


# Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits

Shingo Miyauchi  et al.<sup>#</sup>

Mycorrhizal fungi are mutualists that play crucial roles in nutrient acquisition in terrestrial ecosystems. Mycorrhizal symbioses arose repeatedly across multiple lineages of Mucoromycotina, Ascomycota, and Basidiomycota. Considerable variation exists in the capacity of mycorrhizal fungi to acquire carbon from soil organic matter. Here, we present a combined analysis of 135 fungal genomes from 73 saprotrophic, endophytic and pathogenic species, and 62 mycorrhizal species, including 29 new mycorrhizal genomes. This study samples ecologically dominant fungal guilds for which there were previously no symbiotic genomes available, including ectomycorrhizal Russulales, Thelephorales and Cantharellales. Our analyses show that transitions from saprotrophy to symbiosis involve (1) widespread losses of degrading enzymes acting on lignin and cellulose, (2) co-option of genes present in saprotrophic ancestors to fulfill new symbiotic functions, (3) diversification of novel, lineage-specific symbiosis-induced genes, (4) proliferation of transposable elements and (5) divergent genetic innovations underlying the convergent origins of the ectomycorrhizal guild.

<sup>#</sup>A list of authors and their affiliations appears at the end of the paper.

**M**ycorrhizal fungi are central to the evolution, biology, and physiology of land plants because they promote plant growth by facilitating the acquisition of scarce and essential nutrients, such as phosphorus and nitrogen<sup>1–3</sup>. They are also major drivers of carbon sequestration and they have a well-documented impact on the composition of microbial and plant communities<sup>1,2</sup>. The most ubiquitous classes of mycorrhizal symbioses are ectomycorrhiza, arbuscular mycorrhiza, orchid mycorrhiza, and ericoid mycorrhiza<sup>4,5</sup>. Each class is classified based on host plant and characteristic symbiotic structures. Although mycorrhizal fungi are highly diverse in terms of their evolutionary history, the independent evolution of similar symbiotic morphological structures, and physiological traits in divergent fungal taxa provides a striking example of convergent evolution<sup>3–5</sup>. Although unique and common traits in mycorrhizal symbioses have recently been reviewed<sup>2</sup>, molecular mechanisms underlying these convergent phenotypes remain largely undetermined<sup>1,3–8</sup>.

Prior comparisons of genomes from ectomycorrhizal, orchid and ericoid mycorrhizal fungi, wood decayers and soil decomposers have elucidated the mechanisms of several transitions from saprotrophy to mutualism in Dikarya<sup>9–16</sup>. These analyses have shown that multiple lineages of ectomycorrhizal fungi have lost most genes encoding lignocellulose-degrading enzymes present in their saprotrophic ancestors, explaining the reduced capacity of ectomycorrhizal fungi to acquire C complexed in soil organic matter (SOM) and plant cell walls<sup>17</sup> and, as a consequence, their increasing dependence on the host plant sugars. The diversity of trophic states in extant ectomycorrhizal fungi may be a consequence of their multiple origins from saprotrophic ancestors with varied decay capabilities, including white and brown rot wood decayers, and soil and litter decomposers<sup>3–5</sup>. However, the extent to which ectomycorrhizal fungi make use of their secreted plant cell wall degrading enzymes (PCWDEs) and microbial cell wall degrading enzymes (MCWDE) to decay or decompose SOM is not well understood<sup>17–24</sup>.

Despite their ecological prominence, much remains to be learned about the evolution and functional diversification of mycorrhizal symbionts and the crucial acquisitions that allow colonization of and nutrient exchange with plants<sup>3,25</sup>. Here, we present a combined analysis of 135 fungal genomes from 73 saprotrophic, endophytic and pathogenic fungal species, and 62 mycorrhizal fungal species, including 29 new mycorrhizal genomes. This study approximately doubles the number of published genomes of mycorrhizal fungi, and it samples major groups, for which there were previously no symbiotic genomes available, including Russulales, Thelephorales, Phallomycetidae, and Cantharellales. These groups are important, because they (1) are often ecologically dominant (Russulales and Thelephorales), (2) represent early diverging clades for which no ectomycorrhizal genomes were previously available (Phallomycetidae and Cantharellales), and (3) include groups that arose before (Cantharellales) or after the origin of ligninolytic peroxidases (class II POD) implicated in white rot<sup>26</sup>.

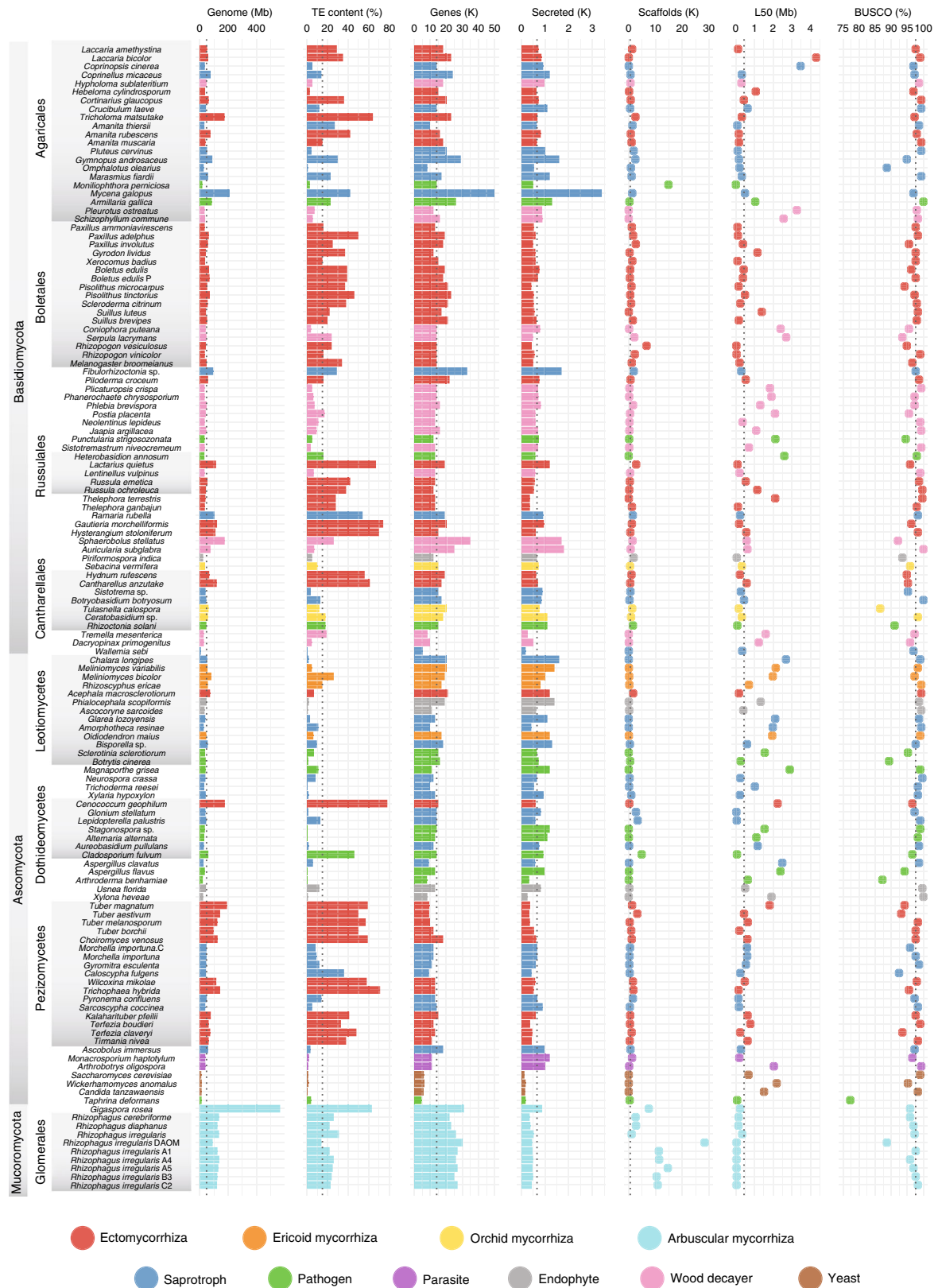
This dataset presents an opportunity to carry out a broader analysis of the evolution of saprotrophic capabilities than that we previously attempted in our large-scale comparative analysis of Agaricomycetidae<sup>12</sup>. In the present study, we hypothesize that the evolutionary mechanisms in action in Agaricomycetidae can be traced back to the early diverging clades of ectomycorrhizal fungi. To assess gains of ectomycorrhizal lifestyle traits, we discuss the fundamental adaptations that underlie convergent evolution of ectomycorrhizal fungi, including the loss of some metabolic functions, such as PCWDEs, and the acquisition of small secreted effector-like proteins that may facilitate the accommodation of symbiotic fungi within their host plants. We also compare

additional symbiosis-related functional traits, such as nitrogen and phosphate acquisition. Finally, we investigate the age distribution of symbiosis-upregulated genes across a phylogenetically representative set of ectomycorrhizal fungi, using a phylostratigraphic approach to resolve lineage-specific and conserved elements of symbiotic transcriptomes. We show that transitions from saprotrophy to ectomycorrhizal symbiosis involve widespread losses of degrading enzymes acting on lignocellulose, co-option of genes present in saprotrophic ancestors to fulfill novel symbiotic functions, diversification of lineage-specific symbiosis-induced genes, proliferation of transposable elements (TEs), and divergent genetic innovations underlying the convergent origins of the ectomycorrhizal guild.

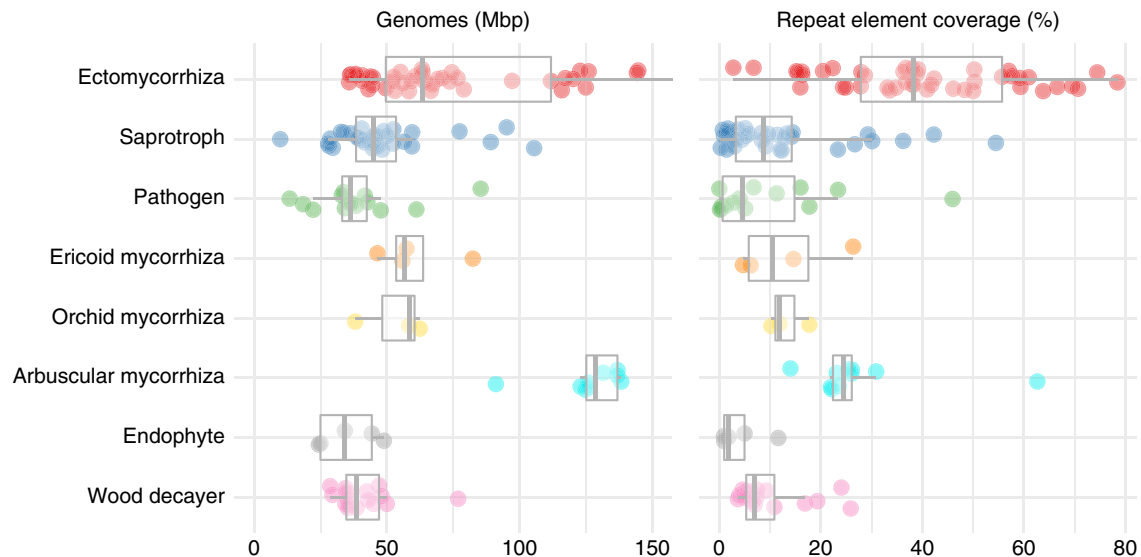
## Results

**Main features of mycorrhizal genomes.** We compared 62 draft genomes from mycorrhizal fungi, including 29 newly released genomes, and predicted 9344–31,291 protein-coding genes per species (see “Methods”, Supplementary Information and Supplementary Data 1). This set includes new genomes from the early diverging fungal clades in the Russulales, Thelephorales, Phallomycetidae, and Cantharellales (Basidiomycota), and Helotiales and Pezizales (Ascomycota). We combined these mycorrhizal fungal genomes with 73 fungal genomes from wood decayers, soil/litter saprotrophs, and root endophytes (Fig. 1 and Supplementary Data 2). There was little variation in the completeness of the gene repertoires, based on Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis (coefficient of variation, c.v. = 7.98), despite variation in assembly contiguity (Fig. 1). Genome size varied greatly within each phylum, with genomes of mycorrhizal fungi being larger than those of saprotrophic species (Figs. 1 and 2, and Supplementary Data 2;  $P < 0.05$ , generalized least squares with the Brownian motion model (GLS)). Glomeromycotina had exceptionally large genomes (Figs. 1 and 2, and Supplementary Data 2), with *Gigaspora rosea* having the largest genome (567 Mb) among the 135 fungi compared<sup>27</sup>.

**TEs in mycorrhizal genomes.** The main driver of genome inflation appeared to be repeat content (Fig. 3 and Supplementary Data 3;  $P < 0.05$  GLS), such as long terminal repeat retrotransposons (LTRs), which ranged from 0.01 to 46.4% of the assembly (Fig. 3 and Supplementary Data 3). The distribution of TE categories varies between ectomycorrhizal fungal taxa even within the same fungal orders or genera (Fig. 3), indicating that invasions by different TE families took place independently in different clades. However, the total TE coverage in genomes of ectomycorrhizal Ascomycota and Glomeromycota was significantly higher than in Basidiomycota ( $P < 0.05$ , GLS; Supplementary Data 3). In Ascomycota, the most abundant TE families are LTRs, such as *Gypsy* and *Copia*, and non-LTR I, whereas in Basidiomycota, *Gypsy* and *Copia* LTRs, *Tad1*, helitrons, and Zizuptons are abundant (Fig. 3). Lifestyle has a higher impact than phylogeny on TE coverage with a significantly higher TE content in ectomycorrhizal symbionts compared to other lifestyles for several TE categories (14–35% of total variances,  $P$  value  $< 0.05$ ; PERMANOVA) (Supplementary Fig. 1a, b, and Supplementary Data 3 and 4). In Agaricales, a large proportion of LTR insertions occurred during the past 5 million years (Mya) in *Amanita rubescens*, *Laccaria bicolor*, *Laccaria amethystina*, and *Cortinarius glaucopus*, while the major bursts are more ancient in *Tricholoma matsutake* (Supplementary Fig. 1c). These LTRs have been proliferating over the past 5–6 Mya for *Tuber melanosporum*, *Tuber aestivum*, and *Choiromyces venosus*, while LTR accumulation in *Tuber magnatum* occurred between 6 and 14 Mya (ref. 16).



**Fig. 1** General features of the 135 fungal genomes analyzed. Genome and assembly features. The different lifestyles (e.g., ectomycorrhiza) are color coded, see bottom panel. The dotted lines show median values. Genome (Mb): genome size in Mb, TE content (%): coverage (%) of transposable elements (TE) in assemblies, Genes (K): number of predicted genes, Secreted (K): number of predicted secreted proteins, Scaffolds (K): number of scaffolds, L50 (Mb): N50 length, BUSCO (%): genome completeness in % based on BUSCO survey. Source data Fig. 1.



**Fig. 2** The distribution of genome size (in Mb) and repeat element coverage (%) for each lifestyles. The boxes represent median, upper, and lower quartiles with the whiskers showing minimal and maximal values, and outliers in circle. The small dots show single observations. The number of species per lifestyle is as follows: ectomycorrhizal fungi ( $n = 45$ ), arbuscular mycorrhizal fungi ( $n = 10$ ), ericoid mycorrhizal fungi ( $n = 4$ ), orchid mycorrhizal fungi ( $n = 3$ ), endophytes ( $n = 5$ ), pathogens ( $n = 14$ ), soil/litter saprotrophs ( $n = 32$ ), and wood decayers ( $n = 17$ ). Source data Fig. 1.

**Phylogeny, orthologous, and paralogous genes.** Reconstructed phylogenetic relationships and estimated divergence dates are generally consistent with the dates recovered by previous studies. For example, we estimated the age of the most recent common ancestor (MRCA) of the Agaricomycetidae to be 132 Mya (Fig. 4a), while it was estimated to be 149 or 125 Mya in phylogenomic analyses by Floudas et al.<sup>26</sup> and Kohler et al.<sup>12</sup>, but 191–176 Mya in a multigene megaphylogeny by Varga et al.<sup>28</sup>. We estimated the ages of the MRCAs of Agaricomycetes, Agaricales, Polyporales, Russulales, and Boletales to be 280, 116, 104, 88, and 82 Mya (Fig. 4a), respectively. We estimated the MRCA of Pezizomycotina at 326 Mya (Fig. 4b), while Floudas et al.<sup>26</sup> obtained a mean age of 344 Mya.

We inferred gene families from the predicted proteomes using gene family clustering. A total of 68,923 and 129,258 gene families were identified in Ascomycota and Basidiomycota, respectively, and were used to infer orthology and paralogy (Supplementary Data 5). The species in our datasets contained substantial novelty in gene content. Species-specific genes ranged from 994 in *T. melanosporum* to 39,410 in *Mycena galopus*, for a total of 251,392 and 606,378 taxon-specific genes in Ascomycota and Basidiomycota, respectively. The large number of species-specific genes in *M. galopus* is the result of a series of striking expansion of gene families.

#### Functional gene categories encoded by mycorrhizal genomes.

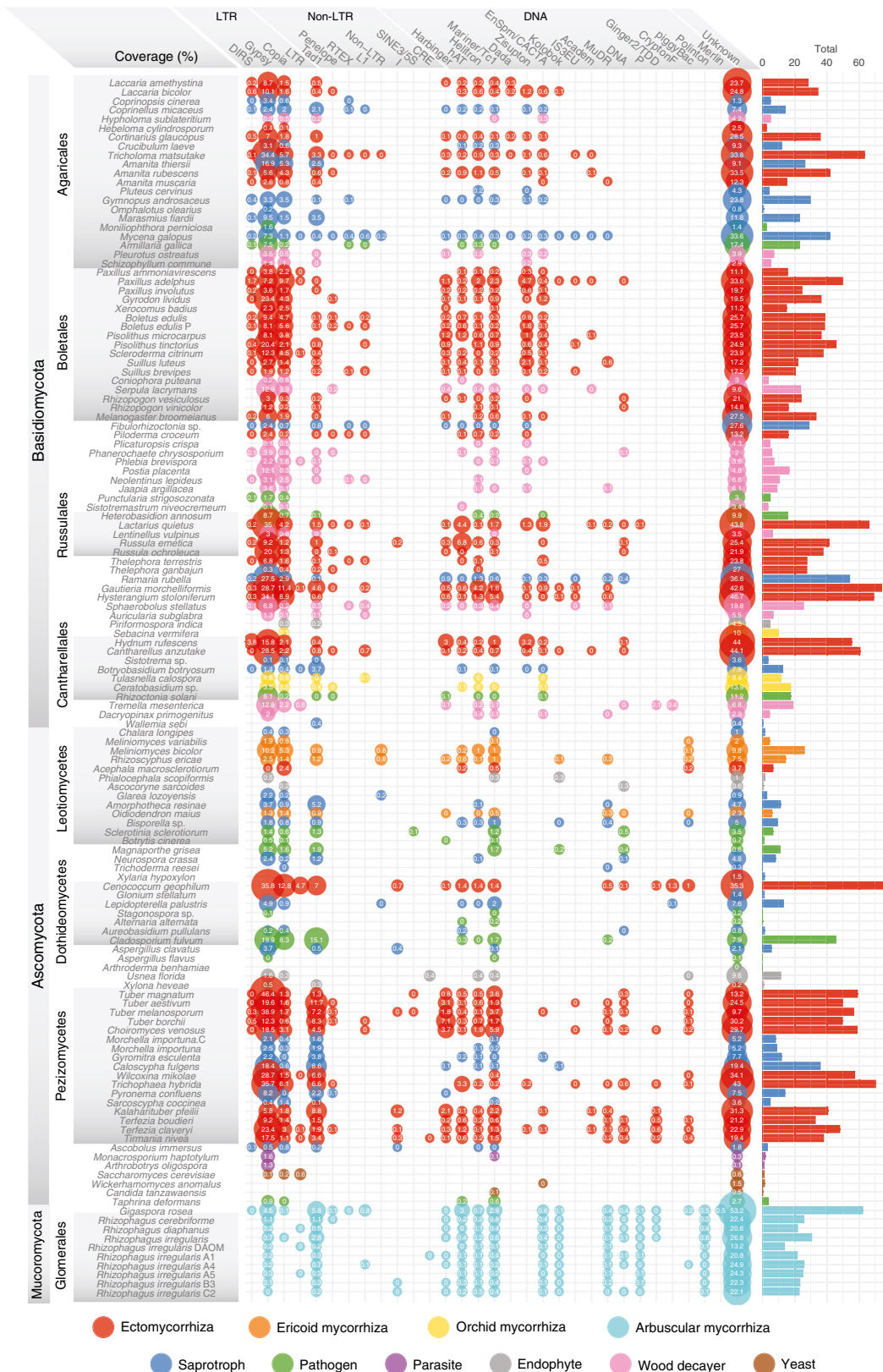
Hierarchical clustering of the presence and abundance of different Pfam protein domains identified genome-wide patterns of functional domain content among some of these fungi (Supplementary Fig. 2a). Arbuscular mycorrhizal fungi clustered together with Pfam categories showing a substantial differential abundance in genes encoding proteins with NUDIX, tetratricopeptide repeat, BTB/POZ, Sel1 repeat, ubiquitin, and high-mobility group box domains, corroborating our previous study<sup>27</sup>. This comparison also emphasizes some of the unique aspects of the genomes from the ectomycorrhizal ascomycete *Acephala macrosclerotiorum* (Helotiaceae) and ericoid mycorrhizal fungi (e.g., *Oidiodendron maius*, *Meliniomyces* species), such as the expansion of genes encoding proteins with FAD and AMP-binding domains, aldehyde dehydrogenases and sugar transporters. They share their

Pfam pattern with soil saprotrophic ascomycete species, such as *Chalara longipes* and the dark septate endophyte (DSE) *Phialocephala scopiformis*. Ectomycorrhizal fungi in Basidiomycota are grouped with soil saprotrophs and wood decayers, and displayed no specific pattern in their primary and secondary metabolism gene repertoires that may explain their symbiosis-related ability. Similarly, hierarchical clustering of the presence and abundance of the different membrane transporters and transcriptional regulators revealed no specific pattern(s) for ectomycorrhizal fungi (Supplementary Fig. 2b, c).

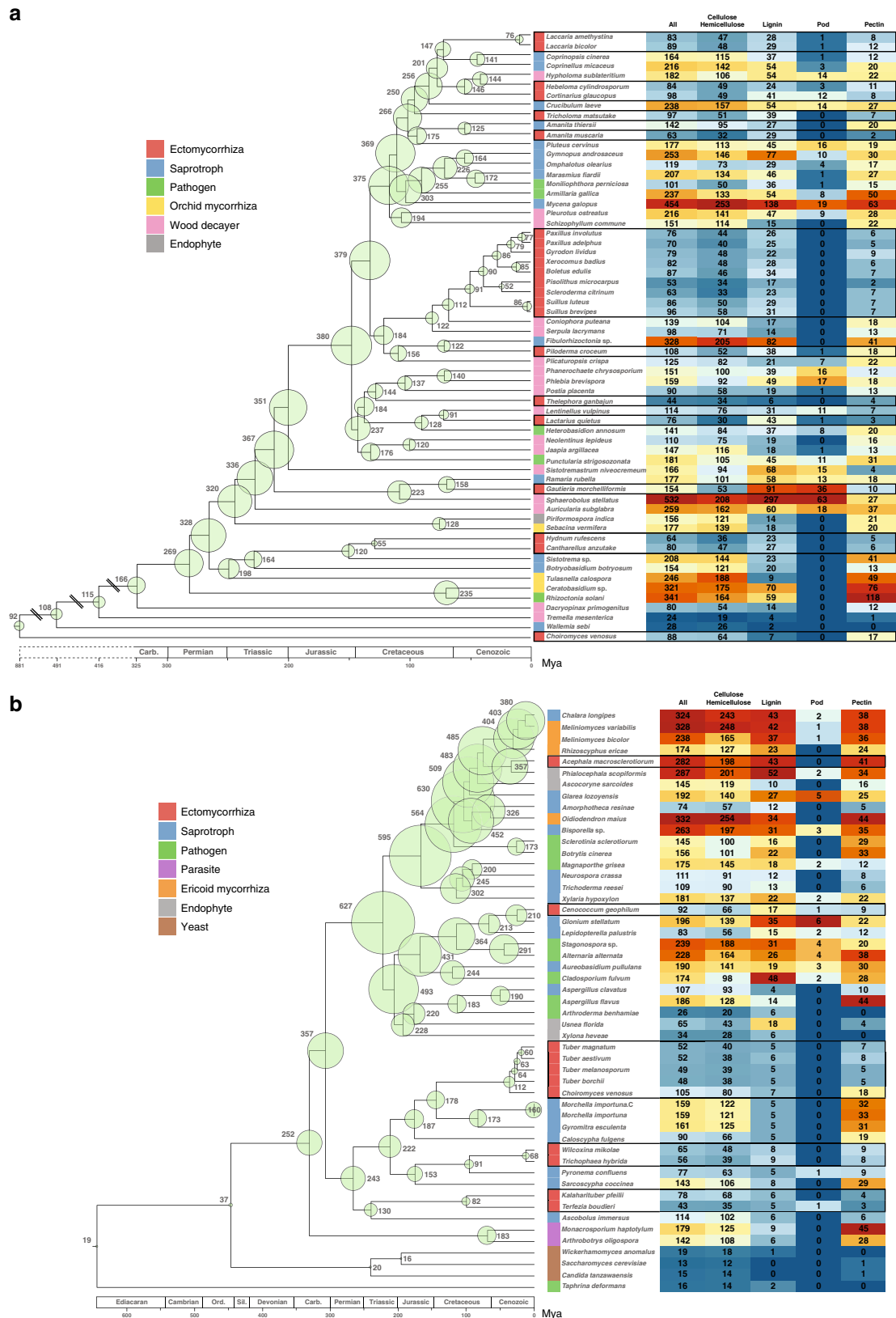
#### Predicted secretomes of saprotrophs and symbiotrophs.

As secreted proteins play a key role in SOM decomposition and symbiosis development, we compared predicted secretomes, including carbohydrate-active enzymes (CAZymes), lipases, proteases, and other secreted proteins, such as effector-like small secreted proteins (SSPs; Supplementary Data 6). The number of genes encoding secreted proteins represents 4.6–5.7% of the total protein repertoire for Basidiomycota and Ascomycota, respectively (Supplementary Fig. 3). The proportions of secreted proteins within the different protein categories were consistent in both phyla (Supplementary Fig. 3, and Supplementary Data 6 and 7). About half of the secreted proteins were SSPs (Supplementary Fig. 3a and Supplementary Data 7c). Only secreted CAZymes and SSPs showed significant differences in relative abundance among lifestyles ( $P < 0.01$ ; generalized Campbell and Skillings procedure; Supplementary Fig. 3 and Supplementary Data 8), with orchid and ericoid mycorrhizal symbionts possessing the largest CAZyme repertoires. On the other hand, the average number of secreted proteases and lipases is similar in saprotrophs and symbiotrophs. No expansion of gene families coding for secreted proteases, secreted phosphatases, and phytases were found in ectomycorrhizal fungi (Supplementary Fig. 3c).

**Losses of PCWDEs.** In Basidiomycota, ectomycorrhizal species contain significantly fewer secreted CAZymes acting on cellulose, hemicellulose, pectins, lignin, suberins, and tannins ( $P < 0.01$ , the generalized Campbell and Skillings procedure; Fig. 5, Supplementary Fig. 3c and Supplementary Data 6) than all other ecological guilds, i.e., lifestyles. They are also reduced in secreted



**Fig. 3 Distribution and coverage (%) of transposable elements families in genome assemblies.** The bubble size is proportional to the coverage of each of TE family (% indicated inside bubbles). The bars show the total coverage per genome. See also Supplementary Data 2-4. Source data are provided as a Source data Fig. 1.



auxiliary activity (AA) enzymes (including class II PODs, laccases, lytic polysaccharide monoxygenases (LPMO)), carbohydrate esterases (CE), polysaccharide lyases (PL), and associated cellulose-binding modules (e.g., CBM1; Supplementary Figs. 3–5, and Supplementary Data 7 and 8). Notably, there is no invertase GH32 gene in their genome, except for

*Cantharellus anzutake*, implying that ectomycorrhizal basidiomycetes are unable to use sucrose directly from the plant. Except for *Acrophthalma macrosclerotiorum*, the number of PCWDEs in ectomycorrhizal Ascomycota is lower than those of most of other groups, excluding yeasts and arbuscular mycorrhizal fungi (Figs. 4 and 5, and Supplementary Figs. 3–5). According to

**Fig. 4 Evolution of gene families encoding PCWDEs in Basidiomycota and Ascomycota.** **a** Basidiomycota. **b** Ascomycota. Fungal taxa are displayed according to their phylogeny (left panel). Bubbles with numbers at the tree nodes represent the total number of genes coding for total PCWDEs (intracellular and secreted) for ancestral nodes determined by COMPARE. The bubble size is proportional to the number of PCWDE genes. Heat maps contain the number of genes coding for substrate-specific PCWDEs for the extant species. Relative abundance of genes is represented by a color scale, from the minimum (blue) to maximum (red) number of copies per species. Ectomycorrhizal fungi are framed with a black line. See also Supplementary Data 8 and Supplementary Fig. 5. The tree is a chronogram estimated with r8s on the basis of a maximum likelihood phylogeny inferred with RAxML. The geological timescale (in million year) is indicated at the bottom. Source data are provided as a Source data Fig. 2.

RedoxiBase (<http://peroxibase.toulouse.inra.fr>), none of the class II PODs of ectomycorrhizal species are ligninolytic (i.e., ligninolytic POD or LiP), except those of *Gautieria morchelliformis* (see below; Supplementary Fig. 4 and Supplementary Data 6p). Below, we discuss the PCWDEs of ectomycorrhizal fungi, emphasizing newly sampled lineages with previously unknown decomposition capacity.

Cantharellales arose before the origin of ligninolytic class II POD<sup>26,28–30</sup>. Both saprotrophs (*Botryobasidium botryosum*, *Sistotrema* sp.) and orchid symbionts (*Tulasnella calospora*, *Ceratobasidium* sp.) in Cantharellales possess large sets of enzymes acting on cellulose, hemicellulose, and pectins, including LPMOs (AA9; Figs. 4 and 5, and Supplementary Figs. 3–5). We sequenced the first genomes of ectomycorrhizal Cantharellales, *Hydnum rufescens*, and *C. anzutake*, and found that they have highly reduced repertoires of secreted PCWDEs. In this regard, the ectomycorrhizal symbionts of Cantharellales are more similar to ectomycorrhizal Agaricales and Boletales than to orchid symbionts of Cantharellales. These results provide another independent example of convergent loss of PCWDEs in ectomycorrhizal lineages, and highlight the different decomposition capacities of two guilds of plant symbionts.

In contrast to Cantharellales, Phallomycetidae diverged shortly after the evolution of ligninolytic class II POD. The one previously published genome from this group, the saprotrophic “cannonball fungus”, *Sphaerobolus stellatus*, has an astonishingly large repertoire of PCWDEs, including 63 class II PODs. The two genomes of ectomycorrhizal Phallomycetidae reported here, *G. morchelliformis* (Gomphales) and *Hysterangium stoloniferum* (Hysterangiales)<sup>31</sup>, have diverse PCWDEs acting on cellulose, hemicellulose, pectins, and lignin, with 31 and five genes encoding ligninolytic class II PODs, respectively (Figs. 4 and 5, Supplementary Figs. 3–5 and Supplementary Data 8). Although many *Ramaria* species (Gomphales) form ectomycorrhiza, the newly sequenced *Ramaria rubella* (subgenus *Lentoramaria*) is likely a litter decomposer, which is consistent with its possession of 13 class II PODs, 6 cellulose-acting LPMOs, and GH6 and GH7 cellobiohydrolases. Thus, a robust suite of PCWDEs appears to be a characteristic of both saprotrophs and ectomycorrhizal species in Phallomycetidae.

Ectomycorrhizal Russulales and Thelephorales, for which we report the first genomes, both have highly reduced complements of PCWDEs. The two Thelephorales species (*Thelephora terrestris* and *T. ganbajun*) resemble Boletales in having a highly reduced suite of the major enzymes acting on crystalline cellulose and lignin (three cellulose-acting LPMOs, no class II POD, and no GH6 or GH7). Ectomycorrhizal Russulales (*Lactarius quietus*, *Russula emetica*, and *Russula ochroleuca*) also have small repertoires of PCWDEs (no CBM1, one or two cellulose-acting LPMOs, no GH6 or GH7, no CE1, and no LiP), but they retain one atypical Mn POD gene (Supplementary Data 6p). In contrast, saprotrophic Russulales (e.g., *Lentinellus vulpinus*) have a typical white rot array of PCWDEs.

In Ascomycota, species in Tubercularia followed the general symbiotroph trend, except *C. venosus*, which has a large complement of PCWDEs (Figs. 4a and 5, see ref. 16). Other

ectomycorrhizal Pezizales, such as the newly sequenced *Wilcoxina mikolae* and *Trichophaea hybrida* (Pyronemataceae), also have few PCWDEs. The desert truffles *Kalahariturber pfeilii* and *Terfezia boudieri* (Pezizaceae) both have ectomycorrhiza-like restricted suites of PCWDEs, but they differ in the numbers of genes encoding cellulose-acting LPMOs (ten and one copies, respectively) and they encode one copy of GH32 invertase. Like other ectomycorrhizal species, *A. macrosclerotiorum* contained no class II PODs, but had a large set of PCWDEs (282 genes, including 17 cellulose-acting LPMOs, and genes encoding GH6, GH7, GH32, GH45, PL1, and PL3; Figs. 4 and 5, Supplementary Figs. 3–5 and Supplementary Data 8). Closely related taxa to *A. macrosclerotiorum* mainly contain DSEs, such as *P. scopiformis*, saprotrophs, and pathogens, which are characterized by a larger set of PCWDEs compared to ectomycorrhizal fungi. To assess whether this large repertoire of PCWDEs is expressed during symbiosis development, we carried out transcript profiling of *Pinus sylvestris*–*A. macrosclerotiorum* ectomycorrhizas using RNA-Seq. The fungal symbiont develops a dense Hartig net within the root cortex, but no mantle sheath (Supplementary Fig. 6a). Transcript profiling of ectomycorrhizas showed that most of the PCWDEs are not induced in symbiotic tissues (Supplementary Fig. 6b and Supplementary Data 16), suggesting a tight transcriptional control of the PCWDE gene expression.

**MCWDEs are retained in ectomycorrhizal fungi.** We explored the genetic capabilities of the sequenced fungi to decompose microbial (i.e., bacterial and fungal) cell walls by comparing PCWDE to MCWDE repertoires (Figs. 4 and 5, Supplementary Figs. 3–5 and Supplementary Data 7). The proportion of MCWDE (acting on chitin, glucans, mannans, and peptidoglycans) in ectomycorrhizal fungi is similar to that in saprotrophs. Thus, ectomycorrhizal fungi have generally retained MCWDE genes (e.g., chitinases,  $\beta$ -1,3-glucanases) although they have lost most PCWDEs (e.g., endo- and exocellulases). Ectomycorrhizal fungi may use secreted MCWDE to scavenge nitrogen compounds trapped in SOM (e.g., chitin) by selectively using these hydrolytic enzymes in addition to oxidative mechanisms<sup>32</sup>. Some of these chitin-, glucan-, and mannan-active enzymes are likely involved in fungal cell wall remodeling during complex developmental processes, such as ectomycorrhiza and sporocarp formation<sup>10,11,33</sup>.

**Mycorrhizal development is driven by gene co-option.** We used a phylostratigraphic approach to characterize the evolutionary origins of ectomycorrhizal lineages on the basis of gene functions in extant organisms. We examined the age distribution of genes induced at different stages of ectomycorrhiza establishment, so-called symbiosis-induced genes, by defining phylogenetic ages (phylostrata), that correspond to internal nodes of the tree along the lineage leading from the root to the symbiotic species for which transcriptomic data are available (Fig. 6, Supplementary Fig. 7 and Supplementary Data 9). In Ascomycota and Basidiomycota, an average of 74% and 67% of the ectomycorrhiza-induced genes predated the evolution of ectomycorrhizal

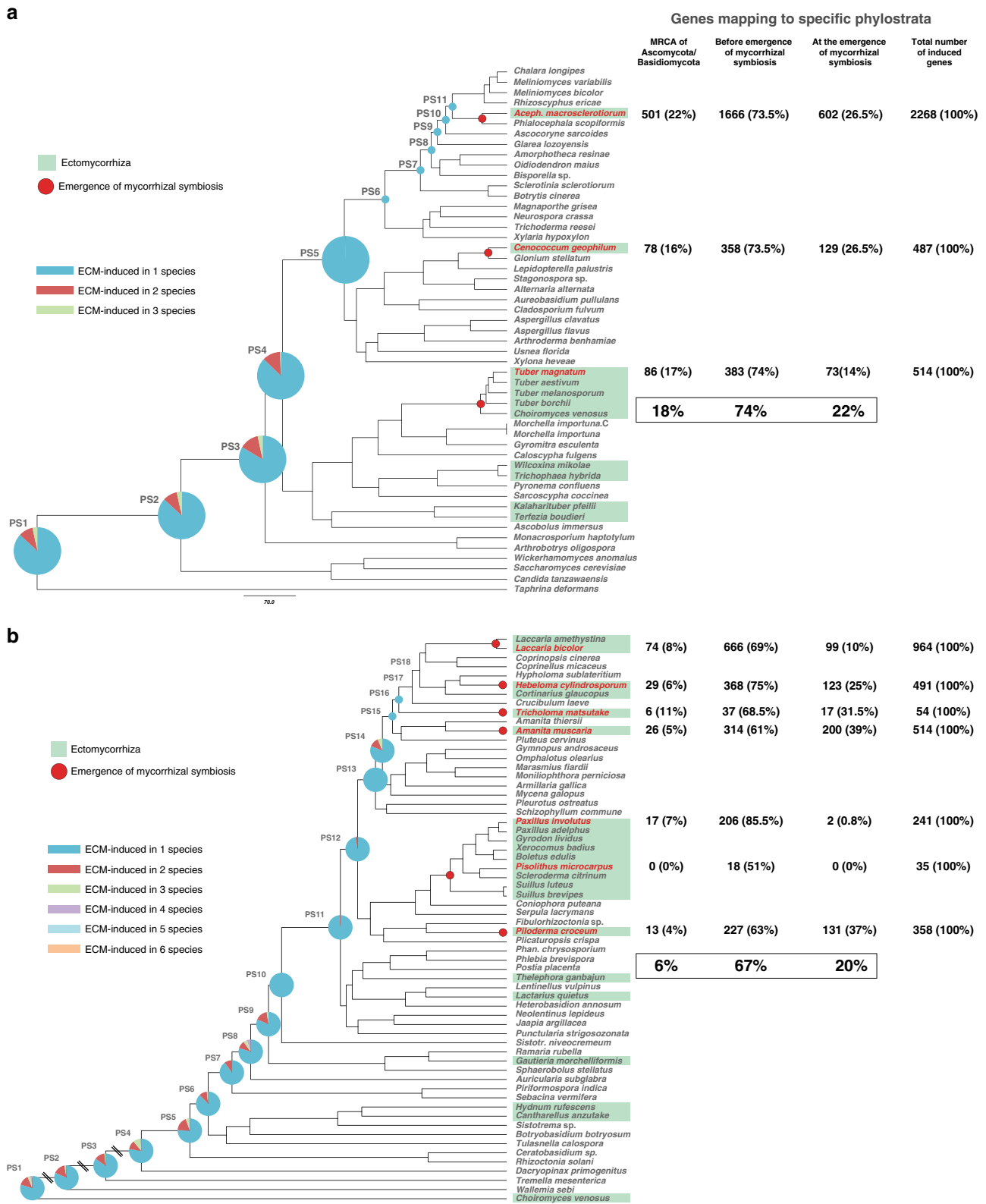


**Fig. 5 Distribution of key secreted PCWDEs in analyzed fungi.** Bubbles with numbers contain the number of genes coding for a series of secreted PCWDEs needed for cellulose and lignin degradation. Taxa are color coded according to their lifestyle (see bottom panel). See also Supplementary Data 6a, j. Left panel, Ascomycota and Mucoromycota; right panel: Basidiomycota. Source data are provided as a Source data Fig. 3.

symbiosis, respectively (Fig. 6). Approximately 6% and 18% of these genes were already present in the MRCAs of the Ascomycota and that of the Basidiomycota (e.g., for *L. bicolor*, hydrophobins, GH131, GH28, and CBM1\_GH5), respectively (Fig. 6 and Supplementary Fig. 7). No specific phylostrata was enriched

in symbiosis-induced genes (Fisher’s exact test,  $P < 0.005$ ; Supplementary Data 9b), suggesting that there was no distinguished period during ectomycorrhiza evolution characterized by an excess number of gene birth events. These observations imply that symbiosis-induced genes have mostly been co-opted





**Fig. 6 Phylostratigraphy for ectomycorrhiza-specific upregulated genes.** Pie charts on the time-calibrated trees represent the number of gene clusters containing ectomycorrhiza-specific upregulated genes from the species compared. Numbers and percentage of genes mapping to phylostrata are shown on the right of the trees. Ectomycorrhizal species used for the comparison are in green boxes. Upregulated genes were selected according to FDR adjusted  $P$  value  $< 0.05$ . **a** Ascomycota, the fold change (FC) used for defining ectomycorrhiza-specific upregulated genes was  $FC \geq 5$ . **b** Basidiomycota, the FC used for defining ectomycorrhiza-specific upregulated genes was  $FC \geq 5$ . See Supplementary Data 9 for the identified phylostrata containing ectomycorrhiza-upregulated genes and enrichment statistics. Source data are provided as a Source data Fig. 4.

for ectomycorrhiza development during evolution from saprotrophic ancestors. However, the phylostratigraphic analysis also suggested that an average of 9–22 % (in Ascomycota; Fig. 6a) and 19–20 % (in Basidiomycota; Fig. 6b) of ectomycorrhiza-induced genes are restricted to specific mycorrhizal lineages. Given our current taxon sampling, it is difficult to distinguish genes that coincidentally evolved with mycorrhiza formation from species-specific orphan genes, except in the case of three clades that comprise more than a single ectomycorrhizal species, Boletales, the genus *Laccaria* and Tuberales, which comprise more than a single ectomycorrhizal species. In these clades, the proportion of symbiosis-induced genes that map to the origin of ectomycorrhiza showed a large variation, from 0 to 0.8% in the Boletales (two species), to 10% in *Laccaria*, and 14% in the Tuberales (Fig. 6).

To test our phylostratigraphic results, we assessed the evolutionary conservation of 5917 ectomycorrhiza-induced transcripts identified in ten different ectomycorrhizal interactions among the 135 studied fungal genomes (Fig. 7a). We found that 15.2 % of the 5917 symbiosis-induced genes are shared by all species in our dataset (clusters IV and V). Most encode proteins involved in core metabolic or signaling functions. Genes from cluster III are conserved within Basidiomycota only, while genes from cluster VII are shared by Ascomycota only. Most of them have no known function (i.e., no KOG domain). Altogether, 31.2% of these symbiosis-induced genes are species specific (cluster VI). Most of these genes encode proteins with unknown KOG functions or mycorrhiza-induced small secreted proteins (MiSSPs). The proportion of species-specific, ectomycorrhiza-induced genes is therefore substantial, as reported earlier<sup>10,12,16</sup> and suggested by our phylostratigraphic analysis.

Next, we examined whether the same or different gene families were co-opted in independent ectomycorrhizal lineages, i.e., whether co-option was convergent or divergent during fungal evolution. We quantified convergence by the extent of overlap among conserved ectomycorrhiza-induced gene families within independent lineages. Conserved ectomycorrhiza-induced genes showed little or no overlap among the analyzed species (Fig. 6 and Supplementary Data 9). For example, there were only eight clusters of ectomycorrhiza-induced genes common to at least four Basidiomycota symbionts (Supplementary Data 9). Similarly, only 12 gene clusters were shared by at least three Ascomycota symbionts (Supplementary Data 9). This low overlap between co-opted genes suggests that independently evolved ectomycorrhizal lineages recruited different ancestral gene families for symbiosis, in addition to a likewise unique set of novel genes, that evolved after the origins of symbiosis<sup>12</sup>.

**Symbiosis-induced secreted proteins in symbionts.** The expression of genes encoding secreted and symbiosis-induced secreted proteins from ectomycorrhizal fungi was measured by RNA-Seq profiling in ten ectomycorrhizal associations (Supplementary Methods). By using a BLASTP-based analysis, we assessed the evolutionary conservation of the expressed 6669 secreted proteins (Fig. 7b) and 1028 symbiosis-induced secreted proteins (Fig. 7c) among 135 fungal species (Supplementary Datas 12, 13 and 14). A substantial proportion of the secreted proteins are species-specific SSPs (cluster III, Fig. 7b). In addition, we found that 38.1% of symbiosis-induced secreted proteins are shared by all species of fungi (clusters I and III). Most code for core metabolic or signaling functions and CAZymes. Genes from cluster II (9.3%) are conserved within Basidiomycota only. Genes from cluster IV (21.4%) are conserved mainly in Ascomycota, showed a lower similarity in Basidiomycota and are poorly conserved in Glomeromycota. They encode

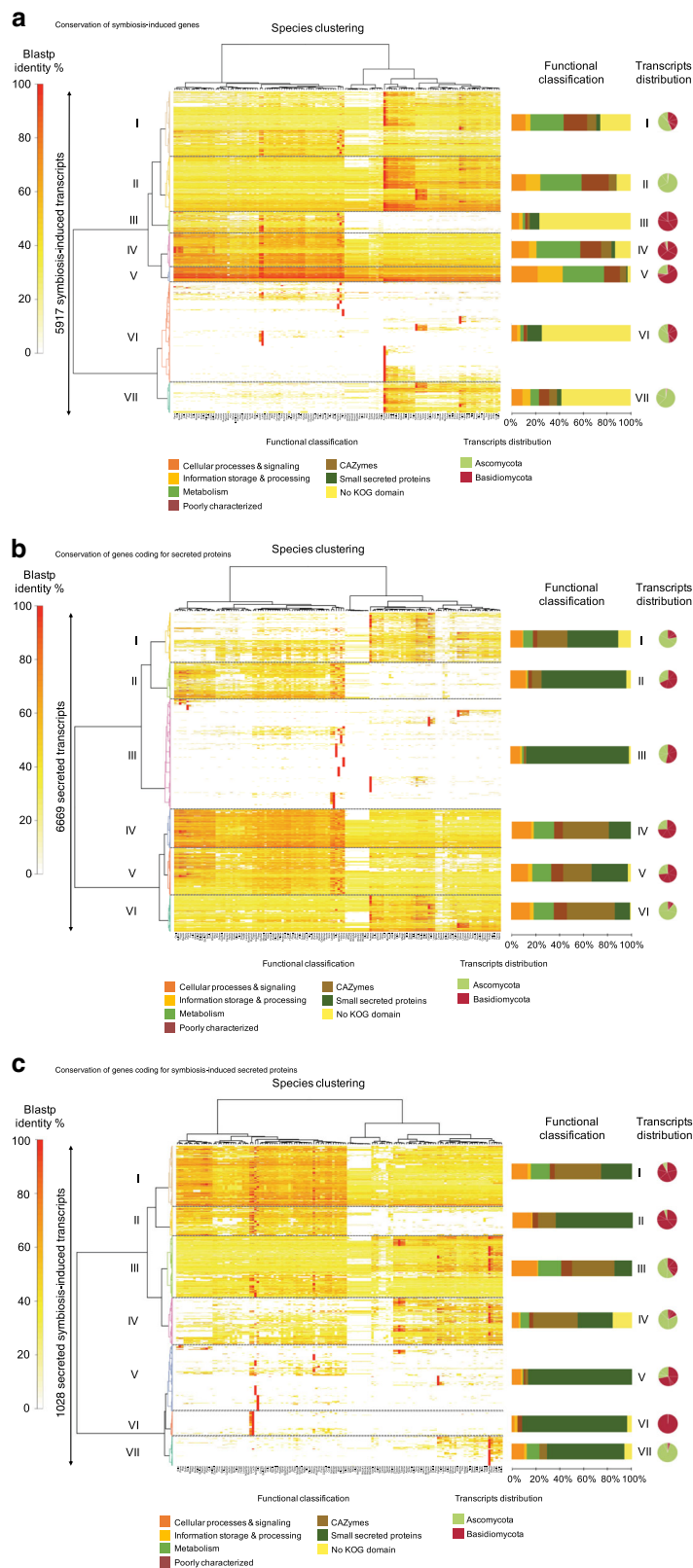
MiSSPs, CAZymes, proteins with unknown KOG function and proteins of signaling and metabolic pathways. Altogether, 31.2% of these 1028 symbiosis-induced genes are mostly species specific (clusters V–VII). Most of these genes encode MiSSPs, proteins with unknown KOG functions and proteins of signaling pathways. The phylostratigraphic analysis of secreted proteins and MiSSPs corroborated these results (Supplementary Fig. 7).

## Discussion

After the origin of Pinaceae (ca. 200 Mya), ectomycorrhizal symbiosis arose repeatedly, perhaps 80 times or more, across multiple lineages of Mucoromycotina (Endogonales), Ascomycota, and especially Basidiomycota<sup>2–4,8,34,35</sup>. The ancestors of ectomycorrhizal fungi are genetically and ecologically diverse, making this a superb example of convergent evolution<sup>3,7,9,12</sup>. To assess the general shared properties of the lifestyle evolution and functional biology of ectomycorrhizal symbioses, we conducted a comparative analysis of 62 mycorrhizal and 73 non-mycorrhizal fungal species. Our dataset, the most comprehensive so far, includes several major fungal clades that have not been sampled previously for ectomycorrhizal genomes, i.e., Cantharellales, Phallomycetidae, Thelephorales, and Russulales. For the sake of comparison, it also includes genomes of arbuscular mycorrhizal fungi, orchid mycorrhizal fungi, and ericoid mycorrhizal fungi providing a unique opportunity to highlight differences and similarities between the major types of mycorrhizal symbioses.

Our analyses of early diverging clades of ectomycorrhizal fungi (e.g., Cantharellales) support the general view that transitions from saprotrophy to ectomycorrhizal symbiosis involve (1) widespread losses of PCWDEs acting on lignin, cellulose, hemicellulose, pectins, suberins, and tannins, (2) co-option of metabolic and signaling genes present in saprotrophic ancestors to fulfill new symbiotic functions, (3) diversification of novel, lineage-specific symbiosis-induced orphan genes, and (4) massive proliferation of TEs. In addition, they corroborated and extended our previous analyses of the genomes of arbuscular mycorrhizal fungi<sup>27</sup> and ericoid mycorrhizal fungi<sup>15</sup>. Despite the general trend toward losses of PCWDEs in ectomycorrhizal lineages in both Ascomycota and Basidiomycota, there is considerable diversity in the apparent decay capacities of ectomycorrhizal fungi. For example, in Agaricales (Agaricomycetidae), *C. glaucopus* possesses 12 recently duplicated copies of atypical class II POD genes, which may confer some ability to obtain nutrients from soil phenolic compounds, as previously suggested<sup>36</sup>. However, secreted PODs can play a variety of additional roles, such as biosynthesis of cell wall melanins, detoxifying the immediate hyphae environment, or/and converting plant polymers into more oxidized, recalcitrant components of SOM. Similarly, the genome of *T. matsutake* (also Agaricales) encodes two GH7 cellobiohydrolases, in agreement with its known facultative saprotrophic ability<sup>37</sup>. Ectomycorrhizal fungi that evolved from brown rot fungi in the Boletales (also Agaricomycetidae), such as *Paxillus involutus*, appear to have adapted the oxidative decomposition system from their saprotrophic ancestors to liberate N entrapped in decaying SOM<sup>38–41</sup>. Their secreted proteases, LPMOs, and laccases may act in concert to decay available SOM compounds. Alternative hypotheses for the role of the remaining PCWDEs in ectomycorrhizal fungi include the remodeling of root cell walls during host colonization<sup>42</sup>.

Among the newly sampled fungal groups, ectomycorrhizal Phallomycetidae and Cantharellales present highly variable suites of PCWDEs. Most striking is the ectomycorrhizal *G. morchelliformis* (Gomphales, Phallomycetidae), which has 25 class II manganese PODs (MnP) that may be involved in the decay of SOM and/or detoxification of soil polyphenolic compounds, as



well as numerous enzymes active on cellulose and hemicellulose. The class II PODs of *G. morchelliformis* are members of a gene family that underwent expansion after the divergence of Auriculariales and other Agaricomycetes. In contrast, Cantharellales evolved prior to the diversification of class II LiP<sup>30</sup>, and none of its species—saprotrophic or mycorrhizal—possesses these

ligninolytic enzymes. In contrast to their saprotrophic cousin *Sistotrema* sp., *C. anzutake* and *H. rufescens* are ectomycorrhizal species of Cantharellales that lack not only class II LiP, but also many glycoside hydrolases acting on cellulose, hemicellulose and pectins, LPMOs, and CBM1 motifs that are needed for plant cell wall and SOM decomposition. In this regard, they resemble some

**Fig. 7 Phylogenetic conservation of ectomycorrhiza-induced genes.** We analyzed genes coding for ectomycorrhiza-induced proteins, secreted proteins and ectomycorrhiza-induced secreted proteins among 135 fungi. Heat maps show BLASTP sequence similarity of 5917 ectomycorrhiza-induced proteins (a), 6669 genes coding for secreted proteins (b), and 1028 ectomycorrhiza-induced secreted proteins (c) from ten ectomycorrhizal fungi, with available symbiotic transcriptomes among 135 genomes of Basidiomycota and Ascomycota. The heat maps depict a double-hierarchical clustering of protein sequences encoded by symbiosis-upregulated genes, genes coding for secreted proteins and genes coding for symbiosis-upregulated secreted proteins (rows, fold change  $\geq 5$  in symbiotic tissues compared to free-living mycelium, false discovery rate-corrected  $P \leq 0.05$ ); right panel, functional categories (KOG) are given for each cluster of sequences in % as bargrams. Clusters I–VIII correspond to group of sequences sharing the same level of protein sequence similarity based on BlastP. The distribution of transcripts belonging to Ascomycota and Basidiomycota in each cluster is shown as pie charts. Data were visualized and clustered using R (package HeatPlus<sup>97</sup>). The hierarchical clustering was done by using a Euclidian distance and Ward clustering method. Color scale on the left (white to red) shows the % of sequence identity according to BLASTP. The symbiosis-induced genes were retrieved from the ectomycorrhizal transcriptomes of ten species: *A. macrosclerotium* EW76-UTF0540, *A. muscaria* Koide, *C. geophilum* 1.58, *H. cylindrosporum* h7, *L. bicolor* S238N, *P. involutus* ATCC 200175, *P. croceum* F1598, *P. microcarpus* 441, *T. matsutake* 945, and *T. magnatum* (see Supplementary Data 10, 13 and 14). GEO accession codes are provided in the “Methods/Phylostratigraphy” section. Source data are provided as a Source data Fig. 5.

of the most derived ectomycorrhizal Agaricomycetidae with highly reduced saprotrophic capabilities, such as species of Boletales. The ectomycorrhizal Russulales and Thelephorales that we sampled also conform to this model of a greatly reduced saprotrophic apparatus in prominent symbiotic fungi playing major roles in forest ecosystems.

Ectomycorrhizal fungi are not all depauperate in PCWDEs. Species such as *A. macrosclerotium* in the Leotiomycetes, may represent transitional steps from pure saprotrophy toward pure ectomycorrhizal symbiosis. Such taxa could be capable of facultative saprotrophy, i.e., SOM decomposition. As suggested for ericoid mycorrhizal fungi, which exhibit a large PCWDE repertoire<sup>15</sup>, a dual saprotrophic/mutualistic habit may provide greater ecological flexibility and fitness under specific environmental conditions. Moreover, in such lineages, evolutionary reversal to saprotrophy might be possible (although that has not been demonstrated). We showed that the PCWDEs of *A. macrosclerotium* are downregulated at the transcriptional level during the symbiotic interaction to avoid triggering plant defense reactions and/or digesting host root tissues. This suggests that loss of genes encoding PCWDEs is a consequence of, but not a requirement for, the evolution of ectomycorrhizal mutualisms. What remains to be determined is why ectomycorrhizal fungi do not retain their ancestral dual lifestyle, and how the ectomycorrhizal associations become so intimate that saprotrophic capabilities are irretrievably lost. The lack of invertase and sucrose transporter genes in most ectomycorrhizal fungi means that they are unable to use apoplastic sucrose released by their host plant. They fully rely on their partner for their glucose supply, a mechanism reinforcing their dependence on the plant partner.

Our results, and those of prior studies<sup>3,14</sup>, suggest that the predominant mechanism for the transition from saprotrophy to ectomycorrhizal symbiosis appears to involve restricted secretion of hydrolytic enzymes acting on plant cell walls (whether via regulatory shifts or gene loss), which enables the symbiont to be accommodated in roots. An example of an ultimate adaptation to symbiosis is illustrated by the arbuscular mycorrhizal fungi that are obligate symbionts. They display the lowest repertoire of PCWDEs of the current set of sequenced mycorrhizal fungi. Unfortunately, no genomic data exist yet to support the origin of arbuscular mycorrhizal fungi from saprotrophic CAZyme-rich Mucoromycotina. In contrast, ericoid mycorrhizal fungi and orchid mycorrhizal fungi have the higher set of PCWDEs supporting their dual saprotrophic/symbiotic lifestyles<sup>3,15</sup>. We thus propose that the PCWDE repertoire is reflective of the age of the symbiosis along the saprotrophy to symbiosis continuum with the arbuscular mycorrhizal symbiosis emerging in the early Devonian (393–419 Mya), the ectomycorrhizal symbiosis during the Jurassic (ca. 200 Mya) and mycorrhizal associations with Ericaceae species to the Cretaceous (ca. 117 Mya)<sup>4</sup>.

Further insight into the differences in gene composition between saprotrophic and symbiotic fungi was obtained by focusing on genes involved in adaptations to the host plant habitat in which symbionts thrive. Novel genes, such as effector-like MiSSPs, are likely required for symbiosis development (e.g., dampening of plant defense reactions<sup>43,44</sup>) and metabolism<sup>45</sup>, and have evolved in every ectomycorrhizal lineage. Transcript profiling of several ectomycorrhizal interactions has shown that 7–38% of genes that are upregulated during symbiosis are species-specific genes, i.e., they are restricted to a single mycorrhizal species. Among all symbiosis-upregulated genes, 8–28% encode candidate secreted effector-like MiSSPs. Their functional analysis is hampered by the the lack of high-throughput genetic engineering techniques for ectomycorrhizal fungi, but some of these MiSSPs have been functionally characterized, including LbMiSSP7, LbMiSSP8, LbMiSSP7.6, and PaMiSSP10 (refs. 44,46–48). The evolution of innovations, such as the control of the host plant immunity, development, and metabolism by these lineage-specific symbiotic effectors was likely a necessary step to colonized the new niches represented by tree roots. Some of the effector-like SSPs, but also structural SSPs like hydrophobins, may have evolved from SSPs used by the saprotrophic ancestors to communicate or compete in the soil and wood environments<sup>46,49</sup>.

Novel and recently evolved genes, including MiSSPs, are undoubtedly responsible for the particular attributes of individual mycorrhizal lineages. Indeed, arrays of ectomycorrhiza-induced genes are unique to each specific clade with between 14 and 39% of symbiosis-upregulated genes being species-specific genes with no known function<sup>12</sup> (present study). At the same time, our phylostratigraphic analysis showed that a large proportion of genes that are upregulated in ectomycorrhizas arose long before the evolution of the mutualistic associations. In other words, numerous genes used by saprotrophic ancestors were co-opted for the symbiotic lifestyle. Important ecological traits, such as N and P acquisition traits (e.g., organic N- and P-degrading secreted enzymes) were already present in free-living saprotrophic ancestors of ectomycorrhizal symbionts. Despite the well-documented ability of ectomycorrhizal fungi to hydrolyze organic phosphate compounds and scavenge nitrogen through the degradation of litter proteins accumulating in soil litter, we found that ectomycorrhizal genomes generally contain a similar number or less copies of genes coding for secreted N- and P-targeting hydrolases than saprotrophs, pathogens, or ericoid mycorrhizal fungi.

One of the most striking genomic features of mycorrhizal fungi is their high content in TEs. A large proportion of the genomes of mycorrhizal fungi can be ascribed to past transposition, providing a major contribution to their genome landscape. It has been proposed that the lack of sexual recombination (i.e., asexual

reproduction) can favor the uncontrolled proliferation of TEs<sup>47</sup>. It may be true for arbuscular mycorrhizal fungi with no known sexual reproduction, but this hypothesis cannot stand for the ectomycorrhizal fungi investigated in the present study as they are known to produce ascocarps or basidiocarps. In biotrophic pathogens, such as *Blumeria graminis* f. sp. *hordei* and *B. graminis* f. sp. *tritici*, it has been proposed that burst(s) of TEs offer a template for rapid evolution of virulence genes by duplications, small-scale rearrangements, or deletions<sup>48</sup>. It is tempting to speculate that the observed TE proliferation(s) play a key role in promoting the rapid evolution of symbiosis-related factors, such as MiSSPs.

Collectively, our results suggest that each independently evolved ectomycorrhizal lineage uses species-specific novel genes, as well as members of ancestral gene families to develop symbiotic interactions, indicating that divergent genetic innovations underlie the convergent origins of the ectomycorrhizal guild. Thus, the functional heterogeneity in symbiosis-related gene networks in ectomycorrhizal fungi is not only found among species-specific genes as evidenced for MiSSPs<sup>3,14,43–45,50,51</sup>, but also in conserved genes (e.g., hydrophobins) that predate the gains of mycorrhizal symbiosis.

The large number of independently evolved ectomycorrhizal lineages raises questions as to why the emergence of this guild is so common in nature. Furthermore, the absence of known reversals to the saprotrophic lifestyle suggests that genetic traits underlying the transitions to ectomycorrhizal lifestyles represent fixed evolutionary transitions. Future studies should focus on conditions that may predispose some groups to evolve ectomycorrhizal symbioses, or other mechanisms that increase the likelihood of convergent evolution<sup>52–55</sup>. The present study roughly doubles the number of published ectomycorrhizal genomes. Even greater sampling of key lineages, particularly early divergent clades, is needed to elucidate the macro- and microevolutionary mechanisms that were responsible for the diversification of ectomycorrhizal fungi.

## Methods

**Strains and fungal material used for genome sequencing.** Fungal strains used for genome sequencing are described in Supplementary Data 1. Assignment of individual taxa to functional guilds was based on common categorizations in the literature<sup>34,56,57</sup>. DNA was extracted with a modified cetyltrimethylammonium bromide protocol<sup>12</sup>.

**Genome sequencing, assembly, and annotation.** All sequencing, assembly, and annotation was performed at JGI. Genome sequencing was done with either Illumina or Pacific Biosciences (PacBio) technology, or both (“hybrid”). Genomes were assembled from Illumina reads using ALLPATHS-LG<sup>58</sup>, followed by patching using PacBio with PBjelly<sup>59</sup> in the case of hybrid assemblies. All other genomes were assembled from PacBio reads using Falcon<sup>60</sup> or Celera Assembler<sup>61</sup>. Mitochondrial genomes were assembled separately. All transcriptome sequencing was done with Illumina only, and subsequently assembled into putative transcripts using either Rnnotator<sup>11</sup> or Trinity<sup>62</sup>. Each genome was annotated using the JGI Annotation Pipeline<sup>63,64</sup>, aided by the transcriptome when available. Detailed methods are described in the Supplementary Methods.

**Organismal phylogeny.** We assembled two datasets, one of Ascomycota that included 51 genomes and one of Basidiomycota that included 62 genomes. We opted for analyzing Ascomycota and Basidiomycota genomes separately, and only fungal genomes to improve computational efficiency and because the origins of the ectomycorrhizal symbioses are shallow evolutionary events, so we reasoned that additional resolution cannot be added by including non-fungal outgroup taxa. The computational step of similarity-based clustering, multiple sequence alignments, and tree inference are particularly sensitive to the phylogenetic breadth that the species cover. Analyzing Ascomycota and Basidiomycota separately improved the accuracy of our predictions, by not having to analyze as deep nodes of the tree as the Dikarya (MRCA of Ascomycota and Basidiomycota). We performed all-vs-all Blast with mpiBlast 1.6.0 (ref. <sup>65</sup>) and clustered proteins into gene families by Hifix 1.0.5 (ref. <sup>66</sup>), using default parameters. Clusters having  $0.5n-2n$  proteins ( $n$  = number of species) were aligned by PRANK140603 (ref. <sup>67</sup>) with default settings, and trimmed with trim-AL 1.4.rev15 (ref. <sup>68</sup>) to remove ambiguously aligned regions (with the argument -gt 0.1). Maximum likelihood (ML) gene trees

were inferred using FastTree 2.1.10 (refs. <sup>69,70</sup>) under the Whelan and Goldman (WAG) model of protein evolution with gamma-distributed rate heterogeneity. We excluded gene trees with deep paralogs to identify single-copy gene. Gene trees of not strictly single-copy gene families were checked if the duplications are deep in the tree, or species specific. In those cases where the duplications were species specific, the gene was marked as suitable for phylogenetic inference and the protein that was closest to the root (in terms of patristic distance) for each species was chosen for inclusion in the phylogenetic analyses.

Single-copy genes were realigned using PRANK v. 150803 (ref. <sup>67</sup>) with the default settings and one round of alignment improvement. Poorly aligned regions were removed using trim-AL v1.2 (ref. <sup>68</sup>) with the -strict setting. The alignments were concatenated into a supermatrix using a custom Perl script, excluding single alignments that were <50 amino acids long. We included genes in the supermatrix only if they were present in >40 and >50 Ascomycota and Basidiomycota species, respectively. ML phylogenies were inferred using RAxML 8.2.4 (ref. <sup>71</sup>) with 100 bootstrap replicates and a partitioned model, where each gene was treated as a separate partition. The PROTGAMMAWAG model was used for each partition. The resulting phylogenies were used in the molecular dating analyses.

In addition, we constructed an organismal phylogeny restricted to the 62 mycorrhizal fungi. Orthologous genes among the fungi were identified using FastOrtho with the parameters set to 50% identity, 50% coverage, and inflation 3.0 (ref. <sup>72</sup>). Protein sequences were downloaded from MycoCosm ([mycocosm.jgi.doe.gov](http://mycocosm.jgi.doe.gov)). Clusters with single-copy genes were identified and aligned with MAFFT 7.221 (ref. <sup>73</sup>), ambiguous regions (containing gaps and poorly aligned) were eliminated, and single-gene alignments were concatenated with Gblocks 0.91b (ref. <sup>74</sup>). A phylogenetic tree was constructed with RAxML 7.7.2 (ref. <sup>75</sup>), the standard algorithm, the PROTGAMMAWAG model of sequence evolution, and 1000 bootstrap replicates.

**Molecular dating.** In order to time calibrate the phylogenies, we used the penalized likelihood algorithm as implemented in r8s (ref. <sup>76</sup>) with the POWELL optimization. To identify the appropriate smoothing parameters, we performed a cross-validation analysis for each dataset. For the Basidiomycota phylogeny, the fossil *Archaeomarasmius legetti* from the mid-Cretaceous (90–94 Ma)<sup>77</sup> was used to calibrate the node containing the suborder Marasmiineae, and the fossil of a suilloid ectomycorrhiza from the middle Eocene (50 Mya)<sup>78</sup> was used to calibrate the node containing the suborders Boletineae, Paxillineae, Sclerodermatineae, and Suillineae. For the Ascomycota phylogeny, the fossil *Paleopyrenomyces devonicus* from the early Devonian (~400 Ma)<sup>79</sup> was used to calibrate the split between Pezizomycotina and Saccharomycotina. The Basidiomycota supermatrix consisted of 940,059 sites from 1042 genes. All nodes were recovered with a bootstrap support of 96% or higher. The Ascomycota supermatrix consisted of 1,474,883 sites from 1432 genes. All nodes were recovered with bootstrap support of 100%.

**Analysis of PCWDE evolution.** A comprehensive list of CAZyme annotations can be found in Supplementary Data 11 and at the Mycorrhizal Fungi page at the JGI MycoCosm database by following this link: [https://mycocosm.jgi.doe.gov/mycocosm/annotations/browser/cazy/summary;-6jBQE?P=Mycorrhizal\\_fungi](https://mycocosm.jgi.doe.gov/mycocosm/annotations/browser/cazy/summary;-6jBQE?P=Mycorrhizal_fungi).

We investigated the evolution of gene families encoding CAZymes with a particular interest in PCWDEs, which include enzymes active on cellulose (various GH families, AA3\_1, LPMO, and CBM1), lignin (AA1\_1, AA1\_3, AA2\_class II POD, AA5\_1, DyP, HTP, and OXO), and pectin (CE8, PL, and GH families; see also Supplementary Data 8). We first performed all-vs-all Blast using mpiBlast 1.6.0 (ref. <sup>80</sup>) for 51 Ascomycota and 61 Basidiomycota proteomes, and clustered proteins into gene families using the Hifix 1.0.5 clustering method, with default parameters resulting in 243,784 and 408,904 clusters, respectively. Identification of PCWDE families was performed based on InterPro (IPR) domains in the similarity-based clustering (see above). IPR data from the JGI website were used to annotate proteins with IPR domains for all species. We considered clusters as PCWDE families if at least 50% of the proteins had the appropriate IPR domain. Further classification of PCWDE families was supported by BLASTP 2.6.0+ search<sup>80</sup>, and validation by the CAZy annotation pipeline<sup>81</sup>. Multiple sequence alignments were inferred using PRANK 140,603 with default settings, and trim-AL 1.4.rev15 was used for removing ambiguously aligned regions (with the argument -gt 0.1). ML gene trees of the identified PCWDE clusters were inferred using FastTree 2.1.10 under the WAG model of protein evolution with gamma-distributed rate heterogeneity. TreeFix 1.1.10 (ref. <sup>82</sup>) was used for gene tree/species tree reconciliation (-m PROTGAMMAWAG, -alpha 0.001, -niter 300) with the genome-based ML species trees. Gene clusters containing less than four proteins were excluded from the analysis. For the reconstruction of gene duplication and loss history of PCWDEs, we used the COMPARE pipeline<sup>54</sup>. Orthogroups were identified, and duplications and losses were inferred for each enzyme group across species trees based on Dollo parsimony. The ancestral gene copy numbers for every internal node were calculated by summing the mapped gains and losses over the species tree.

Class I and class II PODs were also curated by the RedoxiBase database (<http://peroxibase.toulouse.inra.fr>) to distinguish between LiP, versatile PODs, MnP, or atypical Class II PODs. The latter PODs have been annotated as basidiomycete sub-class B (CIIBB) or sub-class C (CIIBC)<sup>83</sup>.

**Phylostratigraphy.** Phylogenetic ages of ectomycorrhiza-induced genes were assigned based on the most phylogenetically distant orthologous gene in the containing gene family, using a phylostratigraphic approach<sup>84</sup>. Ectomycorrhiza-induced transcripts were retrieved from RNA-seq-based transcriptome profilings: *Amanita muscaria*, GSE63867; *Cenococcum geophilum*, GSE83909; *Hebeloma cylindrosporum*, GSE63868; *P. involutus*, GSE63924; *Piloderma croceum*, GSE63925; *T. magnatum*, GSE116692; *A. macrosclerotium*, SRP130276 and SRP130279–82; *L. bicolor*, SRP164436–38, SRP164526, SRP164559, and SRP164564; *Pisolithus microcarpus*, SRP122806, SRP122812, SRP122818, SRP122826, SRP122829, and SRP122850; and *T. matsutake*, SRP103258). Details of the RNA-seq experiments are described on Gene Expression Omnibus (GEO), NCBI. Note that this pooled set of ectomycorrhiza-induced genes represent a heterogeneous set of symbiosis-induced transcripts from ectomycorrhizal roots from different fungal-tree associations sampled at different stages of development, from early to late stage of symbiosis establishment. We selected those having an expression fold change (FC) > 5 and FC > 2 (Supplementary Data 9). The phylogenetic trees of the Asco- and Basidiomycota were divided into phylostrata, based on the phylogenetic distance in terms of internal nodes from the species being examined. Each phylostratum corresponded to an ancestral species; genes were assigned to phylostrata based on the species distribution of orthogroups the studied genes belong to. Using this method, the ages of upregulated genes were defined and mapped to species trees. We thoroughly selected a subset of the sequenced genomes of species belonging to either the Ascomycota or the Basidiomycota. Running COMPARE on the present set of 135 Glomeromycotina, Ascomycota, and Basidiomycota genomes separately would have been prohibitively time consuming involving disproportionate costs.

**Comparative analyses and annotation of functional categories.** Statistics of JGI genome assemblies (i.e., N50, number of genes and scaffolds, and genome size) were obtained from MycoCosm ([mycoCosm.jgi.doe.gov](http://mycoCosm.jgi.doe.gov)). Genome completeness with single-copy orthologues was calculated using BUSCO v3.0.2, with odb9 using default parameters<sup>85</sup>. Secretomes were predicted as described in Pellegrin et al. We compared the different categories of secreted proteins (e.g., CAZymes, lipases, and proteases) according to their lifestyle. Note that the comparisons of genomes are based on analyses of haploid genomes. *P* values were estimated from nonparametric multiple comparisons based on the generalized Campbell and Skillings' procedure with the function *gao\_cs* in R package *nparcomp*<sup>86</sup>. Distributions with the mean of the total count and statistically significant ecological groups (lifestyles; *P* < 0.05) were determined. Genomic features (e.g., genome size and TE coverage) were statistically compared between fungal lifestyles based on the GLS procedure with the Brownian motion model for random evolution, using R package *nlme* with the function *gls*<sup>87</sup>. Output files generated were combined and visualized with a series of custom R scripts, Proteomic Information Navigated Genomic Outlook (PRINGO), incorporating R graphic packages *ggplot2*, *ggtree*, *egg*, and *ggpubr*<sup>88–91</sup>. The coverage of TEs in genomes was calculated and visualized, using a custom pipeline named Transposon Identification Nominative Genome Overview (TINGO)<sup>27</sup>.

We determined the proportion of variances in genomic features explained by fungal ecological groups, and phylogenetic distances of 61 and 51 species from Basidiomycota and Ascomycota, respectively. Phylogenetic trees were converted into evolutionary distances of species with R package *ape*<sup>92</sup>. Genomic features (i.e., TE coverage, genome size, and predicted secreted proteins) were converted into numerical distances for PERMANOVA according to the workflow using R package *vegan*<sup>93,94</sup>. Significantly different variables were examined (*P* value < 0.05; PERMANOVA; variables ~ phylogeny + ecology). Significant differences among the fungal ecological groups were tested using pair-wise PERMANOVA with R package RVAideMemoire<sup>95</sup>.

Since the 5' and 3' LTR of LTR retrotransposons are identical upon insertion, we estimated their time since insertion using the number of substitutions that occur between the two LTRs as described in Castanera et al.<sup>96</sup>. The list of softwares and R packages is provided in Supplementary Dataset 15.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Genome assemblies and annotations for the organisms used in this study are available via the JGI fungal genome portal MycoCosm ([mycoCosm.jgi.doe.gov](http://mycoCosm.jgi.doe.gov)). In addition, the newly sequenced genome assemblies and annotations have been deposited to GenBank (see Supplementary Data 1 for accession codes/BioProjects). The complete transcriptome datasets are available at the GEO at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/geo/>). The accession codes for accessing the data deposited at GEO database are provided in the "Methods/Phylostratigraphy" section and in the caption of Supplementary Fig. 6. All other data supporting the findings of this study are available within the article and its Supplementary information files or are available from the corresponding authors upon request. Source data are provided with this paper.

## Code availability

PRINGO R scripts are deposited at GitHub: <https://github.com/ShingoMiyachi/PRINGO>. The program COMPARE is also available at Github: <https://github.com/laszlognag/COMPARE>.

Received: 22 January 2020; Accepted: 16 September 2020;

Published online: 12 October 2020

## References

1. Van der, Heijden, M. G. A., Martin, F. M., Selosse, M.-A. & Sanders, I. R. Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol.* **205**, 1406–1423 (2015).
2. Genre, A., Lanfranco, L., Perotto, S. & Bonfante, P. Unique and common traits in mycorrhizal symbioses. *Nat. Rev. Microbiol.* <https://doi.org/10.1038/s41579-020-0402-3> (2020).
3. Martin, F., Kohler, A., Murat, C., Veneault-Fourrey, C. & Hibbett, D. S. Unearthing the roots of ectomycorrhizal symbioses. *Nat. Rev. Microbiol.* **14**, 760–773 (2016).
4. Strullu-Derrien, C., Selosse, M.-A., Kenrick, P. & Martin, F. M. The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New Phytol.* **220**, 1012–1030 (2018).
5. Brundrett, M. C. & Tedersoo, L. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytol.* **220**, 1108–1115 (2018).
6. Matheny, P. B. et al. Out of the Palaeotropics? Historical biogeography and diversification of the cosmopolitan ectomycorrhizal mushroom family Inocybaceae. *J. Biogeography* **36**, 577–592 (2009).
7. Tedersoo, L. & Smith, M. E. Ectomycorrhizal fungal lineages: detection of four new groups and notes on consistent recognition of ectomycorrhizal taxa in high-throughput sequencing studies. *Ecol. Stud.* **230**, 125–142 (2017).
8. Lutzoni, F. et al. Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat. Commun.* **9**, 5451 (2018).
9. Chang, Y. et al. Phylogenomics of Endogonaceae and evolution of mycorrhizas within Mucoromycota. *New Phytol.* **222**, 511–525 (2019).
10. Martin, F. et al. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**, 88–92 (2008).
11. Martin, F. et al. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**, 1033–1038 (2010).
12. Kohler, A. et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat. Genet.* **47**, 410–415 (2015).
13. Peter, M. et al. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nat. Commun.* **7**, 12662 (2016).
14. Hess, J. I. et al. Rapid divergence of genome architectures following the origin of an ectomycorrhizal symbiosis in the genus *Amanita*. *Mol. Biol. Evol.* **35**, 2786–2804 (2018).
15. Martino, E. et al. Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. *New Phytol.* **217**, 1213–1229 (2018).
16. Murat, C. et al. Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle. *Nat. Ecol. Evol.* **2**, 1956–1965 (2018).
17. Pellitier, P. T. & Zak, D. R. Ectomycorrhizal fungi and the enzymatic liberation of nitrogen from soil organic matter: why evolutionary history matters. *New Phytol.* **217**, 68–73 (2018).
18. Frey, S. D. Mycorrhizal fungi as mediators of soil organic matter dynamics. *Annu. Rev. Ecol. Syst.* **50**, 237–259 (2019).
19. Clemmensen, K. E. et al. Roots and associated fungi drive long-term carbon sequestration in boreal forest. *Science* **339**, 1615–1618 (2013).
20. Lindahl, B. D. et al. Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. *New Phytol.* **173**, 611–620 (2007).
21. Akroume, E. et al. First evidences that the ectomycorrhizal fungus *Paxillus involutus* mobilizes nitrogen and carbon from saprotrophic fungus necromass. *Environ. Microbiol.* **21**, 197–208 (2019).
22. Maillard, F., Schilling, J., Andrews, E., Schreiner, K. M. & Kennedy, P. Functional convergence in the decomposition of fungal necromass in soil and wood. *FEMS Microbiol. Ecol.* **96**, fuz209 (2020).
23. Koide, R. T., Sharda, J. N., Herr, J. R. & Malcolm, G. M. Ectomycorrhizal fungi and the biotrophy–saprotrophy continuum. *New Phytol.* **178**, 230–233 (2008).
24. Sipos, G. et al. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungus *Armillaria*. *Nat. Ecol. Evol.* **1**, 1931–1941 (2017).
25. Tedersoo, L. & Bahram, M. Mycorrhizal types differ in ecophysiology and alter plant nutrition and soil processes. *Biol. Rev.* **94**, 1857–1880 (2019).
26. Floudas, D. et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**, 1715–1719 (2012).

27. Morin, E. et al. Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. *New Phytol.* **222**, 1584–1598 (2019).
28. Varga, T. et al. Megaphylogeny resolves global patterns of mushroom evolution. *Nat. Ecol. Evol.* **3**, 668–678 (2019).
29. Zak, D. R. et al. Exploring the role of ectomycorrhizal fungi in soil carbon dynamics. *New Phytol.* **223**, 33–39 (2019).
30. Nagy, N. G. et al. Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities. *Mol. Biol. Evol.* **33**, 959–970 (2016).
31. Churchland, C. & Grayston, S. J. Specificity of plant-microbe interactions in the tree mycorrhizosphere biome and consequences for soil C cycling. *Front. Microbiol.* **5**, 261 (2014).
32. Shah, F. et al. Ectomycorrhizal fungi decompose soil organic matter using oxidative mechanisms adapted from saprotrophic ancestors. *New Phytol.* **209**, 1705–1719 (2016).
33. Krizsán, K. et al. Transcriptomic atlas of mushroom development reveals conserved genes behind complex multicellularity in fungi. *Proc. Natl Acad. Sci. USA* **116**, 7409–7418 (2019).
34. Tedersoo, L., May, T. W. & Smith, M. E. Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza* **20**, 217–263 (2010).
35. Tedersoo, L. & Brundrett, M. Evolution of ectomycorrhizal symbiosis in plants. *Ecol. Stud.* **230**, 407–467 (2017).
36. Bödeker, I. T. M. et al. Ectomycorrhizal *Cortinarius* species participate in enzymatic oxidation of humus in northern forest ecosystems. *New Phytol.* **203**, 245–256 (2014).
37. Kusuda, M., Ueda, M., Miyatake, K. & Terashita, T. Characterization of the carboxylase productions of an ectomycorrhizal fungus, *Tricholoma matsutake*. *Mycoscience* **49**, 291–297 (2008).
38. Rineau, F. et al. Carbon availability triggers the decomposition of plant litter and assimilation of nitrogen by an ectomycorrhizal fungus. *ISME J.* **7**, 2010–2022 (2013).
39. Lindahl, B. D. & Tunlid, A. Ectomycorrhizal fungi—potential organic matter decomposers, yet not saprotrophs. *New Phytol.* **205**, 1443–1447 (2015).
40. Op De Beeck, M., Troein, C., Peterson, C., Persson, P. & Tunlid, A. Fenton reaction facilitates organic nitrogen acquisition by an ectomycorrhizal fungus. *New Phytol.* **218**, 335–343 (2018).
41. Nicolás, C. et al. The soil organic matter decomposition mechanisms in ectomycorrhizal fungi are tuned for liberating soil organic nitrogen. *ISME J.* **13**, 977–988 (2019).
42. Zhang, F. et al. The ectomycorrhizal basidiomycete *Laccaria bicolor* releases a secreted  $\beta$ -1,4 endoglucanase that plays a key role in symbiosis development. *New Phytol.* **220**, 1309–1321 (2018).
43. Plett, J. M. et al. Effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the *Populus* JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proc. Natl Acad. Sci. USA* **111**, 8299–8304 (2014).
44. Kang, H. et al. The small secreted effector protein MiSSP7.6 of *Laccaria bicolor* is required for the establishment of ectomycorrhizal symbiosis. *Environ. Microbiol.* **22**, 1435–1446 (2020).
45. Plett, J. M. et al. Mycorrhizal effector PaMiSSP10b alters polyamine biosynthesis in *Eucalyptus* root cells and promotes root colonization. *New Phytol.* (2020, in the press).
46. Almási et al. Comparative genomics reveals unique wood-decay strategies and fruiting body development in the Schizophyllaceae. *New Phytol.* **224**, 902–915 (2019).
47. Schurko, A. M., Neiman, M. & Logsdon, J. M. Signs of sex: what we know and how we know it. *Trends Ecol. Evol.* **24**, 208–217 (2009).
48. Frantzeskakis, L. et al. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC Genomics* **19**, 27 (2018).
49. Pellegrin, C., Morin, E., Martin, F. M. & Veneault-Fourrey, C. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Front. Microbiol.* **6**, 1278 (2015).
50. Pellegrin, C. et al. *Laccaria bicolor* MiSSP8 is a small-secreted protein decisive for the establishment of the ectomycorrhizal symbiosis. *Environ. Microbiol.* **21**, 3765–3779 (2019).
51. Liao, H. L., Chen, Y. & Vilgalys, R. Metatranscriptomic study of common and host-specific patterns of gene expression between pines and their symbiotic ectomycorrhizal fungi in the genus *Suillus*. *PLOS Genet.* **14**, e1007742 (2018).
52. Soltis, D. E. et al. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl Acad. Sci. USA* **92**, 2647–2651 (1995).
53. Werner, G. D. A., Cornwell, W. K., Sprent, J. I., Kattge, J. & Kiers, E. T. A single evolutionary innovation drives the deep evolution of symbiotic N<sub>2</sub>-fixation in angiosperms. *Nat. Commun.* **5**, 4087 (2014).
54. Nagy, L. G. et al. Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat. Commun.* **5**, 4471 (2014).
55. Nagy, L. G., Kovács, G. M. & Krizsán, K. Complex multicellularity in fungi: evolutionary convergence, single origin, or both? *Biol. Rev.* **93**, 1778–1794 (2018).
56. Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
57. Nguyen, N. H. et al. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).
58. Gnerre, S. et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl Acad. Sci. USA* **108**, 1513–1518 (2011).
59. English, A. C. et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE* **7**, e47768 (2012).
60. Chin, C.-S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
61. Koren, S. et al. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693–700 (2012).
62. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
63. Grigoriev, I. V. et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**, D699–D704 (2014).
64. Kuo, A., Bushnell, B. & Grigoriev, I. V. in *Ecological Genomics of Fungi* (ed Martin, F.), *Advances In Botanical Research*, 1–52 (Elsevier Academic Press, Cambridge, United Kingdom, 2014).
65. Darling, A. E. et al. The design, implementation, and evaluation of mpiBLAST. *ClusterWorld Conference and Expo and the 4<sup>th</sup> International Conference on Linux Clusters* (The HPC Revolution, CA, USA, 2003).
66. Miele, V. et al. High-quality sequence clustering guided by network topology and multiple alignment likelihood. *Bioinformatics* **28**, 1078–1085 (2012).
67. Löytynoja, A. & Goldman, N. Phylogeny-aware Gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632–1635 (2008).
68. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
69. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
70. Löytynoja, A. In *Methods in molecular biology*, Vol. 1079, 155–170 (Human Press, Clifton, N.J., 2014).
71. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
72. Wattam, A. R. et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42**, D581–D591 (2014).
73. Katoh, K. & Standley, D. M. MAFFT: multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
74. Cruickshank, R. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
75. Stamatakis, A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688e2690 (2006).
76. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
77. Hibbett, D. S., Grimaldi, D. & Donoghue, M. J. Fossil mushrooms from Miocene and Cretaceous ambers and the evolution of Homobasidiomycetes. *Am. J. Bot.* **84**, 981–991 (1997).
78. Lepage, B. A., Currah, R. S., Stockey, R. A. & Rothwell, G. W. Fossil ectomycorrhizae from the Middle Eocene. *Am. J. Bot.* **84**, 410–412 (1997).
79. Taylor, T. N., Hass, H., Kerp, H., Krings, M. & Hanlin, R. T. Perithecial ascomycetes from the 400 million year old Rhynie chert: an example of ancestral polymorphism. *Mycologia* **97**, 269–285 (2005).
80. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
81. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
82. Wu, Y.-C., Rasmussen, M. D., Bansal, M. S. & Kellis, M. TreeFix: statistically informed gene tree error correction using species trees. *Syst. Biol.* **62**, 110–120 (2013).
83. Mathé, C., Fawal, N., Roux, C. & Dunand, C. In silico definition of new ligninolytic peroxidase sub-classes in fungi and putative relation to fungal life style. *Sci. Rep.* **9**, 20373 (2019).
84. Domazet-Lošo, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
85. Sima, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. Genome analysis BUSCO: assessing genome assembly and annotation

- completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
86. Konietzschke, F. Simultane Konfidenzintervalle fuer nichtparametrische relative Kontrasteffekte. PhD thesis, University of Goettingen (2009).
  87. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. *R package version 3.1-137* <https://CRAN.R-project.org/package=nlme> (2018).
  88. Auguie, B. egg: <https://cran.r-project.org/web/packages/egg/index.html> (2017).
  89. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
  90. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2009).
  91. Kassambara, A. ggpubr R package: ggplot2-based publication ready plots. *R package version 0.2.5.999* <https://rpkgs.datanovia.com/ggpubr/> (2020).
  92. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2018).
  93. Mensy, F. Detecting the effect of biological categories on genome composition. <https://github.com/fantin-mesy/Effect-Of-Biological-Categories-On-Genomes-Composition> (2020).
  94. Oksanen, J. et al. Vegan: community ecology package. *R package version 2.5-6* <https://CRAN.R-project.org/package=vegan> (2019).
  95. Hervé, M. RVAideMemoire: testing and plotting procedures for biostatistics. *R package version 0.9-75* <https://CRAN.R-project.org/package=RVAideMemoire> (2020).
  96. Castanera, R. et al. Transposable elements versus the fungal genome: Impact on whole-genome architecture and transcriptional profiles. *PLoS Genet.* **12**, e1006108 (2016).
  97. Ploner, A. Heatplus: Heatmaps with row and/or column covariates and colored clusters. *R package version 2.34.0*, <https://github.com/alexploner/Heatplus> (2020).

## Acknowledgements

The project was funded by the U.S. Department of Energy (DOE) Joint Genome Institute, a DOE Office of Science User Facility, and supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-05CH11231 within the framework of the Mycorrhizal Genomics Initiative (CSP # 305), Metatranscriptomics of Forest Soil Ecosystems project (CSP # 570), and the 1000 Fungal Genome projects (CSP # 662 and 1974). This research was also supported by the Laboratory of Excellence ARBRE (ANR-11-LABX-0002-01), the Region Lorraine, the European Regional Development Fund, and the Plant–Microbe Interfaces Scientific Focus Area in the Genomic Science Program, the Office of Biological and Environmental Research in the US DOE Office of Science. L.G.N. is funded by the Momentum Program of the Hungarian Academy of Sciences (Grant No. LP2019/13-2019) and the National Research, Development and Innovation office (Contract No: GINOP-2.3.2-15-2016-00052). G.S. received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 714774 and the grant GINOP-2.3.2.-15-2016-00057. We also thank the many colleagues whom provided fungal strains and material, including Jan Colpaert, Pierre-Emmanuel Courty, Joanna F. Dames, Patricia Jargeat, Björn Lindahl, Asuncion Morte, Jose-Eduardo Marque-Galvez, Andy Taylor, and Yukari Kuga.

## Author contributions

F.M.M. conceived and coordinates the Mycorrhizal Genomics Initiative. F.M.M., L.G.N., D.H., and I.V.G. designed the project. I.V.G. coordinated genome sequencing and annotation at JGI. A.Kuo, B.A., K.W.B., A.Carver, A.C.L., C.C., A.L., K.L., M.N., R.A.O., R.D.H., R.R., J.P., and M.Y. performed genome and transcriptome sequencing, assembly and annotation at J.G.I. E.D., V.L., and B.H. performed CAZyme annotations. S.M., F.M., F.M.M., and M.B. analyzed CAZyme distribution. E.K., A.Kohler, L.G.N., and F.M.M. performed analyses of PCWDE evolution, phylostratigraphic studies and analyses of convergence. C.D. annotated peroxidases. S.M. performed comparative genome analyses with the help of E.M. and C.M., and developed scripts for the visual display of data. M.S.G. and D.H. constructed dated trees. M.F.P., M.P., S.P., and L.M.V. performed mycorrhizal experiments for RNA-Seq. A.Kohler and E.M. analyzed transcriptome data. Z.K. and G.S. performed clustering analyses using HiFix and Silix. A.Y., B.S., C.M., D.C., F.M., L.F., G.B., J.B.S., J.M., J.W.S., J.X., L.T., L.Z., M.B., M.G., M.P., M.S., M.Y., N.C., P.B., P.W., P.W.C., R.V., S.P., T.B. and Y.S. provided biological material, DNA and RNA for sequencing or genomes. F.M.M., and S.M. wrote the manuscript with the help of D.H. and L.G.N.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-18795-w>.

**Correspondence** and requests for materials should be addressed to F.M.M.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.







**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Shingo Miyauchi<sup>1,2,5</sup>, Enikő Kiss<sup>2,25</sup>, Alan Kuo<sup>3,25</sup>, Elodie Drula<sup>4</sup>, Annegret Kohler<sup>1</sup>, Marisol Sánchez-García<sup>5</sup>, Emmanuelle Morin<sup>1</sup>, Bill Andreopoulos<sup>3</sup>, Kerrie W. Barry<sup>3</sup>, Gregory Bonito<sup>6</sup>, Marc Buée<sup>1</sup>, Akiko Carver<sup>3</sup>, Cindy Chen<sup>3</sup>, Nicolas Cichocki<sup>1</sup>, Alicia Clum<sup>3</sup>, David Culley<sup>7</sup>, Pedro W. Crous<sup>8</sup>, Laure Fauchery<sup>1</sup>, Mariangela Girlanda<sup>9</sup>, Richard D. Hayes<sup>3</sup>, Zsófia Kéri<sup>2</sup>, Kurt LaButti<sup>3</sup>, Anna Lipzen<sup>3</sup>, Vincent Lombard<sup>4</sup>, Jon Magnuson<sup>7</sup>, François Maillard<sup>1</sup>, Claude Murat<sup>1</sup>, Matt Nolan<sup>3</sup>, Robin A. Ohm<sup>3</sup>, Jasmyn Pangilinan<sup>3</sup>, Maíra de Freitas Pereira<sup>1</sup>, Silvia Perotto<sup>9</sup>, Martina Peter<sup>10</sup>, Stephanie Pfister<sup>10</sup>, Robert Riley<sup>3</sup>, Yaron Sitrit<sup>11</sup>, J. Benjamin Stielow<sup>8</sup>, Gergely Szöllősi<sup>2</sup>, Lucia Žifčáková<sup>12</sup>, Martina Štursová<sup>12</sup>, Joseph W. Spatafora<sup>13</sup>, Leho Tedersoo<sup>14</sup>, Lu-Min Vaario<sup>15</sup>, Akiyoshi Yamada<sup>16</sup>, Mi Yan<sup>3</sup>, Pengfei Wang<sup>17</sup>, Jianping Xu<sup>18</sup>, Tom Bruns<sup>19</sup>,



Petr Baldrian <sup>12</sup>, Rytas Vilgalys <sup>20</sup>, Christophe Dunand<sup>21</sup>, Bernard Henrissat <sup>4,22,23</sup>,  
Igor V. Grigoriev <sup>3,19,26</sup>, David Hibbett<sup>5,26</sup>, László G. Nagy<sup>2,26</sup> & Francis M. Martin<sup>1,24,26</sup>✉

<sup>1</sup>Université de Lorraine, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, UMR Interactions Arbres/Microorganismes, Centre INRAE Grand Est-Nancy, 54280 Champenoux, France. <sup>2</sup>Synthetic and Systems Biology Unit, Biological Research Centre, 6726 Szeged, Hungary. <sup>3</sup>US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>4</sup>INRAE, USC1408 Architecture et Fonction des Macromolécules Biologiques, 13009 Marseille, France. <sup>5</sup>Biology Department, Clark University, Lasry Center for Bioscience, 950 Main Street, Worcester, MA 01610, USA. <sup>6</sup>Plant Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA. <sup>7</sup>Chemical & Biological Processes Development Group, Pacific Northwest National Laboratory, Richland, WA, USA. <sup>8</sup>Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, Netherlands. <sup>9</sup>Department of Life Sciences and Systems Biology, University of Torino, Viale Mattioli 25, 10125 Torino, Italy. <sup>10</sup>Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zuercherstrasse 111, 8903 Birmensdorf, Switzerland. <sup>11</sup>The Jacob Blaustein Institutes for Desert Research, Bergman Campus, Ben-Gurion University of The Negev, Beer-Sheva, Israel. <sup>12</sup>Laboratory of Environmental Microbiology, Institute of Microbiology of the Czech Academy of Sciences, Videnska 1083, 14220 Praha 4, Czech Republic. <sup>13</sup>Department Botany & Plant Pathology, Oregon State University, Corvallis, OR, USA. <sup>14</sup>Natural History Museum, University of Tartu, 14a Ravila, 50411 Tartu, Estonia. <sup>15</sup>Department of Forest Sciences, University of Helsinki, Helsinki, Finland. <sup>16</sup>Institute of Mountain Science, Faculty of Agriculture, Shinshu University, Minami-minowa, Kami-ina, Nagano 399-4598, Japan. <sup>17</sup>Department of Key Laboratory, The 2nd Affiliated Hospital of Kunming Medical University, 374 Dian Mian Road, Kunming 650101 Yunnan, China. <sup>18</sup>Department of Biology, McMaster University, 1280 Main St. West, Hamilton, ON L8S 4K1, Canada. <sup>19</sup>Department of Plant and Microbial Biology, University of California - Berkeley, Berkeley, CA, USA. <sup>20</sup>Department of Biology, Duke University, Durham, NC 27708, USA. <sup>21</sup>Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Toulouse, France. <sup>22</sup>Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, Aix-Marseille Univ., 13009 Marseille, France. <sup>23</sup>Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. <sup>24</sup>Beijing Advanced Innovation Centre for Tree Breeding by Molecular Design (BAIC-TBMD), Institute of Microbiology, Beijing Forestry University, Tsinghua East Road Haidian District, Beijing, China. <sup>25</sup>These authors contributed equally: Shingo Miyauchi, Enikő Kiss, Alan Kuo. <sup>26</sup>These authors jointly supervised this work: Igor V. Grigoriev, David Hibbett, László G. Nagy, Francis M. Martin. ✉email: [francis.martin@inrae.fr](mailto:francis.martin@inrae.fr)