



















# Evolutionary innovations through gain and loss of genes in the ectomycorrhizal Boletales

Gang Wu<sup>1,2,3\*</sup> , Shingo Miyauchi<sup>2\*</sup> , Emmanuelle Morin<sup>2</sup> , Alan Kuo<sup>4</sup> , Elodie Drula<sup>5</sup> , Torda Varga<sup>6</sup>,  
Annegret Kohler<sup>2</sup> , Bang Feng<sup>1,3</sup> , Yang Cao<sup>7</sup> , Anna Lipzen<sup>4</sup>, Christopher Daum<sup>4</sup>, Hope Hundley<sup>4</sup>,  
Jasmyn Pangilinan<sup>4</sup>, Jenifer Johnson<sup>4</sup>, Kerrie Barry<sup>4</sup> , Kurt LaButti<sup>4</sup> , Vivian Ng<sup>4</sup>, Steven Ahrendt<sup>4</sup> ,  
Byoungnam Min<sup>4,8</sup>, In-Geol Choi<sup>8</sup>, Hongjae Park<sup>9</sup>, Jonathan M. Plett<sup>10</sup> , Jon Magnuson<sup>11</sup> ,  
Joseph W. Spatafora<sup>12</sup>, László G. Nagy<sup>6,13</sup> , Bernard Henrissat<sup>5,14,15</sup> , Igor V. Grigoriev<sup>4,16</sup> ,  
Zhu-Liang Yang<sup>1,3</sup>, Jianping Xu<sup>17</sup> , and Francis M. Martin<sup>2,18</sup> 

<sup>1</sup>CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China; <sup>2</sup>Université de Lorraine, INRAE, UMR Interactions Arbres/Microorganismes, Centre INRAE Grand Est-Nancy, Champenoux 54 280, France; <sup>3</sup>Yunnan Key Laboratory for Fungal Diversity and Green Development, Kunming, Yunnan 650201, China; <sup>4</sup>Lawrence Berkeley National Laboratory, US Department of Energy (DOE) Joint Genome Institute (JGI), Berkeley, CA 94720, USA; <sup>5</sup>Architecture et Fonction des Macromolécules Biologiques (USC1408), INRAE, Marseille 13009, France; <sup>6</sup>Synthetic and Systems Biology Unit, Biological Research Centre, Szeged 6726, Hungary; <sup>7</sup>Yunnan Institute of Tropic Crops, Jinghong, Yunnan 666100, China; <sup>8</sup>Department of Biotechnology, College of Life Sciences and Biotechnology, Korea University, 02841, Seoul, Korea; <sup>9</sup>Department of Aquatic Microbial Ecology, Institute of Hydrobiology, Biology Centre of the Czech Academy of Sciences, 370 05, České Budějovice, Czech Republic; <sup>10</sup>Hawkesbury Institute for the Environment, Western Sydney University, Richmond, NSW 2753, Australia; <sup>11</sup>Chemical and Biological Processes Development Group, Pacific Northwest National Laboratory, Richland, WA 99354, USA; <sup>12</sup>Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA; <sup>13</sup>Department of Plant Anatomy, Institute of Biology, Eötvös Loránd University, Budapest 1117, Hungary; <sup>14</sup>Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, Marseille 13009, France; <sup>15</sup>Department of Biological Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia; <sup>16</sup>Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA; <sup>17</sup>Department of Biology, McMaster University, Hamilton, ON L8S 4K1, Canada; <sup>18</sup>Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing 100083, China

## Summary

Authors for correspondence:  
Gang Wu  
Email: wugang@mail.kib.ac.cn

Francis M. Martin  
Email: francis.martin@inrae.fr

Received: 6 September 2021  
Accepted: 4 November 2021

New Phytologist (2021)  
doi: 10.1111/nph.17858

**Key words:** Boletales, brown-rot fungi, CAZymes, comparative genomics, ectomycorrhizal fungi, trait evolution.

- We aimed to identify genomic traits of transitions to ectomycorrhizal ecology within the Boletales by comparing the genomes of 21 symbiotrophic species with their saprotrophic brown-rot relatives.
- Gene duplication rate is constant along the backbone of Boletales phylogeny with large loss events in several lineages, while gene family expansion sharply increased in the late Miocene, mostly in the Boletaceae.
- Ectomycorrhizal Boletales have a reduced set of plant cell-wall-degrading enzymes (PCWDEs) compared with their brown-rot relatives. However, the various lineages retain distinct sets of PCWDEs, suggesting that, over their evolutionary history, symbiotic Boletales have become functionally diverse. A smaller PCWDE repertoire was found in Sclerodermatineae. The gene repertoire of several lignocellulose oxidoreductases (e.g. laccases) is similar in brown-rot and ectomycorrhizal species, suggesting that symbiotic Boletales are capable of mild lignocellulose decomposition. Transposable element (TE) proliferation contributed to the higher evolutionary rate of genes encoding effector-like small secreted proteins, proteases, and lipases. On the other hand, we showed that the loss of secreted CAZymes was not related to TE activity but to DNA decay.
- This study provides novel insights on our understanding of the mechanisms influencing the evolutionary diversification of symbiotic boletes.

## Introduction

In forest soils, wood decayers, litter/soil decomposers, and ectomycorrhizal fungi form entangled networks of mycelia that

compete for nutrient resources. A major part of soil carbon (C) and nitrogen (N) organic compounds is stored in forest biomes as wood and soil organic matter (SOM). Decomposition of these resources is thought to be mainly driven by fungal wood decayers and soil/litter saprotrophs (Baldrian *et al.*, 2012). These

\*These authors contributed equally as first authors to this work.

functional guilds of fungi play a major role in C recycling and sequestration, as well as of other nutrients, from plant litter and dead bacterial, fungal, and animal materials. In addition, ectomycorrhizal symbionts impact SOM decomposition by competing with saprotrophs and regulate the input of plant-related C compounds into soil microbial communities. Deciphering how N acquisition and SOM decomposition potentials differ in saprotrophic and symbiotrophic fungi can provide insight into the underlying mechanisms driving fungal ecological processes and ecosystem functioning (Peay *et al.*, 2016). Describing how trait variation and the gene copy number of key proteins involved in SOM decomposition and mycorrhizal symbiosis (e.g. signaling pathways) vary across functional guilds is shedding light on potential evolutionary trajectories of life-history traits in forest fungi (Lebreton *et al.*, 2021).

In the most comprehensive phylogenetic analysis on evolution of ectomycorrhizal fungi in Agaricomycetes carried out to date, 36 origins of ectomycorrhizal lineages were found (Sánchez-García *et al.*, 2020). Large-scale studies of mycorrhizal genomes have shown that ancestors of ectomycorrhizal fungi were genetically and ecologically diverse, being white rots, brown rots, or soil/litter saprotrophs (Kohler *et al.*, 2015; Miyauchi *et al.*, 2020; Lebreton *et al.*, 2021). Though tens to hundreds of millions of years separate the independent evolution of ectomycorrhizal symbioses in Endogonales, Ascomycota, and Basidiomycota, they share remarkable phenotypic and metabolic similarities. The polyphyletic evolution of the ectomycorrhizal lifestyle is marked by the convergent, near complete loss of core cellulose and lignin-acting carbohydrate-active enzymes (CAZymes), such as ligninolytic peroxidases (class II PODs), cellobiohydrolases of glycosyl hydrolase (GH) families 6 and 7, and cellulose-binding modules found in the ancestral decomposition apparatus of their saprotrophic ancestors. Nevertheless, many of the ectomycorrhizal symbionts investigated have kept a unique array of plant cell-wall-degrading enzymes (PCWDEs), including endoglucanases (GH5), pectinases (GH28), and oxidoreductases/laccases (AA1, AA9), thus suggesting that they possess diverse abilities to scavenge plant and microbial detritus from the soil/litter. Finally, transposable elements (TEs) and mycorrhiza-induced small secreted proteins (MiSSPs) tend to be enriched in ectomycorrhizal genomes (Pellegrin *et al.*, 2015; Miyauchi *et al.*, 2020; Lebreton *et al.*, 2021).

With the increasing number of fungal genomes available, it is becoming more feasible to trace the evolutionary events defining the origin of ectomycorrhizal symbiosis within a specific fungal lineage. Functional traits – such as gene copy numbers for PCWDEs, N and phosphate acquisition pathways, MiSSPs, and TEs – within a fungal family across functional guilds (e.g. saprotrophs vs symbiotrophs) have been investigated within the Amanitaceae, Endogonaceae, Russulaceae, Suillaceae, and Tubercaceae (Hess *et al.*, 2018; Murat *et al.*, 2018; Chang *et al.*, 2019; Lofgren *et al.*, 2021; Looney *et al.*, 2021). These studies have shown that the aforementioned hallmarks of the ectomycorrhizal ecology are recapitulated at the family level, although idiosyncrasies have been identified in each of these fungal families, e.g. a significant enrichment of terpene synthetase and nonribosomal peptide

synthetase (NRPS)-like secondary metabolism clusters in ectomycorrhizal *Suillus* species.

The order Boletales is among the most species-rich orders in the Agaricomycetes and includes five suborders (i.e. Boletineae, Sclerodermatineae, Suillineae, Coniophorineae, and Tapinellineae) and 16 families. Among them, nearly two-thirds of the species are affiliated to the iconic ectomycorrhizal Boletaceae family, as a result of their rapid evolutionary diversification with their coevolving angiosperm hosts (Grubisha *et al.*, 2001; Binder & Hibbett, 2006; Drechsel *et al.*, 2008; Kirk *et al.*, 2008; Dentinger *et al.*, 2010; Sato & Toju, 2019). Species in the suborders Boletineae, Sclerodermatineae, and Suillineae establish symbiotic associations with diverse host plants (Newman & Reddell, 1987; Bougher, 1995; Henkel *et al.*, 2002; den Bakker *et al.*, 2004; Sato *et al.*, 2007; Hosen *et al.*, 2013; Nuhn *et al.*, 2013; Wu *et al.*, 2014, 2016). On the other hand, species in the suborders Coniophorineae and Tapinellineae are brown-rot fungi that grow on dead wood. No white-rot species are known in Boletales, and the most recent common ancestor (MRCA) of Boletales was presumably a brown-rot fungus (Ruiz-Dueñas *et al.*, 2020). However, the mode of nutrition in several Boletaceae genera, such as *Phlebopus* and *Boletinellus*, is still elusive (Binder & Hibbett, 2006; Tedersoo & Smith, 2013; Sato & Toju, 2019). For example, both the  $^{13}\text{C}/^{15}\text{N}$  isotopic signature of the basidiocarps from *Phlebopus portentosus* (Berk. & Broome) Boedijn and the ability of its soil mycelium to establish ectomycorrhiza-like structures with *Pinus kesiya* support a symbiotrophic mode of nutrition (Pham *et al.*, 2012; Kumla *et al.*, 2016). On the other hand, the free-living mycelium of *P. portentosus* is able to produce basidiocarps in the absence of any host tree (e.g. in pots) (Ji *et al.*, 2011), suggesting a substantial saprotrophic ability to sustain the massive demand for C compounds required by fruiting body construction. These contrasted ecological features suggest that these Boletaceae have not fully transitioned to the symbiotrophic lifestyle. Symbiotic Boletales bear several ecologically relevant attributes that warrant study in a genomic context, such as their abundance in temperate, subtropical, and tropical ecosystems (Heinemann, 1951; Corner, 1972; Bessette *et al.*, 2000; Muñoz, 2005; Zang, 2006), their late-stage fruiting in forest successions (Ortega-Martínez *et al.*, 2011), the production of unique volatile organic compounds (Rapier *et al.*, 1997), and an accelerated evolutionary rate of speciation, morphological transition, and host expansion (Wu *et al.*, 2016; Sato & Toju, 2019).

To address a number of hypotheses on the evolution of ectomycorrhizal symbioses from brown-rot ancestors and trophic ecology in the various Boletales families, we sequenced, annotated, and compared the genomes of 28 Boletales species, including seven newly sequenced genomes from species found in subtropical and tropical forests. This study had four main objectives. First, to generate a robust phylogenomic framework for the Boletales order. Second, to determine whether the genomes of symbiotrophic Boletales show signatures of the ectomycorrhizal ecology, including large genome size due to TE expansions, reduction in the diversity of PCWDEs, and diversification of small secreted protein (SSPs) that might function in the mycorrhizal symbiosis. Third, to infer the contribution of TE

expansion to gene innovation/decay of PCWDEs and SSPs; and fourth, to characterize the gene repertoire and transcript profiling of the Boletineae *P. portentosus*, having a dual saprotrophic/symbiotrophic lifestyle. By comparing genomes of saprotrophic and symbiotic species, we reveal the genetic basis for their contrasted lignocellulose and protein-degrading abilities. We also identify major differences in their repertoire of secondary metabolism enzymes, and we assess the conservation of symbiotic-related traits in this fungal order. Finally, we showed that the genomes of *Phlebopus* species have a symbiotrophic signature.

## Materials and Methods

### Fungal material, genome sequencing, assembly, and annotation

Fungal strains used for genome sequencing are described in Supporting Information Table S1. Genomic DNA was extracted with a modified cetyltrimethylammonium bromide protocol, as described in Kohler *et al.* (2015).

The genome of *Boletus reticuloceps* (M. Zang *et al.*) Q.B. Wang & Y.J. Yao (strain BR01), *Butyriboletus roseoflavus* (Hai B. Li & Hai L. Wei) D. Arora & J.L. Frank (strain LA02), *Lanmaoa asiatica* G. Wu & Zhu L. Yang (strain LA01), and *Chiuia virens* (W.F. Chiu) Y.C. Li & Zhu L. Yang (strain LA06) were sequenced for this study using the Pacific Biosciences (PacBio) Sequel platform at Nextomics Biosciences (Wuhan, China). Library construction, genome sequencing, genome assembly, and gene annotation were carried out using the company standardized pipeline. Detailed information is provided in Methods S1-1. The genome of *Phlebopus* sp. (strain FC\_14), *Leucogyrophana mollusca* (Fr.) Pouzar (strain KUC20120723A-06), and *Hygrophoropsis aurantiaca* (Wulfen) Maire (strain ATCC 28755) were sequenced for this study at the Joint Genome Institute (JGI) using the PacBio Sequel or Illumina HiSeq sequencing platforms.

Additional information on methods used for genome sequencing, genome assembly, and gene prediction is provided in Methods S1-1. Note that we performed a stringent quality control to avoid any spurious contaminating sequences in the genome assemblies (see Methods S1-1).

### Organismal phylogeny

We retrieved protein sequences from the Boletales portal at the JGI MycoCosm database (<https://mycocosm.jgi.doe.gov/boletales/boletales.info.html>) and Nextomics Biosciences (Wuhan, China). A phylogenomic tree was constructed using 28 species of Boletales, two of Atheliales, one of Amylocorticiales, and one of Agaricales. Two Polyporales and allied species were chosen as an outgroup: *Polyporus brumalis* (Pers.) Fr and *Cristinia sonora* Nakasone & Gilb. We used 434 highly conserved protein coding genes that have previously proved useful for higher level phylogenetic analysis of fungi (Beaudet *et al.*, 2018; [https://github.com/1KFG/Phylogenomics\\_HMMs](https://github.com/1KFG/Phylogenomics_HMMs), JGI\_1086 set), for phylogenomic analyses

with the PHYLING pipeline ([https://github.com/stajichlab/PHYling\\_unified](https://github.com/stajichlab/PHYling_unified)) with the default settings. Out of the 434 conserved orthologous markers, 430 were identified in our data set with *hmmsearch* (cutoff:  $1 \times 10^{-10}$ ). The protein sequence homologues identified in each species, for each phylogenetic marker, were aligned with *hmmalign* to the marker profile hidden Markov model. Both programs belong to the HMMER package (v.3.3, Eddy, 2011). The protein alignments were concatenated into a superalignment with 430 partitions defined by each gene marker. To select the best-fit amino acid substitution models for each partition, we used MODELTEST-NG (v.0.1.6; Darriba *et al.*, 2019). A maximum likelihood inference for our phylogenomic data set was achieved with RAXML-NG (v.0.9.0; Kozlov *et al.*, 2019) using a partitioned analysis and 1000 bootstraps replicates.

### Time-calibrated phylogeny

To calibrate the organismal phylogeny, we used the *mcmctree* method implemented in PAML (v.4.8; Yang, 2007) with the independent-rates clock model, a Whelan and Goldman (WAG) substitution model, and approximate likelihood calculation. For each gene we estimated the substitution rate with *codeml* using the corresponding gene alignment and the following parameters: *clock* = 1, *model* = 2, *aaRateFile* = *wag.dat*, *getSE* = 0, and *Small\_Diff* = 1e7. We then set the time unit to 100 million years (Myr) and applied uniform priors on two fossil calibrations with lower and upper hard bounds. The fossil of a suilloid ectomycorrhizal root tip from the middle Eocene (40–60 Ma (million years ago); LePage *et al.*, 1997; Varga *et al.*, 2019) was used to calibrate the node containing the suborder Suillineae, and a quite relaxed secondary calibration point was set to the Boletales MRCA using the estimated stem age of Boletales (mean: 218 Ma; 95% highest posterior density (HPD): 84–279 Ma) from Han *et al.* (2018)). We also constrained the age of the root to be < 500 Ma.

### Genome synteny and rearrangement analysis

For assessing the genome synteny, we used the 10 largest scaffolds of the genome assemblies with the highest contiguity and completeness, mainly from Boletineae species. We excluded the four Boletineae genomes having the highest assembly fragmentation: *Xerocomus badius* ( $\equiv$  *Imleria badia*) (Xarba1), *Paxillus ammoniavirescens* (Paxam1), *Paxillus adelphus* (Paxru1), and *Melanogaster broomeianus* (Melbro1; Fig S2). Then, we identified syntenic blocks using a custom script, Synteny Governance Overview (SynGO), incorporating the R packages DECIPHER and CIRCLIZE (Gu *et al.*, 2014; Wright, 2015; Hage *et al.*, 2021). We measured the mean TE-to-gene distances with statistical support with the R package REGIONER (Gel *et al.*, 2016). Macrosynteny with combined genomic features among the selected species was visualized and evaluated with a custom script, Visually Integrated Numerous Genres of Omics (VINGO), incorporating the R package KARYOPLOTTER (Gel & Serra, 2017; Looney *et al.*, 2021). We assessed and displayed associations between TEs and genes using the R packages VEGAN (Oksanen *et al.*, 2019) and PCATOOLS (Blighe & Lun, 2020).

## Analysis of gene evolution using COMPARE pipeline

An all-vs-all search of whole proteomes of 34 species (Table 1) was performed using MPIBLAST 1.6.0 (Darling *et al.*, 2003) with 50% bidirectional coverage filter and an *e*-value cutoff of  $10^{-5}$ . Proteins were then clustered using the Markov cluster (MCL) algorithm (van Dongen, 2000) with an inflation parameter 2.0. MCL clusters of protein sequences were aligned using MAFFT 7.4.07 (Kato & Standley, 2013) with the *-LiNSI* algorithm and trimmed using TRIMAL 1.4 with parameter *-gt* 0.2. Maximum likelihood gene trees for each cluster and Shimodaira–Hasegawa-like branch support values were inferred using the PTHREADS version of RAxML 8.2.12 (Stamatakis, 2014) under the PROTGAMMAWAG model. Next, we reconciled the rooted gene trees with the species tree using NOTUNG 2.9 (Chen *et al.*, 2000) with an edge-weight threshold of 0.95. We reconstructed the duplication/loss history of all protein clusters across the species tree using the COMPARE pipeline (Nagy *et al.*, 2014, 2016). Duplication and loss rates were computed by dividing the number of inferred duplication and loss events by the length of the respective branch of the species tree. Rates through time plots were created as follows: the time interval of the tree was divided into bins (100 sections) and in each bin the average rate was calculated. Gene families were functionally characterized using InterPro annotations. We performed an InterPro search using INTERPROSCAN-5.36-75.0 (Jones *et al.*, 2014) with the ‘goterms’ argument. Cluster-based Gene Ontology (GO) enrichment analysis was carried out by using the TOPGO v.2.38.1 package (Alexa & Rahnenfuhrer, 2020) with the *weight01* algorithm (Alexa *et al.*, 2006) and Fisher’s exact test. Graphical maps of gene duplication/loss histories were generated using custom scripts in R.

## Results

### Genome features and phylogenetic analysis

The nuclear genome of four symbiotrophic Boletaceae (*B. reticuliceps*, *B. roseoflavus*, *L. asiatica*, and *C. virens*), a dual symbiotrophic/saprotrophic Boletiniellaceae (*Phlebopus* sp. FC\_14), and two saprotrophic Hygrophoropsidaceae (*H. aurantiaca* and *L. mollusca*) were newly sequenced, *de novo* assembled, and annotated (Table S1) for this study. These genomes were then compared with available sequenced genomes of Boletales in Paxillaceae, Sclerodermatineae, Suillineae, Coniophoraceae, and Serpulaceae (Tables 1, S2). In Boletales, the genome size ranges from 30 Mbp for *P. portentosus* to 71 Mbp for *Pisolithus tinctorius* (Fig. 1b; Table 1). The completeness of these genome assemblies varies from 83.1% to 99.7% according to their benchmarking universal single-copy orthologues (BUSCO) scores with <5% missing BUSCO genes (Fig. 1b; Table 1). By combining homologue-based, *ab initio*, and transcriptome-based approaches, 9749 (*P. portentosus*) to 22 701 (*Pi. tinctorius*) protein-coding genes were predicted for these genomes (Fig. 1b; Table 1).

A maximum likelihood phylogenetic analysis, based on the concatenated amino acid sequence alignment of 430 conserved,

orthologous proteins of Boletales and related Atheliales/Amylocorticiales species (Fig. 1a) corroborated and extended previous phylogenetic topologies (Binder *et al.*, 2010; Kohler *et al.*, 2015; Sato & Toju, 2019). We showed that *Hydnomerulius pinastri*, a brown-rot species, is clustered with ectomycorrhizal Boletineae and Boletiniellaceae, a family with a dual symbiotrophic/saprotrophic ecology. The Hygrophoropsidaceae was confirmed as the sister lineage of the main clade comprising Boletineae, Sclerodermatineae, and Suillineae (Fig. 1a). The estimated age of the MRCA of Boletales was 151 Ma (95% HPD: 112–183 Ma) in the Late Jurassic (Fig. 1a) according to our Bayesian molecular clock dating. The inferred ages were similar to previous estimates (146 Ma, Zhao *et al.*, 2017; 142 Ma, Varga *et al.*, 2019).

### Genomes of ectomycorrhizal Boletales have a higher load in transposable elements

Ectomycorrhizal fungi have the largest genomes with a significantly higher TE content compared with their brown-rot relatives (Figs 1b, 2; Tables S3–S5;  $P < 0.05$  for TEs; pairwise permutational multivariate ANOVA (PERMANOVA)). TE content ranges from 3.8% (*Coniophora olivacea*) to 50.3% (*Pa. adelphus*) of the genome assembly (Fig. 1b; Tables 1, S2). The genome size variation was largely explained by the fungal lifestyle (29.3%; Fig. 2;  $P < 0.05$ ; PERMANOVA; Table S4). Among the known TEs, *Gypsy* and *Copia* long terminal repeat (LTR) retrotransposons are widely distributed in Boletales and massively expanded in ectomycorrhizal species with lineage-specific features (Figs 3, S1; Table S6). Notably, there is an unexpectedly high copy number of *Gypsy*, *Copia*, and *EnSpm/CACTA* in *Gyrodon lividus*, an alder-specific symbiont. *Lanmaoa asiatica* genome encodes a very high number of *Copia*, *Gypsy*, *hAT*, and *Mariner* elements, whereas *C. virens* has a high content in *Copia*, *Harbinger*, *Academ*, and *Kolobok* families. In contrast to other brown-rot Boletales, *Serpula himantioidea* and *Serpula lacrymans* genomes encode expanded TE repertoires, enriched in *Gypsy* and *Copia* LTR retrotransposons (Figs 1b, S1; Table S6). The Kimura distance-based copy divergence showed that TE copies proliferated in the recent historical period in most Boletales species, especially in the ectomycorrhizal symbionts (Fig. 3). Furthermore, a various set of TE families accumulated in very recent time in *Bu. roseoflavus*, *C. virens* and *Suillus luteus*, whereas *L. asiatica* and *Paxillus involutus* showed two bursts of TE proliferation (Fig. 3). Although TE invasion, proliferation, and decay took place at a different pace in the various species, it appears that the symbiotic lifestyle is correlated to a higher TE accumulation.

### Genomic synteny and rearrangements

We restricted the genome synteny analysis to the top 10 largest scaffolds, with the highest contiguity and completeness, including >50% of the genome assembly and capturing more than one-third of secreted CAZyme coding genes (Fig. S2; Table S7). As expected, the genomes of the European isolates Prilba and BED4 of *Boletus edulis* share very similar gene repertoires (18 722 and 18 123 genes, Figs 1b, 4). Together with the phylogenetically related



**Table 1** Taxonomic affiliation, brief genomic features and statistics for the 34 included genomes.

Species name	JGI assembly ID/species abbreviation	Ecology	Genome size (nt)	Total genes	Total scaffolds	Scaffold N50	Scaffold L50 (Mbp)	BUSCO (%) <sup>a</sup>
<i>Boletus edulis</i>	Boledp1	EcM	69 185 680	18 123	478	40	0.38	0.855
<i>Boletus edulis</i>	Boled5	EcM	66 525 535	18 722	594	29	0.42	0.895
<i>Boletus reticuloceps</i>	Bolret1	EcM	55 835 099	15 643	127	18	1.11	0.846
<i>Xerocomus badius</i> (≡ <i>Imleria badia</i> )	Xerba1	EcM	38 390 415	14 938	1390	110	0.09	0.891
<i>Lanmaoa asiatica</i>	Lanmao1	EcM	46 638 294	14 118	112	8	2.45	0.870
<i>Butyriboletus roseoflavus</i>	Butyri1	EcM	33 976 981	12 051	67	6	2.7	0.861
<i>Chiuva virens</i>	Chivi1	EcM	48 496 802	13 144	149	10	1.93	0.845
<i>Paxillus ammoniavirescens</i>	Paxam1	EcM	35 762 581	13 473	1194	84	0.09	0.909
<i>Paxillus involutus</i>	Paxin1	EcM	58 301 126	17 968	2681	29	0.38	0.881
<i>Paxillus adelphus</i>	Paxru2	EcM	64 458 135	18 999	1671	165	0.11	0.905
<i>Gyrodon lividus</i>	Gyrli1	EcM	43 048 674	11 779	369	14	1.16	0.903
<i>Melanogaster broomeianus</i>	Melbro1	EcM	49 614 189	14 167	1176	63	0.2	0.819
<i>Hydnomerulius pinastri</i>	Hydpi1	Brown rot	38 278 792	13 270	603	16	0.69	0.905
<i>Phlebobus portentosus</i>	Phlpo1	Dual symbiotrophic/saprotrophic	30 352 192	9467	108	8	1.45	0.907
<i>Phlebobus</i> sp.	PhlFC14_2	Dual symbiotrophic/saprotrophic	37 508 415	9249	354	8	1.57	0.897
<i>Pisolithus tinctorius</i>	Pisti1	EcM	71 007 534	22 701	610	36	0.49	0.894
<i>Pisolithus microcarpus</i>	Pismi1	EcM	53 027 657	21 064	1064	89	0.15	0.863
<i>Scleroderma citrinum</i>	Sclici1	EcM	56 144 862	21 012	938	63	0.24	0.909
<i>Rhizopogon vinicolor</i>	Rhivi1	EcM	36 102 320	14 469	2310	218	0.04	0.905
<i>Rhizopogon vesiculosus</i>	Rhives1	EcM	43 809 644	14 218	6700	388	0.03	0.900
<i>Suillus brevipes</i>	Suibr2	EcM	52 027 859	21 458	1550	84	0.16	0.913
<i>Suillus luteus</i>	Suilu4	EcM	44 486 502	16 588	67	10	1.39	0.903
<i>Hygrophoropsis aurantiaca</i>	Hygaur1	Brown rot	36 975 565	14 285	2305	169	0.05	0.888
<i>Leucogyrophana mollusca</i>	Leumo1	Brown rot	35 193 237	14 619	1262	108	0.09	0.915
<i>Coniophora olivacea</i>	Conol1	Brown rot	39 071 688	14 928	863	80	0.14	0.914
<i>Coniophora puteana</i>	Conpu1	Brown rot	42 968 544	13 761	210	7	2.4	0.912
<i>Serpula himantoides</i>	Serla_varsha1	Brown rot	45 978 785	13 805	4893	554	0.02	0.831
<i>Serpula lacrymans</i>	SerlaS7_9_2	Brown rot	42 734 799	12 789	36	6	2.95	0.922
<i>Fibulorhizoctonia</i> sp.	Fibsp1	Other saprotrophic	95 125 689	32 946	1918	98	0.29	0.923
<i>Piloderma croceum</i>	Pilcr1	EcM	59 326 866	21 583	715	33	0.53	0.914
<i>Plicaturopsis crispa</i>	Plicr1	White rot	34 498 416	13 626	316	7	1.83	0.926
<i>Cristinia sonora</i>	Crison1	White rot	30 130 434	11 356	114	16	0.65	0.918
<i>Polyporus brumalis</i>	Polbr1	White rot	45 718 511	18 244	621	34	0.36	0.928
<i>Agaricus bisporus</i>	Agabi_varbisH97_2	Soil saprotrophic	30 233 745	10 438	29	6	2.3	0.915

EcM, ectomycorrhizal; JGI, Joint Genome Institute.

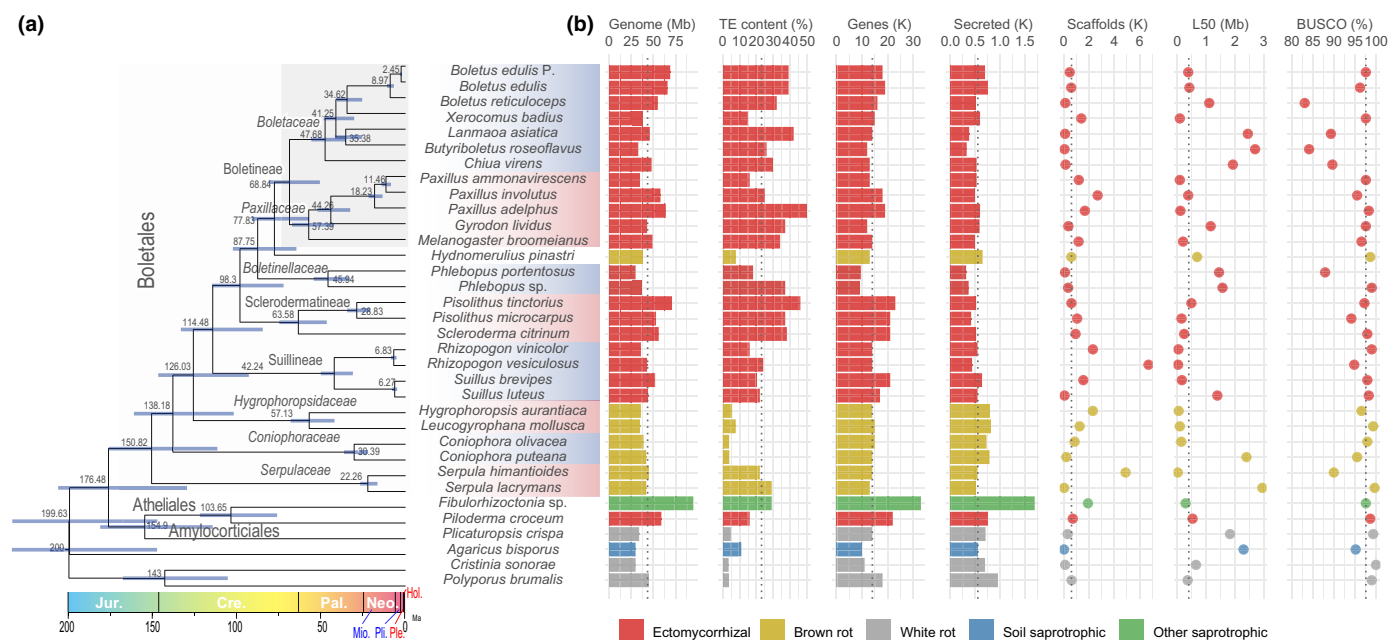
<sup>a</sup>Benchmarking universal single-copy orthologues, refers to genome completeness.

*B. reticuloceps*, they present the highest proportion of syntenic regions (Figs 5a,b, S3, S4). The genome synteny is, however, disrupted by numerous TE sequences, suggesting that TE activity reshuffled large parts of these closely related *Boletus* genomes (Figs 5a,b, S5–S9). The length of syntenic regions decreased with increased phylogenetic distances (Fig. S4). Scaffolds 2 and 5 of *B. edulis* (Boled5) still share many syntenic regions with other Boletales genomes (Figs S6, S7, S9). The CAZyme gene order is well conserved amongst species, but several inversions took place (e.g. between *Pa. involutus* (Paxin1), *H. pinastri* (Hydpi1), and *P. portentosus* (Phlpo1)). Intriguingly, gene insertions or inversions were often observed in TE-rich regions (e.g. the genes coding for AA3\_2, EXPN, GH128, and GT2 labelled in orange (Figs S6–S8), suggesting that TE insertions played a key role in these events).

Of note, we found that genes coding for proteases, lipases, and SSPs are closer to the nearest TEs than the random expectation (Fig. 5c; Table S8). Multivariate and ordination analyses further

indicated that unclassified TE sequences are significantly associated with gene counts for chitin synthase (GT2) and SSP genes in ectomycorrhizal genomes (Tables 2, S9; Fig. S10).

A hallmark of the genome evolution in ectomycorrhizal species is the dramatic loss of genes encoding CAZymes acting on plant cell wall polysaccharides in most species. We thus investigated the microsynteny of several CAZyme genes to track down the underlying mechanism(s) leading to gene loss. We found that the conservation of the nucleotide sequences of protein-coding genes framing the CAZyme genes – such as the cellobiohydrolase GH6 (Fig. S11), but also cellobiohydrolase (GH7), endoglucanases (CBM1-GH5\_5), and xyloglucanase (GH74-CBM1) (data not shown) – is highly conserved with no apparent disruption by TE sequences. It appears that the loss of these genes in ectomycorrhizal Boletales involved the accumulation of discrete nucleotide mutations, the so-called DNA decay process also observed in Tuberaceae (Murat *et al.*, 2018).



**Fig. 1** Time-calibrated phylogeny and genome features of 28 Boletales species and six outgroup species. (a) The time-calibrated phylogenetic tree in which all nodes received 100% bootstrap support based on maximum likelihood analysis, except the node comprising *Fibulorhizoctonia* sp., *Piloderma croceum*, and *Plicaturopsis crispa* with 99% support. The divergence time of each node with confidence interval is shown on/beside the branch. The named clades on the branches of the tree indicate the different taxonomic levels in different fonts (italic font: families, regular font: suborders, regular font in larger size: order). (b) Genome features. Colors in plots represent different lifestyles. Genome: Genome size; TE content, coverage of transposable elements in genomes; Genes, number of genes; Secreted, number of secreted proteins; Scaffolds, number of scaffolds; L50, N50 length; BUSCO, benchmarking universal single-copy orthologues score indicating genome completeness. The geological timescale (in million years, Ma) is indicated at the bottom.

## Core, dispensable, and species-specific gene families in Boletales

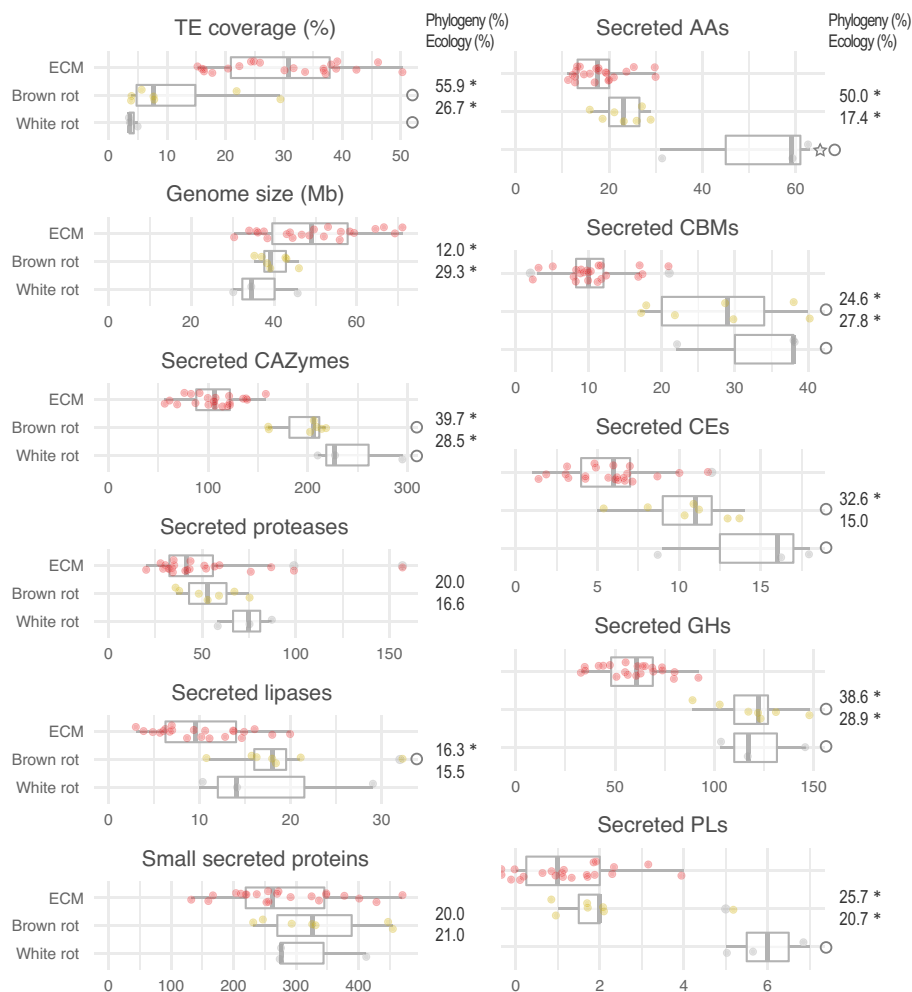
To compare the protein-coding gene repertoires encoded by the sequenced Boletales and identify species-specific gene families that might contribute to diversification of symbiosis-related traits in ectomycorrhizal Boletales, we clustered the predicted proteins to infer orthologous gene groups (orthogroups), including core genes (i.e. occurring in the 28 Boletales species), dispensable genes (i.e. found in at least two species), and species-specific genes (i.e. unique to a taxon). Whereas the core set of conserved genes ranged from 2291 to 2757, the repertoire of species-specific genes varies widely, ranging from 1364 in *Phlebobius* sp. FC\_14 to 12 816 in *Pi. tinctorius* (Fig. S12a; Table S10). The proportion of species-specific genes is much higher in Sclerodermatineae than in Boletineae. Species-specific genes, which are also referred to as taxonomically restricted genes, have recently become associated with the evolution of novelty, as numerous studies across the tree of life have now linked expression of taxonomically restricted genes with novel phenotypes (Johnson, 2018). As expected, they mostly encode proteins with no known function in the Boletales. We further identified the orthogroups, allowing the discrimination between saprotrophs and ectomycorrhizal symbionts. The number of symbiont-specific orthogroups ranges from 1347 in *G. lividus* to 5231 in *B. edulis* BED1, and that of brown-rot-specific genes varies from 842 in *H. pinastri* to 3755 in *Coniophora puteana* (Fig. S12b; Table S11). The number of symbiont/brown-rot-specific orthogroups tended to be

proportional to the genome size of the corresponding fungi (Figs 1, S12b).

## Diversification of the gene repertoires

The newly sequenced genomes, and more specifically the Boletaceae and Boletinellaceae genomes, offer a unique opportunity to examine the evolution of bolete-related and symbiosis-related genes and gene families across the Boletales order. Our reconstructions of genome-wide duplication and contraction events in Boletales revealed a considerable heterogeneity in the temporal dynamics of genome diversification between the different clades (Fig. 6a). We reconstructed similar ancestral copy numbers (12 000–13 000 genes) in ancestral nodes, with a moderate increase from the root of the tree (Fig. 6a). In accordance, net gene duplications (duplications minus losses) were inferred to be more or less constant along the tree backbone, with large loss events (2000–5000 gene losses, Fig. 6a) on the branches leading to the families. The mean gene duplication rate in Boletales increased abruptly around 10 Ma in the Late Miocene (Fig. 6b). This seems to be driven mostly by the family Boletaceae, and to a smaller extent by Suillineae (Fig. S13), and was not caused by whole-genome duplication (WGD) events (Fig. S14).

To gain insights into the unique features of the bolete genomes, we compared GO annotation frequencies of genes having duplications at ancestral nodes of Boletaceae. We performed gene-family-based enrichment analysis using GO annotations

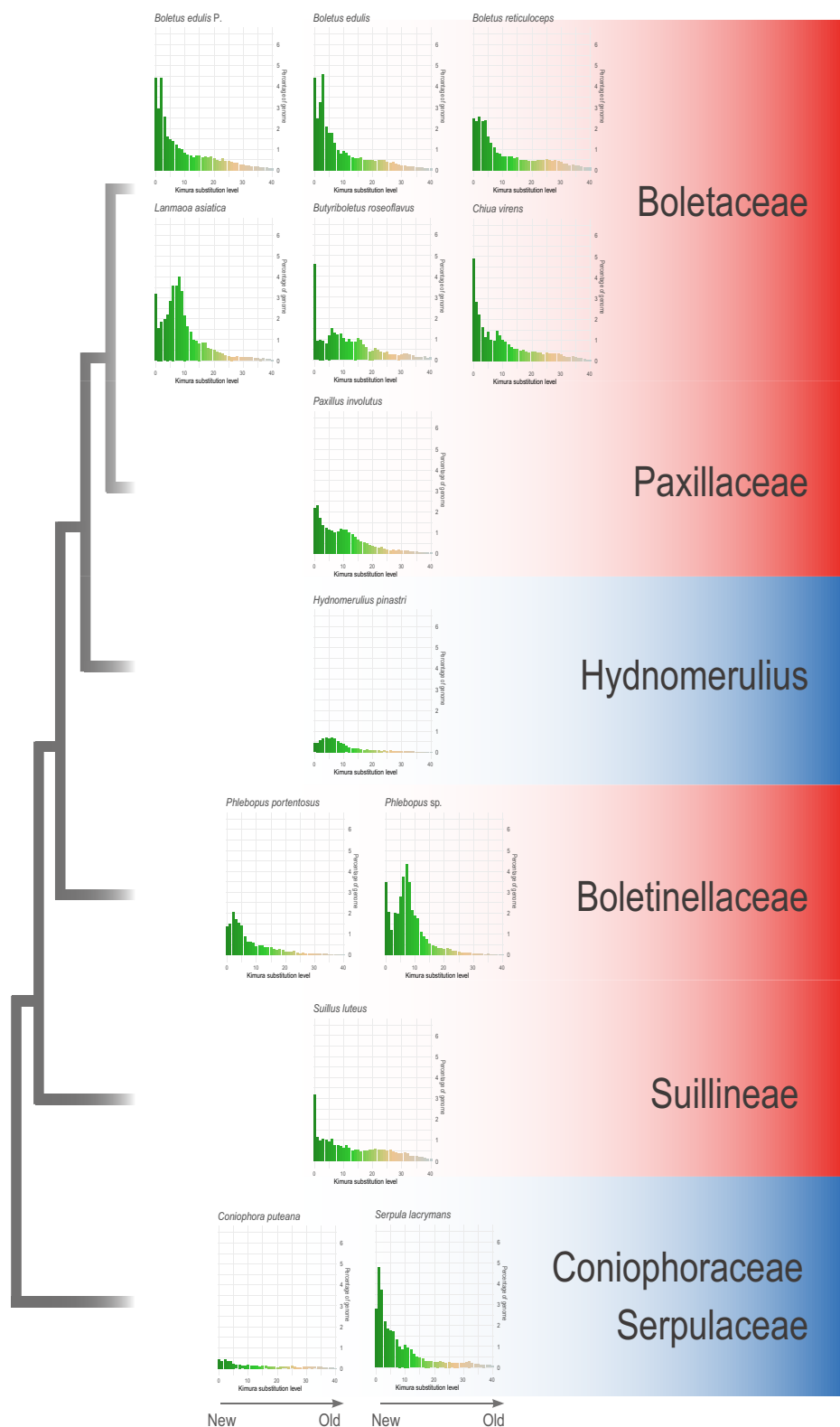


**Fig. 2** Distribution of the number of secreted proteins with percentage of variance explained by phylogeny and ecological groups. TE coverage, percentage of transposable elements (TEs) in genomes; Genome size, size of genomes in megabase pairs; CAZymes, number of secreted carbohydrate-active enzymes (CAZyme) domains; Small secreted proteins; number of small secreted proteins (< 300 amino acids); Proteases, number of secreted proteases; CAZyme families (AA, CBM, CE, GH, PL), number of secreted proteins with CAZyme domains. Asterisks indicate significantly different phylogenetic distances of species and ecological groups ( $P < 0.05$ ; permutational multivariate ANOVA (PERMANOVA) model, Genomic features ~ Phylogeny + Ecology). Circle and star shapes indicate significantly different ecological groups compared with ectomycorrhizal (ECM) symbionts and brown rots, respectively ( $P < 0.05$ ; pairwise PERMANOVA). See detailed information in Supporting Information Table S4 (PERMANOVA; pairwise PERMANOVA).

and found 23 molecular function, 13 biological process, and seven cellular component terms, which were significantly over-represented (Fisher's exact test,  $P < 0.05$ ) in gene families that showed duplications in early Boletaceae (Table S12). The most significant terms were 'superoxide dismutase activity', 'ubiquitin-like modifier activating enzyme activity', 'protein kinase activity', 'proteolysis', 'organophosphate catabolic process', 'myosin complex' and 'integral component of plasma membrane'. We found 169 InterPro terms that were significantly enriched (hypergeometric test,  $P < 0.05$ ) among gene families that showed duplications in ancestors of the Boletaceae. Among these InterPro terms, some exhibited specific functions, such as those related to protein kinases (e.g. fungal-type protein kinase, serine/threonine-protein kinase), GHs (lytic polysaccharide monooxygenases (LPMOs), GH20  $\beta$ -hexosaminidase, and GH18 chitinase) and transposases (*Kyakuja-Dileera-Zisupton* transposase, *Plavaka* transposase, and *Tc1*-like transposase).

## Lignocellulose and protein degradation abilities in Boletales

The repertoires of predicted secreted proteins in Boletales species were compared to identify possible lineage-specific features of the machinery involved in SOM and plant/microbial cell wall degradation (Figs 4, S15). Ectomycorrhizal Boletales have a significantly smaller mean number of secreted CAZymes (e.g. GHs) than brown-rot and white-rot species do, whereas no significant difference was found for SSPs (Fig. 2; Table S4;  $P < 0.05$ ; pairwise PERMANOVA). We sorted gene families coding for secreted cell-wall-degrading enzymes into three main functional categories; that is, enzymes involved in the degradation of lignin, cellulose/hemicellulose (including pectin), and microbial polysaccharides (i.e. chitin, fungal and bacterial glucans). The PCWDE gene copy numbers showed striking differences between brown-rot and ectomycorrhizal species (Fig. 4). The number of secreted PCWDEs in ectomycorrhizal species ranged from 16 for

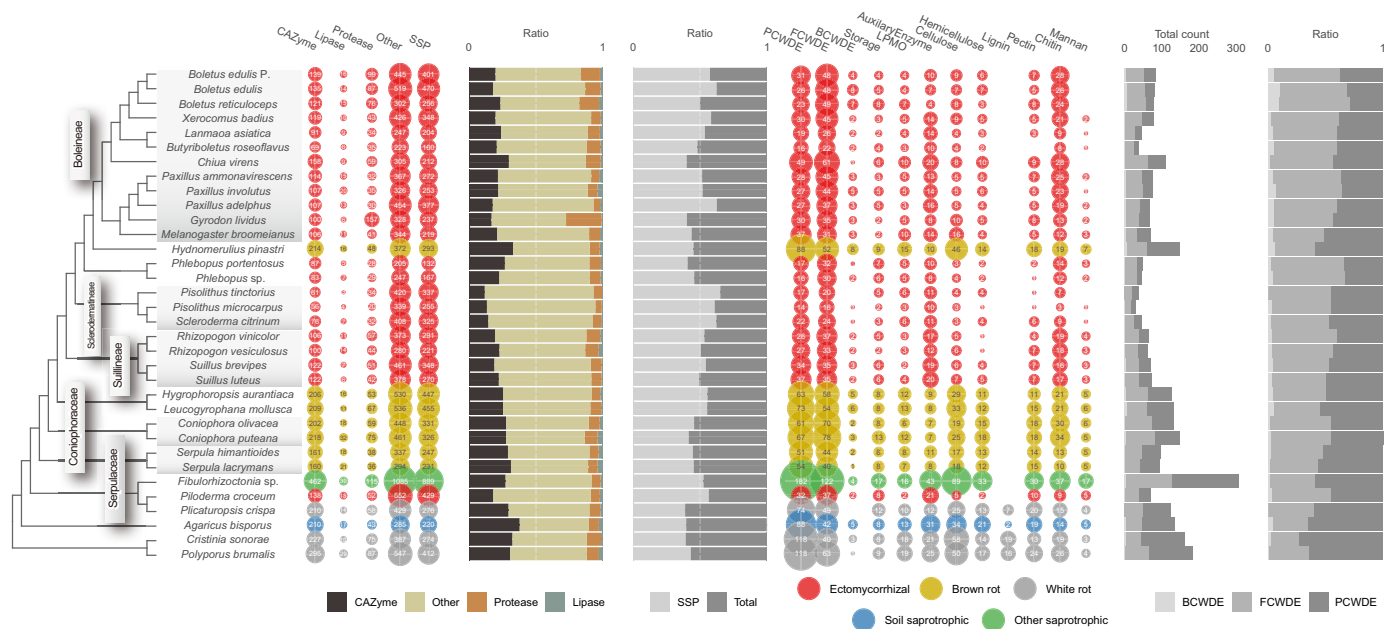


**Fig. 3** The landscape of transposable elements (TEs). The bar charts exhibit the genome coverage (%) of all TEs identified on the Y axis. Kimura substitution level (%) on the X axis indicates TE expansion events. Gradient colors (green to gray) show recent to ancient events. The selected representative fungi are sorted in the approximate evolutionary order. The taxonomic family of species is labeled. The fungal lifestyles are in color. Ectomycorrhizal (red), brown rot (blue).

*Bu. roseoflavus* to 49 for *C. virens*. The nonmetric multidimensional scaling analysis grouped the sequenced Boletales species according to their known nutritional modes, brown-rot vs symbiosis (Fig. S16). Of note, the two *Phlebopus* species clustered with the other symbiotic species. Phylogenetic relatedness of the

species (39.7%) and fungal ecology (28.5%) significantly accounted for the variation in secreted CAZyme genes (Fig. 2; Table S4;  $P < 0.05$ ; PERMANOVA model, Genomic features ~ Phylogeny + Ecology). The ectomycorrhizal Sclerodermatineae (*Pi. tinctorius*, *Pisolithus microcarpus*, and *Scleroderma*





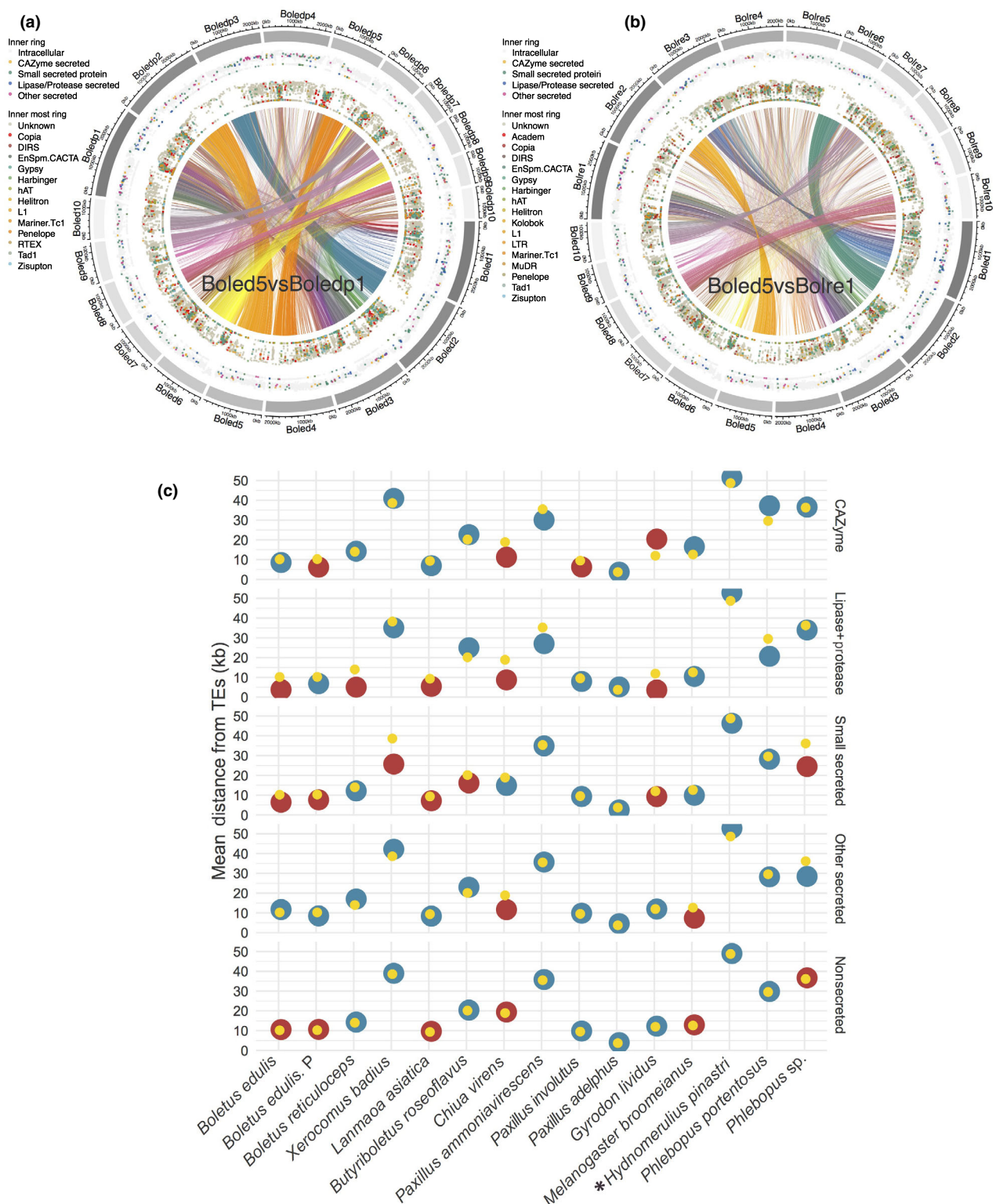
**Fig. 4** Secretome profiles of 28 Boletales and six outgroup species. The first bubble plot (on the left) shows the number of secreted genes for carbohydrate-active enzymes (CAZymes), lipases, proteases and others (i.e. all secreted proteins not in these first three groups). The small secreted protein (SSP) group is a subcategory showing the number of SSPs (< 300 aa). The size of the bubbles corresponds to the number of genes. Fungal species are color coded according to their known ecology. The first bar plots (in the middle) represent the ratio of CAZymes, lipases, and proteases to all secreted proteins (left) and the ratio of SSPs among the entire secretome (right). The second bubble plot (on the right) shows the number of plant cell-wall-degrading enzymes (PCWDEs) and microbial cell-wall-degrading enzymes (MCWDEs, including bacterial cell-wall-degrading enzymes (BCWDEs) and fungal cell-wall-degrading enzymes (FCWDEs)), lytic polysaccharide monooxygenases (LPMO), substrate-specific enzymes for cellulose, hemicellulose, lignin, and pectin (plant cell walls); chitin, glucan, mannan (fungal cell walls). The second bar plots (far right) show the total count of genes including PCWDE and MCWDE (left) and the ratios of PCWDE, BCWDE, and FCWDE (right).

*citrinum*) were further separated from other symbiotic species because of their even lower content in secreted CAZymes (Figs 4, S16).

Brown-rot and ectomycorrhizal Boletales and Atheliales/Amylocorticiales lack class II lignin-modifying PODs (AA2) (Fig. S15). In addition, the loss of synergistically acting GH7 and GH6 cellobiohydrolases, operating on cellulose reducing and nonreducing ends, is a hallmark of the ectomycorrhizal Boletales (Fig. S15). The near complete loss of core cellulose-acting CAZymes is underlined by the restricted set of CBM1 binding modules, which are often attached to key PCWDEs to mediate the targeting of enzymes to cellulose. However, a single copy of the endoglucanase of families GH9 and GH45, digesting cellulose into cellobiosaccharides, is found in the symbiotrophic species, suggesting a limited capacity to degrade cellulose. The endoglucanase GH12 is missing from *Paxillus*, *Phlebopus*, and *Scleroderma* species but is present in the Boletaceae. Concerning hemicellulose degradation, neither *endo*- $\beta$ -1,4-xylanases (GH11) acting on the xyloglucan backbone nor the debranching enzymes working synergistically with GH11 (families GH54, GH62, GH67, and CE3) were found. Enzymes involved in lignocellulose oxidation (LOX) in brown-rot species, such as cellobiohydrolases (AA3\_1), iron reductase domain containing proteins (AA8), and LPMOs (AA16), are also missing from the symbiotic Boletales genomes, except for one copy of pyranose-2-oxidase (AA3\_4) in *B. edulis*, *B. reticuloceps*, *L. asiatica*, and *M. broomeianus*. By contrast, gene

copy number in several LOX gene families – including laccases (AA1\_1), aryl alcohol/glucose oxidases (AA3\_2), alcohol oxidases (AA3\_3) (except for *Pisolithus* species), glyoxal oxidases (AA5), benzoquinone reductases (AA6), glucooligosaccharide oxidases (AA7), cellulose-acting LPMOs (AA9), and xylan-acting LPMOs (AA14) – is similar in brown-rot and symbiotic species. *Chiua virens*, a basal species of Boletaceae, has a very high number of phenoloxidases (AA1\_1), even higher than that of Boletales brown-rot species. Similarly, the copy number of genes coding for enzymes acting on fungal and bacterial polysaccharides is similar for ectomycorrhizal and brown-rot Boletales (Fig. 4).

The number of genes coding for secreted proteases ranges from 36 to 75 in brown-rot species, whereas it varies from 20 to 157 in ectomycorrhizal symbionts (Fig. 4), suggesting they have a similar capacity to degrade proteins accumulating in SOM. However, the set of secreted proteases is much higher in *Boletus* species (*B. edulis*, 87 and 99 copies; *B. reticuloceps*, 76 copies) and *G. lividus* (157 copies) (Fig. 4). We observed a striking gene duplication of the gene encoding pepstatin-insensitive carboxyl proteinases (family G01, formerly A4) in the *Boletus* clade with 25–34 copies (Fig. S17). This gene duplication occurred in a syntenic region in *B. reticuloceps* (Bolret1) and *B. edulis* Prilba (Boledp1) (Fig. S8; Table S13). Gene copy number for secreted lipases is similar in the genomes of ectomycorrhizal and brown-rot Boletales (Fig. 4), except for Sclerodermatineae with the lower repertoire of lipases.



**Fig. 5** (a, b) Syntenic analysis of *Boletus* species and (c) genomic distance investigation between genes and transposable elements (TEs). (a, b) Plots showing genes and TEs present only in syntenic regions of *Boletus* species. Outer circle, scaffold size; first inner circle, genes coding for secreted carbohydrate-active enzymes (CAZymes), small secreted proteins (SSPs), lipases, or proteases (see legend); second inner circle, TEs (see legend); vertical axis of inner circles, mean distances of between neighboring genes and TEs. Shorter distances between genes and TEs result in dots toward plot centers, whereas longer distances result in dots toward the outer circle. Boled5, *Boletus edulis* (BED1); Boledp1, *Boletus edulis* P (Prilba); Bolre1, *Boletus reticuliceps* (BR01). (c) Mean distances between TEs and genes coding for CAZymes, lipases/proteases, SSPs, other secreted proteins, and nonsecreted proteins. Yellow: mean distances of 10 000 randomly reshuffled genome models (to generate a null hypothesis). Blue: mean distances observed in genomes with no statistical significance ( $P > 0.05$ ). Red: mean distances observed in genomes with statistical significance ( $P < 0.05$ ). Brown rot (asterisk), ectomycorrhizal (unmarked). See Supporting Information Table S8 for detailed mean distances between repeat elements and genes.

**Table 2** Multivariate analysis<sup>a</sup> of transposable elements and genes.

Variation explained by unclassified repeats (%)	FDR-adjusted <i>P</i> values	Count of genes
21.90	<i>P</i> < 0.01	GT2
11.70	<i>P</i> < 0.05	Total GTs
8.75	<i>P</i> < 0.05	SSPs

FDR, false discovery rate; SSPs, small secreted proteins.

<sup>a</sup>See details of permutational multivariate ANOVA results in Supporting Information Table S9.

### *Phlebopus portentosus* saprotrophic capability

As already mentioned, *P. portentosus* PP33 and *Phlebopus* sp. FC\_14 are amongst the ectomycorrhizal Boletales having the lower set of PCWDEs (Fig. 4), suggesting a limited saprotrophic ability. Still, *P. portentosus* PP33 is able to produce basidiocarps in the absence of any host plant (Ji *et al.*, 2011; Cao *et al.*, 2015), raising the question of the catabolic pathways used to provide the high load of simple carbohydrates required for sustaining basidiocarp development. We thus performed RNA sequencing (RNA-seq) profiling of *P. portentosus* PP33 free-mycelium grown either on potato dextrose agar (PDA) medium or the rubber tree sawdust organic medium used to trigger basidiocarp development and measured the expression of the few genes coding for secreted PCWDEs (Table S14). We found that genes coding for laccase (AA1\_1), xyloglucan : xyloglucosyltransferase (GH16), chitinase (GH18), chitin deacetylase (CE4),  $\beta$ -1,3-glucanase (GH128), *exo*- $\beta$ -1,3-glucanase (GH55), and cellulose-acting LPMO (AA9) are expressed at a high level on PDA and sawdust/rice seed organic media (Fig. S18), suggesting the corresponding enzymes are involved in the release of soluble carbohydrates from plant organic matter. The higher expression of the laccase and chitinase genes (Fig. S18) suggests that sawdust polyphenols and chitin from fungal walls are potentially used to sustain the increased C demand required by fruiting body development.

### Conservation of symbiosis-induced genes

To test the conservation of ectomycorrhiza-related genes within Boletales species, we characterized the protein sequence similarity of available symbiosis-induced genes from ectomycorrhizal Boletales species amongst the nine saprotrophic and 19 symbiotrophic Boletales genomes. In the absence of transcriptomic data sets from bolete ectomycorrhizas, we queried the protein-coding gene repertoires using BLASTP and 275 known symbiosis-induced genes retrieved from *S. luteus*, *Pa. involutus*, *Pi. microcarpus*, and *Pi. tinctorius* transcript profiles obtained from ectomycorrhizal root tips of *Betula*, *Eucalyptus*, and *Pinus* (Kohler *et al.*, 2015) (Fig. S19). Most symbiosis-induced genes (85%) are conserved amongst saprotrophic and symbiotic Boletales (clusters II to V). These conserved symbiosis-induced genes encode for proteins involved in cellular processes and signaling, core metabolic pathways and CAZymes, or proteins with no eukaryotic orthologous groups (KOG) annotation. The 80 genes (10%) from cluster I

are mainly specific to *Paxillus*, *Pisolithus*, or *Suillus* species. They mainly encode MiSSPs or genes with no known function (no KOG domain).

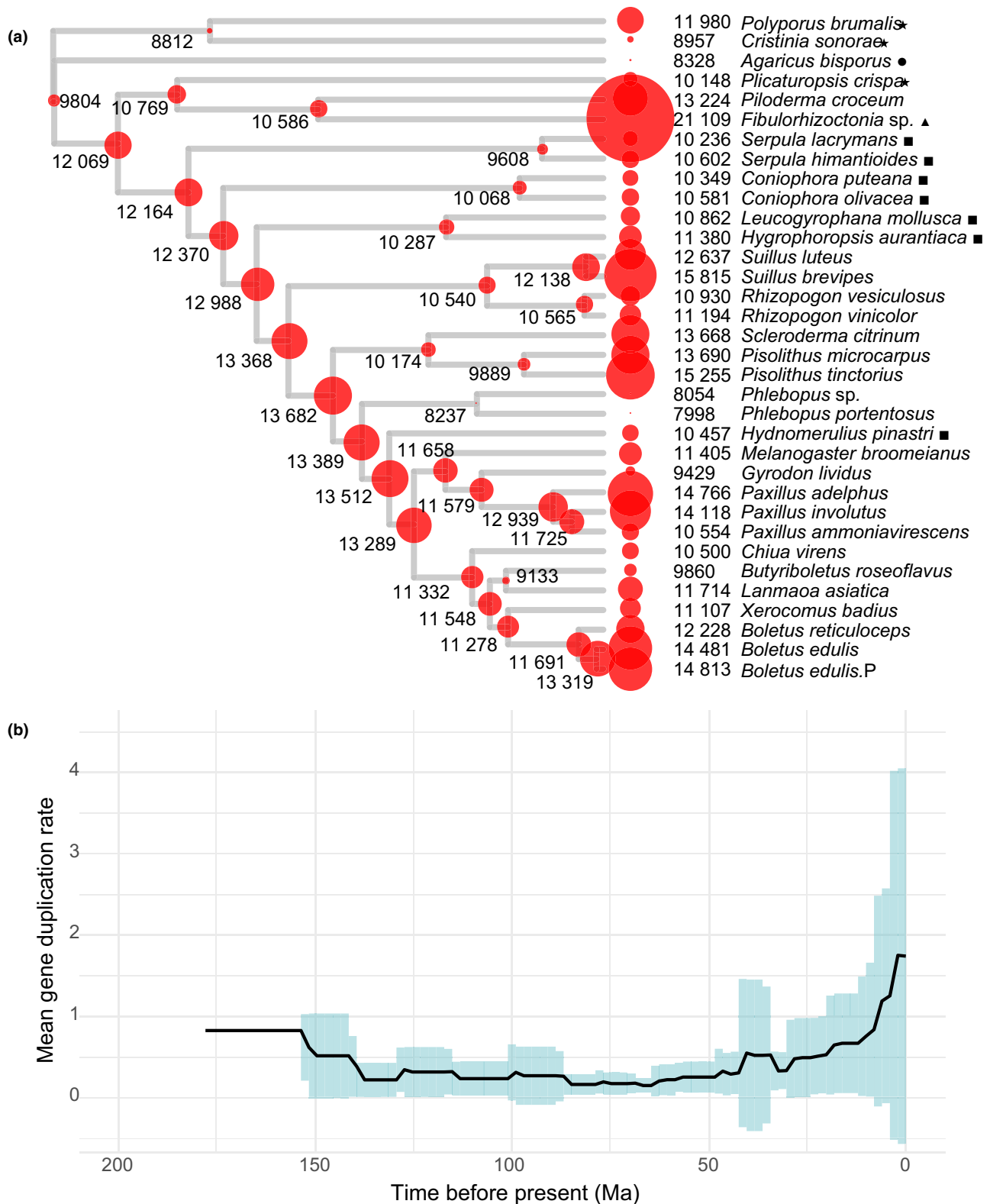
### Transporters, transcription factors, and secondary metabolism enzymes

Although the genomes of the Boletales species investigated have a similar set of terpene cyclases, symbiotrophic boletes have a smaller number of genes encoding NRPS and NRPS-like enzymes than brown-rot species do (Fig. S20), indicating a restricted capability to synthesize secondary metabolites (SMs). The gene copy numbers for the different membrane transporters and transcriptional regulators revealed no specific pattern(s) for ectomycorrhizal fungi in comparison with brown-rot species (Fig. S20).

## Discussion

The Boletales order includes *c.* 1300 described species (Kirk *et al.*, 2008). As either brown-rot decomposers or ectomycorrhizal symbionts, they play a key role in C cycling and sequestration in boreal, temperate, and montane forests (Eastwood *et al.*, 2011). Understanding the evolution and ecology of this major fungal group can be achieved only if coupled with a comprehensive description of their genomic traits. Deciphering the evolution of gene families related to the modes of nutrition (i.e. saprotrophy or symbiotrophy) is of particular interest to understand the ecology of these plant-associated fungi. The number of published Boletales genomes is now large enough to permit a comparative analysis of genome evolution between and within suborders and families exhibiting a wide range of ecological traits (e.g. mode of nutrition, habitat, host specificity). The 28 genomes of Boletales compared in the present study cover both the saprotrophic and symbiotic lifestyles found in this order. The evolutionary relationships between the selected Boletales and related Atheliales/ Amylocorticiales species were assessed by constructing a time-calibrated phylogenetic tree. Our phylogeny is in agreement with the recent megaphylogenies of Agaricomycetes (Krah *et al.*, 2018; Varga *et al.*, 2019). In these phylogenies, the emergence of the MRCA of Agaricomycetidae, Agaricales, Polyporales, Russulales, and Boletales was dated at 185 Ma, 173 Ma, 150 Ma, 152 Ma, and 142 Ma, respectively. In our analysis, the MRCA of the Boletales was dated at 151 Ma in the Late Jurassic, which is close to previous estimates. Many of the orders within Agaricomycetes (Agaricales, Boletales, Polyporales) originated before, but diversified after, the angiosperm and gymnosperm origins (Krah *et al.*, 2018). Hibbett & Matheny (2009) suggested that the ectomycorrhizal Boletales are younger than angiosperms and conifers but slightly older than the rosids, which contain many symbiotic partners of extant Boletales.

In our analysis, most brown-rot-producing Boletales species are placed as a paraphyletic group at the base of Boletales. However, we confirmed the placement of *H. pinastri* among ectomycorrhizal species in the 'core Boletales' (Sclerodermatineae, Suillineae, and Boletineae) (Kohler *et al.*, 2015; Krah *et al.*,



**Fig. 6** COMPARE analysis of Boletales. (a) The reconstructed gene copy numbers of Boletales species and outgroup species. (b) Gene duplication rates through time plot. Solid black line represents the mean of the gene duplication rate across lineages and time. Blue bars depict the SD of the gene duplication rates in a given timeframe. Timeframes were calculated by dividing the time interval of Boletales evolution by 100. White rot (asterisk), soil saprotrophic (circle), brown rot (quadrangle), other saprotrophic (triangle), ectomycorrhizal (unmarked).



2018). This brown-rot species retains numerous PCWDE genes lost in the lineages leading to the ectomycorrhizal members of the 'core Boletales' (this study; Kohler *et al.*, 2015), suggesting that the ancestor of the core Boletales had substantial saprotrophic ability, which has been retained in *H. pinastri*. Our work reveals that the overall pattern of gene loss and diversification differs amongst the ectomycorrhizal Boletales suborders, supporting the contention that these lineages originated from different ecologically diverse brown-rot precursors.

In their analysis of the Agaricomycetes evolution, Sánchez-García *et al.* (2020) showed that across their 8400-species phylogeny, diversification rates of ectomycorrhizal lineages were no greater than those of saprotrophic lineages. However, some ectomycorrhizal lineages have elevated diversification rates in comparison with their nonsymbiotic sister clades, suggesting that the evolution of ectomycorrhizal symbioses may act as a key innovation at local phylogenetic scales. In the Boletales, the relative time course of genome innovation in each of the family-level clades is distinct. PCWDE and SM gene losses is a well-known pattern associated with ectomycorrhizal evolution (Wolfe *et al.*, 2012; Martin *et al.*, 2016; Hess *et al.*, 2018; Lebreton *et al.*, 2021). Here, we showed that extensive gene loss took place in the early ectomycorrhizal Boletales clades. A substantial proportion of protein orthogroups contracted within the MRCA of ectomycorrhizal Boletales and appeared purged from the genome completely, suggesting that the early loss of a wide range of genes accompanied the formation of ectomycorrhizal associations. However, the present analysis also suggests a gradual increase in gene duplication rates in Boletaceae that is especially apparent through the Late Miocene to the present. This finding may be explained by at least three hypotheses. First, constraint on gene duplicability may have been lifted in these clades, leading to a surge of gene duplications. Second, single events with large impacts (e.g. large segmental duplications) may have shaped Boletaceae (and Suillineae) genomes, leading to an increase in duplication rates. This hypothesis was mostly excluded by our WGD analysis (Fig. S14). Third, it is also possible that the observed gene duplications represent adaptive events that contributed to the evolutionary success of boletes. Which, or whether a mixture, of these can best explain the observed rate increase will need to be examined with more in-depth analyses. It should also be noted that gene duplication/loss rates may be underestimated in our analyses due to the inherent inability of all comparative approaches to account for events in extinct lineages. However, it is unlikely that this has an effect on the detection rate acceleration in the Boletaceae.

During the period from the Late Miocene to the present, TEs also proliferated in most Boletales genomes, although Boletaceae genomes experienced the highest accumulation. Of note, three protein domains related to transposases (IPR040521, IPR038717, IPR041078) are enriched among ancestral nodes of Boletaceae. Our findings support the view that most Boletales species are undergoing a period of genome expansion (Castanera *et al.*, 2017). Intriguingly, TE-mediated genome amplification coincides with the estimated origins of ectomycorrhizal symbiosis in Boletales (this study; Kohler *et al.*, 2015).

The higher copy number of chitin synthase (*GT2*) genes in Boletaceae could be explained by the activity of TEs, as suggested by the close location of several *GT2* genes to TEs. In addition, the genomic location of genes coding for SSPs, proteases, and lipases tended to be closer to the nearest TEs than the random expectation. This suggests that TEs have contributed to the observed higher evolutionary rate of genes encoding effector-like SSPs, proteases, and lipases. These genome innovations may be related to the dramatic increase in species richness of Boletaceae (Wu *et al.*, 2016; Sato & Toju, 2019). Although the extrinsic factors that contributed to this Miocene gene innovation remain to be investigated, the substantial differences in the intrinsic biology and host/habitat preferences may be relevant to the different histories of Suillineae, Sclerodermatineae, and Boletineae.

In Agaricales, Russulales, Thelephorales, and Pezizales, the transition from saprotrophy to the symbiotic ecology coincided with the loss of most hydrolytic enzymes acting on lignocellulose (Kohler *et al.*, 2015; Hess *et al.*, 2018; Murat *et al.*, 2018; Miyauchi *et al.*, 2020; Looney *et al.*, 2021; Marqués-Gálvez *et al.*, 2021). Our analyses of the largest set of ectomycorrhizal Boletales genomes to date support the view that transitions from brown-rot to symbiosis entailed the widespread losses of PCWDEs acting on lignin, cellulose, hemicellulose, and pectin. We found striking genomic signatures related to (hemi)cellulose and lignin-degradation genes that separate ectomycorrhizal species from their brown-rot relatives. Although brown-rot species are depauperate in PCWDEs compared with white rots, their PCWDE repertoire, such as secreted GHs and polysaccharide lyases, is still substantially higher than the phylogenetically related symbiotrophic species. Compared with their saprotrophic cousins, all ectomycorrhizal Boletales lack the set of enzymes required for efficient cellulose degradation consisting of cellobiohydrolases from families GH6 and GH7, and cellulose-binding module CBM1 appended to cellulases and endoglucanases. The absence of the synergistically acting GH7 and GH6 enzymes, acting on cellulose reducing and nonreducing ends, is a hallmark of ectomycorrhizal species in Agaricales and Russulales (Wolfe *et al.*, 2012; Kohler *et al.*, 2015; Looney *et al.*, 2021). Concerning hemicelluloses, neither the *endo*- $\beta$ -1,4-xylanases (GH11) acting on the xyloglucan backbone nor the debranching enzymes working synergistically with GH11 were found. Within the general evolutionary trend of PCWDE loss, there are, however, more specific dynamics at play. Different ectomycorrhizal fungi retain distinct sets of PCWDEs, suggesting that, over their evolutionary history, symbiotic Boletales have become functionally diverse. For example, *C. virens* has a set of PCWDEs and fungal cell-wall-degrading enzymes (FCWDEs) closer to its brown-rot cousins. This likely reflects their ecological niche and variable dependence on their host plants. A smaller PCWDE repertoire was found in ectomycorrhizal Sclerodermatineae, such as *Pi. tinctorius* and *Pi. microcarpus*, reinforcing their dependence on the plant host. This may be related to their preferred ecological niche (i.e. sandy soils) with a scarce content in organic matter.

Despite their ecology, genomes of most symbiotic Boletales encode polygalacturonases (GH28, GH43), an endoglucanase GH5\_5 with a CBM1 module, endoglucanases GH12, and one

or two copies of xyloglucan-specific *endo*- $\beta$ -1,4-glucanases (GH45) in *B. edulis*, *X. badius*, *Phlebopus*, and *Suillus* species. The  $\beta$ -1,4 endoglucanase (GH5\_5-CBM1) is the orthologue of the *Laccaria bicolor* LbGH5-CBM1 involved in cell wall remodeling during the formation of the Hartig net and is an important determinant for successful symbiotic colonization of the *L. bicolor*–*Populus tremula*  $\times$  *alba* association (Zhang *et al.*, 2018). The remaining set of GHs may thus play a role in host root colonization (Zhang *et al.*, 2018, 2021). In addition, the gene copy number of several CAZyme families involved in LOX (laccases, alcohol oxidases, glyoxal oxidases, benzoquinone reductases, LPMOs (AA9, AA14)) is similar in brown-rot and ectomycorrhizal species, suggesting that some symbiotic Boletales are capable of mild lignocellulose decomposition (e.g. litter bleaching) to scavenge N trapped in SOM, as suggested by Floudas *et al.* (2020).

Ectomycorrhizal fungi are not all depauperate in PCWDEs. A few species, such as *Acephala macrosclerotium* (Leotiomyces), have kept a substantial set of PCWDEs that are repressed in ectomycorrhizal root tips (Miyachi *et al.*, 2020). They may represent transitional steps from pure saprotrophy toward pure ectomycorrhizal symbiosis (Lebreton *et al.*, 2021). Within the Boletales studied here, *Phlebopus* species in the Boletinellaceae appear to have a dual lifestyle. These boletes are able to establish ectomycorrhizal root tips with *P. kesiya* (Sanmee *et al.*, 2010; Kumla *et al.*, 2016), their basidiocarp  $^{13}\text{C}/^{15}\text{N}$  isotopic signature is typical of ectomycorrhizal species (Kumla *et al.*, 2016), and we showed in this study that their PCWDE repertoire is amongst the lower of the symbiotic boletes, supporting their adaptation to the symbiotic lifestyle. On the other hand, they are known to produce basidiocarps in Asian, African, and Australian grasslands (Ji *et al.*, 2011; Wilson *et al.*, 2012). The formation of fruiting bodies from soil free-living mycelium away from any known ectomycorrhizal host plants is in favor of a substantial saprotrophic ability of the mycelium to sustain the development of the sexual organs. Here, we confirmed that an ectomycorrhiza-forming *P. portentosus* mycelium can produce fruiting bodies on a mixed sawdust/seed substrate, and we proposed that the activity of their limited set of PCWDEs and FCWDEs, such as laccases and chitinases, is sufficient to release the carbohydrates required for basidiocarp development in the absence of glucose from the plant. This mild decay mechanism may play a role in litter decomposition in natural settings in the absence of host trees.

Novel and recently evolved genes, including effector-like MiSSPs, are thought to be responsible for the specific attributes of individual mycorrhizal lineages (Kohler *et al.*, 2015; Martin *et al.*, 2016). However, we have shown by using a phylostratigraphic analysis that a large proportion of symbiosis-related genes are orthologous to genes encoded by saprotrophic lineages that arose long before the evolution of the mutualistic associations (Miyachi *et al.*, 2020). These genes encoded by saprotrophic ancestors have been co-opted for the symbiotic lifestyle. They are related to key ecological traits, such as N and P acquisition (e.g. organic N and P-degrading secreted enzymes, nutrient transporters) already present in free-living saprotrophic ancestors of symbiotrophic species. Unfortunately, the identification of

symbiosis-related genes in boletes, such as *B. edulis*, is precluded by our inability to produce ectomycorrhizas *in vitro* for boletes, with these late-stage symbionts being poor colonizers of tree seedlings. We therefore used symbiosis-induced genes identified in transcript profiling from *Pa. involutus*, *Pi. tinctorius*, *Pi. microcarpus*, and *S. luteus* (Kohler *et al.*, 2015) and assessed their sequence conservation/divergence within the Boletales genomes. We confirmed that most of the symbiosis-induced genes are conserved amongst saprotrophic and symbiotic Boletales, and only a small set (*c.* 10%) of symbiosis-induced transcripts (encoding MiSSPs or proteins of unknown function) are species specific. This proportion is much higher (>30%) in the Agaricales (Miyachi *et al.*, 2020).

In conclusion, the genomes of symbiotrophic Boletales species have a greatly increased content in TEs and a strikingly reduced set of genes coding for PCWDEs in comparison with their brown-rot relatives. In addition, several species in the Boletaceae, Paxillaceae, and Boletinellaceae have kept a substantial set of endoglucanases and LPMOs acting on cellulose/hemicellulose and fungal polysaccharides, indicating that they may partly decompose SOM by a combined activity of oxidative and hydrolytic enzymes, as shown for *Pa. involutus* and *L. bicolor* (Nicolás *et al.*, 2019). In the latter ectomycorrhizal symbionts, the host tree actively controls the decomposing activities of the associated fungi by controlling the amount of photosynthetic C provided to the fungus. These genomic features are shared with ectomycorrhizal Agaricales and Russulales (Miyachi *et al.*, 2020; Looney *et al.*, 2021). Based on our findings, it is tempting to speculate that TEs accelerated the evolutionary rate of genes encoding effector-like SSPs, proteases, and lipases. On the other hand, we showed that the loss of secreted CAZymes was not related to TE activity, but to DNA decay. Here, we also showed that ectomycorrhizal Boletaceae, and to a smaller extent Suillineae, experienced an obvious expansion of gene families in the Late Miocene, an evolutionary event possibly contributing to the evolutionary success of boletes. This study produced the most inclusive phylogenomic and genomic analyses for the Boletales order to date, including characterization of several transitions from brown-rot to ectomycorrhizal lifestyles.

## Acknowledgements

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (grant no. XDB31000000 to ZLY), the Laboratory of Excellence ARBRE (grant no. ANR-11-LABX-0002-01), the Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, the Region Lorraine, and the European Regional Development Fund (to FMM). We also acknowledge grants from the International Partnership Program of Chinese Academy of Sciences (grant no. 151853KYSB20170026 to ZLY), the National Natural Science Foundation of China (grant no. 31970015 to GW), the Yunnan Ten Thousand Talents Program Plan Young & Elite Talents project (to GW), the Youth Innovation Promotion Association of the Chinese Academy of Sciences (grant no. 2017436 to GW), the Yunling Scholars Funds of the Ten-Thousand-Talents Program of Yunnan Provincial

Government (to ZLY), the Momentum Program of the Hungarian Academy of Sciences (grant no. LP2019/13-2019 to LGN), the National Research, Development and Innovation Office (contract no. GINOP-2.3.2-15-2016-00052 to LGN), and the Natural Science Foundation of Yunnan Province (grant no. 2017FB025 to YC). This work was also funded by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, and supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231 within the framework of the Mycorrhizal Genomics Initiative (CSP no. 305), Metatranscriptomics of Forest Soil Ecosystems project (CSP no. 570), and the 1000 Fungal Genome project (CSP no. 1974). GW would like to thank the China Scholarship Council for supporting his research stay at INRAE and thank Dr Hong Luo for his advice in performing genomic DNA isolation.













## Author contributions







FMM conceived and coordinated the Mycorrhizal Genomics Initiative. FMM and GW designed the project. GW, SM and FMM wrote the manuscript with the help of LGN, TV and ZLY. IVG coordinated genome sequencing and annotation at JGI. GW, BF, ZLY, YC and JX coordinated genome and transcriptome sequencing, and annotation at the Kunming Institute of Botany. A Kuo, AL, BM, CD, JJ, JP, HH, KB, KL, SA and VN performed transcriptome sequencing, assembly, and gene annotation at JGI. ED and BH performed CAZyme annotations. SM, GW, EM and FMM performed comparative genome analyses. GW, SM and A Kohler carried out the RNA-seq analysis. TV and LGN reconstructed and analyzed genome-wide duplication and contraction events. IGC, HP, JMP, JM and JWS provided some of the unpublished genomes used in this study. GW and SM contributed equally as first authors to this work.

## Code availability

Visual omics SHINGOTools – PRINGO, TINGO, SynGO, and VINGO – are available at GitHub: <https://github.com/ShingoMiyauchi>. The program COMPARE is at GitHub: <https://github.com/laszlognag/COMPARE>.

## ORCID

Steven Ahrendt  <https://orcid.org/0000-0001-8492-4830>  
 Kerrie Barry  <https://orcid.org/0000-0002-8999-6785>  
 Yang Cao  <https://orcid.org/0000-0002-1449-7244>  
 Elodie Drula  <https://orcid.org/0000-0002-9168-5214>  
 Bang Feng  <https://orcid.org/0000-0001-7237-2899>  
 Igor V. Grigoriev  <https://orcid.org/0000-0002-3136-8903>  
 Bernard Henrissat  <https://orcid.org/0000-0002-3434-8588>  
 Annegret Kohler  <https://orcid.org/0000-0002-9575-9567>  
 Alan Kuo  <https://orcid.org/0000-0003-3514-3530>  
 Kurt LaButti  <https://orcid.org/0000-0002-5838-1972>  
 Jon Magnuson  <https://orcid.org/0000-0001-7712-7024>  
 Francis M. Martin  <https://orcid.org/0000-0002-4737-3715>

Shingo Miyauchi  <https://orcid.org/0000-0002-0620-5547>  
 Emmanuelle Morin  <https://orcid.org/0000-0002-7268-972X>  
 László G. Nagy  <https://orcid.org/0000-0002-4102-8566>  
 Jonathan M. Plett  <https://orcid.org/0000-0003-0514-8146>  
 Gang Wu  <https://orcid.org/0000-0002-0076-5957>  
 Jianping Xu  <https://orcid.org/0000-0003-2915-2780>

## Data availability

Genome assemblies and gene annotations for the JGI-sequenced Boletales used in this study are available via the JGI fungal genome portal MycoCosm (see the Boletales portal at <https://mycocosm.jgi.doe.gov/boletales/boletales.info.html>). Sequenced genomes are also available at the National Center for Biotechnology Information (NCBI) GenBank (see Table S1 for accession codes/BioProjects). The complete transcriptome data sets are available at NCBI Sequence Read Archive (SRA). The accession codes for accessing the data deposited at NCBI-SRA are provided in the caption of Fig. S18. All other data supporting the findings of this study are included within the article and its additional files.

## References

- Alexa A, Rahnenführer J. 2020. *TOPGO: enrichment analysis for gene ontology*. [WWW document] URL <https://www.bioconductor.org/packages//2.10/bioc/html/topGO.html> [accessed 15 October 2020].
- Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22: 1600–1607.
- den Bakker HC, Zuccarello GC, Kuyper TW, Noordeloos ME. 2004. Evolution and host specificity in the ectomycorrhizal genus *Leccinum*. *New Phytologist* 163: 201–215.
- Baldrian P, Kolařík M, Stursová M, Kopecký J, Valášková V, Větrovský T, Zifčáková L, Snajdr J, Řídl J, Vlček Č *et al.* 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME Journal* 6: 248–258.
- Beaudet D, Chen ECH, Mathieu S, Yildirim G, Ndikumana S, Dalpé Y, Séguin S, Farinelli L, Stajich JE, Corradi N. 2018. Ultra-low input transcriptomics reveal the spore functional content and phylogenetic affiliations of poorly studied arbuscular mycorrhizal fungi. *DNA Research* 25: 217–227.
- Bessette A, Roody WC, Bessette AR. 2000. *North American Boletes: a color guide to the fleshy pored mushrooms*. New York, NY, USA: Syracuse University Press.
- Binder M, Hibbett DS. 2006. Molecular systematics and biological diversification of *Boletales*. *Mycologia* 98: 971–981.
- Binder M, Larsson KH, Matheny PB, Hibbett DS. 2010. Amylocorticiales ord. nov. and Jaapiiales ord. nov.: early diverging clades of Agaricomycetidae dominated by corticioid forms. *Mycologia* 102: 865–880.
- Blighe K, Lun A. 2020. *PCATools: everything principal components analysis*. R package v.2.2.0. [WWW document] URL <https://github.com/kevinblighe/PCATools> [accessed 10 April 2021].
- Bougher NL. 1995. Diversity of ectomycorrhizal fungi associated with eucalypts in Australia. In: Brundett M, Dell B, Malajczuk N, Gong MQ, eds. *Mycorrhizas for plantation forestry in Asia*. Canberra, ACT, Australia: Australian Centre Int Agricultural Research, 8–15.
- Cao Y, Zhang Y, Yu Z, Wang P, Tang X, He X, Mi F, Liu C, Yang D, Xu J. 2015. Genome sequence of *Phlebopus portentosus* strain PP33, a cultivated bolete. *Genome Announcements* 3: e00326–15.
- Castanera R, Pérez G, López-Varas L, Anselem J, LaButti K, Singan V, Lipzen A, Haridas S, Barry K, Grigoriev IV *et al.* 2017. Comparative genomics of *Coniophora olivacea* reveals different patterns of genome expansion in Boletales. *BMC Genomics* 18: e883.



- Chang Y, Desirò A, Na H, Sandor L, Lipzen A, Clum A, Barry K, Grigoriev IV, Martin FM, Stajich JE *et al.* 2019. Phylogenomics of Endogonaceae and evolution of mycorrhizas within Mucoromycota. *New Phytologist* 222: 511–525.
- Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology* 7: 429–447.
- Corner EJJH. 1972. *Boletus in Malaysia*. Singapore City, Singapore: Botanic Gardens.
- Darling AE, Carey LB, Feng W-C. 2003. The design, implementation, and evaluation of mpiBLAST. In: *4<sup>th</sup> International Conference on Linux Clusters and ClusterWorld 2003*. San Jose, CA, USA.
- Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2019. MODELTEST-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution* 37: 291–294.
- Dentinger BTM, Ammirati JF, Both EE, Desjardin DE, Halling RE, Henkel TW, Moreau P-A, Nagasawa E, Soyong K, Taylor AF *et al.* 2010. Molecular phylogenetics of porcini mushrooms (*Boletus* section *Boletus*). *Molecular Phylogenetics and Evolution* 57: 1276–1292.
- van Dongen S. 2000. *Graph clustering by flow simulation*. PhD thesis, University of Utrecht, Utrecht.
- Drehmel D, James T, Vilgalys R. 2008. Molecular phylogeny and biodiversity of the boletes. *Fungi* 1: 17–23.
- Eastwood DC, Floudas D, Bander M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M *et al.* 2011. The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* 333: 762–765.
- Eddy SR. 2011. Accelerated profile HMM searches. *Plos Computational Biology* 7: e1002195.
- Floudas D, Bentzer J, Ahrén D, Johansson T, Persson P, Tunlid A. 2020. Uncovering the hidden diversity of litter-decomposition mechanisms in mushroom-forming fungi. *The ISME Journal* 14: 2046–2059.
- Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. 2016. REGIONER: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* 32: 289–291.
- Gel B, Serra E. 2017. KARYOPLLOT: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33: 3088–3090.
- Grubisha LC, Trappe JM, Molina R, Spatafora JW. 2001. Biology of the ectomycorrhizal genus *Rhizopogon*. V. Phylogenetic relationships in the Boletales inferred from LSU rDNA sequences. *Mycologia* 93: 82–89.
- Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. CIRCIZE implements and enhances circular visualization in R. *Bioinformatics* 30: 2811–2812.
- Hage H, Miyauchi S, Virágh M, Drula E, Min B, Chaduli D, Navarro D, Favel A, Norest M, Lesage-Meessen L *et al.* 2021. Gene family expansions and transcriptome signatures uncover fungal adaptations to wood decay. *Environmental Microbiology* 23: 5716–5732.
- Han L-H, Feng B, Wu G, Halling RE, Buyck B, Yorou NS, Ebika STN, Yang Z-L. 2018. African origin and global distribution patterns: evidence inferred from phylogenetic and biogeographical analyses of ectomycorrhizal fungal genus *Strobilomyces*. *Journal of Biogeography* 45: 201–212.
- Heinemann P. 1951. Champignons récoltés au Congo belge par Madame M. Goossens-Fontana I. Boletineae. *Bulletin du Jardin Botanique de l'État a Bruxelles* 21: 223–346.
- Henkel TW, Terborgh J, Vilgalys RJ. 2002. Ectomycorrhizal fungi and their leguminous hosts in the Pakaraima Mountains of Guyana. *Mycological Research* 106: 515–531.
- Hess J, Skrede I, Chaib De Mares M, Hainaut M, Henriessat B, Pringle A. 2018. Rapid divergence of genome architectures following the origin of an ectomycorrhizal symbiosis in the genus *Amanita*. *Molecular Biology and Evolution* 35: 2786–2804.
- Hibbett DS, Matheny PB. 2009. The relative ages of ectomycorrhizal mushrooms and their plant hosts estimated using Bayesian relaxed molecular clock analyses. *BMC Biology* 7: e13.
- Hosen MI, Feng B, Wu G, Zhu XT, Li Y-C, Yang ZL. 2013. *Borofutius*, a new genus of Boletaceae from tropical Asia: phylogeny, morphology and taxonomy. *Fungal Diversity* 58: 215–226.
- Ji K-P, Cao Y, Zhang C-X, He M-X, Liu J, Wang W-B, Wang Y. 2011. Cultivation of *Phlebopus portentosus* in southern China. *Mycological Progress* 10: 293–300.
- Johnson B. 2018. Taxonomically restricted genes are fundamental to biology and evolution. *Frontiers in Genetics* 9: e407.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al.* 2014. INTERPROSCAN 5: genome-scale protein function classification. *Bioinformatics* 30: 1236–1240.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kirk PM, Cannon PF, Minter D, Stalpers JA. 2008. *Dictionary of the fungi*, 10<sup>th</sup> edn. Wallingford, UK: CAB International.
- Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A *et al.* 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics* 47: e410.
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35: 4453–4455.
- Krah F-S, Bässler C, Heibl C, Soghigian J, Schaefer H, Hibbett DS. 2018. Evolutionary dynamics of host specialization in wood-decay fungi. *BMC Evolutionary Biology* 18: e119.
- Kumla J, Hobbie EA, Suwannarach N, Lumyong S. 2016. The ectomycorrhizal status of a tropical black bolete, *Phlebopus portentosus*, assessed using mycorrhizal synthesis and isotopic analysis. *Mycorrhiza* 26: 333–343.
- Lebreton A, Zeng Q, Miyauchi S, Kohler A, Dai Y-C, Martin FM. 2021. Evolution of the mode of nutrition in symbiotic and saprotrophic fungi in forest ecosystems. *Annual Review of Ecology, Evolution and Systematics* 52: 385–404.
- LePage BA, Currah RS, Stockey RA, Rothwell GW. 1997. Fossil ectomycorrhizae from the middle Eocene. *American Journal of Botany* 84: 410–412.
- Lofgren LA, Nguyen NH, Vilgalys R, Ruytinx J, Liao H-L, Branco S, Kuo A, LaButti K, Lipzen A, Andreopoulos W *et al.* 2021. Comparative genomics reveals dynamic genome evolution in host specialist ectomycorrhizal fungi. *New Phytologist* 230: 774–792.
- Looney B, Miyauchi S, Morin E, Drula E, Courty PE, Kohler A, Kuo A, LaButti K, Pangilinan J, Lipzen A *et al.* 2021. Evolutionary transition to the ectomycorrhizal habit in the genomes of a hyper-diverse lineage of mushroom-forming fungi. *New Phytologist*, in press.
- Marqués-Gálvez JE, Miyauchi S, Paolucci F, Navarro-Ródenas A, Arenas F, Pérez-Gilbert M, Morin E, Auer L, Barry KW, Kuo A *et al.* 2021. Desert truffle genomes reveal their reproductive modes and new insights into plant–fungal interaction and ectomycorrhizal lifestyle. *New Phytologist* 229: 2917–2932.
- Martin F, Kohler A, Murat C, Veneault-Fourrey C, Hibbett DS. 2016. Unearthing the roots of ectomycorrhizal symbioses. *Nature Reviews Microbiology* 14: 760–773.
- Miyauchi S, Kiss E, Kuo A, Drula E, Kohler A, Sánchez-García M, Morin E, Andreopoulos B, Barry KW, Bonito G *et al.* 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nature Communications* 11: e5125.
- Muñoz JA. 2005. *Fungi Europaei 2: Boletus s.l.: Strobilomycetaceae*. Alassio, Italy: Edizioni Causso.
- Murat C, Payen T, Noel B, Kuo A, Morin E, Chen J, Kohler A, Krizsán K, Balestrini R, Da Silva C *et al.* 2018. Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle. *Nature Ecology & Evolution* 2: 1956–1965.
- Nagy LG, Ohm RA, Kovács GM, Floudas D, Riley R, Gácsér A, Sipiczki M, Davis JM, Doty SL, de Hoog GS *et al.* 2014. Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nature Communications* 5: e4471.
- Nagy LG, Riley R, Tritt A, Adam C, Daum C, Floudas D, Sun H, Yadav JS, Pangilinan J, Larsson K-H *et al.* 2016. Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities. *Molecular Biology and Evolution* 33: 959–970.



- Newman EI, Reddell P. 1987. The distribution of mycorrhizas among families of vascular plants. *New Phytologist* 106: 745–751.
- Nicolás C, Martin-Bertelsen T, Floudas D, Bentzer J, Smits M, Johansson T, Troein C, Persson P, Tunlid A. 2019. The soil organic matter decomposition mechanisms in ectomycorrhizal fungi are tuned for liberating soil organic nitrogen. *The ISME Journal* 13: 977–988.
- Nuhn ME, Binder M, Taylor AFS, Halling RE, Hibbett DS. 2013. Phylogenetic overview of the Boletineae. *Fungal Biology* 117: 479–511.
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P *et al.* 2019. *VEGAN: community ecology package*. R package v.2.5-6. [WWW document] URL <https://CRAN.R-project.org/package=vegan> [accessed 25 April 2019].
- Ortega-Martínez P, Águeda B, Fernández-Toirán LM, Martínez-Peña F. 2011. Tree age influences on the development of edible ectomycorrhizal fungi sporocarps in *Pinus sylvestris* stands. *Mycorrhiza* 21: 65–70.
- Peay KG, Kennedy PG, Talbot JM. 2016. Dimensions of biodiversity in the earth mycobiome. *Nature Reviews Microbiology* 14: 434–447.
- Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. 2015. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Frontiers in Microbiology* 6: e1278.
- Pham N-D, Suzuki A, Pham N-D, Yamada A, Shimizu K, Noda K, Dang L-A, Suzuki A. 2012. A sheathing mycorrhiza between the tropical bolete *Phlebopus spongiosus* and *Citrus maxima*. *Mycoscience* 53: 347–353.
- Rapier S, Fruchier A, Bessière J-M. 1997. Volatile aroma constituents of agarics and boletes. In: Pandalai SG, ed. *Recent research developments in phytochemistry*. Trivandrum, India: Research Signpost, 567–584.
- Ruiz-Dueñas FJ, Barrasa JM, Sánchez-García M, Camarero S, Miyauchi S, Serrano A, Linde D, Babiker R, Drula E, Ayuso-Fernández I *et al.* 2020. Genomic analysis enlightens Agaricales lifestyle evolution and increasing peroxidase diversity. *Molecular Biology and Evolution* 38: 1428–1446.
- Sánchez-García M, Ryberg M, Kalsoom Khan F, Varga T, Nagy LG, Hibbett DS. 2020. Fruiting body form, not nutritional mode, is the major driver of diversification in mushroom-forming fungi. *Proceedings of the National Academy of Sciences, USA* 117: 32528–32534.
- Sanmee R, Lumyong S, Lumyong P, Dell B. 2010. *In vitro* cultivation and fruit body formation of the black bolete, *Phlebopus portentosus*, a popular edible ectomycorrhizal fungus in Thailand. *Mycoscience* 51: 15–22.
- Sato H, Toju H. 2019. Timing of evolutionary innovation: scenarios of evolutionary diversification in a species-rich fungal clade, Boletales. *New Phytologist* 222: 1924–1935.
- Sato H, Yumoto T, Murakami N. 2007. Cryptic species and host specificity in the ectomycorrhizal genus *Strobilomyces* (Strobilomycetaceae). *American Journal of Botany* 94: 1630–1641.
- Samataakis A. 2014. RAXML v.8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Tedersoo L, Smith ME. 2013. Lineages of ectomycorrhizal fungi revisited: foraging strategies and novel lineages revealed by sequences from belowground. *Fungal Biology Reviews* 27: 83–99.
- Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllősi GJ, Szarkándi JG, Papp V, Albert L *et al.* 2019. Megaphylogeny resolves global patterns of mushroom evolution. *Nature Ecology & Evolution* 3: 668–678.
- Wilson AW, Binder M, Hibbett DS. 2012. Diversity and evolution of ectomycorrhizal host associations in the Sclerodermatineae (Boletales, Basidiomycota). *New Phytologist* 194: 1079–1095.
- Wolfe BE, Tulloss RE, Pringle A. 2012. The irreversible loss of a decomposition pathway marks the single origin of an ectomycorrhizal symbiosis. *PLoS ONE* 7: e39597.
- Wright ES. 2015. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics* 16: e322.
- Wu G, Feng B, Xu J, Zhu X-T, Li Y-C, Zeng N-K, Hosen MI, Yang Z-L. 2014. Molecular phylogenetic analyses redefine seven major clades and reveal 22 new generic clades in the fungal family Boletaceae. *Fungal Diversity* 69: 93–115.
- Wu G, Li Y-C, Zhu X-T, Zhao K, Han L-H, Cui Y-Y, Li F, Xu J-P, Yang Z-L. 2016. One hundred noteworthy boletes from China. *Fungal Diversity* 81: 25–188.
- Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.
- Zang M. 2006. *Flora fungorum sinicorum. Vol. 22, Boletaceae (I)*. Beijing, China: Science Press.
- Zhang F, Anasontzis GE, Labourel A, Champion C, Haon M, Kemppainen M, Commun C, Deveau A, Pardo A, Veneault-Fourrey C *et al.* 2018. The ectomycorrhizal basidiomycete *Laccaria bicolor* releases a secreted  $\beta$ -1,4 endoglucanase that plays a key role in symbiosis development. *New Phytologist* 220: 1309–1321.
- Zhang F, Labourel A, Haon M, Kemppainen M, Silva Machado ED, Brouilly N, Veneault-Fourrey C, Kohler A, Rosso M-N, Pardo A *et al.* 2021. The ectomycorrhizal basidiomycete *Laccaria bicolor* releases a GH28 polygalacturonase that plays a key role in symbiosis establishment. *bioRxiv*. doi: 10.1101/2021.09.24.461608.
- Zhao R-L, Li G-J, Sánchez-Ramírez S, Stata M, Yang Z-L, Wu G, Dai Y-C, He S-H, Cui B-K, Zhou J-L. 2017. A six-gene phylogenetic overview of Basidiomycota and allied phyla with estimated divergence times of higher taxa and a phyloproteomics perspective. *Fungal Diversity* 88: 43–74.

## Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

**Fig. S1** The coverage and copy number of transposable elements identified in Boletales and selected genomes.

**Fig. S2** Bar plot showing the scaffold size of selected fungi in their phylogenetic order.

**Fig. S3** The percentage of syntenic blocks in selected Boletales genomes.

**Fig. S4** Genome macrosynteny in Boletineae and allied Boletinelaceae.

**Fig. S5** Locations of CAZyme-coding genes in scaffold 1 of *Boletus edulis* (Boled5) with corresponding syntenic regions in other allied Boletales fungi.

**Fig. S6** Locations of CAZyme-coding genes on scaffold 2 of *Boletus edulis* (Boled5) with corresponding syntenic regions in other allied Boletales fungi.

**Fig. S7** Locations of CAZyme-coding genes on scaffold 5 of *Boletus edulis* (Boled5) with corresponding syntenic regions in other allied Boletales fungi.

**Fig. S8** Locations of SSP genes on scaffold 2 of *Boletus edulis* (Boledp1) with corresponding syntenic regions in other allied Boletales fungi.

**Fig. S9** Presence of CAZyme-coding genes in scaffolds of *Boletus edulis* (Boled5) aligned with other Boletales fungi.

**Fig. S10** Ordination analysis assessing the correlation between TE coverage and selected protein-coding gene counts.

**Fig. S11** Conservation of the protein-coding genes framing the CBM1-GH6 cellobiohydrolase gene located on scaffold 2 of the saprotrophic *Hydnomerulius pinastri* (Hydpi2) and missing from the symbiotrophic Boletales species.

**Fig. S12** Gene conservation and innovation in Boletales fungi.

**Fig. S13** Diversification of the protein-coding gene repertoires in Boletales families.

**Fig. S14** Ks distributions of the species in Boletaceae and Suillineae which drove the high gene duplication rate in Boletales.

**Fig. S15** Number of secreted and total CAZyme-coding genes in the genomes.

**Fig. S16** Non-metric multidimensional scaling (NMDS) analysis of total and secreted CAZyme-coding gene repertoires in Boletales and other species.

**Fig. S17** Number of genes coding for proteases and protease inhibitors (according to the MEROPS database) in the genomes of 28 Boletales and selected outgroup species.

**Fig. S18** The transcription level of *Ph. portentosus* PP33 genes for plant and fungal cell wall degrading enzymes under three conditions.

**Fig. S19** Phylogenetic conservation of symbiosis-induced genes in Boletales.

**Fig. S20** Number of genes coding for secondary metabolism, transcription factors and membrane transporters.

**Methods S1** Genome sequencing, assembly, annotation and other comparative genomic analyses.

**Table S1** Strain information of the newly sequenced Boletales species.

**Table S2** Taxonomic affiliation, genomic features and statistics for the 34 included genomes.

**Table S3** Data for PERMANOVA on genome features of Boletales species.

**Table S4** Pair-wise PERMANOVA and PERMANOVA of genome features.

**Table S5** The statistics of generalized least squares analysis with the Brownian motion model.

**Table S6** Identified transposable elements in the Boletales species and outgroupss.

**Table S7** CAZy count in top 10 scaffolds and total scaffolds of the 11 species with high-quality genomes out of 15 target fungi.

**Table S8** Mean distances between repeat elements and genes.

**Table S9** PERMANOVA results among TEs and SSPs, GT2, GT.

**Table S10** Core, dispensable, specific genes of 34 fungi.

**Table S11** Count of EcM-specific and brown-rot-specific genes in 28 Boletales species.

**Table S12** COMPARE function enrichment.

**Table S13** Data for Kirsame (drizzle) plot illustrating SSPs in syntenic regions with SSP annotations.

**Table S14** *Phlebopus portentosus* PP33 transcription levels for genes coding for PCWDEs+FCWDEs and all CAZymes.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.