Title

Comparative genomics reveals a dynamic genome evolution in the ectomycorrhizal milk-cap (Lactarius) mushrooms

Journal

New Phytologist, 235(1)

ISSN

0028-646X

Authors

Lebreton, Annie

Tang, Nianwu

Kuo, Alan

et al.

Publication Date

2022-07-01

DOI

10.1111/nph.18143

Comparative genomics reveals a dynamic genome evolution in the ectomycorrhizal milk-cap (*Lactarius*) mushrooms

Annie Lebreton[1,2#], Nianwu Tang[2,3#*], Alan Kuo[4], Kurt LaButti[4], Bill Andreopoulos[4], Elodie Drula[5], Shingo Miyauchi[6], Kerrie Barry[4], Alicia Clum[4], Anna Lipzen[4], Daniel Mousain[7], Vivian Ng[4], Ran Wang[3], Yucheng Dai[1], Bernard Henrissat [8,9], Igor V. Grigoriev[4,10], Alexis Guerin-Laguette[11], Fuqiang Yu[3*], Francis M. Martin[1, 2*]

[1] *Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, 100083 Beijing, China.*

[2] *Université de Lorraine, INRAE, UMR Interactions Arbres/Microorganismes, Centre INRAE-Grand Est-Nancy, 54280 Champenoux, France.*

[3] *Germplasm Bank of Wild Species, Yunnan Key Laboratory for Fungal Diversity and Green Development, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China.*

[4] *U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.*

[5] *Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, Marseille 13288, France.*

[6] *Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany.*

[7] *Retired INRAE scientist, 11 rue Demians, Nîmes 30000, France.*

[8] *Department of Biotechnology and Biomedicine (DTU Bioengineering), Technical University of Denmark, 2800 Kgs Lyngby, Denmark.*

[9] *Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.*

[10] *Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA 94720, USA.*

[11] *Mycotree C/- Southern Woods Nursery, 1002 Robinsons Road, RD8, Christchurch 7678, New Zealand.*

[#] These authors contributed equally to this work.

*Authors for correspondence:

Francis M. Martin (francis.martin@inrae.fr)

Fuqiang Yu (fqyu@mail.kib.ac.cn)

Nianwu Tang (tangnianwu@hotmail.com)

Total word count: 5611.

Introduction: 721; Materials and Methods: 1696; Results: 1903; Discussion: 1291.

Figures: 6; Tables: 1; Supporting Information elements: 14.

Summary

- Ectomycorrhizal fungi play a key role in forests by establishing mutualistic symbioses with woody plants. Genome analyses have identified conserved symbiosis-related traits among ectomycorrhizal fungal species, but the molecular mechanisms underlying host-specificity remain poorly known.
- We sequenced and compared the genomes of seven species of milk-cap fungi (*Lactarius*, Russulales) with contrasted host-specificity. We also compared these genomes with those of symbiotic and saprotrophic Russulales species aiming to identify genes involved in their ecology and host-specificity.
- The size of *Lactarius* genomes is significantly larger than other Russulales species, owing to a massive accumulation of transposable elements and duplication of dispensable genes. As expected, their repertoire of genes coding for plant cell wall degrading enzymes is restricted, but they retained a substantial set of genes involved in microbial cell wall degradation. Notably, *Lactarius* species showed a striking expansion of genes encoding proteases, such as secreted ectomycorrhiza-induced sedolisins. A high copy number of genes coding for small secreted LysM proteins and *Lactarius*-specific lectins were detected; they may be linked to host-specificity.
- This study revealed a large diversity in the genome landscapes and gene repertoires within Russulaceae. The known host specificity of *Lactarius* symbionts may be related to mycorrhiza-induced species-specific genes, including secreted sedolisins.

Key words: Russulales, comparative genomics, ectomycorrhizal fungi, trait evolution, proteases

Introduction

Fungi perform essential ecological functions in terrestrial ecosystems, whether as saprotrophs feeding on dead organic matters or as biotrophs (parasites or symbionts) acquiring nutrients from living hosts. Soil-borne ectomycorrhizal (EcM) fungi establish symbiotic relationships with 60% of tree stems on Earth, and mediate the exchange of plant carbohydrates for soil minerals (Brundrett & Tedersoo, 2018; Steidinger *et al.*, 2019). They evolved independently, at least 80 times, from diverse saprotrophic ancestors (Tedersoo *et al.*, 2010; Martin *et al.*, 2016; Lebreton *et al.*, 2021b). These multiple emergences of EcM lineages involved lineage-specific genomic innovations, such as effector-like mycorrhiza-induced small secreted proteins (MiSSPs), but also loss of gene families, such as plant cell wall degrading enzymes (PCWDEs). Each lineage however retains unique set of PCWDEs , likely reflecting their specific evolutionary history and ecological roles (Kohler *et al.*, 2015; Miyauchi *et al.*, 2020; Lebreton *et al.*, 2021b). Species-specific changes in gene repertoires have also been observed within a single lineage, i.e., Amanitaceae, including expansion of clade-specific small secreted proteins (SSPs) (Hess *et al.*, 2018). The loss of PCWDE genes in a few species of saprotrophic ancestors suggest a possible preadaptation to EcM symbiosis in some lineages (Hess *et al.*, 2018; Looney *et al.*, 2022).

Lactarius is an EcM fungal genus belonging to Russulaceae (Russulales), a lineage that is rich in EcM species and widely distributed in temperate and subtropical forests (Looney *et al.*, 2016, 2018). The specific traits, such as host specificity and defense-related latex exudation makes this genus an ideal group to investigate the evolution of ectomycorrhizal fungi at the genomic level (Nuytinck *et al.*, 2007; Verbeken & Nuytinck, 2013; Looney *et al.*, 2018; Wang *et al.*, 2019). Given the contrasting patterns of host-specificity between *Lactarius* and *Russula* symbionts, the latter being mostly generalists, a comparison of their gene repertoires may provide novel insights on the molecular mechanisms underlying the specific interactions between EcM fungi and their host(s). It has been suggested that lectins, carbohydrate-binding proteins that are highly specific for sugar groups, could be involved in the recognition between *L. deterrimus* and spruce roots during the early stage of symbiosis (Guillot *et al.*, 1991; Giollant *et al.*, 1993), but definitive demonstration is lacking. A metatranscriptomic study of host-specific patterns of gene expression between *Pinus* species and their symbiotic EcM fungi in the genus *Suillus* revealed that the host plant and EcM fungal symbiont each express unique gene sets during incompatible vs. compatible pairings. These genes code for proteins involved in signaling pathways, including G-protein coupled

receptors (GPCRs), secretory pathways, leucine-rich repeat proteins, and pathogen resistance proteins that are similar to those associated with host-pathogen interactions (Liao *et al.*, 2016). In contrast, a large-scale comparative study of *Suillus* and other less specific EcM fungal genomes found that only terpene- and nonribosomal polyketide synthases (NRPS), but not GPCRs or small secreted proteins (SSPs), expanded in host-specific *Suillus* (Lofgren *et al.*, 2021).

In order to link gene repertoires to ecological traits in Russulaceae, we sequenced and analyzed the genome of seven *Lactarius* species in section *Deliciosi*. These milk-cap species were collected from various geographical regions and are known for their host specificity toward Pinaceae (Wang *et al.*, 2019; Tang *et al.*, 2021). The section *Deliciosi* contains at least 38 taxa worldwide, including many well known edible species (Nuytinck *et al.*, 2007). Most species in this section form ectomycorrhizas with *Pinus*, but they can also associate with other conifers (*Picea*, *Abies*, etc.), while a few species, i.e., *L. indigo* and *L. subindigo*, have been reported to interact with broadleaved trees, such as *Quercus* and *Castanopsis*. The host switch between Pinaceae and Fagaceae seems to have occurred a few times throughout evolution (Nuytinck *et al.*, 2007). Moreover, European species have a well-documented host specificity, e.g. *L. salmonicolor* on *Abies* and *L. deterrimus* on *Picea*. We hypothesize that a comparison of the available gene repertoires of Russulaceae and *Lactarius* species would provide new information on (1) the evolution of the symbiotic lifestyle within the Russulaceae and (2) the molecular mechanisms underlying host selection in a major group of ectomycorrhizal symbionts. By comparing genomes of saprotrophic and symbiotic Russulaceae species, we revealed the genetic basis for their contrasted lignocellulose- and protein-degrading abilities. We also identified major differences in their repertoires of dispensable genes and secreted proteases. Finally, we assessed the conservation of symbiotic-related traits in this fungal order.

Materials and Methods

DNA and RNA extraction for genome sequencing

Seven *Lactarius* strains belonging to the section *Deliciosi*, namely *L. akahatsu* QP, *L. deliciosus* 48*, L. hatsudake* 109, *L. hengduanensis* 84, *L. pseudohatsudake* 88, *L. sanguifluus* B21 and *L. vividus* 141 were selected for genome sequencing (Supporting Information Table S1). The dikaryotic (diploid) mycelia were originally isolated from fresh fruiting bodies. To produce adequate material for DNA and RNA extraction, mycelial pieces were cultured for 4

to 6 weeks on solid ½ MMN + ½ PDA agar media covered with cellophane membranes at 23 °C in the dark (Wang *et al.*, 2019). Mycelia were harvested and snap frozen in liquid nitrogen and kept at - 80 °C until DNA and RNA extractions. High molecular weight genomic DNA was extracted from 2 g of mycelia following the Joint Genome Institute (JGI) genomic DNA extraction protocol (http://1000.fungalgenomes.org/home/wp-content/uploads/2013/02/genomicDNAProtocol-AK0511.pdf, accessed in 2017), and purified with the AMPure XP magnetic beads (Beckman Coulter, Cat.No A3881) according to the manufacturer's instructions. The quality of genomic DNA (size >23 kbp) was confirmed by pulsed field gel electrophoresis (PFGE). Mycelial total RNA was extracted using 100 mg of mycelium and the RNeasy Plant Mini Kit (Qiagen, Cat.No 74904) following the manufacturer's instructions. DNA and RNA samples were shipped to the JGI in DNAstable/RNAstable (Biomatrica) for library construction and sequencing.

Genome assembly and annotation

Genomic DNA of the *Lactarius* species was sequenced using PacBio platform, then assembled using the software Falcon v1.8.8 (Chin *et al.*, 2016) and annotated at JGI following standard pipelines (Grigoriev *et al.*, 2014, Supporting Information Methods S1). This dataset was supplemented with genomes and corresponding annotations of 24 additional Russulales, one Polyporales, one Phallales and one Geastrales (the latter three being used as the outgroup) downloaded from the JGI MycoCosm database (Supporting Information Table S1). As the DNA was extracted from diploid mycelium, the gene annotation was "haploidized" by using only the catalogue of primary alleles. The quality of all these genome assemblies and annotations was evaluated by Benchmarking Universal Single-Copy Orthologs (BUSCO, v.3.0.2) (Simão *et al.*, 2015) using the Basidiomycota set (busco.ezlab.org/datasets/basidiomycota_odb9.tar.gz).

Identification and annotation of transposable elements (TEs) was carried out as described by Payen *et al.* (2016) and Morin *et al.* (2019) using RepBase v.24.02 (Bao *et al.*, 2015). Functional annotations of Eukaryotic Orthologous Groups of Proteins (KOG), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) and InterPro (IPR) domains were performed using JGI pipelines and datasets are available on the genome portal for each species. Carbohydrate-active enzymes (CAZymes) were identified using the annotation pipeline described in Lombard *et al.* (2014) with the CAZy database (www.cazy.org) and subsequent manual curation by CAZyme team (version of December

2020). Secreted proteins were identified using the pipeline described by Pellegrin *et al.* (2015). Lectins were detected by *hmmscan* v3.3 using the Unilectin3D database (www.unilectin.eu, version of January 2020) (Lebreton *et al.*, 2021a). G protein-coupled receptor (GCPR) annotation was carried out as described by Lofgren *et al.* (2021). Candidate genes involved in the latex rubber biosynthesis were identified by BLASTp v2.10 searches (e-value <1E-5, query coverage >50%), using the protein homologs identified in the rubber tree (*Hevea brasiliensis*) as queries (Tang *et al.*, 2016; Liu *et al.*, 2020), based on the conservation of building unit (isopentenyl diphosphate, IPP) and biosynthetic pathway (Yamashita & Takahashi, 2020).

Peptidases from the subtilase family are composed of subtilisin (S8) and sedolisin (S53) families. Subtilases were initially identified in the MycoCosm gene repertoires by searching predicted proteins with one of the following annotations/keywords: S8, S53, PF00082, PF00089, PF09286, EC3.4.21.4 or EC3.4.14.9. Additional subtilases or subtilase-like proteins were further identified by BLASTp (evalue <1E-3) queries against the Russulales proteomes using the 904 putative functional subtilase identified by Li *et al.*, (2017), hereafter called reference subtilases. CLANS (Frickey & Lupas, 2004), a software allowing to visualize pair-wise sequence similarities, was then used to remove sequences that did not cluster with the reference subtilases and to assign the remaining ones to subtilase subfamilies. In order to keep only functional subtilase candidates, amino acid sequences of each subfamily were aligned using MUSCLE in MEGA X software (Kumar *et al.*, 2018) with default parameters. Subtilase sequences lacking two of the canonical regions were discarded of any further analysis; sequences lacking only one canonical region were annotated as partial. When the three regions matched the expected conserved subtilase pattern, the subtilase candidate was annotated as containing canonical regions. If at least one of the regions lack a perfect match to the known pattern, the subtilase sequence was annotated as containing non-canonical regions.

Protein orthology

The orthology among the 31 Russulales proteomes was assessed using OrthoFinder v2.3.3 (-M msa -S diamond -A mafft -I 1.5, Emms & Kelly, 2015). Based on this clustering, we determined the set of proteins shared by the 31 Russulales species (i.e., core genes/proteins), sets of proteins encoded in at least two genomes (i.e., dispensable genes/proteins) and sets of proteins unique to a genome (i.e., species-specific genes/proteins). For each protein set,

duplicated sequences were identified. Using the same clustering, the core, dispensable and species-specific genes/proteins of the nine *Lactarius* species were also identified. In addition, orthogroups containing proteins of all *Lactarius* species sharing a similar host tree, namely pine, oak and spruce were identified. *In silico* functional annotation was assigned to an orthogroup only if this annotation was present in at least half of the protein members of this orthogroup.

Phylogenomic analysis

The 934 single-copy gene orthogroups predicted with OrthoFinder were used for the phylogenomic analysis. Protein sequences of each orthogroup were aligned using MAFFT v7.471 (Yamada *et al.*, 2016). After removing the ambiguous regions (containing gaps and poorly aligned) with trimAl v1.4.rev15 (Capella-gutiérrez *et al.*, 2009), the resulting 934 alignments were concatenated into a super-alignment. ModelTest-NG v0.1.6 (Darriba *et al.*, 2020) was then used to identify the best protein substitution model for each partition of this super-alignment corresponding to an orthogroup. The species tree was then reconstructed from this super-alignment using RAxML-NG v.0.9.0 (Kozlov *et al.*, 2019) with partitions and 500 bootstrap replicates. The species tree was then calibrated on a time scale with MCMCtree available in PAML v4.8 (Yang, 2007), using three estimated time points identified by Varga *et al.* (2019), namely the divergence between *Heterobasidion annosum* and *Stereum hirsutum* 45 millions years ago (Mya), *Auriscalpium vulgare* and *Peniophora* sp. 93 Mya and *A. vulgare* and *Lentinellus vulpinus* 135 Mya. One calibrated tree per batch of 10 single copy genes was performed. The final tree was reconstructed based on the 50% median values obtained (mean values for branch length and extreme values for highest posterior density 95% confidence intervals). The obtained tree was plotted using MCMCtreeR v1.1 (Puttick, 2019).

Comparison of gene families between saprotrophic and EcM fungi

The protein orthogroups with different number of proteins between saprotrophs and EcM species or between *Lactarius* and other EcM species were identified with a BM test, using the R packages brunnermunzel v1.4.1 (Neubert & Brunner, 2007) and stats v4.0.1. Figures were displayed using the R packages ggplot2 and pheatmap v1.0.12 (Kolde, 2019). A PCA based on CAZymes families genes count was performed with the factoextra v1.0.7 package (Kassambara & Mundt, 2017). For this analyse, genes families with spearman correlation >0.8 (corrr v0.4.3 package, Kuhn *et al.*, 2020) were binned together.

Gains and losses in gene families

Expansion and contraction of *Lactarius* gene families were predicted with CAFÉ v.5 (Zenodo https://doi:10.5281/zenodo.3625141, as developed on GitHub). Singletons were removed from orthologs reconstructions. The previously identified species tree was pruned at the last common ancestor of *Lactarius* species with iTOL v5 (Letunic & Bork, 2019).

Gene tree reconstruction

The sedolisin gene tree was reconstructed from the protein sequences identified within Russulales (1951) and in outgroups (136, see Li *et al.*, 2017). They were aligned with MAFFT v7.471 and trimed with trimAL v1.4.rev15, which resulted in 124 sites. The best model, JTT+I+G4, identified with ModelTest-NG (v0.1.6) was used for the phylogeny reconstruction by RAxML-NG (v.0.9.0, Kozlov *et al.*, 2019). Similarly, the GH25 family tree was reconstructed based on the alignment of the 87 proteins (184 sites) and VT model. 500 bootstrap replicates were performed for the tree of GH25 genes.

Insertion age of LTR-retrotransposons

Full-length long terminal repeat (LTR) retrotransposons were identified in genome assemblies using LTRharvest with default parameters. This tool belongs to the GenomeTools genome analysis software (v1.5.10, Ellinghaus *et al.*, 2008). LTRs belonging to the *Gypsy* and *Copia* families were used for molecular dating of their genome invasion; selection was based on a BLASTx against Repbase v24.02 (Bao *et al.*, 2015). The 3'- and 5'-LTR nucleotide sequences were extracted and aligned with MAFFT v7.471. Alignments were used to calculate Kimura's 2P distances. The insertion age was determined using the formula $T = K / 2r$, with K being the distance between the two LTR sequences and r, the estimated substitution rate of 1.05×10^{-9} nucleotides per site per year for fungi (Dhillon *et al.*, 2014; Castanera *et al.*, 2016).

Repeat element-gene distance analysis

We statistically measured the mean repeat-gene distances with the first ten largest scaffolds by comparing the locations of observed genes and repeat elements and 10000 null hypothesis genome models made by randomly reshuffling the locations of genes. The probability (p-value) of mean repeat-gene distances was calculated with R package, regioneR v1.26.1 (Gel *et al.*, 2016). We calculated distances of all genes to the nearest repeat regions and examined significant differences among the fungi by performing Kruskal-Wallis with

Dunn's test using the R package agricolae v1.3-5 (De Mendiburu, 2014). The process was orchestrated with the visual omics pipeline, Syntey Governance Overview (SynGO; Looney *et al*., 2022).

Identification of differentially expressed genes in ectomycorrhizas

Data on differential gene expression in ectomycorrhizal roots were obtained from Tang *et al.*, (2021). In that study, RNA sequencing datasets were produced from the free-living mycelia and ectomycorrhizal roots of *L. akahatsu, L. deliciosus, L. sanguifluus* and *L. vividus*. Filtered RNAseq reads were mapped onto their corresponding *Lactarius* genomes, and differentially expressed genes (DEGs) were identified using DEseq2 v1.28.1 (Love *et al.*, 2014) by comparing normalized gene expression levels in transcriptomes from ectomycorrhizas and free-living mycelia. Genes with a $\log_2$(fold-change) >2 or <-2, and FDR p-value <0.05 were considered to be differentially expressed.

Results

***Lactarius* genome features and species tree phylogeny of Russulales**

The nuclear genomes of seven *Lactarius* strains, namely *L. akahatsu* QP, *L. deliciosus* 48, *L. hatsudake* 109, *L. hengduanensis* 84, *L. pseudohatsudake* 109, *L. sanguifluus* B21 and *L. vividus* 141, were sequenced, assembled and annotated at JGI and are available at the MycoCosm database (Grigoriev *et al.*, 2014). The quality and completeness of these genomes were confirmed by BUSCO analysis (Supporting Information Table S2). The size of the genome assemblies ranged from 62 to 100 Mb and contained 11612 to 20824 protein-coding genes (Fig. 1a, b). By including the published genomes from *L. quietus* (116 Mb, 18943 genes) (Miyauchi *et al.*, 2020) and *L. psammicola* (70 Mb, 13442 genes) (Looney *et al.*, 2022), we noticed a nearly two-fold variation in the genome size and gene content for *Lactarius* species. EcM species (n=19) displayed a significantly larger genome size and TE content than the saprotrophic species (n=12), and among EcM fungi, *Lactarius* species (n=9) presented a larger genome size and TE content than the others (*Russula*, *Lactifluus* and *Multifurca* species, n=10) (Fig. 1a). The gene content of *Lactarius* species is also higher compared to other EcM species (permuted BM test, p-value <0.01), but similar to saprotrophic species. Genome structural analysis (i.e., synteny) showed that no whole-genome duplication occurred in *Lactarius*. Instead, analysis of protein orthology indicated that the higher gene/protein content in *Lactarius* species is mainly due to duplication of

dispensable genes, while conserved- and species-specific genes are less prone to this duplication event (Fig. 1b).

The species tree phylogeny of the Russulales, reconstructed from an alignment of 934 single-copy orthologous genes, confirmed the monophyletic origin of *Lactarius* sect. *Deliciosi* after the earlier divergence from *L. quietus* and *L. psammicola* (Fig. 1c). *Lactifluus* and *Multifurca*, the two other genera producing milky latex, clustered with non-milk-cap *Russula* species, rather than with *Lactarius* species. Time calibration estimated the origin of the Russulales order at ~260 Mya, and common ancestor of EcM species at ~70 Mya, which is consistent with the recent estimation by Looney *et al*. (2022).

TE profiles and evolution within Russulales

Since TE accumulation accounts for the larger size of *Lactarius* genome assemblies, we further investigated the composition and evolution of these repeated elements, keeping in mind that a substantial proportion of TEs might have not been assembled owing to their high number of repetition. In this study, we identified more TE in EcM genomes than in saprotroph genomes (BM test, Bonferroni p-value <0.01, Fig. 2a). For instance, the *Harbinger* and *hAt* found in most EcM species, were absent in the Russulales saprotrophs. *Lactarius* also contains some TE categories, such as *Academ*, *Zisupton* and *Penelope* that were hardly found in other EcM species (Fig. 2a). Other TE such as Mariner, Gypsy and Copia were also largely expanded in EcM species (BM test, Bonferroni p-value <0.01). We estimated that the accumulation of the most abundant *Gypsy* and *Copia* LTRs started at ~70 Mya. The TE invasion coincided with the estimated origin of the symbiotic Russulales, while the massive LTR expansion in *Lactarius* species took place in the last 10 Mya after their speciation (Fig. 2b). We observed a striking heterogeneity in TE expansion rate among *Lactarius* species. For instance, *L. hengduanensis* presents a much lower TE expansion rate than the others species, a profile resembling the non-*Lactarius* EcM fungi (Fig. 2b).

***Lactarius* genomes encode expanded gene families coding for proteases**

The sequenced *Lactarius* genomes displayed the highest content in protease genes among Russulales species. This is in sharp contrast with other EcM Russulales species which display a reduced protease gene set compared to saprotrophic species. This enrichment in proteases is mainly associated to a drastic gene expansion of the sedolisin family (S53), one of the two subtilase families (PF09286, EC3.4.14.9) (Fig. 3a, Supporting Information Table S3).

Comparison of the sedolisin protein sequences indicated that most of the sedolisins in *Lactarius* species lack at least one of the three canonical sedolisin regions (Fig. 3a). Several sedolisin genes are clustered (tandem duplications) in the genome. Protein orthology analysis classified all Russulales sedolisins (1951) into 46 multiple-gene families and 123 singletons. Although nearly all these families (159) contained only *Lactarius* genes, they were not evolved newly in *Lactarius*, but expanded from a more ancestral sedolisin clade (Fig. 3b). Beside sedolisins, fungalysin family (M36) also expanded largely in *Lactarius* (13.1 copies, as compared to 1.4 copies in other EcM species, Supporting Information Table S3). However, fungalysin and cytophagalysin (M43B) genes are scarcely detected in *L. quietus*, the oak-associated species.

Sedolisins are rapidly evolving in Lactarius species

Gene family expansion and contraction analysis within *Lactarius* species identified 229 rapidly evolving gene families (Supporting Information Table S4). For each of them, significant expansion/contraction was observed on multiple nodes of the phylogenomic tree (Fig. 4). Noteworthy, seven, out of the eight families with annotations were sedolisins. As it could be related to the shift/switch of host specificity, we focused on three ancestral nodes: the closest ancestor of *L. quietus* and *L. psammicola*, the closest ancestor of *L. psammicola* and the species restricted to pines, and the closest ancestor of spruce associated species. Consistently, the sedolisin families were the gene families showing major expansions or contractions.

Lactarius sedolisin genes are co-localized with TEs

As transposable elements are known to duplicate genes through transposing activity, we examined associations between TEs and sedolisin-coding genes by estimating the distance of the genes to the nearest repeat elements. Indeed, the sedolisin genes were found to be significantly closer, with a mean distance of 2.5kb, to the repeats in *Lactarius* than in the rest of Russulales fungi (Kruskal-Wallis with Dunn's test, FDR p-value <0.05; Fig. 5a). Most of the co-localized repeats within a distance of 4.5 Kb, were unclassified categories (Fig. 5b).

Genes coding for secreted sedolisins are upregulated in host-specific symbioses

Transcript profiling using RNA-seq datasets from four compatible *Lactarius-Pinus* pairings revealed that nearly half of the transcripts coding for secreted sedolisins (S53) were strikingly induced during the host-specific interactions (Supporting Information Fig. S1). Importantly, the eight rapidly evolving sedolisin gene families were upregulated upon symbiosis.

Although TEs could influence the regulation of genes nearby, we did not detect neither significant proximity of these mycorrhiza-induced sedolisins to any TE category, compared with the non-induced ones (Supporting Information Fig. S2a), nor clear association between the regulation amplitude and distance to repeats (Pearson correlation coefficient with 95% confidence; Supporting Information Fig. S2b).

Secreted CAZymes

As expected from previous EcM genome analyses (Kohler et al. 2015; Miyauchi et al. 2020), the arsenal of enzymes involved in lignocellulose decomposition is strikingly reduced in EcM Russulales species compared to saprotrophic species (30 CAZyme families; BM test, FDR p-value <0.01; Fig. 6a; Supporting Information Table S5). The number of secreted genes containing the chitin-binding domain CBM5 is also reduced in EcM species. However, *Lactarius* species have retained a larger polysaccharide degrading potential than other EcM species, since they encode more genes acting on fungal glucan (GH16,GH17, GH152), chitin (GH20, CBM5, CBM50), plant cellobiose (AA3) and cellulose (GH3, GH131) (Supporting Information Table S5). Besides, secreted GH25, which acts on bacterial peptidoglycan, appeared to be expanded specifically in *Lactarius* species, especially in the two spruce-specific species (Fig. 6a; Supporting Information Fig. S3). These differences in secreted CAZymes clearly separate *Lactarius* from the other EcM fungi within Russulales (Fig. 6b).

Effector-like SSPs

Regarding effector-like SSPs, we found sixteen subgroups with known Pfam domains showing differential distribution either between saprotrophic and EcM fungi, or between *Lactarius* and other EcM species (BM test, FDR p-value <0.01; Table 1). In accordance with previous results, four of them belong to CAZymes including three acting on cellulose (CBM1, GH12 and AA9) depleted in ECM and one on bacterial peptidoglycan (GH25) specifically enriched in *Lactarius* species. Two domains (PF01476: LysM and PF01522: polysaccharide deacetylase) involving chitin binding and modification were also found to be significantly enriched in *Lactarius*. Another domain (PF00314: Thaumatin), possibly acting on the beta-1,3-glucans in fungal cell walls (Sakamoto *et al.*, 2006), was enriched in *Lactarius* species as well. In consistence with the overrepresentation of protease gene, we detected three protease-associated domains (PF09286, PF13582 and PF13688) that were enriched in *Lactarius* SSPs. However, it should be noticed that there is a clear difference for some domains among these *Lactarius* species. For instance, *L. vividus* and *L. hatsudake*

contain no pro-kumamolisin activation domain (PF09286), and the oak-specific *L. quietus* harbors the lowest number of SSPs containing LysM domain. When the distance between effector-like SSPs and TEs was investigated, these genes appeared significantly closer to TEs than other genes. However, those TEs were mainly unclassified.

Lectins

Given the potential role of lectins in determining host-specificity in several plant-fungus interactions, including ectomycorrhizal symbiosis (Guillot *et al.*, 1991; Giollant *et al.*, 1993; Varrot *et al.*, 2013), we surveyed the lectin gene distribution in Russulales. We found six lectin families with differential gene content between saprotrophic and EcM fungi, or between *Lactarius* and other EcM species within Russulales (BM test, Bonferroni p-value <0.01; Supporting Information Fig. S4). Among these, PVL-like family was only detected within *Lactarius* species and they restricted to species associating to pine and spruce hosts. The H-type lectin genes, rarely found in non-*Lactarius* genera, are mainly expanded in *Lactarius* species with six copies in *L. sanguifluus* to 18 copies in *L. psammicola*.

GPCRs

Given their high upregulation during EcM colonization in *Laccaria bicolor*, *Tuber melanosporum* and *Suillus* species (Voiblet *et al.*, 2001; Martin *et al.*, 2010; Plett *et al.*, 2012; Liao *et al.*, 2016), GPCRs were considered as candidates related to host-specificity or associated with EcM colonization more generally. In the present Russulales genome dataset, no specific expansion was detected in EcM species, with a mean of 14 ± 2 copies, and the host-specific genus *Lactarius* contained the lowest number of GPCR (BM test, p-value <0.01; Supporting Information Table S6). During ectomycorrhizal development involving *L. akahatsu, L. sanguifluus, L. deliciosus* or *L. vividus* with a compatible host, only one GPCR gene was significantly upregulated in *L. akahatsu* and another one downregulated in *L. deliciosus*.

Secondary metabolism pathways

Based on the possible relevance of secondary metabolites (SMs) in determining host-specificity in *Suillus* species (Lofgren *et al.*, 2021), we compared the repertoire of SM-related genes among Russulales species. Compared with other EcM fungi, *Lactarius* species harbor a higher number of terpene synthase (TPS) genes (BM test, bonferroni p-value=0.018; Supporting Information Fig. S5). However, the TPS gene content varies among *Lactarius* species (from nine copies in *L. quietus* to 20 copies *in L. pseudohatsudake*). Besides, TPS

genes were also enriched in the basal EcM fungus *Multifurca ochricompacta* (21 copies) and some of the most related saprotrophic species such as *Clavicorona pyxidate* and *Auriscalpium vulgare* (17 and 12 copies respectively). Among these genes, three were identified as upregulated during mycorrhiza formation: one in *L. sanguifluus* and two in *L. deliciosus*.

Biosynthesis of latex rubber

Latex production is a well-known feature of milk-cap fungi including species in *Lactarius*, *Lactifluus* and *Multifurca* genera. Considering its ecological importance, such as the resistance to fungivorous predation (Taskirawati & Tuno, 2016), genes potentially involved in fungal latex rubber biosynthesis were surveyed. Genes of the cytosolic mevalonate (MVA) pathway, rubber initiation and elongation, were found in all Russulales genomes (Supporting Information Fig. S6, Table S7). No genes coding for the plastidial methylerythritol phosphate (MEP) pathway were found in these fungi. We observed a slight enrichment of latex biosynthesis genes in *Lactarius* spp. compared to other Russulales species (BM test, bonferroni p-value=0.034).

Discussion

The shift from the saprophytic to symbiotic lifestyle of ancestral Russulales species took place at ~70 Mya, during the third wave of plant root diversification (Strullu-Derrien *et al.*, 2018). It has been suggested that this event was linked to a global climate change, as well as an increase in potential habitats and soil complexity, which presumably resulted in a competitive advantage for more specialized root types. In association to this root diversification, multiple saprophytic fungi in various fungal lineages shifted to an EcM lifestyle (Looney *et al.*, 2018). As a result of convergent evolution, EcM lineages in most fungal orders share similar genomic features, including a larger genome size resulting from TE proliferation, a restricted set of PCWDEs and a specific suite of effector-like SSPs (Kohler *et al.*, 2015; Miyauchi *et al.*, 2020; Lebreton *et al.*, 2021b). These convergent sets of genetic traits are the hallmarks of the EcM lifestyle and they are shared by the symbiotic Russulales. However, we found a series of idiosyncrasies that distinguish *Lactarius* species from other EcM lineages as discussed below.

Divergent evolution of the symbiotic lifestyle within Russulaceae

Despite descending from a single lineage in Russulaceae, *Lactarius* and *Russula* species display divergent genomic traits that may impact the development and functioning of their

ectomycorrhizal associations. The large expansion of sedolisin proteases is unique to *Lactarius* species and was not reported in other EcM fungal lineages sequenced so far (Kohler *et al.*, 2015; Peter *et al.*, 2016; Murat *et al.*, 2018; Miyauchi *et al.*, 2020; Lofgren *et al.*, 2021). Their co-localization with TEs suggests that the large expansion was probably caused by the recent TE proliferation occurred in the last 10 Mya. These proteases may play a role in releasing organic N nutrients (i.e., amino acids or oligopeptides) from soil organic matter (SOM) via protein cleavage. However, their extensive induction during EcM symbioses suggests that these sedolisins are more likely to be involved in the interaction with host plants (Tang *et al.*, 2021). Secreted proteases in several plant pathogens could dampen the host defense reactions via cleaving immunity-related proteins, such as chitinases, secreted by the host roots (Naumann *et al.*, 2011; Jashni *et al.*, 2015; Sanz-Martín *et al.*, 2016; Ökmen *et al.*, 2018). This protease-based strategy is supported by our finding that several other proteases, such as fungalysins were also strongly induced during the symbiosis (Tang *et al.*, 2021).

Within the Russulaceae family, *Russula* species are known for their broad range of hosts, i.e., most of them are known as host generalists. Their species diversification has been linked to frequent host switching between angiosperms and Pinaceae with subsequent host expansion (Looney *et al.*, 2016). On the contrary, many *Lactarius* species, such as the ones in the section of *Deliciosi*, have long been considered as host specialists (Nuytinck *et al.*, 2007; Verbeken & Nuytinck, 2013; Wang *et al.*, 2019). This divergent host selection provides a unique opportunity to explore the molecular determinants involved in host-specificity. In pathogenic fungi, a restricted host range is often accompanied by gene losses (Spanu *et al.*, 2010; Baroncelli *et al.*, 2016). However, a recent study comparing the gene repertoires of host-specific species in *Suillus* (Boletales) and other less host-specific fungal symbionts reported no significant gene loss, but suggested that secondary metabolites synthesized by terpene- and nonribosomal polyketide synthases (NRPS) may play a role in the host-specificity determination (Lofgren *et al.*, 2021). Interestingly, we also found a slight enrichment of terpene synthase (TPS) genes, but not of NRPS genes in *Lactarius* species. Strikingly, a dramatic expansion of sedolisin proteases was observed in *Lactarius.* Moreover, our analysis of the expansion and contraction of gene families indicated that several sedolisin gene families were rapidly evolving in multiple phylogenetic nodes where host switches likely occurred. This evidence, together with their unique regulation in various EcM symbioses (Tang *et al.*, 2021), support an important role of sedolisins in the host-specificity

of *Lactarius* ectomycorrhizal associations. In addition, other protein categories showing a significant enrichment in *Lactarius* species or a unique symbiotic regulation, such as the LysM-domain-containing SSPs and lectins may also be involved in the interaction with specific host species, owing to their biochemical role in ligand-binding mechanisms (Guillot *et al.*, 1991; Giollant *et al.*, 1993; Kombrink & Thomma, 2013; Labbé *et al.*, 2019; Bozsoki *et al.*, 2020).

Heterogeneity among *Lactarius* genomes

We found a substantial heterogeneity in genome size and gene composition among the sequenced *Lactarius* species, even though they belong to a single section. A nearly two-fold variation in genome size and gene content was observed among *Lactarius* species, which contrasts to the homogeneity of *Russula* genomes (Fig. 1a, b). Since comparable BUSCO completeness was reported for all these genomes, this variation is not related to differential quality scores in genome assemblies or gene annotations among *Lactarius* species. The absence of whole-genome duplication indicates that this variation mainly results from duplication of specific gene families. Indeed, the protein orthology analysis revealed that the oak-specific species *L. quietus* and the other four Pinaceae-specific species (*L. sanguifluus, L. deliciosus, L. hatsudake* and *L. pseudohatsudake*) present higher rates of duplications of dispensable genes than the others; *L. quietus* itself having a higher content of species-specific genes than the others. The smaller genome of *L. psammicola, L. vividus* and *L. akahatsu* can also be explained by large gene reductions (Fig. 1). There is also a large difference in the genome size and gene content among the species associated with a single genus of hosts (i.e., *Pinus* or *Picea*). Interestingly, a large variation within a single EcM genus has also been observed in *Suillus* and *Amanita*, the latter in which both EcM and non-EcM species have evolved (Hess *et al.*, 2018; Lofgren *et al.*, 2021). However, unlike the large amplification of species-specific gene families in *Amanita*, *Lactarius* presents more duplication of dispensable families shared by at least two species whereas its specific families have a very limited amplification. This difference highlights the diversity of genome evolution in different EcM fungal lineages. The high heterogeneity found in both specialistic lineages, i.e., *Suillus* and *Lactarius*, may on the other hand suggest an important role of host specialization in shaping EcM fungal genomes, as observed frequently in plant pathogens (Vries *et al.*, 2020).

Concluding remarks

To better understand the evolution of EcM symbiotic lifestyle and host-specificity, we sequenced several milk-cap fungal species and performed genomic comparisons with their ancestral saprotrophs and symbiotic sister genera (*Russula*, *Lactifluus* and *Multifurca*) within Russulaceae. *Lactarius* species encode significant larger genomes than the other clades, as a result of TE proliferation. They also convergently lost PCWDEs, but retained a number of CAZymes acting on microbial cell wall components, especially the bacterial peptidoglycan. Most remarkably, *Lactarius* harbors a drastically expanded sedolisin protease family, a feature absent from any other EcM fungal lineages sequenced so far, including its sister genera within the same family. The expansion and rapid evolution of sedolisin genes, together with their extensive symbiotic upregulation thus strongly suggest that milk-cap fungi use a protease-based toolkit to dialogue with their specific host species, a strategy adopted by some plant pathogens yet not reported in plant symbionts. On-going functional analysis of symbiosis-induced sedolisins will provide the needed information on the substrate of these proteases and their role in EcM development. Besides, other gene products with known high ligand-binding specificity may also play a role in the host specialization. Meanwhile, this long-term host specialization/adaptation may have in turn reshaped fungal genomes, causing large interspecific difference in their size and gene repertoire. Taken together, this study casts a new light on the evolution of EcM lifestyle and highlights an important role of secreted proteases in host-specific *Lactarius* symbioses. The uniqueness of *Lactarius* revealed here thus warrants diverse lineages to be investigated in the future for a full view of ectomycorrhizal evolution.

Acknowledgments

This work was supported by Chinese Scholarship Council (CSC) (to NT); Yunnan Key Program of Science and Technology (202102AE090034) (to FY); Laboratory of Excellence ARBRE (ANR-11-LABX-0002-01), the Region Lorraine and the European Regional Development Fund (to F.M.M.); Beijing Advanced Innovation Center funding for Tree Breeding by Molecular Design, Beijing Forestry University (to AL, YD and FMM). This work was also funded by the U.S. Department of Energy, Joint Genome Institute, a DOE Office of Science User Facility, and supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 within the framework of the Mycorrhizal Genomics Initiative (10.46936/10.25585/60000842), Metatranscriptomics of Forest Soil Ecosystems project (10.46936/10.25585/60000986) and the 1000 Fungal Genome

project (10.46936/10.25585/60001060). We would like to thank Xianghua Wang for her assistance in the identification of some *Lactarius* isolates, and Annegret Kohler, Emmanuelle Morin, François Bonnardel, Anne Imberty for helpful discussions and suggestions. Joey Spatafora, Jon Magnusson, Daniel Lindner, Laszlo Nagy and Pedro Crous are acknowledged for the access to unpublished JGI genomes.

Author Contribution

FMM conceived and coordinates the Mycorrhizal Genomics Initiative. NT, FY, AGL and FMM designed the present project. ALe, NT and FMM wrote the manuscript with input from FY and YD. RW and DM isolated and identified the fungi. NT extracted the high-quality DNAs and RNAs. IVG coordinated genome sequencing and annotation at JGI. AK, KL, BA, KB, AC, ALi and VN performed transcriptome sequencing, assembly and gene annotation at JGI. ED and BH performed CAZyme annotations. ALe, NT and SM performed comparative genome analyses. ALe and NT contributed equally to this work.

Data Availability

Genome assemblies and gene annotations used in this study are available via the JGI fungal genome portal MycoCosm (see the Russulales page: https://mycocosm.jgi.doe.gov/Russulales/Russulales.info.html) and NCBI Genome database under the BioProject of PRJNA500114 to PRJNA500118, PRJNA500120 and PRJNA500123 (Accession No. JAKELG000000000, JAKELH000000000, JAKELI000000000, JAKELK000000000, JAKELL000000000, JAKEYE000000000 and JAKEYF000000000). RNA-seq read data are available at the NCBI Sequence Read Archive (SRA) under the BioProject of PRJNA706172. All other data supporting the findings of this study are included within the article and its additional files.

References

Bao W, Kojima KK, Kohany O. **2015**. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**: 11.

Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, Le Floch G, Harrison RJ, Holub E, Sukno SA, Sreenivasaprasad S, *et al.* 2016. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC genomics* **17**: 555.

Bozsoki Z, Gysel K, Hansen SB, Lironi D, Krönauer C, Feng F, de Jong N, Vinther M, Kamble M, Thygesen MB, *et al.* 2020. Ligand-recognizing motifs in plant LysM receptors are major determinants of specificity. *Science* **369**: 663–670.

Brundrett MC, Tedersoo L. 2018. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist* **220**: 1108–1115.

Capella-gutiérrez S, Silla-martínez JM, Gabaldón T. 2009. trimAl : a tool for automated alignment trimming in large-scale phylogenetic analyses. **25**: 1972–1973.

Castanera R, Lopez-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, Grimwood J, Perez G, Pisabarro AG, Grigoriev I V, *et al.* 2016. Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. *PLoS genetics* **12**: e1006108.

Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, *et al.* 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nature methods* **13**: 1050–1054.

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution* **37**: 291–294.

Dhillon B, Gill N, Hamelin RC, Goodwin SB. 2014. The landscape of transposable elements in the finished genome of the fungal wheat pathogen *Mycosphaerella graminicola*. *BMC genomics* **15**: 1132.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC bioinformatics* **9**: 18.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**: 157.

Frickey T, Lupas A. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics (Oxford, England)* **20**: 3702–3704.

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.

Giollant M, Guillot J, Damez M, Dusser M, Didier P, Didier É. 1993. Characterization of a lectin from *Lactarius deterrimus*. *Plant Physiology* **101**: 513–522.

Grigoriev I V, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, *et al.* 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic acids research* **42**: D699–D704.

Guillot J, Giollant M, Damez M, Dusser M. **1991**. Isolation and characterization of a lectin from the mushroom, *Lactarius deliciosus*. *Journal of Biochemistry* **109**: 840–845.

Hess J, Skrede I, De Mares MC, Hainaut M, Henrissat 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.

Jashni MK, Dols IHM, Iida Y, Boeren S, Beenen HG, Mehrabi R. **2015**. Synergistic action of a metalloprotease and a serine protease from *Fusarium oxysporum* cleaves chitin-binding tomato chitinases. *Molecular Plant Microbes Interaction* **28**: 996–1008.

Kassambara A, Mundt F. **2017**. Factoextra: extract and visualize the results of multivariate data analyses. *R package version* **1**: 337–354.

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**: 410–415.

Kolde R. **2019**. pheatmap: Pretty Heatmaps. R package version 1.0.12. *R Packag. version 1.0* **8**.

Kombrink A, Thomma BPHJ. **2013**. LysM Effectors: Secreted Proteins Supporting Fungal Life. *PLOS Pathogens* **9**: e1003769.

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 (Oxford, England)* **35**: 4453–4455.

Kuhn M, Jackson S, Cimentada J. **2020**. corrr: Correlations in R. *R Package version 0.4* **2**.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. **2018**. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* **35**: 1547–1549.

Labbé J, Muchero W, Czarnecki O, Wang J, Wang X, Bryan AC, Zheng K, Yang Y, Xie M, Zhang J, *et al.* 2019. Mediation of plant–mycorrhizal interaction by a lectin receptor-like kinase. *Nature Plants* **5**: 676–680.

Lebreton A, Bonnardel F, Dai Y-C, Imberty A, Martin FM, Lisacek F. **2021a**. A Comprehensive Phylogenetic and Bioinformatics Survey of Lectins in the Fungal Kingdom. *Journal of Fungi* **7**.

Lebreton A, Zeng Q, Miyauchi S, Kohler A, Dai Y-C, Martin FM. **2021b**. Evolution of the Mode of Nutrition in Symbiotic and Saprotrophic Fungi in Forest Ecosystems. *Annual Review of Ecology, Evolution and Systematics* **52**.

Letunic I, Bork P. **2019**. Interactive Tree Of Life (iTOL) v4: recent updates and new

developments. *Nucleic Acids Research* **47**: W256–W259.

Li J, Gu F, Wu R, Yang J, Zhang K-Q. **2017**. Phylogenomic evolutionary surveys of subtilase superfamily genes in fungi. *Scientific Reports* **7**: 45456.

Liao HL, Chen Y, Vilgalys R. **2016**. Metatranscriptomic Study of Common and Host-Specific Patterns of Gene Expression between Pines and Their Symbiotic Ectomycorrhizal Fungi in the Genus *Suillus*. *PLoS Genetics* **12**: 1–24.

Liu J, Shi C, Shi CC, Li W, Zhang QJ, Zhang Y, Li K, Lu HF, Shi C, Zhu ST, *et al.* 2020. The Chromosome-Based Rubber Tree Genome Provides New Insights into Spurge Genome Evolution and Rubber Biosynthesis. *Molecular Plant* **13**: 336–350.

Lofgren LA, Nguyen NH, Vilgalys R, Ruytinx J, Liao H, 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.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. **2014**. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**: 490–495.

Looney BP, Meidl P, Piatek MJ, Miettinen O, Martin FM, Matheny PB, Labbé JL. **2018**. Russulaceae: a new genomic dataset to study ecosystem function and evolutionary diversification of ectomycorrhizal fungi with their tree associates. *New Phytologist* **218**: 54–65.

Looney B, Miyauchi S, Morin E, Drula E, Courty PE, Kohler A, Kuo A, LaButti K, Pangilinan J, Lipzen A, *et al.* 2022. Evolutionary transition to the ectomycorrhizal habit in the genomes of a hyperdiverse lineage of mushroom-forming fungi. *New Phytologist* **233**: 2294–2309.

Looney BP, Ryberg M, Hampe F, Sánchez-García M, Matheny PB. **2016**. Into and out of the tropics: Global diversification patterns in a hyperdiverse clade of ectomycorrhizal fungi. *Molecular Ecology* **25**: 630–647.

Love MI, Huber W, Anders S. **2014**. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**: 550.

Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, *et al.* 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033–1038.

Martin F, Kohler A, Murat C, Veneault-Fourrey C, Hibbett DS. **2016**. Unearthing the roots of ectomycorrhizal symbioses. *Nature Reviews Microbiology* **14**: 760–773.

De Mendiburu F. **2014**. Agricolae: statistical procedures for agricultural research. *R package version* **1**: 1–4.

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**: 5125.

Morin E, Miyauchi S, San Clemente H, Chen ECH, Pelin A, de la Providencia I, Ndikumana S, Beaudet D, Hainaut M, Drula E, *et al.* 2019. Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. *New Phytologist* **222**: 1584–1598.

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 and Evolution* **2**: 1956–1965.

Naumann TA, Wicklow DT, Price NPJ. 2011. Identification of a chitinase-modifying protein from *Fsarium verticillioides*: Truncation of a host resistance protein by a fungalysin metalloprotease. *Journal of Biological Chemistry* **286**: 35358–35366.

Neubert K, Brunner E. 2007. A studentized permutation test for the non-parametric Behrens–Fisher problem. *Computational Statistics & Data Analysis* **51**: 5192–5204.

Nuytinck J, Verbeken A, Miller SL. 2007. Worldwide phylogeny of *Lactarius* section Deliciosi inferred from ITS and glyceraldehyde-3-phosphate dehydrogenase gene sequences. *Mycologia* **99**: 820–832.

Ökmen B, Kemmerich B, Hilbig D, Wemhöner R, Aschenbroich J, Perrar A, Huesgen PF, Schipper K, Doehlemann G. 2018. Dual function of a secreted fungalysin metalloprotease in *Ustilago maydis*. *New Phytologist* **220**: 249–261.

Payen T, Murat C, Martin F. 2016. Reconstructing the evolutionary history of gypsy retrotransposons in the Périgord black truffle (*Tuber melanosporum* Vittad.). *Mycorrhiza* **26**: 553–563.

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**.

Peter M, Kohler A, Ohm RA, Kuo A, Krützmann J, Morin E, Arend M, Barry KW, Binder M, Choi C, *et al.* 2016. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nature Communications* **7**: 12662.

Plett JM, Kohler A, Martin F. 2012. 6 De-Constructing a Mutualist: How the Molecular Blueprints of Model Symbiotic Fungi Are Changing Our Understanding of Mutualism. In: Fungal Associations. Berlin, Heidelberg: Springer Berlin Heidelberg, 93–117.

Puttick MN. **2019**. MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* **35**: 5321–5322.

Sakamoto Y, Watanabe H, Nagai M, Nakade K, Takahashi M, Sato T. **2006**. Lentinula edodes tlg1 encodes a thaumatin-like protein that is involved in lentinan degradation and fruiting body senescence. *Plant Physiology* **141**: 793–801.

Sanz-Martín JM, Pacheco-Arjona JR, Bello-Rico V, Vargas WA, Monod M, Díaz-Mínguez JM, Thon MR, Sukno SA. **2016**. A highly conserved metalloprotease effector enhances virulence in the maize anthracnose fungus *Colletotrichum graminicola*. *Molecular plant pathology* **17**: 1048–1062.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. **2015**. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, Ver Loren van Themaat E, Brown JKM, Butcher SA, Gurr SJ, *et al.* 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science (New York, N.Y.)* **330**: 1543–1546.

Steidinger BS, Crowther TW, Liang J, Van Nuland ME, Werner GDA, Reich PB, Nabuurs G, de-Miguel S, Zhou M, Picard N, *et al.* 2019. Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* **569**: 404–408.

Strullu-Derrien C, Selosse MA, Kenrick P, Martin FM. **2018**. The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New Phytologist* **220**: 1012–1030.

Tang N, Lebreton A, Xu W, Dai Y, Yu F, Martin FM. **2021**. Transcriptome Profiling Reveals Differential Gene Expression of Secreted Proteases and Highly Specific Gene Repertoires Involved in *Lactarius–Pinus* Symbioses. *Frontiers in Plant Science* **12**: 1775.

Tang C, Yang M, Fang Y, Luo Y, Gao S, Xiao X, An Z, Zhou B, Zhang B, Tan X, *et al.* 2016. The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature Plants* **2**: 16073.

Taskirawati I, Tuno N. **2016**. Fungal defense against mycophagy in milk caps. *Sci. Rep. Kanazawa Univ* **60**: 1–10.

Tedersoo L, May TW, Smith ME. **2010**. Ectomycorrhizal lifestyle in fungi: Global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza* **20**: 217–263.

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

741 of mushroom evolution. *Nature Ecology and Evolution* **3**: 668–678.
742 **Varrot A, Basheer SM, Imberty A**. **2013**. Fungal lectins: Structure, function and potential
743 applications. *Current Opinion in Structural Biology* **23**: 678–685.
744 **Verbeken A, Nuytinck J**. **2013**. Not every milkcap is a Lactarius. *Scripta Botanica Belgica*
745 **51**: 162–168.
746 **Voiblet C, Duplessis S, Encelot N, Martin F**. **2001**. Identification of symbiosis-regulated
747 genes in *Eucalyptus globulus* - *Pisolithus tinctorius* ectomycorrhiza by differential
748 hybridization of arrayed cDNAs. *Plant Journal* **25**: 181–191.
749 **Vries S, Stukenbrock EH, Rose LE**. **2020**. Rapid evolution in plant–microbe interactions –
750 an evolutionary genomics perspective. *New Phytologist* **226**: 1256–1262.
751 **Wang R, Guerin-Laguette A, Huang L-L, Wang X-H, Butler R, Wang Y, Yu F-Q**. **2019**.
752 Mycorrhizal syntheses between *Lactarius* spp. section Deliciosi and *Pinus* spp. and the effects
753 of grazing insects in Yunnan, China. *Canadian Journal of Forest Research* **49**: 616–627.
754 **Yamada KD, Tomii K, Katoh K**. **2016**. Application of the MAFFT sequence alignment
755 program to large data—reexamination of the usefulness of chained guide trees. *Bioinformatics*
756 **32**: 3246–3251.
757 **Yamashita S, Takahashi S**. **2020**. Molecular Mechanisms of Natural Rubber Biosynthesis.
758 *Annual Review of Biochemistry* **89**: 1–31.
759 **Yang Z**. **2007**. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology*
760 *and evolution* **24**: 1586–1591.
761
762

Figure legends:

Fig. 1 Genome and phylogeny of Russulales.

(a) Size of genome assemblies and proportion of TE in the assembly in Russulales; (b) Conserved, dispensable and species-specific genes in Russulales. Counts of duplicated protein sequences are also shown; (c) Species tree phylogeny of Russulales calibrated on a time scale (Mya). Star indicating the transition from saprotrophic to EcM lifestyle.

Fig. 2 TE composition and evolution in Russulales.

(a) Genome coverage of TE categories annotated in the Russulales genomes. *: BM test significance, bonferroni p-value <0.01 between EcM and saprotrophs or *Lactarius* spp. and other EcM species; (b) Estimated ages of Copia and Gypsy LTRs. TE counts per age were binned by 2 Mya.

Fig. 3 Sedolisin gene content and evolution in Russulales.

(a) Sedolisin (S53) gene content in Russulales genomes. Sedolisins missing one of their three catalytic regions were labelled as partial; the sedolisin was labelled as containing a non-canonical region if at least one of the catalytic region lack a perfect match to the pattern described in the literature; (b) Phylogeny of sedolisins in Russulales and outgroup species.

Fig. 4 Expansion and contraction of gene families in *Lactarius* species.

The number of gene families are displayed on the nodes of RAxML species tree with expanding gene families in blue and contracting gene families in red.

Fig. 5 Co-localization of sedolisin genes with TEs in *Lactarius*.

(a) Sedolisin gene-transposable element (TE) distances in Russulales species. Gene-TE distances were plotted for each species and comparisons were performed among all species. Significant differences (Kruskal-Wallis with Dunn's test, p-value <0.05) were indicated by the letters on the right side of each species; (b) Number of each TE category found within a distance of 4.5 Kb, to sedolisin genes.

Fig. 6 Differential distribution of secreted CAZyme genes among Russulales fungi.

The number of genes coding for secreted CAZymes was compared among Russulales fungi (*Lactarius*, other EcM and saprotroph). (a) Secreted CAZyme categories, grouped by their potential substrates, showing differential distributions either between saprotrophic and EcM

fungi, or between *Lactarius* and other EcM species were shown (BM test, FDR p-value <0.01, detailed in Table S5). (b) Principle component analysis (PCA) of secreted CAZyme genes showing differential distributions among the groups of interest. For display purpose, genes families with a spearman correlation >0.8 were bined together.

Table 1 SSP domains differing in abundance among Russulales species.

Pfam domains contained in SSPs showing differential distributions either between saprotrophic and EcM fungi, or between *Lactarius* and other EcM species were listed (BM test, FDR p-value <0.01).