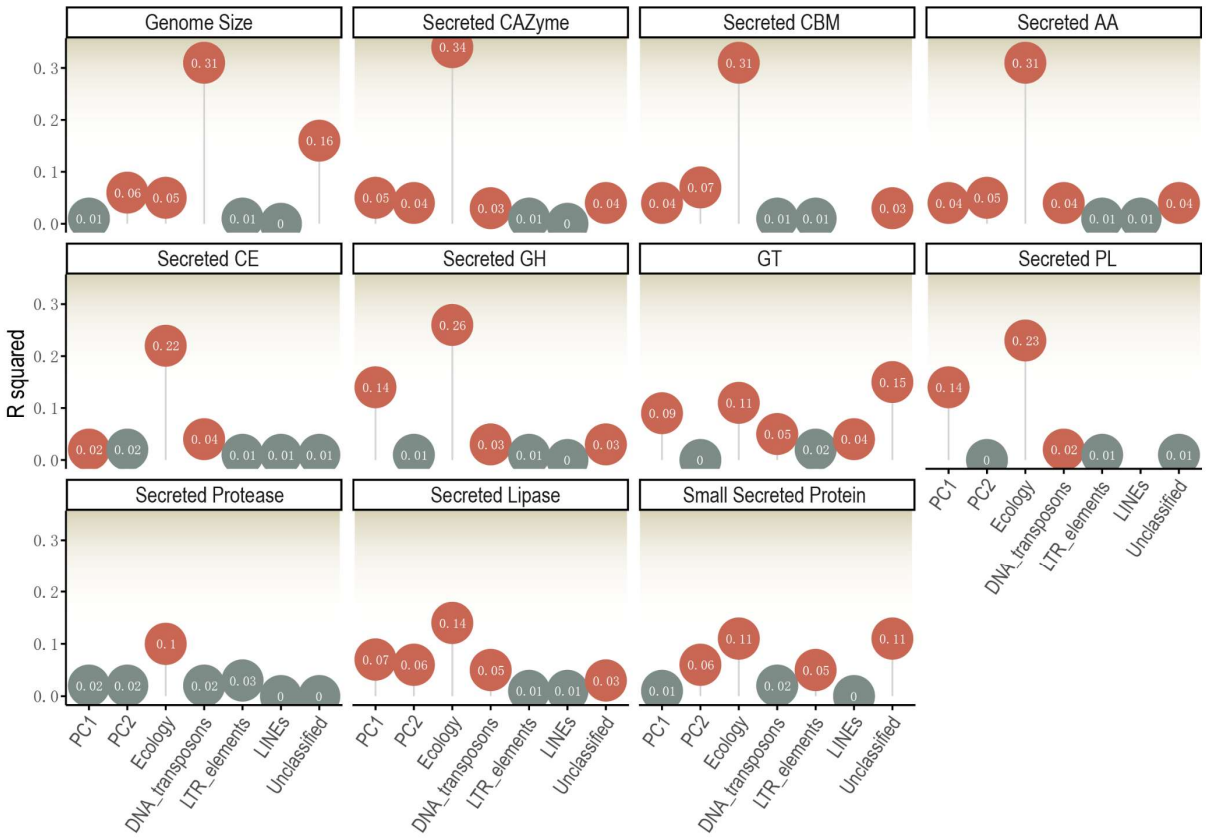

Fig. S5. Genomic features explained by the fungal lifestyle. The proportion of variation explained in selected genomic features is shown. *Mycena* species were excluded. Circles with numbers inside indicate R squared values for the variables tested. Circles in red indicate significant variables (p-value < 0.05; PERMANOVA model, Genomic features ~ Phylogenetic distance + Phenotypic group + Types of transposable elements). Phylo distance is represented as PC1 and PC2: Major principal components covering over 90% of the variation in phylogenetic distances. Ecology: Fungal lifestyle. DNA/LTR/Unclassified: TE family coverage in the genomes. See Table S4.

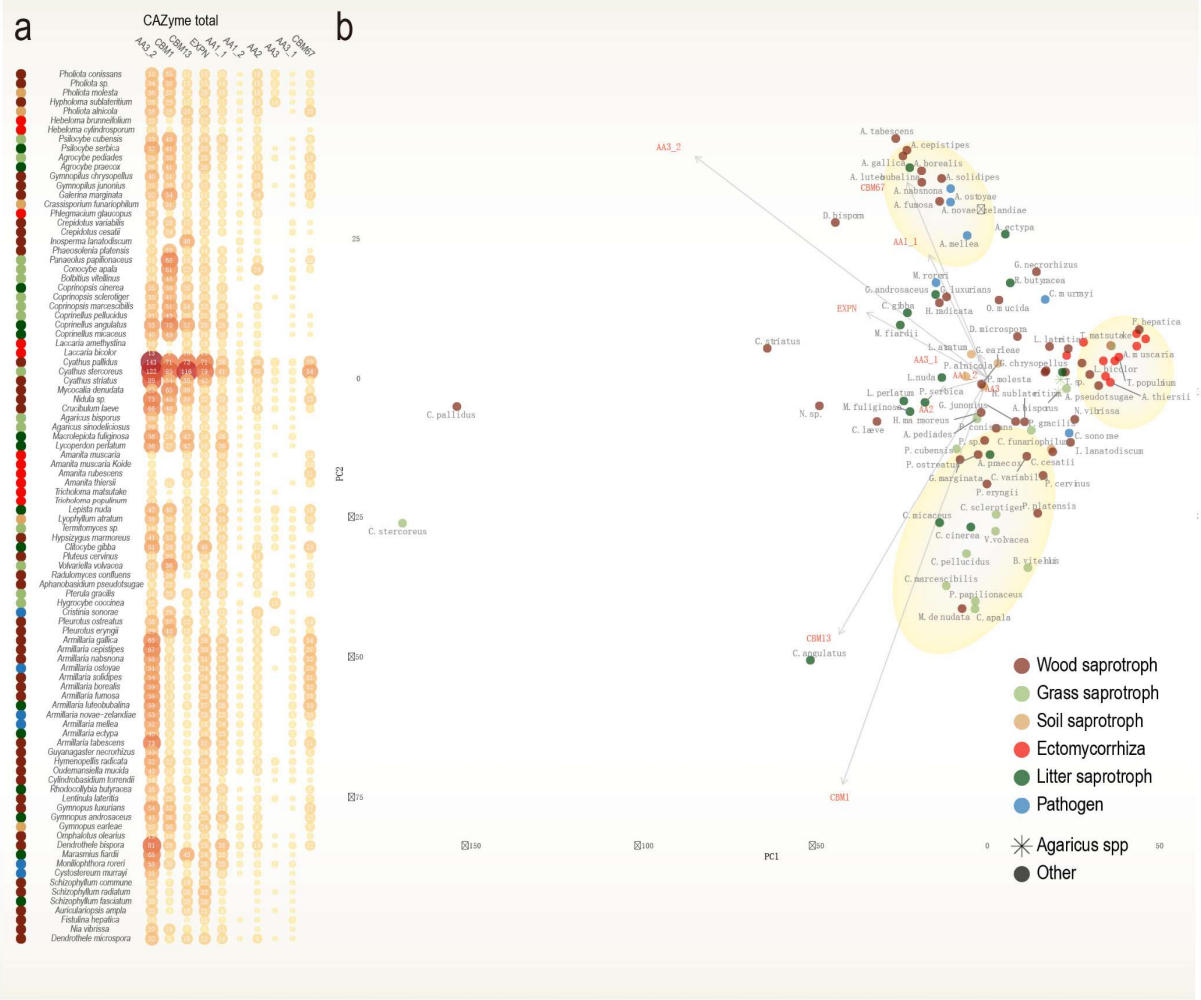

Fig.S6. Trends of fungal lifestyle groups by total CAZymes. a. The count of proteins (including secreted and non-secreted) with high loadings (see b) shows distinguished patterns for the ecological groups. Fungal lifestyles are in color. b. Two dimensional maps of the species. The first and second principal components were calculated from the count of extracellular and intracellular proteins. Phylogenetic distances of the species were taken into account. High loading proteins are shown in red with arrows. The lifestyles are in colour. Clustered groups are encircled in ovals. *Agaricales* species are labelled as a star shape. See Table S5.

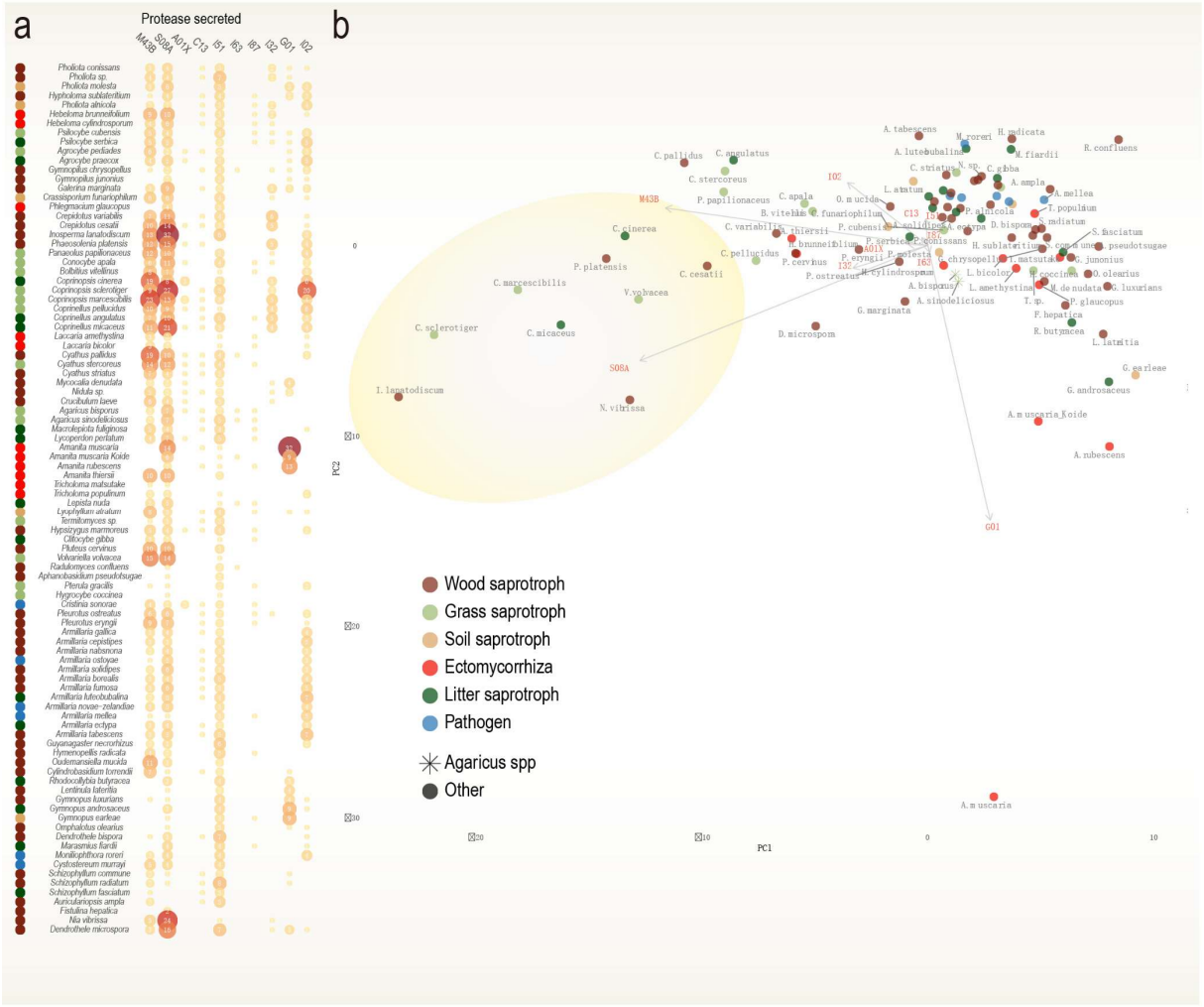

Fig.S7. Trends of fungal lifestyle groups by secreted protease. a. The count of proteins with high loadings (see b) shows distinguished patterns for the ecological groups. Fungal lifestyles are in color. b. Two dimensional maps of the species. The first and second principal components were calculated from the count of extracellular proteins. Phylogenetic distances of the species were taken into account. High loading proteins are shown in red with arrows. The lifestyles are in color. Clustered groups are encircled in ovals. *Agaricales* species are labelled as a star shape. See Table S5.

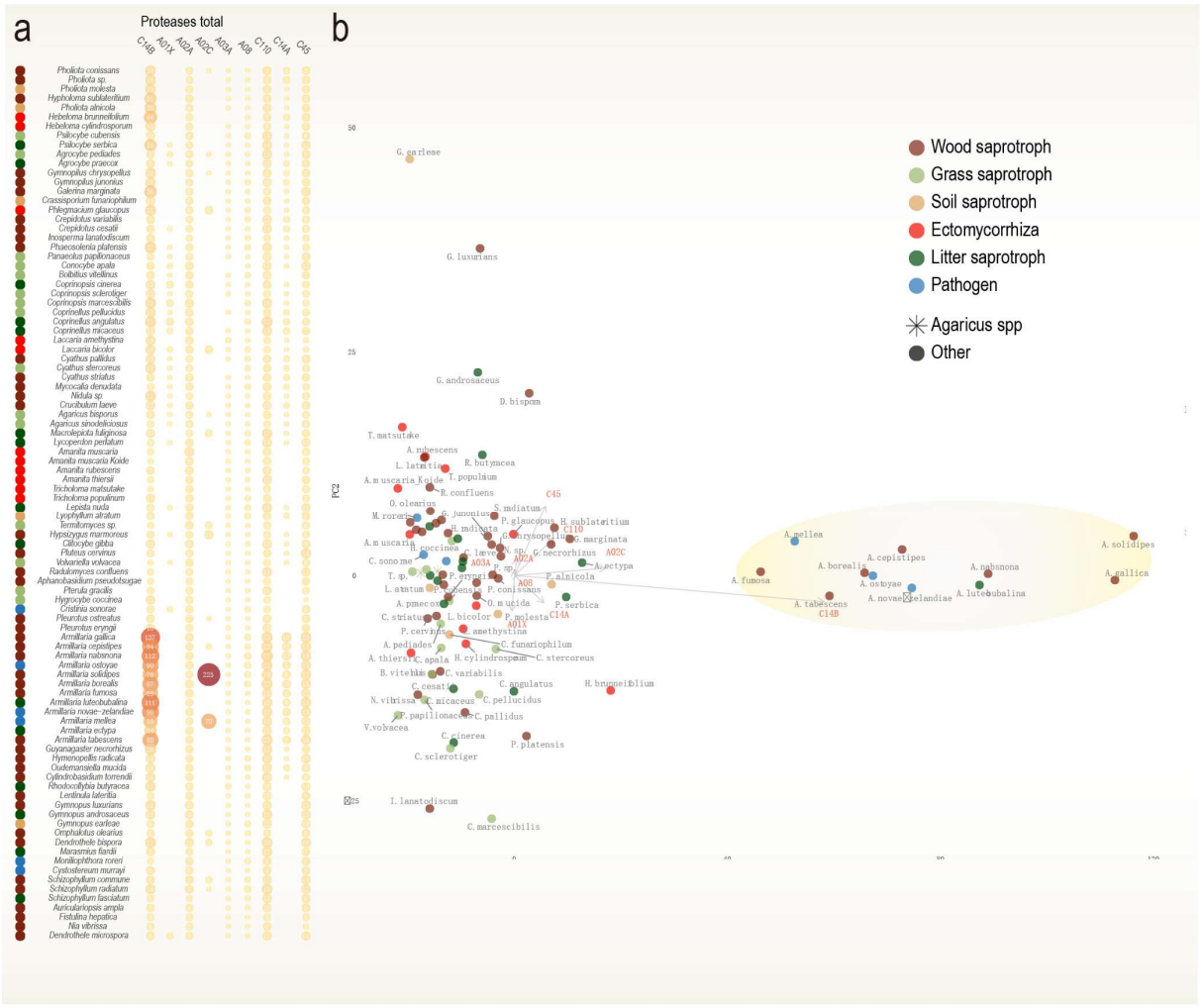

Fig.S8. Trends of fungal lifestyle groups by total proteases. a. The count of proteins with high loadings (see b) shows distinguished patterns for the ecological groups. Fungal lifestyles are in color. b. Two dimensional maps of the species. The first and second principal components were calculated from the count of extracellular proteins. Phylogenetic distances of the species were taken into account. High loading proteins are shown in red with arrows. The lifestyles are in color. Clustered groups are encircled in ovals. *Agaricales* species are labelled as a star shape. See Table S5.

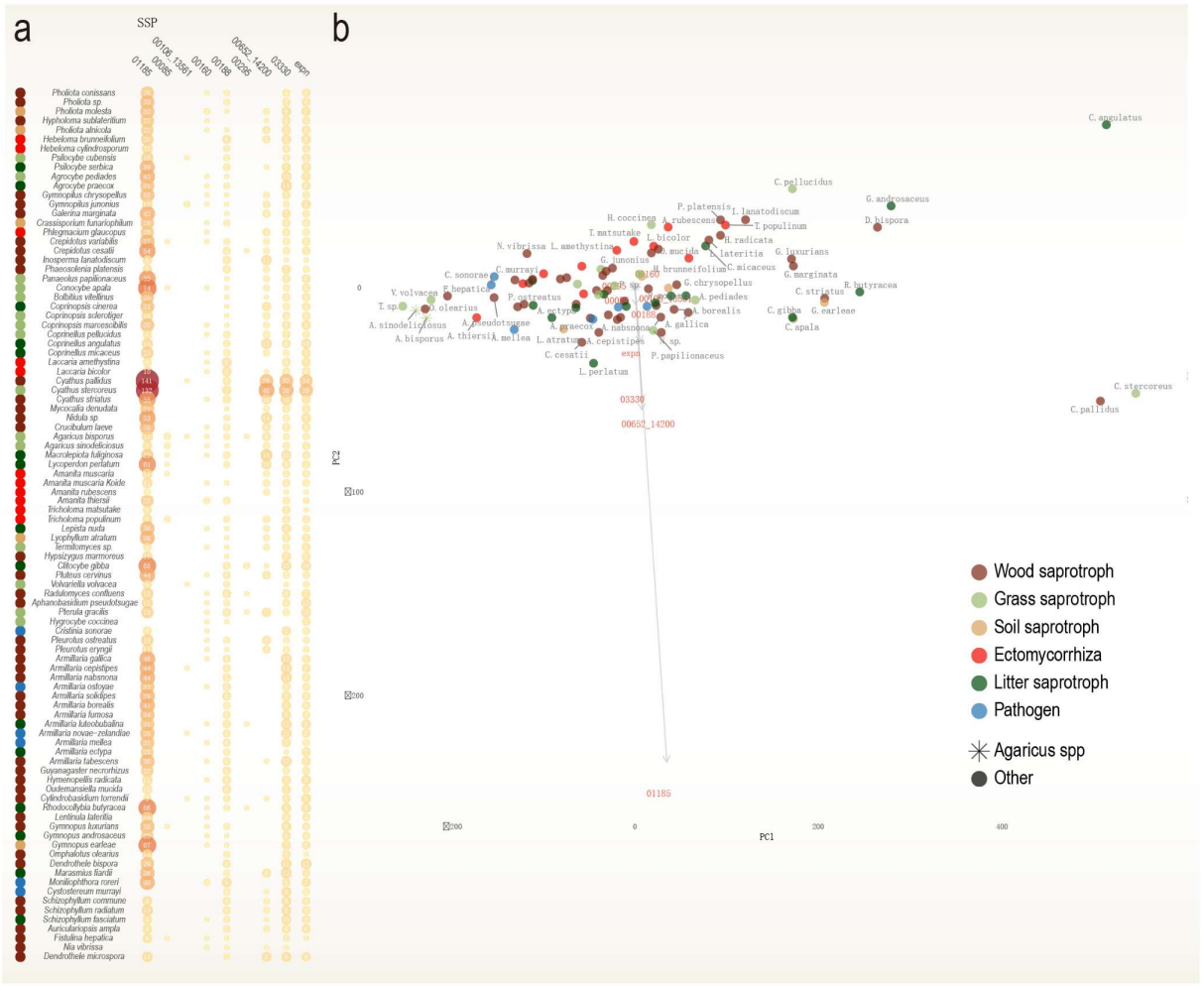

Fig.S9. Trends of fungal lifestyle groups by small secreted proteins. a. The count of proteins with high loadings (see b) shows distinguished patterns for the ecological groups. Fungal lifestyles are in color. b. Two dimensional maps of the species. The first and second principal components were calculated from the count of extracellular proteins. Phylogenetic distances of the species were taken into account. High loading proteins are shown in red with arrows. The lifestyles are in color. Clustered groups are encircled in ovals. *Agaricales* species are labelled as a star shape. See Table S5.

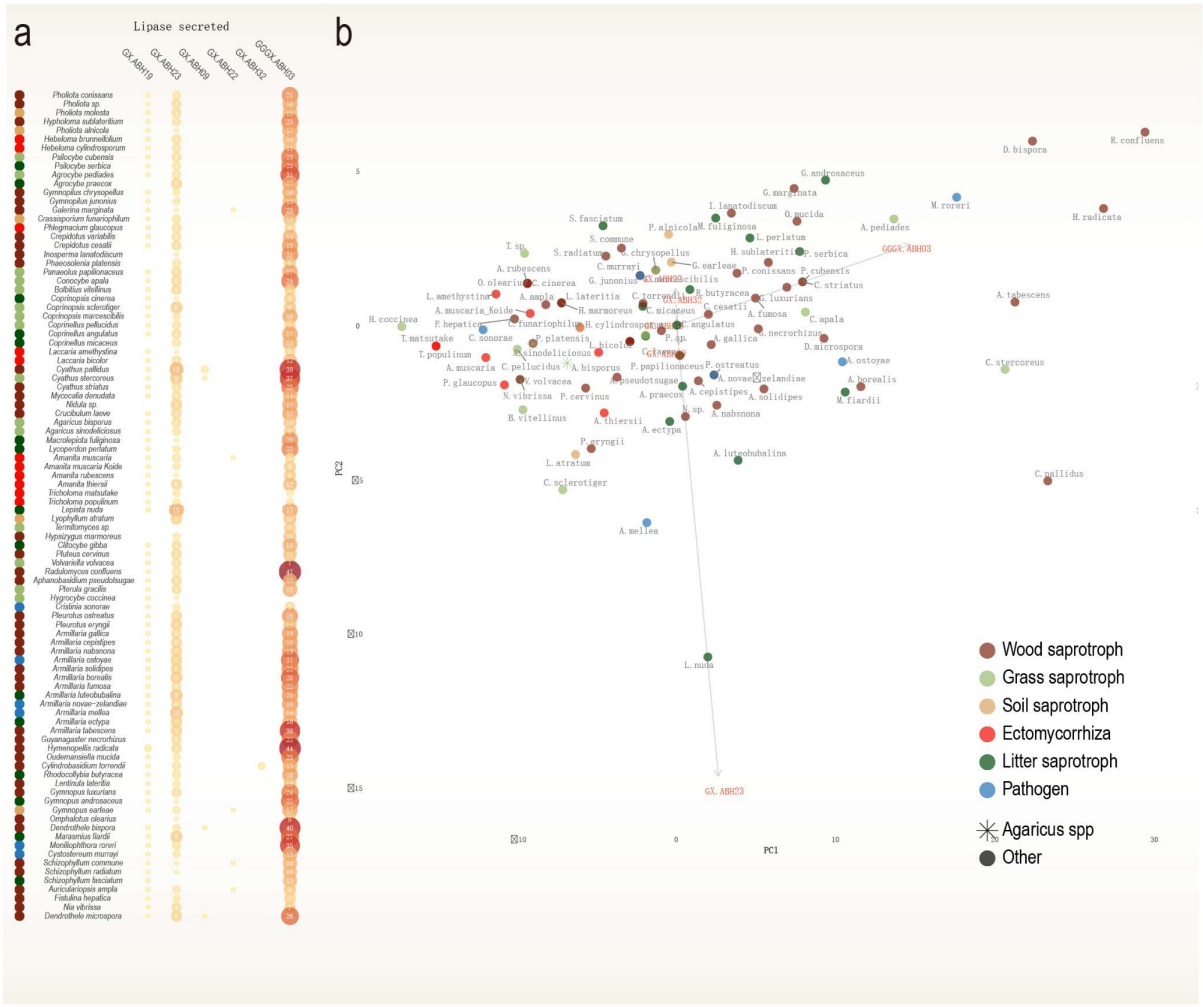

Fig.S10. Trends of fungal lifestyle groups by secreted lipases. a. The count of proteins with high loadings (see b) shows distinguished patterns for the ecological groups. Fungal lifestyles are in color. b. Two dimensional maps of the species. The first and second principal components were calculated from the count of extracellular proteins. Phylogenetic distances of the species were taken into account. High loading proteins are shown in red with arrows. The lifestyles are in color. Clustered groups are encircled in ovals. *Agaricales* species are labelled as a star shape. See Table S5.

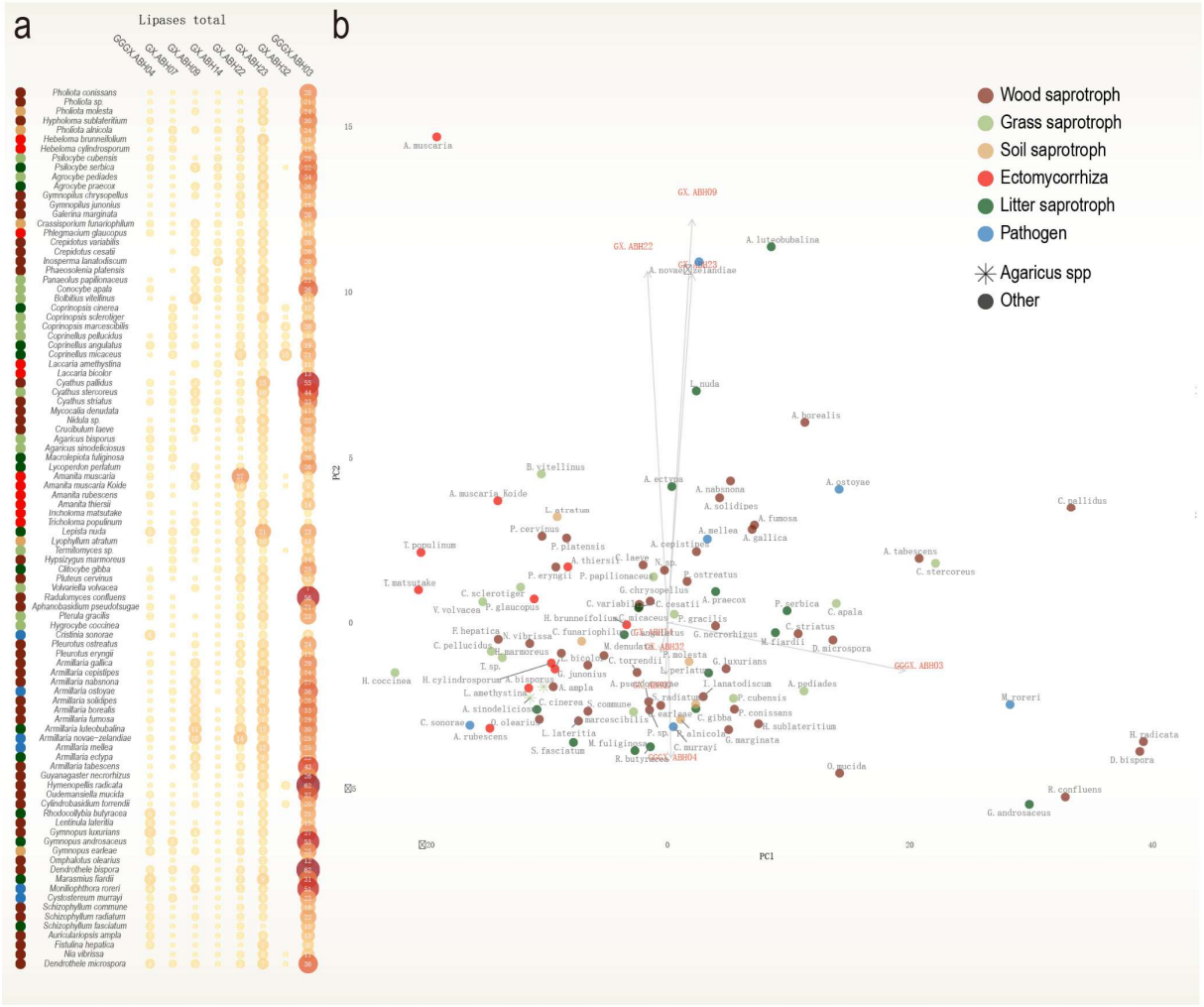

Fig.S11. Trends of fungal lifestyle groups by total lipases. a. The count of proteins (including secreted and non-secreted) with high loadings (see b) shows distinguished patterns for the ecological groups. Fungal lifestyles are in color. b. Two dimensional maps of the species. The first and second principal components were calculated from the count of extracellular and intracellular proteins. Phylogenetic distances of the species were taken into account. High loading proteins are shown in red with arrows. The lifestyles are in color. Clustered groups are encircled in ovals. *Agaricales* species are labelled as a star shape. See Table S5.

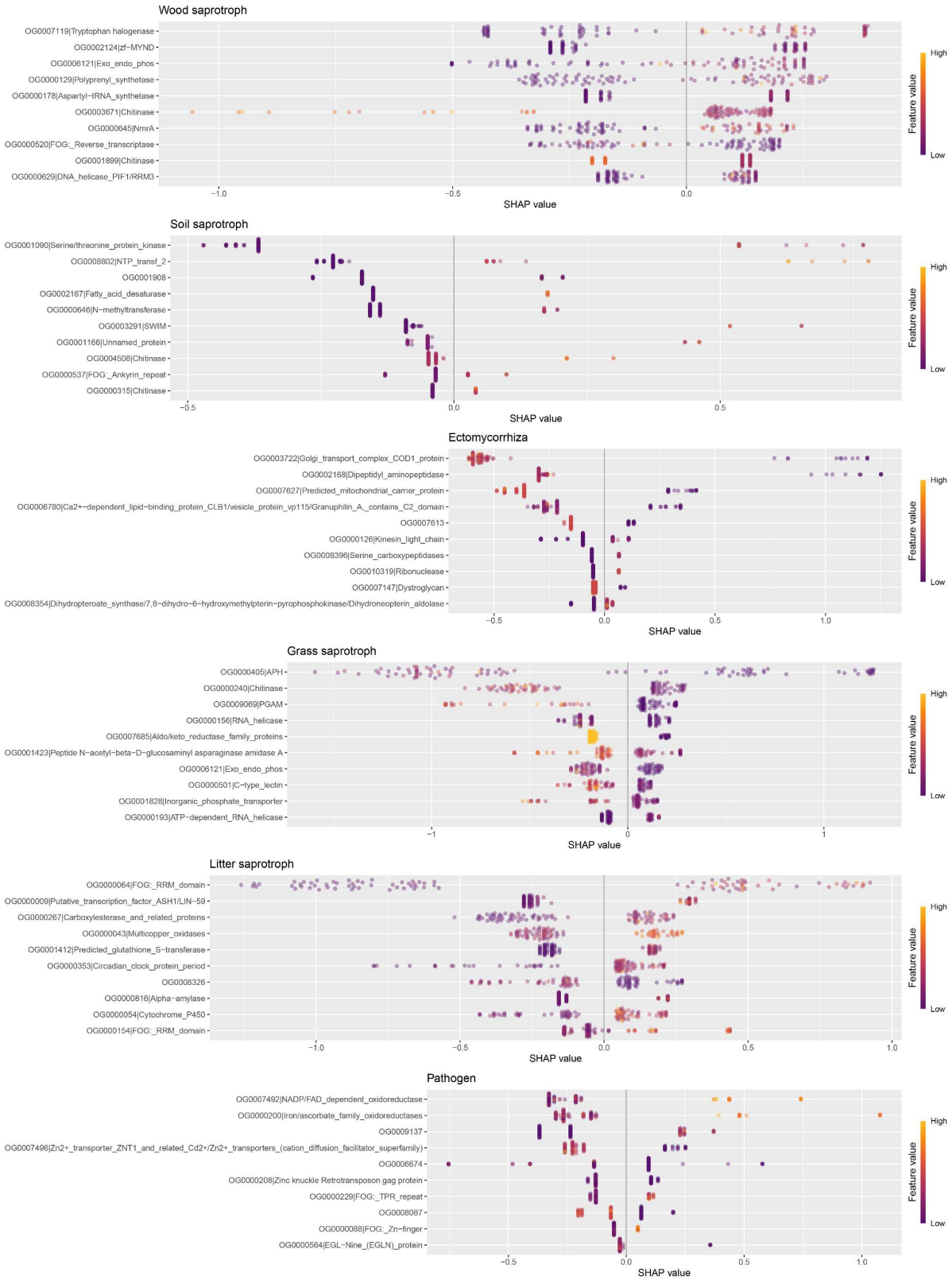

Fig.S12. Top 10 key features identified in the common orthologous genes among the 95 species. Important features influencing the prediction model were constructed using XGBoost. The gradient of colors indicates the feature values (see Methods). Higher values represent more important variables contributing to the prediction of classification by machine learning. Calculated SHAP values are shown on the x-axis. The values are based on the count of the orthologous genes (Table S6).

Fig.S13. Key secreted proteases between wood decayers and litter/soil/grass decomposers. a: Top 10 proteases identified among the 95 species. Decomposers include litter, soil, and grass saprotrophs. Important features influencing the prediction model were constructed using XGBoost (see Methods). The gradient of colors indicates the feature values. Higher values represent more important variables contributing to the prediction of classification by machine learning. Calculated SHAP values are shown on the X axis. The values are based on the count of secreted proteases. b: The number of proteases is shown per fungal group. The species are in evolutionary order. The numbers in the bubbles indicate gene copy number. See Table S5 for the number of secreted proteases.

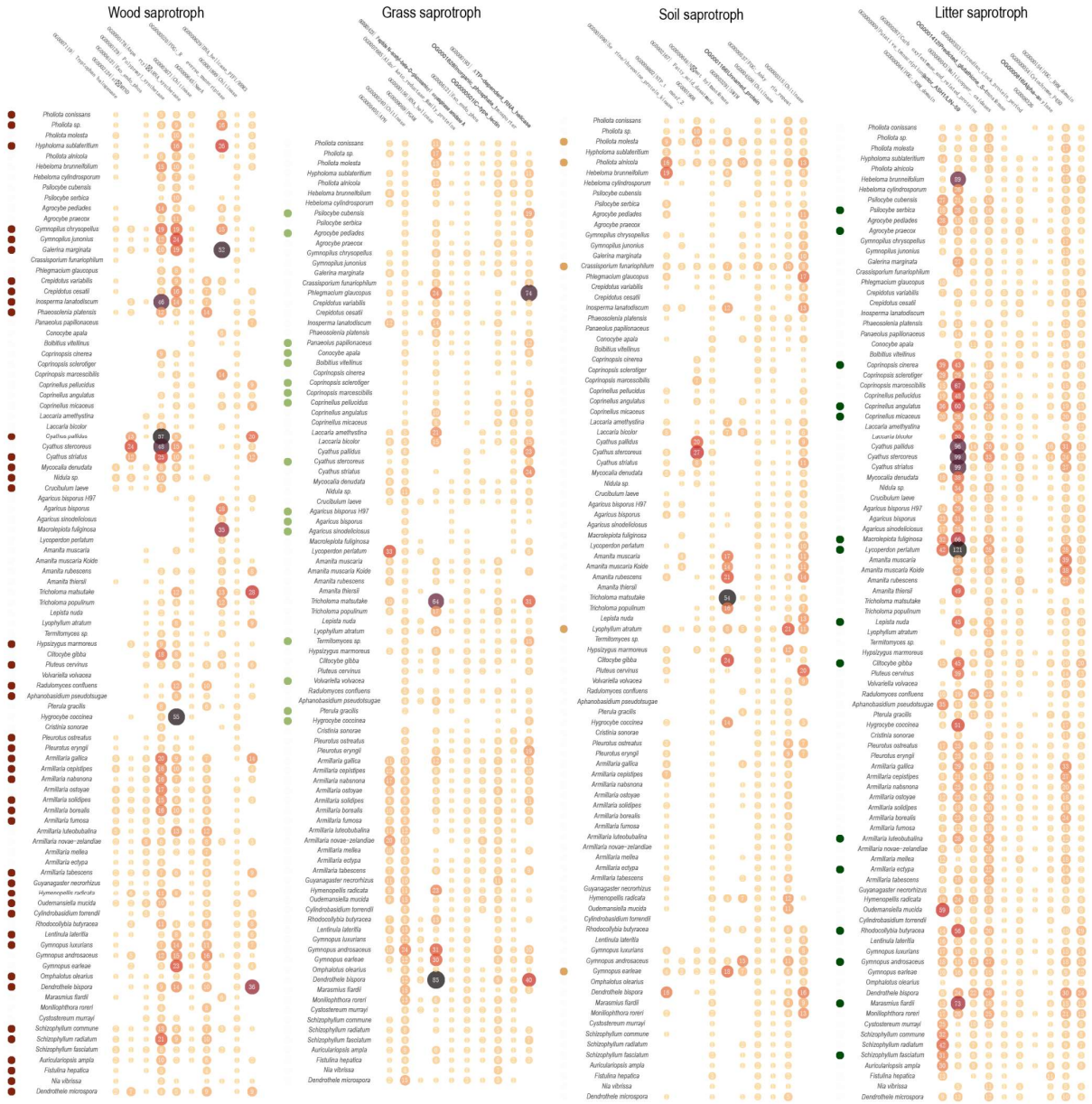

Fig.S14. Lifestyle-specific orthologous genes in the 95 fungi. Influential common orthologues among the 95 fungi were selected using XGBoost (see Methods). Orthologue group IDs and representative annotations are presented at the top. Fungal lifestyles are indicated by colors. The numbers in the bubbles represent the gene copy number. See important features identified in Fig.S8. See Table S6 for the count of orthologues.

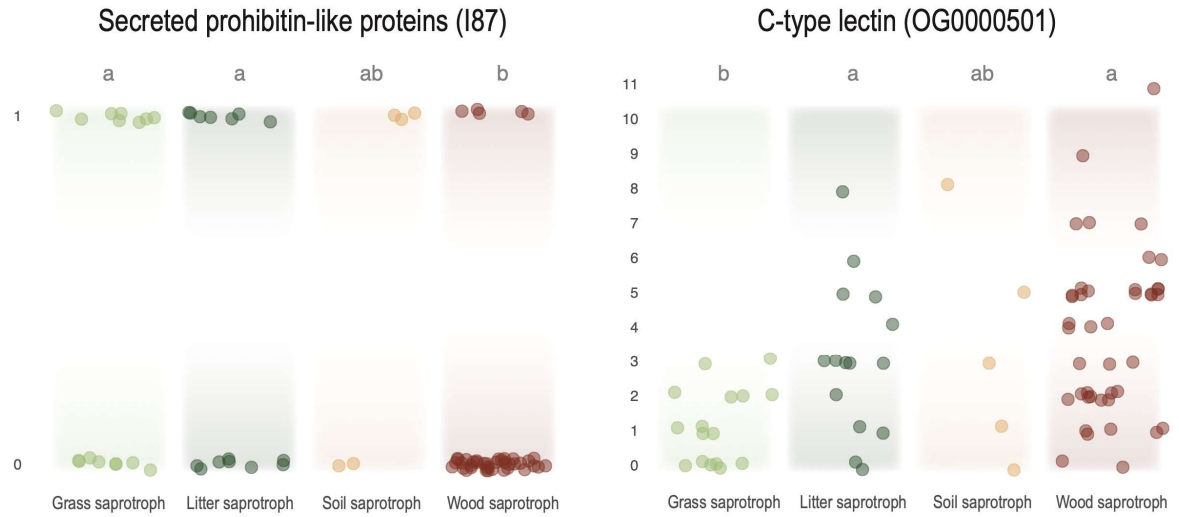

Fig.S15. Identified genes distinguishing wood and grass saprotrophs. Left: The gene copies of secreted prohibitin-like proteins (I87). Right: The count of orthologous gene for growth-arrest-specific protein (C-type lectin; OrthoID: OG0000501). Dots indicate gene copy number per species. The copy number is displayed. Alphabet letters indicate significant differences among the ecological groups (FDR adjusted $p < 0.05$; Kruskal Wallis test with post hoc Dunn test).

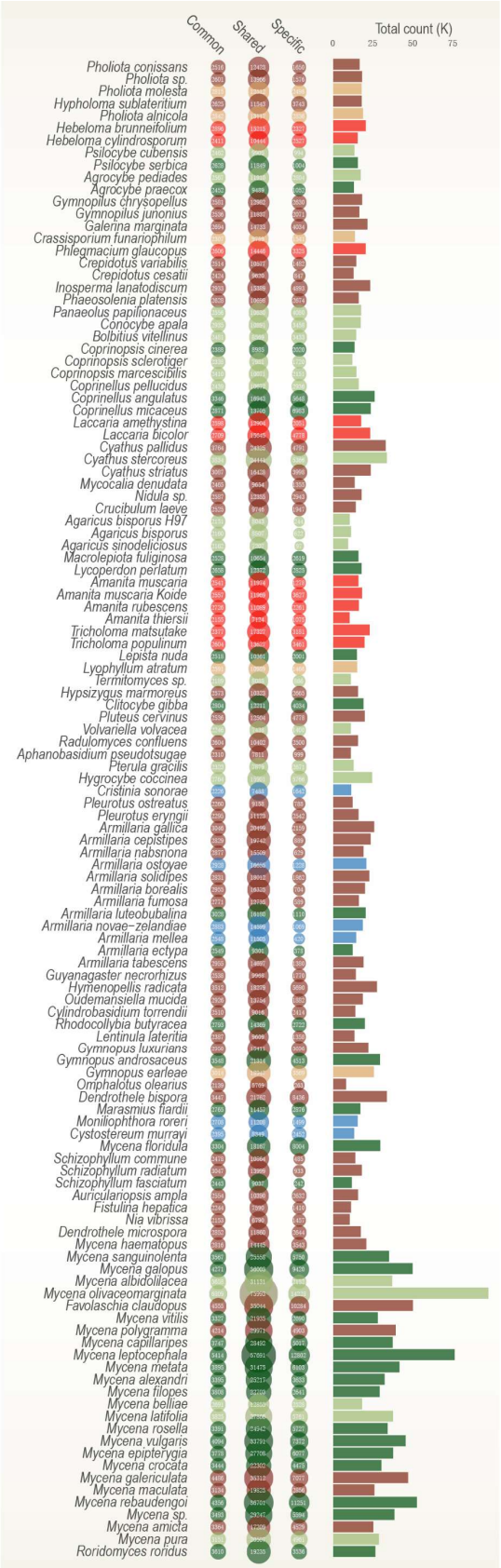

Fig.S16. Overview of orthologous gene groups. There were 106,548 orthologous groups. Of these, there are three categories: Genes are present among all species (common); genes are shared with some species (shared), genes are unique to single species. (specific). The species are in the evolutionary order. See Table S7 for details.

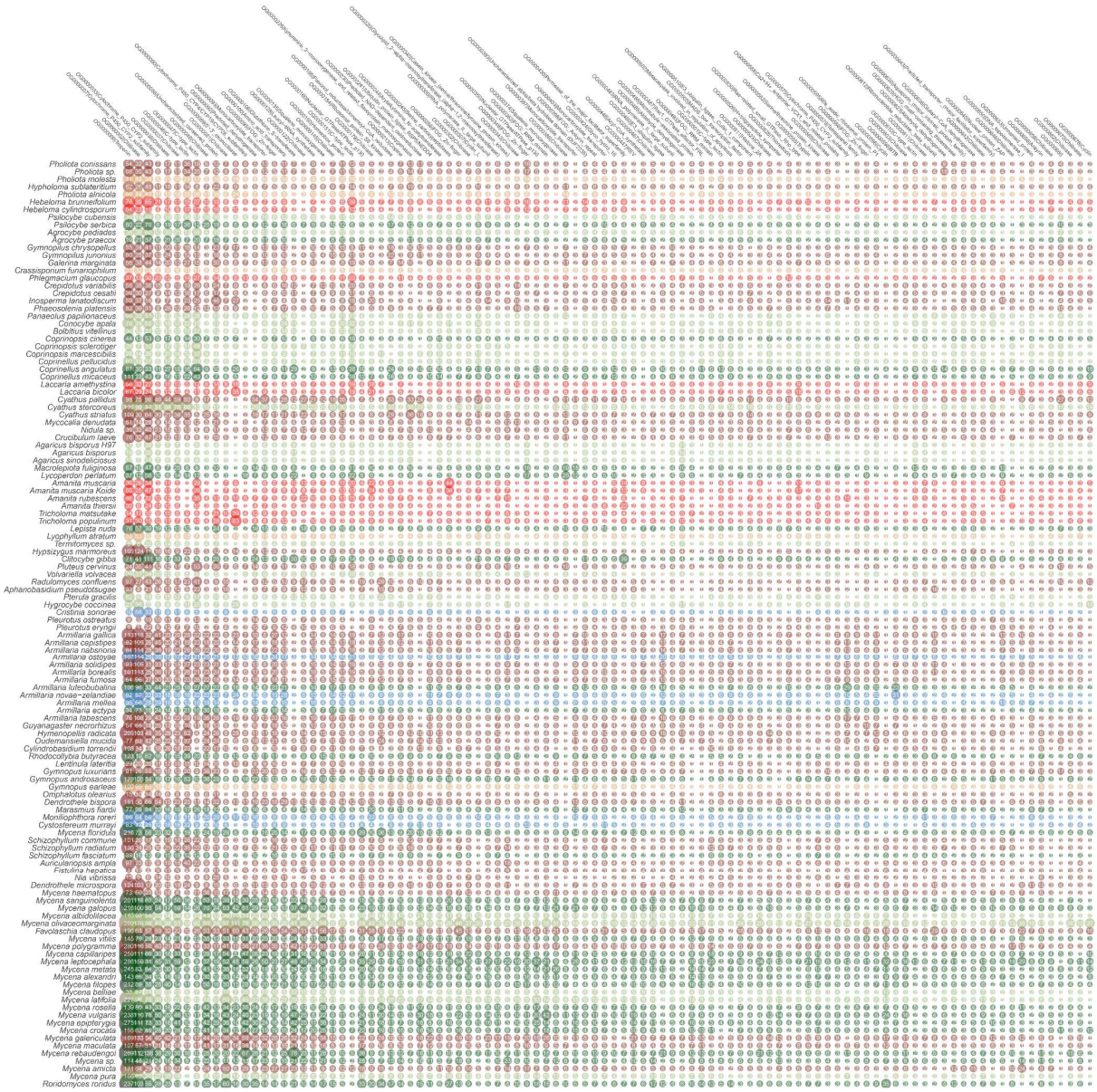

Fig.S17. Orthologous gene groups identified among 122 *Agaricales* fungi. The first largest 100 groups (of 106,548) are shown. The orthogroup ID with annotation information is presented on top. The species are in the evolutionary order. See Table S7 for details.

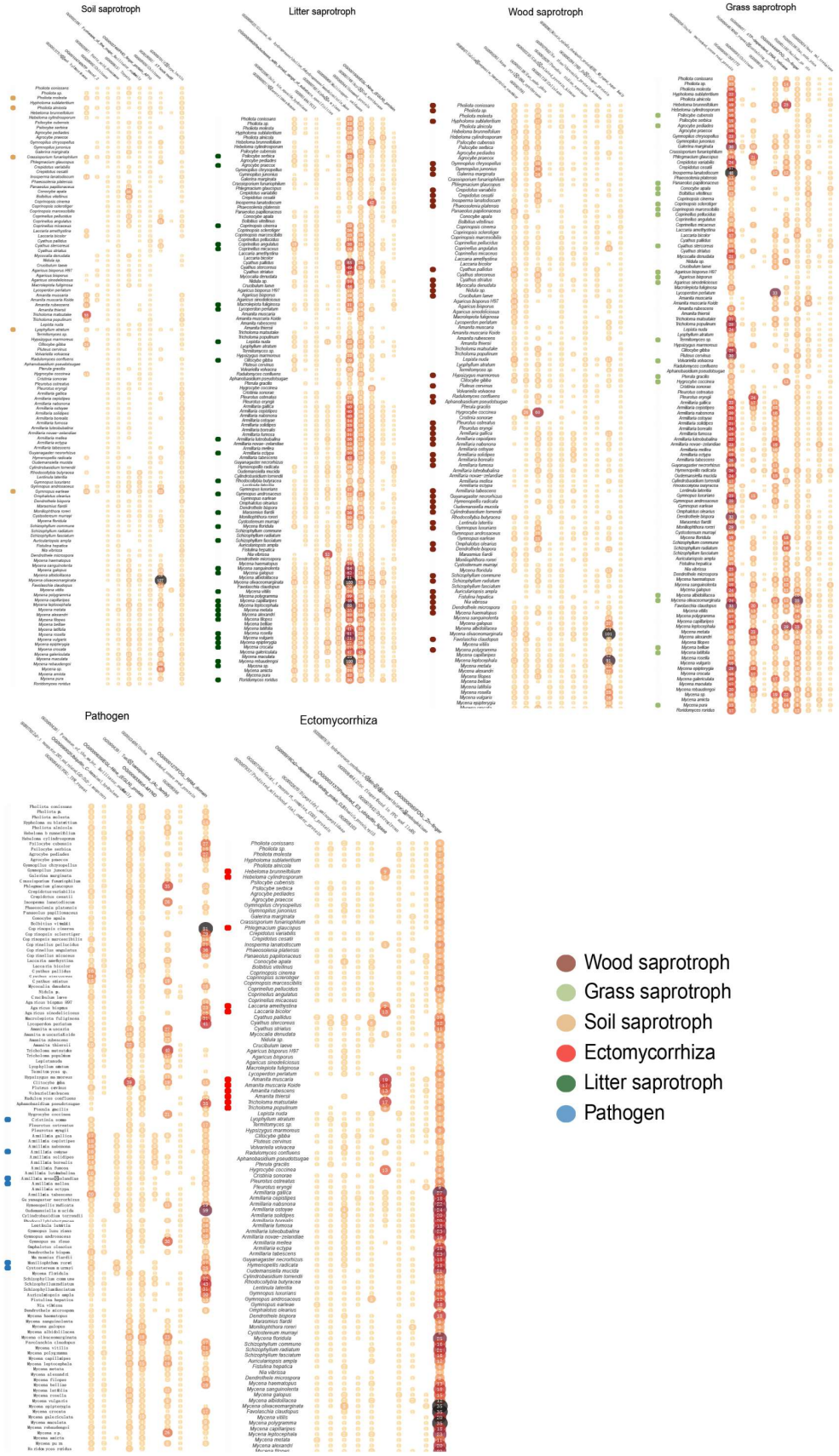

Fig.S18. Fungal lifestyle-wise associated common orthologous genes. The top ten orthologous groups influencing the prediction model are shown per fungal lifestyle group. The species are in the evolutionary order. Machine learning models were made with XGBoost.

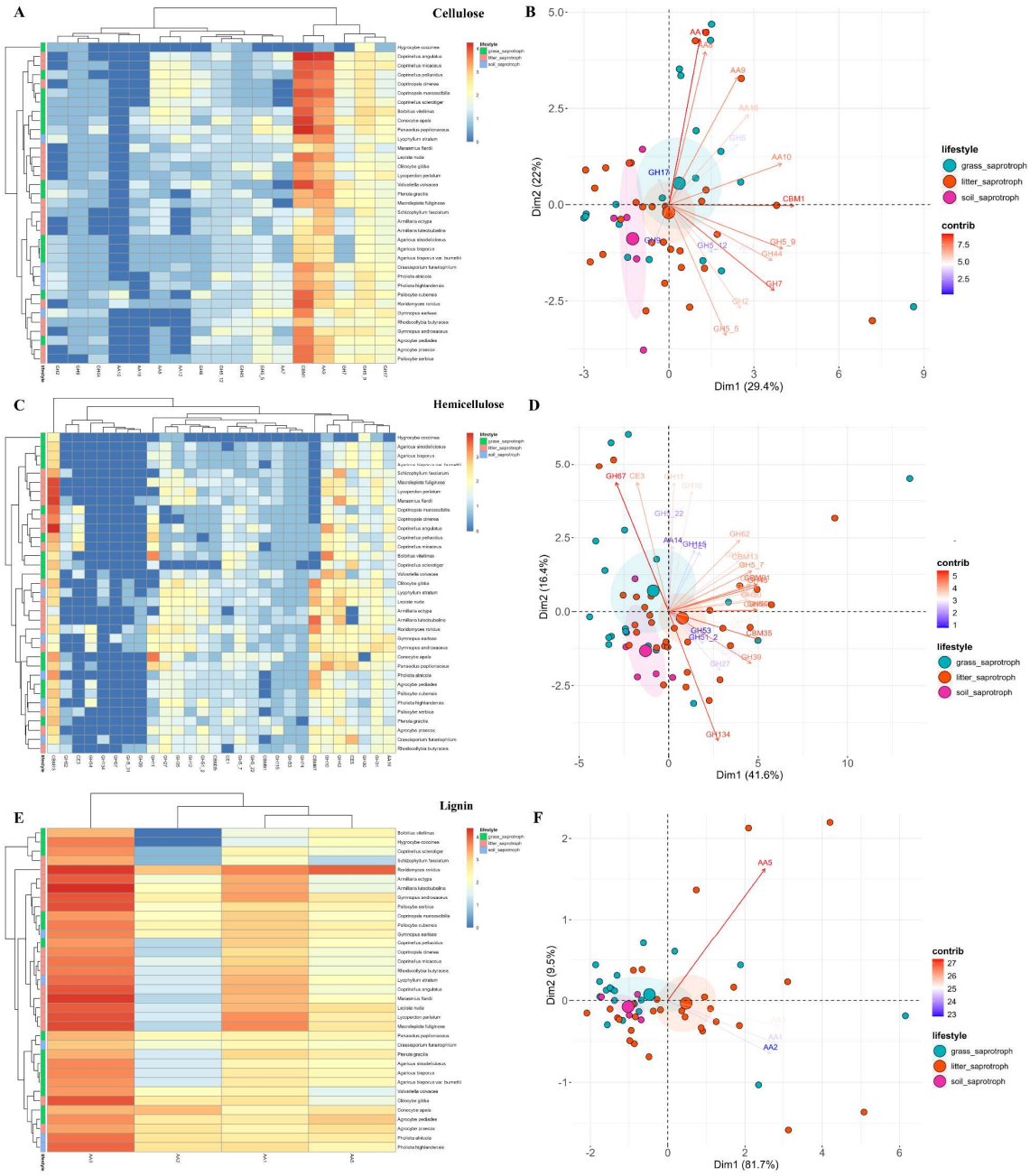

Fig.S19. Heatmap and principal coordinate map of lignin-, hemicellulose-, and cellulose- degrading secretome gene in grass/litter/soil saprotroph. A: The count of cellulose degrading CAZymes. B: PCA showing patterns of fungal lifestyle based on the count of cellulose degrading CAZymes. C: The count of hemicellulose degrading CAZymes. D: PCA showing patterns of fungal lifestyle based on the count of hemicellulose degrading CAZymes. E: The count of lignin degrading CAZymes. F: PCA showing patterns of fungal lifestyle based on the count of lignin degrading CAZymes.

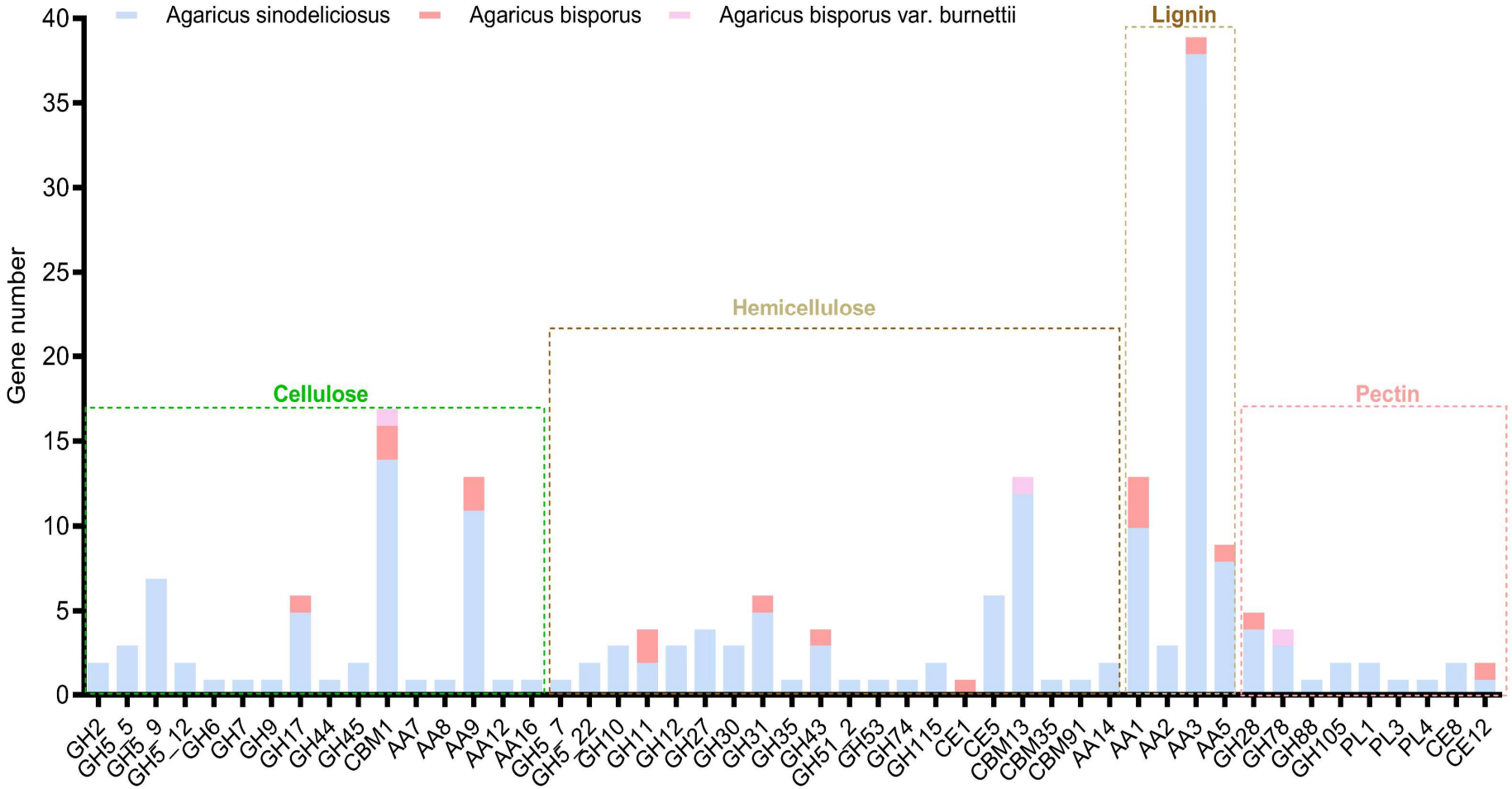

Fig.S20. Overview of CAZymes specific to the substrates among the three *Agaricus* species. Coding gene involved in degradation of cellulose, hemicellulose, lignin and pectin in *A. sinodeliciosus*, *A. bisporus* and *A. bisporus var.burnettii*.

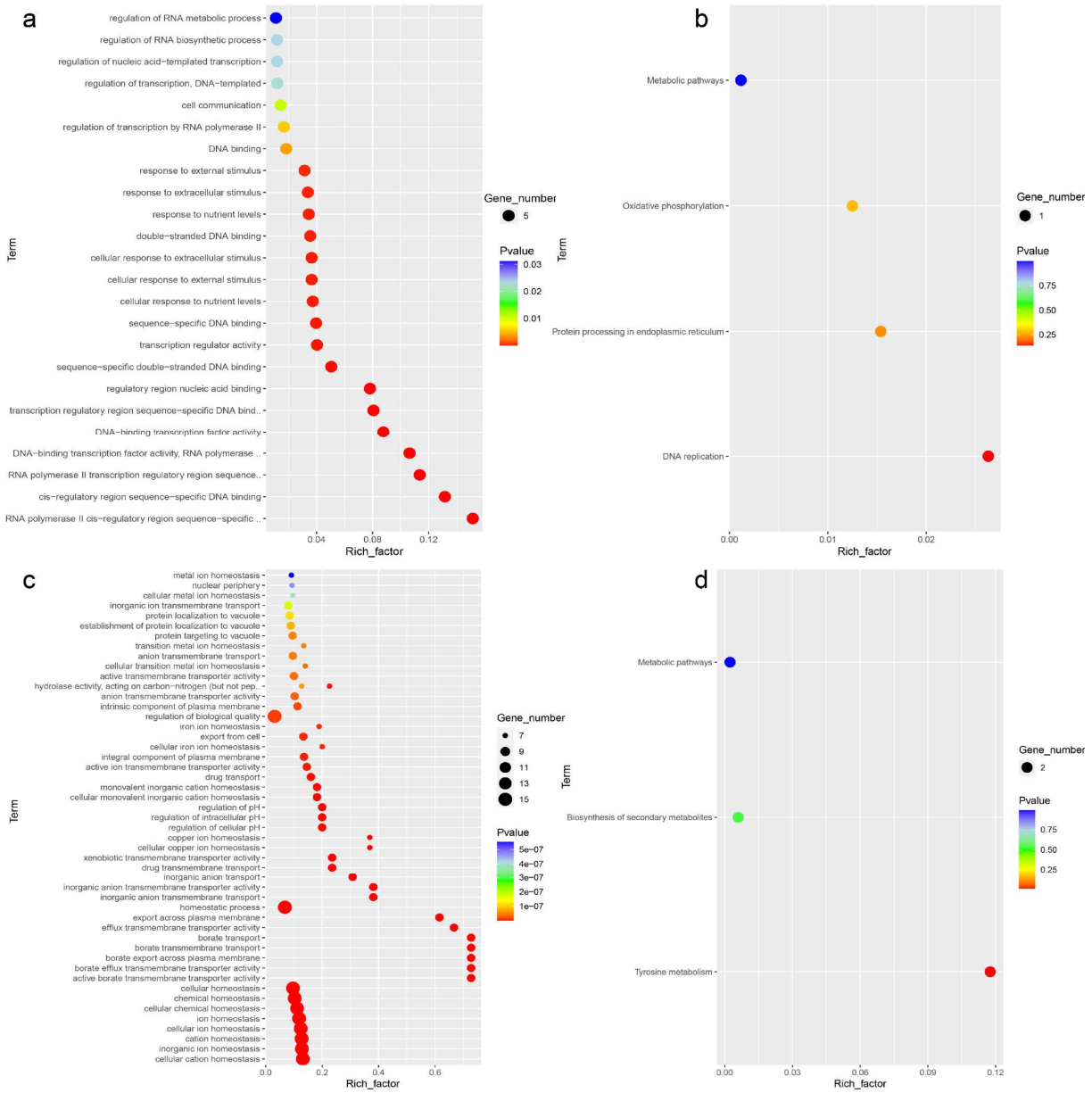

Fig.S21. GO and KEGG enrichment of *A.sinodeliciosus.* a: GO enrichment analysis revealed the 174 contracted genes were mainly enriched in RNA polymerase II cis-regulatory region sequence-specifc, DNA-binding transcription factor activity, cellular response to stimulus and nutrient levels. b: KEGG enrichment analysis revealed the 174 contracted genes were mainly enriched in DNA replication. c: GO enrichment analysis revealed the 136 expanded genes were mainly associated with cation homeostasis, chemical homeostasis, borate export across plasma membrane, pH regulation, protein transport. d: KEGG enrichment analysis revealed the 136 expanded genes were mainly associated with tyrosine metabolism.

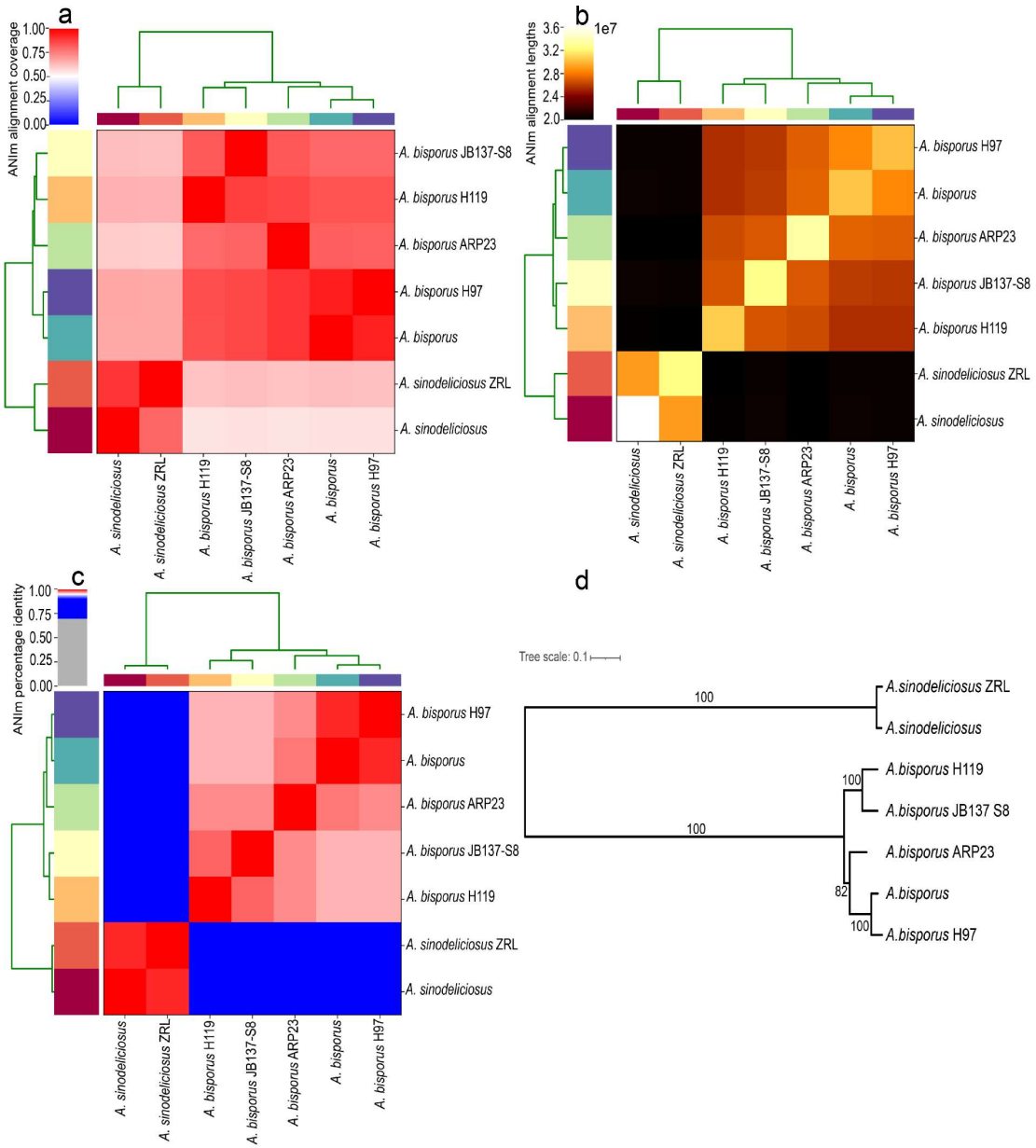

Fig. S22. Average Nucleotide Identity (ANI) among *A. sinodeliciosus*, *A. sinodeliciosus* ZRL, *A. bisporus* JB 137-S8, *A bisporus* ARP23, *A bisporus* and *A. bisporus* H97 based on MUMmer (ANIm). (a) ANIm alignment coverge, (b) ANIm alignment lengths, (c) ANIm percentage identity, and (d) SNP tree of *A. sinodeliciosus*, *A. sinodeliciosus* ZRL, *A. bisporus* JB 137-S8, *A bisporus* ARP23, *A bisporus* and *A. bisporus* H97.

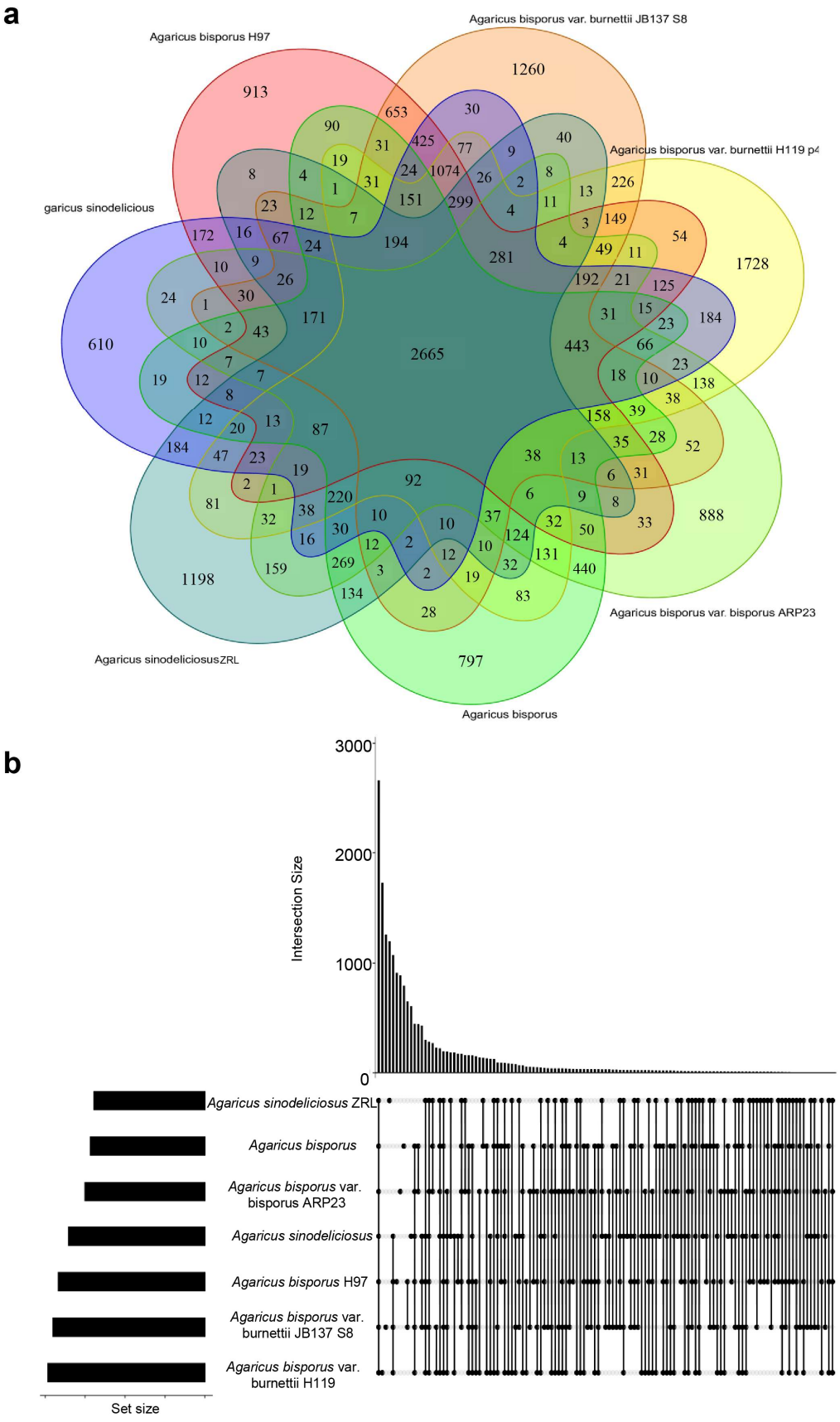

Fig. S23. Collinearity analysis of reported *A. sinodeliciosus*, *A. sinodeliciosus* ZRL, *A. bisporus* JB 137-S8, *A bisporus* ARP23, *A bisporus* and *A. bisporus* H97**.** a. Venn diagram shows the core and specific gene among aboved *A. bisporus* and *A. sinodeliciosus.* **b.** UpSetR plot of the gene distribution within the accessory genome of aboved *A. bisporus* and *A. sinodeliciosus*.

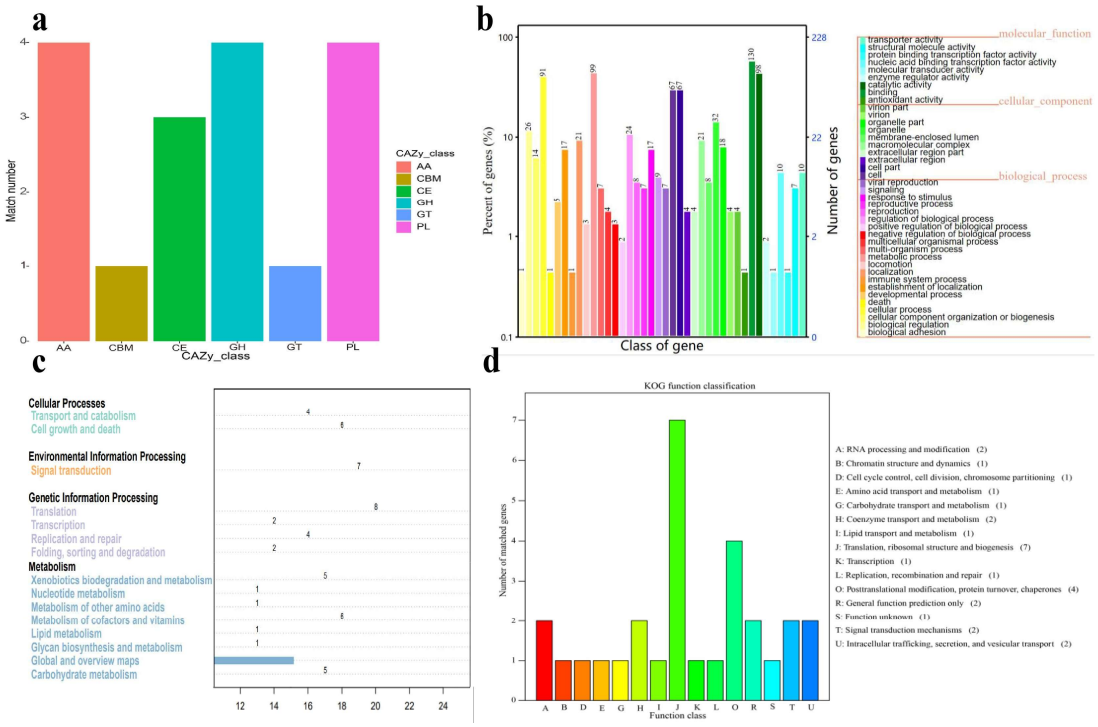

Fig. S24. The functional annotation of the 610 species-specific genes in *A. sinodeliciosus* with various databases. a: CAZymes database annotation. b: GO annotation analysis. c: KEGG annotation analysis. d: KOG annotation analysis.

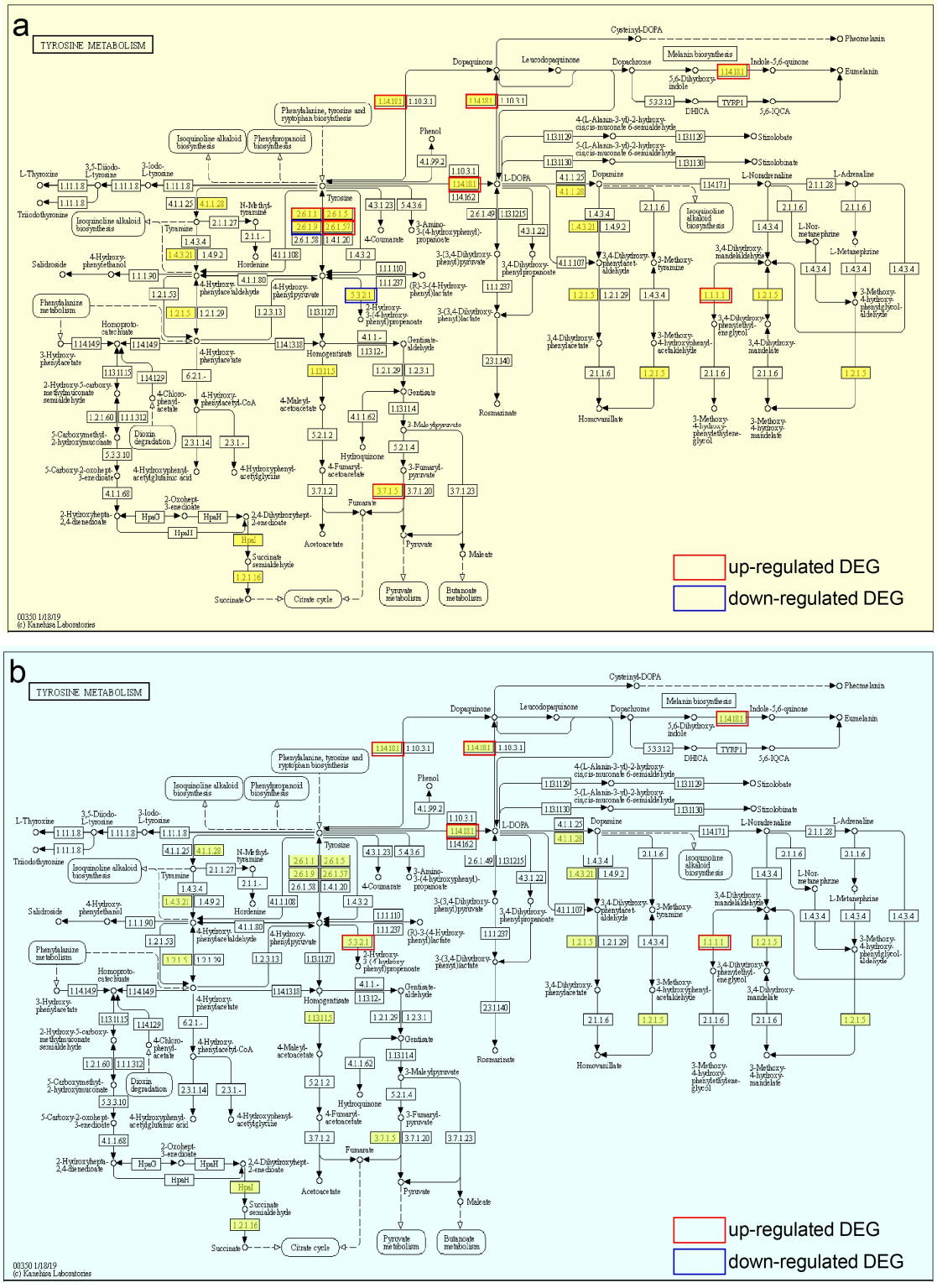

Fig. S25. Overview of tyrosine metabolic pathways in *A. sinodeliciosus* under different development stage. a: DEGs associated with tyrosine metabolism in primordium formation from mycelium. b. DEGs associated with tyrosine metabolism in fruiting body formation from primordium.

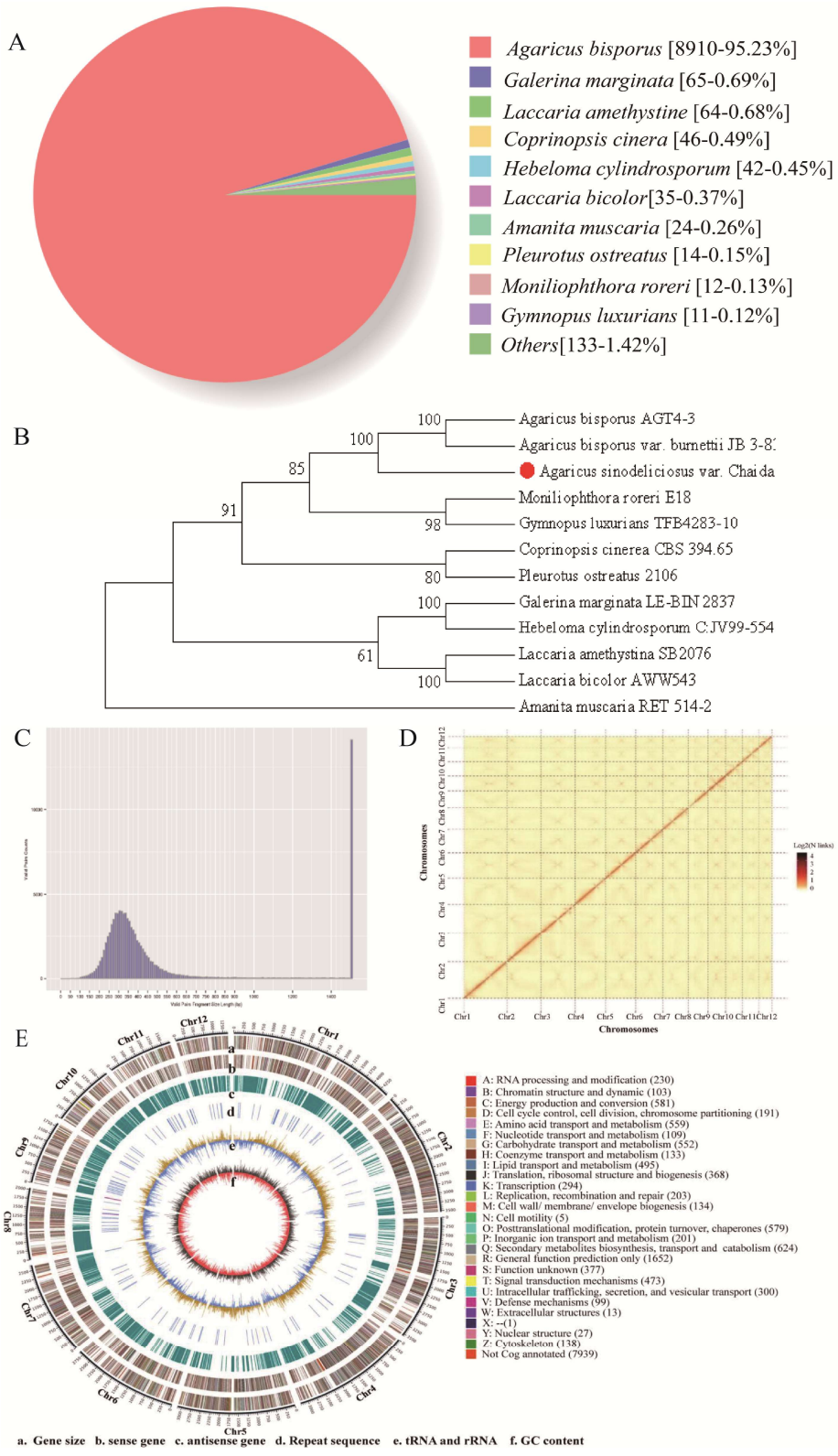

Fig.S26. (A) Species distribution map and (B) construction of developmental tree from NR database alignment to sequence. Chromosome-level gene assembly based on HiC technology. (C) Length distribution of inserted fragments in the library. (D) Hi-C contact map, and genome assembly of pepino genome. (E) High-quality genome of *A. sinodeliciosus* allows integration of genetic and expression data. The layer of rings represents genomic features. a gene size; b sense gene; c antisense gene; d, Repeat sequence; e, tRNA and rRNA; f, GC content

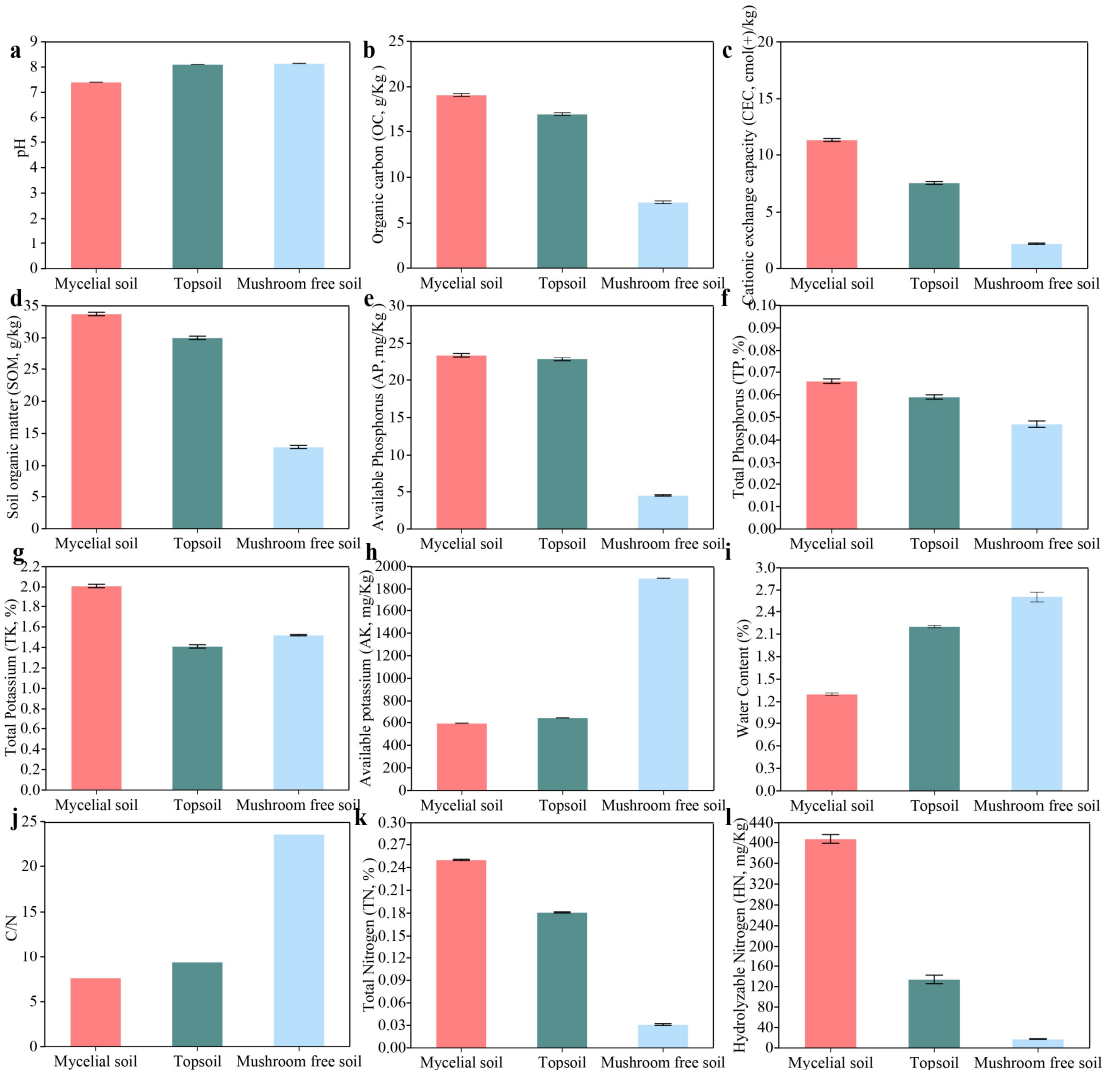

Fig.S27. Soil physicochemical index detection of mycelia soil, topsoil of *A. sinodeliciosus*. and mushroom free soil in Qinghai-Tibet Plateau.