

Evolutionary transition to the ectomycorrhizal habit in the genomes of a hyperdiverse lineage of mushroom-forming fungi

Brian Looney^{1*} , Shingo Miyauchi^{2*} , Emmanuelle Morin², Elodie Drula^{3,4} , Pierre Emmanuel Courty⁵ , Annegret Kohler² , Alan Kuo⁶ , Kurt LaButti⁶ , Jasmyn Pangilinan⁶, Anna Lipzen⁶, Robert Riley⁶ , William Andreopoulos⁶, Guifen He⁶, Jenifer Johnson⁶, Matt Nolan⁶, Andrew Tritt⁶, Kerrie W. Barry⁶ , Igor V. Grigoriev^{6,7} , László G. Nagy^{8,9} , David Hibbett¹ , Bernard Henrissat^{3,10} , P. Brandon Matheny¹¹, Jesse Labbé^{12†}  and Francis M. Martin^{2,13†} 

¹Department of Biology, Clark University, Worcester, MA 01610, USA; ²UMR Interactions Arbres/Microorganismes, Centre INRAE Grand Est-Nancy, INRAE, Université de Lorraine, Champenoux 54000, France; ³Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Univ., Marseille 13009, France; ⁴USC1408 Architecture et Fonction des Macromolécules Biologiques (AFMB), INRAE, Marseille 13009, France; ⁵Agroécologie, AgroSup Dijon, CNRS, INRAE, Université de Bourgogne, Université de Bourgogne Franche-Comté, Dijon 25000, France; ⁶Lawrence Berkeley National Laboratory, US Department of Energy Joint Genome Institute, Berkeley, CA 94720, USA; ⁷Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA 94720, USA; ⁸Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre, Szeged 6726, Hungary; ⁹Department of Plant Anatomy, Institute of Biology, Eötvös Loránd University, Budapest 1053, Hungary; ¹⁰Department of Biological Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia; ¹¹Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA; ¹²Biosciences Division, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN 37830, USA; ¹³Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing 100083, China

Summary

Authors for correspondence:

Brian Looney

Email: blooney@clarku.edu

Francis M. Martin

Email: francis.martin@inra.fr

Received: 16 March 2021

Accepted: 24 November 2021

New Phytologist (2022) 233: 2294–2309

doi: 10.1111/nph.17892

Key words: ectomycorrhizal habit, evolutionary transition, Russulaceae, Russulales, secondary metabolism cluster, synteny, transposable elements.

- The ectomycorrhizal (ECM) symbiosis has independently evolved from diverse types of saprotrophic ancestors. In this study, we seek to identify genomic signatures of the transition to the ECM habit within the hyperdiverse Russulaceae.
- We present comparative analyses of the genomic architecture and the total and secreted gene repertoires of 18 species across the order Russulales, of which 13 are newly sequenced, including a representative of a saprotrophic member of Russulaceae, *Gloeopeniophorella convolvens*.
- The genomes of ECM Russulaceae are characterized by a loss of genes for plant cell wall-degrading enzymes (PCWDEs), an expansion of genome size through increased transposable element (TE) content, a reduction in secondary metabolism clusters, and an association of small secreted proteins (SSPs) with TE ‘nests’, or dense aggregations of TEs. Some PCWDEs have been retained or even expanded, mostly in a species-specific manner. The genome of *G. convolvens* possesses some characteristics of ECM genomes (e.g. loss of some PCWDEs, TE expansion, reduction in secondary metabolism clusters).
- Functional specialization in ECM decomposition may drive diversification. Accelerated gene evolution predates the evolution of the ECM habit, indicating that changes in genome architecture and gene content may be necessary to prime the evolutionary switch.

Introduction

Fungi fulfill diverse and essential functional roles in facilitating ecosystem viability at a multitude of scales, and these roles are directly mediated by their gene content and evolutionary history. Current understandings of functional roles of fungi are closely linked to their nutritional uptake mode because fungi must live in close proximity to nutrient sources for

absorption. Fungal strategies for nutrient acquisition are dynamic across the fungal tree of life. Seemingly redundant trophic strategies have independently evolved numerous times. Within a single order, Sebaciales, we see multiple origins of plant-associated roles including endophytism, ectomycorrhizas, orchid mycorrhizas, ericoid mycorrhizas and liverwort symbiosis, all derived from a saprotrophic ancestry (Weiß *et al.*, 2016). This versatility of nutritional modes, with frequent transitions during evolution can be seen throughout the Agaricomycetes (Hibbett, 2006). Molecular traits contributing to these plant-associated functional roles are largely unexplored,

*These authors are joint first authors.

†These authors are joint senior authors.

especially in ectomycorrhizal (ECM) fungi (Koide *et al.*, 2014; Martin *et al.*, 2016).

The ECM symbiosis is functionally characterized by the transfer of water and nutrients to the plant and photoassimilates to the fungus through a cell-to-cell interface within roots, called the Hartig net (Smith & Read, 2010). The ECM habit has independently evolved in up to 82 lineages of fungi in Endogonomycetes, Pezizomycetes, and Agaricomycetes (Tedersoo & Smith, 2017). In plants, the ability to harbor these fungi have evolved in at least 30 lineages in Gymnospermae and Angiospermae (Brundrett & Tedersoo, 2018). The ECM habit in fungi has evolved from diverse ancestral trophic states, including white-rot saprotrophs, brown-rot saprotrophs, litter decomposers and root endophytes, with each evolutionary history necessitating different selective pressures (Tedersoo & Smith, 2013; Martin *et al.*, 2016; Pellitier & Zak, 2018; Strullu-Derrien *et al.*, 2018; Miyauchi *et al.*, 2020). These evolutionary shifts in trophic strategy have often led to specializations of function that contribute to rapid diversification that are defining for clades of plant-associated fungi (Sánchez-García & Matheny, 2017; Lutzoni *et al.*, 2018; Sánchez-García *et al.*, 2020).

The characterization of the *Laccaria bicolor* genome established a number of attributes that are characteristic for the genome of an ECM fungus, such as high transposable element (TE) content, loss of plant cell wall-degrading enzymes (PCWDEs), and the occurrence of effector-like mycorrhiza-induced small secreted proteins (MiSSPs) that are upregulated during symbiosis (Martin *et al.*, 2008; Labbé *et al.*, 2012; Pellegrin *et al.*, 2015; Plett *et al.*, 2017). These genomic features have since been found in a wide array of mycorrhizal fungi belonging to Ascomycota, Basidiomycota and Mucoromycota (Kohler *et al.*, 2015; Martin *et al.*, 2016; Morin *et al.*, 2019; Miyauchi *et al.*, 2020). To establish the evolutionary events defining the origin(s) of ECM associations more precisely and to discriminate these from lineage-specific evolutionary changes, comparative genomic analyses of densely sampled evolutionary lineages of ECM fungi, all descended from a single origin of symbiosis, are needed. The evolution of ECM genomes within a single lineage has been investigated for the Amanitaceae, which showed a rapid expansion and contraction of functionally relevant genes early in the evolution of the ECM habit (Hess *et al.*, 2018). Another study focusing on the Pinaceae-associated genus *Suillus* found dynamic gene evolution within the lineage, with secondary metabolites linked to host specificity (Lofgren *et al.*, 2021).

Russulaceae contains an iconic lineage of ECM fungi that are dominant in ectotrophic landscapes and are prized for their edible mushrooms (Looney *et al.*, 2018). Russulaceae possesses several ecologically relevant attributes that warrant study in a genomic context, such as a nitrophilic tendency of some members (Lilleskov *et al.*, 2002), the production of unique sesquiterpenoid secondary compounds (Clericuzio *et al.*, 2012), and an accelerated evolutionary rate of speciation, morphological transition and host expansion (Looney *et al.*, 2016). The majority of species in Russulaceae are ECM, forming a single clade. However, the residual group of Russulaceae are made up of wood-decaying, crust fungi (Larsson & Larsson, 2003). Thus, Russulaceae

provides an exceptional opportunity to study an evolutionary transition between nutritional modes. Here, we describe trends in genomic architecture and gene content in 18 representative species of Russulales. Our dataset contains 13 previously unanalyzed genomes, including eight species of ECM Russulaceae and their saprotrophic sister group (*Gloeopeniophorella convolvens*), as well as a number of mainly saprotrophic members of Russulales, including *Amylostereum chailletii*, *Artomyces pyxidata*, *Auriscalpium vulgare* and *Vararia minispora*. Our analysis elucidates patterns of functional diversity that have evolved within the ECM symbiotrophs, including evolution of PCWDEs, retention of genes to scavenge nitrogen compounds in soil organic matter, secondary metabolism, and TE invasion favoring duplication of species-specific genes. We hypothesize that a defined core set of genes derived from the common ancestor of ECM Russulaceae defines a particular niche for this lineage according to the 'family gene conservation' hypothesis (Looney *et al.*, 2018).

Materials and Methods

Taxon sampling and nucleic acid extraction

Newly sequenced genomes and transcriptomes were derived from phylogenetically distinct lineages within the family Russulaceae. Representative species were sampled as mushroom sporocarps from forested habitat in the Great Smoky Mountains National Park and surrounding areas. To retrieve high-molecular-weight DNA and undegraded RNA, the inner flesh of the sporocarps was extracted in the field using a sterilized scalpel and placed in a 50 mg Falcon tube. Material was then flash-frozen in the field in liquid nitrogen. Tissue samples were also attempted on Melin-Norkrans Modified media with collections for experimental applications with a low success rate. A member of the closest related extant outgroup, *G. convolvens*, was also sampled for comparative analyses between different trophic modes. Vouchered specimens are accessioned in the herbarium of the University of Tennessee. Detailed methods for DNA and RNA extraction, quality assessment and taxon sampling are described in the Supporting Information Methods S1–S3; Table S1.

Genome sequencing, assembly and annotation

All sequencing, assembly and annotation were performed at JGI. Genome sequencing was done with Pacific Biosciences (PacBio) technology, except for four genomes sequenced with Illumina technology. All PacBio-sequenced genomes were assembled using FALCON (Chin *et al.*, 2016). Of the Illumina-sequenced genomes, three were assembled serially using VELVET (Zerbino & Birney, 2008) followed by ALLPATHS-LG (Gnerre *et al.*, 2011), and one was assembled with ALLPATHS-LG alone. Mitochondrial genomes were assembled separately. All transcriptome sequencing was done with Illumina only, and subsequently assembled into putative transcripts using TRINITY (Grabherr *et al.*, 2011) or RNNOTATOR (Martin *et al.*, 2010). Each genome was annotated using the JGI Annotation Pipeline (Grigoriev *et al.*, 2014; Kuo *et al.*, 2014)

aided by the transcriptome. Detailed methods are described in Methods S3–S7).

Comparative genomic feature analyses

Statistics of JGI genome assemblies (i.e. N50, number of genes and scaffolds, genome size) were obtained from JGI MYCOCOSM (<https://mycocosm.jgi.doe.gov>). Genome completeness with single-copy orthologues was calculated using BUSCO v.3.0.2 with default parameters (Simão *et al.*, 2015). The TE coverage in genomes was calculated using a custom pipeline, Transposon Identification Nominative Genome Overview (TINGO; Morin *et al.*, 2019). The secreted gene repertoire was predicted as described previously (Pellegrin *et al.*, 2015). An annotated and manually curated dataset of genes coding for carbohydrate-active enzymes (CAZymes) was developed using the CAZy database (Lombard *et al.*, 2014). We calculated, visualized and compared the count and ratio of total (present in the genomes) and predicted secreted CAZymes, lipases, proteases and small secreted proteins (SSPs; < 300 amino acids) as a subcategory. We calculated the total count of the following using all predicted secreted PCWDEs and fungal cell wall-degrading enzymes (FCWDEs). We examined the total and predicted secreted counts of CAZymes/lipases/proteases/SSPs. Statistically significant ecological groups ($P < 0.05$) were determined. Output files generated in this way were combined and visualized with custom R scripts, Proteomic Information Navigated Genomic Outlook (PRINGO; Miyauchi *et al.*, 2020). The percentage of variances (R^2) was estimated for selected genomic features explained by variables including ecological groups, phylogenetic distances and the genome coverage of TEs with permutational multivariate analysis of variance (PERMANOVA). Detailed procedures are as previously described (Miyauchi *et al.*, 2020). We examined associations between TEs and different types of proteins with the R packages PCATOOLS and GGALLY (Blighe & Lun, 2020; Schloerke *et al.*, 2021).

Phylogenomic inference and molecular clock analyses

We constructed a phylogeny based on orthologous genes among the selected fungi using FASTORTHO with the parameters set to 50% identity, 50% coverage, inflation 3.0 (Wattam *et al.*, 2014). Protein sequences used for the process were genome-wide protein assemblies from JGI fungal portal MYCOCOSM. We identified clusters with single-copy genes, aligned each cluster with MAFFT 7.221 (Katoh & Standley, 2013), eliminated ambiguous regions (containing gaps and poorly aligned), and concatenated single-gene alignments with GBLOCKS 0.91b (Castresana, 2000). We constructed a phylogenetic tree with RAXML 7.7.2 (Stamatakis, 2006) using the standard algorithm, the PROTGAMMAWAG model of sequence evolution and 1000 bootstrap replicates.

A set of 38 genomes across the Agaricomycotina were selected for calibrating a molecular clock and dating of the Russulales lineage (Table S2). Gene selection for molecular clock analysis of Russulales was done based on phylogenetic informativeness performed in PHYDESIGN (López-Giráldez & Townsend, 2011).

Molecular clock analysis was performed in BEAST v.1.8.4 (Drummond *et al.*, 2012) using the 20 most phylogenetically informative loci, owing to computational constraints of the program for dealing with large datasets. Three calibrations based on fossils were used: *Archaeomarasmius leggetti*, an agaric fossilized in 90 million-yr-old (Ma) Dominican amber as the minimum age of Agaricales (Hibbett *et al.*, 1997); *Quatsinoporites cranhamii*, a poroid shelf fungus estimated at 113 Ma as the minimum age of the Hymenochaetales (Smith *et al.*, 2004); and *Gastroidea lobata*, a gastroid fruiting body with a double-layered peridium from the Cretaceous (72–66 Ma) (Krassilov & Makulbekov, 2003). The analysis used an uncorrelated lognormal relaxed clock model prior with a constant coalescent tree prior. Markov chain Monte Carlo analysis was run independently three times for 50 million generations, logging every 1000 generations. The runs were checked for convergence and mixing using TRACER v.1.6 (Rambaut *et al.*, 2018). An ultrametric maximum-clade-credibility (MCC) tree was summarized in TREEANNOTATOR 1.8.4 with a burn-in of 25% of trees.

Secondary metabolite analysis

Secondary metabolite gene clusters (SMCs) were predicted with ANTI-SMASH 5.0, using a relaxed strictness through the online dedicated server (Blin *et al.*, 2019). Filtered gene models were used as feature annotations. The resulting .gbz files were analyzed through the BIGSCAPE pipeline using default parameters and the PFAM-A v.30.0 database (Navarro-Muñoz *et al.*, 2020).

Genome rearrangement analysis

Syntenic blocks were identified from pairwise comparisons of genomes with the R package DECIPHER (Wright, 2015). Macrosynteny was determined using the 'FindSynteny' function with default parameters with the argument for masking repeat sequence turned off, whereas mesosynteny was identified using the modified parameters (i.e. maxSep = 1000, maxGap = 1000) suitable for highly similar sequences at the gene level. We determined the genomic coordinates of genes from the JGI genome-wide gene catalog files (Grigoriev *et al.*, 2014). JGI functional gene annotations from the INTERPRO database were used for the description of intracellular and extracellular proteins (<https://mycocosm.jgi.doe.gov>). Then, it was combined with the secreted gene repertoire and repeats data described earlier. The integrated results were used for the circular representation of the genome assemblies with the combined genomic information using the R package CIRCLIZE (Gu *et al.*, 2014). In addition, we measured the mean TE–gene distances with statistical support by comparing the locations of observed genes and TEs and 5000 null hypothesis genome models made by randomly reshuffling the locations of genes. The probability (P -value) of mean TE–gene distances was calculated using the R package REGIONER (Gel *et al.*, 2016). The process was conducted with a set of custom R scripts, Synteny Governance Overview (SYNGO; Hage *et al.*, 2021). Scaffolds containing major syntenic regions among the species were visualized along with the identified predicted secreted gene repertoire

and TEs using the R package KARYOPLOTTER (Gel & Serra, 2017). Data integration was performed with a set of custom R scripts, Visually Integrated Numerous Genes of Omics (VINGO).

Gene evolution analysis

Evolutionary gains or losses of orthologous gene groups were estimated on the basis of the constructed phylogeny using Software for Computational Analysis of gene Family Evolution (CAFE; DeBie *et al.*, 2006). The software uses a random birth and death process to model gene gain and loss across a user specified tree structure. The distribution of family sizes generated under the random model provides a basis for assessing the significance of the observed family size differences among taxa. We selected gene families with P value < 0.001.

Evolutionary rate analysis

Orthologous gene clusters were imported into COUNT (Csurös, 2018) for total and secreted gene repertoire clustering analyses to assess gene evolution rates and reconstruct gene family history. Gene clusters containing fewer than three species were filtered out of the rate optimization. Rate optimization used the gain-loss-duplication model with a Poisson distribution at the root and lineage-specific variation estimated. The analysis was run for 100 rounds with a convergence threshold on the likelihood of 0.1. Gene ancestral reconstruction was inferred using Dollo parsimony and posterior probabilities using a birth-and-death model.

Results

Phylogeny of Russulales

A reconstructed phylogeny of Russulales based on 725 single-copy orthologs recovered *Lactarius* as forming the sister clade to the rest of the ECM Russulaceae and *Lactifluus* forming the sister clade to *Russula* (Figs 1a, S1). Within *Russula*, a clade comprising *R. brevipes* and *R. dissimulans* was inferred as sister to the rest of *Russula* (Fig. S1). The common ancestor of the ECM Russulaceae is inferred to have arisen around the Cretaceous-Paleogene (K-Pg) extinction event (73.6–60.1 Ma), a period of rapid ecological and anatomical innovation in plant communities (Alfaro *et al.*, 2018). The family Russulaceae, including the saprotrophic *G. convolvens*, began diversification around the same time as the saprotrophic neighbor family in the order Russulales, Auriscalpiaceae, during the Cretaceous (*c.* 74 Ma).

Main genomic features of Russulales

A total of 13 genomes were newly sequenced for this study, including nine from Russulaceae (Table S1). Genomes of ECM species within Russulales are larger than saprotrophic and dual saprotrophic/pathogenic species (adjusted false discovery rate: P < 0.05, pairwise PERMANOVA; Fig. 1), with *Lactarius quietus* having the largest genome (115.9 Mb) and other ECM genomes ranging from 43.3 to 63 Mb (Fig. 1b; Table S1). Over 94% of a

benchmark set of conserved fungal genes (BUSCO; Simão *et al.*, 2015) were found in genome assemblies (Fig. 1b), and up to 97% of the RNA-Seq reads mapped to the gene repertoire (see the 'info' page at JGI Russulales portal, <https://mycocosm.jgi.doe.gov/Russulales/Russulales.info.html>), indicating that assembled genomes capture most of the coding gene space.

The total gene repertoire of Russulales comprises over 250 000 predicted genes for the 18 species compared, ranging from 10 514 genes for *Multifurcata ochricompacta* to 18 952 genes for *Peniophora* sp. (Fig. S2). The core gene orthogroups that are found in at least 17 of the 18 fungi make up one-quarter of all genes, averaging *c.* 3500 genes for most species and up to 4023 genes for *L. quietus*. The species-specific gene content varies considerably between species but not between trophic categories, with *Peniophora* sp. and *L. quietus* having the highest number of unique genes (11 721 and 10 313, respectively), and *M. ochricompacta* having only 2832 unique genes. Secondary alleles were identified from sequenced dikaryotic genomes by the PacBio sequencing technology; they comprised 14–39% of all protein models (Table S1).

The major principal components of genomic features captured > 80% of variation and clearly separated the ECM species from the saprotrophic relatives for many of these features (Fig. S3a). Variation in secreted lipases, secreted CAZymes, TE content and secondary metabolites among the sampled genomes was primarily explained by phylogenetic distances (Fig. 1c). However, the total genome size, TE content, secreted CAZymes, and SSPs all showed significant contributions from the fungal ecology, and TEs were primary drivers of total genome size, number of secreted proteases and number of SSPs. The ECM species showed higher TE content with larger genome size and fewer secreted CAZymes than the saprotrophic groups (Fig. 1d; P < 0.05, pairwise PERMANOVA; Table S3). The variation in gene numbers in individual secreted CAZyme classes (i.e. AA, CBM, CE, GH, PL) was largely attributed to the phylogenetic distance (P < 0.05, PERMANOVA), with the exception of secreted glycoside hydrolases (GHs) (Figs 1c, S3b). Therefore, the loss of secreted GH gene content, in particular, is driving the reduction of secreted CAZymes in the ECM Russulaceae.

The evolution of plant cell wall degrading enzymes in ECM Russulales

Ectomycorrhizal Russulales contain a smaller set of CAZymes than do saprotrophic taxa (Figs 2, S4, S5; Table S4). They have lost a core set of genes required for efficient degradation of PCWDEs and FCWDEs. The number of gene copies for many secreted enzymes involved in the decomposition of cellulose, hemicellulose, pectin, lignin, chitin and mannan is restricted or absent in symbiotrophs compared with the related saprotrophs, including *G. convolvens* (Tables 1, S2; Fig. S4). For many orthologous clusters, however, this reduction is also seen in the genome of *G. convolvens*, suggesting that it had already occurred in the ancestor of Russulaceae (Tables S2, S5, S6). These orthogroups include subtilases, aspartic proteases, AA3_2 aryl alcohol oxidases, GH12 endoglucanases and expansin-like proteins, among

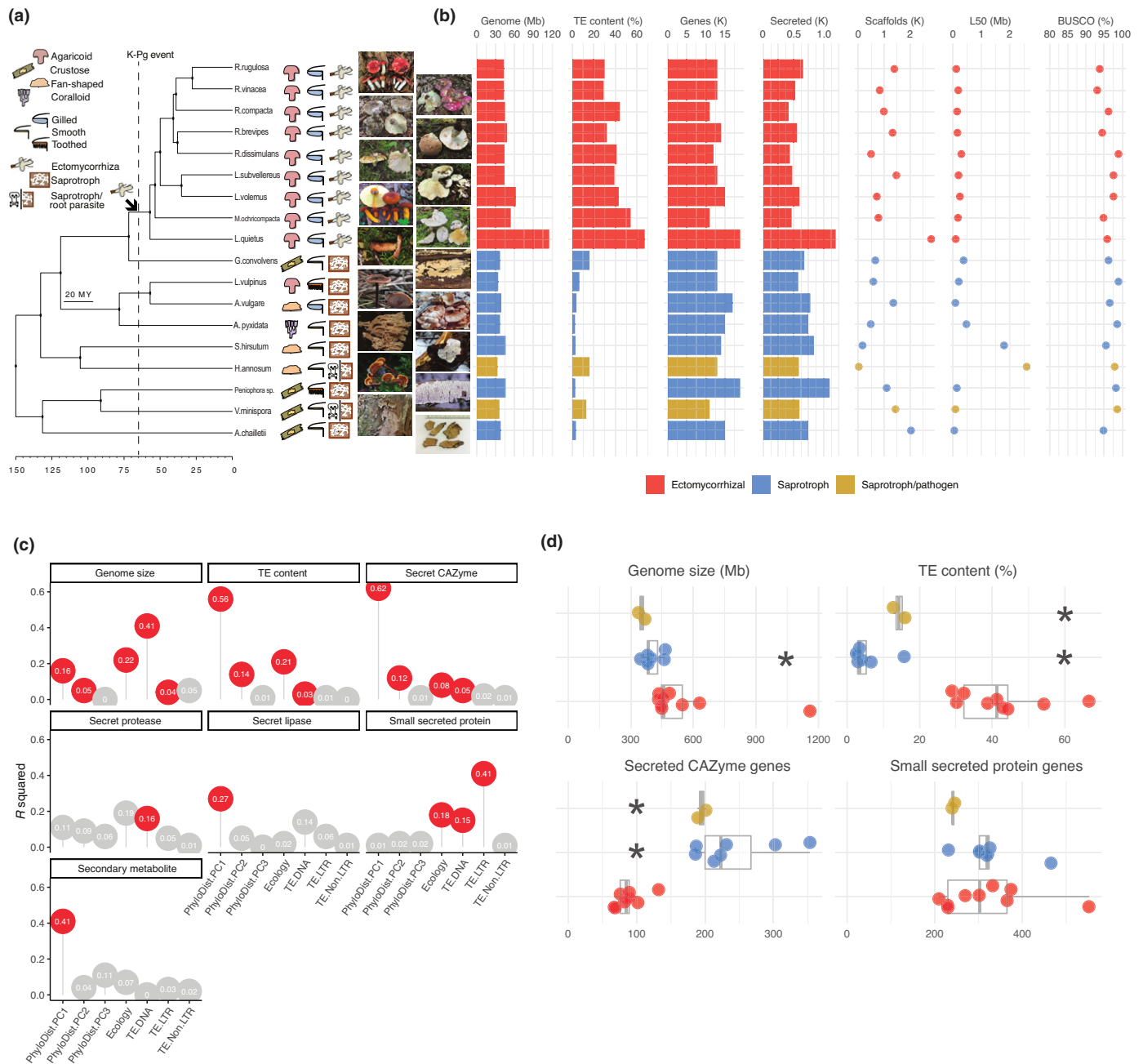


Fig. 1 Overview of genome features of 18 fungi. (a) Phylogenetic reconstruction of 18 Russulales genomes. A total of 2518 single-copy genes were used in RAxML with 1000 bootstrap iterations. Fruitbody form (FB form), hymenium type (Hym) and nutritional strategy (Nutr strat) are given with images of the genome source (see details in Supporting Information Fig. S1). (b) The following parameters are shown: genome (genome size); TE content (coverage of transposable elements (TEs) in the genomes); genes (number of genes); secreted (number of predicted secreted proteins); scaffolds (number of scaffolds); L50 (N50 length); and BUSCO (genome completeness) (Table S1). (c) Percentage of variation explained in selected genomic features. R^2 values indicate contributions of genomic features tested. Abbreviations are as follows: PhyloDist.PC1–3, major principal components covering > 80% of variation in phylogenetic distances; ecology, saprotroph, ectomycorrhiza (ECM) or pathogen; TE.DNA/LTR/NonLTR, TE family coverage in the genomes. Circles with numbers inside show R^2 -values contributing to the genomic features tested. Circles in red indicate significant variables ($P < 0.05$; PERMANOVA model, genomic features ~ phylogeny + ecology + TE types). See detailed information in Tables S11, S12. (d) Comparisons of genome size, TE content and small secreted proteins (SSPs) sorted by fungal ecology as a boxplot. Median values are indicated as gray lines within the box, representing the interquartile range of the distribution. Significantly different groups in comparison to the ECM are indicated with an asterisk (*, $P < 0.05$, pairwise PERMANOVA; Table S13).

others. The ECM Russulaceae have retained a few key orthologous clusters involved in cellulose degradation, such as GHs GH45 and lytic polysaccharide monoxygenases (LPMOs, AA9)

which may be involved in the host root penetration or fungal cell wall remodeling (Table S7) (Veneault-Fourrey *et al.*, 2014; Krizán *et al.*, 2019; Zhang *et al.*, 2019).

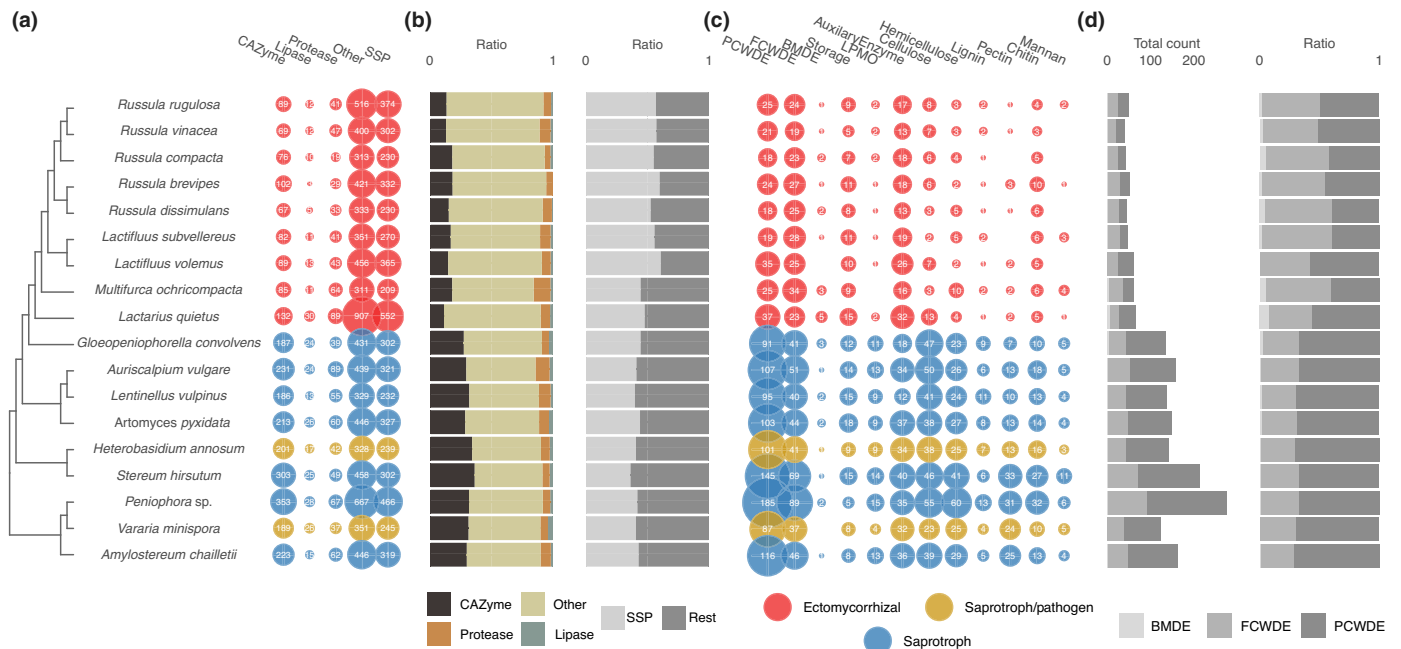


Fig. 2 Predicted secreted gene repertoire of 18 members of Russulales. (a) Number of secreted genes for CAZymes, lipases, proteases and others (i.e. all secreted proteins not in these first three groups). The group of small secreted proteins (SSPs) is a subcategory showing the number of SSPs (< 300 amino acids). The size of bubbles corresponds to the number of genes. The fungi are colored according to their ecology. (b) The ratio of CAZymes, lipases and proteases to all secreted proteins (left); and the ratio of SSPs among the entire secreted gene repertoire (right). (c) The number of plant cell wall-degrading enzymes (PCWDE); fungal cell wall-degrading enzymes (FCWDE); bacterial membrane (i.e. peptidoglycan)-degrading enzymes (BMDE); enzymes for starch and glycogen (storage); lytic polysaccharide monoxygenase (LPMO); AA family CAZymes (auxiliary enzymes); substrate-specific enzymes for cellulose; hemicellulose; lignin; pectin (plant cell walls); chitin (fungal cell walls); and mannan. (d) The total count of genes coding for PCWDE, FCWDE and BMDE (left); and the proportion of PCWDE, FCWDE and BMDE (right) (see Supporting Information Table S14).

At a gene family level, the ECM Russulaceae is enriched in GH45 cellulases and inhibitors of caspases and cysteine endopeptidases (family I32) compared with saprotrophic Russulales (Table 1). An important aspect of ECM Russulaceae specialization is the retention of lignolytic manganese peroxidase (POD) genes as a remnant of a white rot ancestry of Russulaceae (Fig. S6). These species have not retained the same POD genes, as two different clades of POD genes were recovered for ECM Russulaceae. The two paralogous gene clusters are present in the *R. brevipes* and *Multifurca* genomes, indicating that there were probably multiple independent losses in both clades for the other ECM species. In both cases, the same gene has been retained in the closest extant saprotrophic ancestor, *G. convolvens*.

Although ECM Russulales have experienced a concerted loss of CAZymes, members of Russulaceae have experienced species-specific expansions of particular gene orthologs (Table S7). This includes enzymes involved in degradation of cellulose (e.g. GH5_12, GH5_30, GH45), hemicellulose (e.g. CE4, CBM13), chitin (e.g. GH20, CBM18) and mannan (e.g. GH92). The second largest genome, *L. volemus*, possesses the highest copy numbers of AA1_1 laccases in Russulales. The third largest ECM Russulaceae genome, *M. ochricompacta*, possesses the fewest number of genes in Russulales and the second highest TE proportion. In the *M. ochricompacta* genome, the gene content of three groups of subtilisin-like serine proteases is particularly expanded, whereas it is absent in most of the other species in Russulaceae, and the number of genes coding

for secreted CE4 chitin deacetylases and GH47 α -mannosidases is also expanded. *Lactifluus subvellereus* is characterized by a substantial expansion of AA5_1 glyoxal oxidases with moderate expansions in aspartyl proteases and AA1_1 laccases. *Russula brevipes* is the only ECM Russulaceae species to possess PL14_4 β -1,4-glucuronan lyases, and the number of GH72-CBM43 β -1,3-glucanosyltransferases is expanded. Expansion in *Russula rugulosa* includes a group of serine carboxypeptidases and a group of tyrosinases. In *Russula vinacea* a cluster of carboxylesterase lipases is highly expanded. Not all ECM Russulaceae species exhibit substantial expansions in their secreted gene repertoire, with *R. compacta* and *R. dissimulans* mostly lacking gene duplication-mediated expansions. Substantial expansions are less common for nonECM Russulales, with the exception of *Peniophora*, in which a cluster of lipases and AA1_1 laccases are substantially expanded, and *A. vulgare*, in which two subtilase clusters, a GMC oxidoreductase cluster and a cluster of aspartyl proteases, are expanded.

Functional specialization is also evident in the enrichment of unique species-specific secreted gene clusters for ECM Russulaceae. *Lactarius quietus* possesses unique secreted gene clusters, with two clusters of putative fungistatic metabolites as well as a thaumatin-like protein group and a group of unique expansin-like proteins. *Multifurca ochricompacta* possesses a unique cluster of fungistatic metabolite genes. *Lactifluus volemus* possesses a unique cluster of secreted protein genes with a LysM domain. *Lactifluus subvellereus* possesses a unique cluster of GH45

Table 1 Total and secreted gene families enriched for species of ectomycorrhizal (ECM) Russulaceae and functional groups. Red indicates substantial enrichment; purple indicates a unique family; black indicates no depletion among families with high gene count (more than three per genome).

Species	Total gene repertoire	Secreted proteins
<i>Lactarius quietus</i>	AA3_3 (alcohol oxidase), AA5_1 (copper radical oxidase), CBM13 (carbohydrate-binding module), GH45 (cellulase), GH92 (mannosidase), GT20 (α -trehalose-phosphate synthase), GT22 (mannosyltransferase), GT24 (glycoprotein α -glucosyltransferase), C54 (cysteine protease), S53 (sedolisin), T06 (threonine protease)	AA5_1 (copper radical oxidase), EXPN (expansin), GH45 (cellulase), S53 (sedolisin), GH13, GH13_22, I63, M24A, T06
	AA1, AA3, AA3_3, AA5_1, CBM12, CBM13, EXPN, GH13, GH13_22, GH25, GH45, GH92, GT20, GT22, GT24, GT76, C04, C54, C65, M14A, M16C, S53, T06, GGGX ABH03	AA3_2, AA5_1, EXPN, GH13, GH13_22, GH25, GH45, S10, S53, I02, I63, M24A, T06, GGGX ABH03
<i>Lactifluus subvellerus</i>	AA5_1 (copper radical oxidase), GH37 (trehalase)	I25B
	AA5_1, CBM12, GH13_22, GH37, M13, S09C, GX ABH08	AA5_1, S09C, M23B, I25B, I25X
<i>Lactifluus volemus</i>	AA1_1 (laccase), GH20 (β -N-acetylglucosaminidase), GH37 (trehalase), S28 (lysosomal Pro-Xaa carboxypeptidase)	AA1_1 (laccase)
	AA1_1, AA1_2, GH13_22, GH15, GH20, GH37, GH38, GT66, A22B, C46, M03A, M76, S28, S72, T02, T03, T06, C82, GGGX ABH03	AA1_1, GH9, S10, GGGX ABH03
<i>Multifurca ochricompacta</i>	S08A (subtilisin), S28 (lysosomal Pro-Xaa carboxypeptidase)	CE4 (chitin deacetylase), GH47 (α -mannosidase), M43B (cytophagalyisin), S08A (subtilisin), CE14
	CE4, CE14, GH47, GT22, C01B, I25A, M43B, S08A, S28, GX ABH07	CE4, CE14, GH47, M43B, S08A, I25X
<i>Russula brevipes</i>	CBM50 (carbohydrate-binding module), GH47 (α -mannosidase), I43 (serine protease inhibitor), AA5	CBM50 (carbohydrate-binding module)
	AA5, CBM18, CBM43, CBM50, CE14, GH3, GH5_30, GH17, GH47, GH72, GT39, M23B, C82, A31, I32, I43	AA1_1, CBM50, GH9
<i>Russula compacta</i>	None	None
	GH13_1, GH13_5, A31	AA5_1, CBM50
<i>Russula compacta</i>	None	None
	AA1_2, CE9, GH13_1, C39, I32, M48X, S09A	GH13_32, I25X
<i>Russula rugulosa</i>	GH5_30 (cellulase), GT48 (1,3- β -glucan synthase), C19 (ubiquitin-specific protease), S16 (Ion protease), C85	GH5_12 (cellulase), GH5_30 (cellulase)
	CBM12, EXPN, GH5_30, GH30, GH92, GT31, GT48, GT49, GT59, C19, C45, I32, S10, S16, C40, C85	EXPN, GH5_12, GH5_30, S10, C19
<i>Russula vinacea</i>	GT4 (glycosyltransferase), GT48 (glycosyltransferase), C19 (ubiquitin-specific protease), I21	GH5_30 (cellulase), CE9
	AA7, CE9, EXPN, GT4, GT48, GT69, GT76, C19, I25X, I32, C89, A31, I21	EXPN, CE9, GH5_30
ECM Russulaceae	GH45 (cellulase), I32 (IAP)	GH45 (cellulase)
Saprotrophs	AA2, AA3_1, AA3_2, AA3_4, AA8, AA9, CBM1, CBM5, CBM35, CE1, CE8, CE15, CE16, GH1, GH2, GH3, GH5, GH5_5, GH5_7, GH5_12, GH5_22, GH5_50, GH6, GH7, GH10, GH11, GH12, GH16, GH18, GH27, GH28, GH29, GH30_3, GH31, GH43, GH51, GH53, GH55, GH74, GH76, GH78, GH79, GH81, GH95, GH105, GH115, GH128, GH131, GH145, GT41, PL8_4, PL14_4, CO3B, C12, C56, I51, M28E, G01, M77, GGGX ABH04, GX ABH09, GX ABH23	AA2, AA3_2, AA8, AA9, CBM1, CBM5, CBM35, CE1, CE8, CE15, CE16, GH3, GH5_5, GH5_7, GH6, GH7, GH10, GH12, GH18, GH28, GH30_3, GH35, GH43, GH51, GH53, GH55, GH72, GH79, GH81, GH92, GH95, GH115, GH145, PL8_4, PL14_4, G01

endoglucanases. *Russula compacta* possesses a unique cluster of hydrophobin genes. *Russula brevipes* is characterized by seven unique gene clusters, of which the largest is a cluster of serine

carboxypeptidases. *Russula dissimulans* possesses a cluster of glutathione S-transferases. The *R. rugulosa* genome also has many expanded unique gene clusters, with a single large cluster of

expansin-like proteins. *Peniophora* is the only nonECM member with a large number of unique gene clusters, with 26 unique gene clusters of which the largest cluster has nine gene copies.

The repertoire of small secreted proteins

The number of SSPs in the ECM species were similar to those of saprotrophic or pathogenic relatives (Fig. 1d). However, *c.* 73% of variation in the number of SSPs was explained by TEs and ecology (Fig. 1c; $P < 0.05$, PERMANOVA). Small secreted proteins were moderately correlated with the DNA transposon family EnSpm/CACTA (Spearman coefficient = 0.47, $P < 0.05$; Fig. S3c), suggesting evolutionary associations between TEs and SSPs. We identified a single SSP gene that is evolutionarily conserved in all ECM Russulaceae species. It encodes a phosphatidylglycerol/phosphatidylinositol transfer protein with an MD-2-related lipid-recognition domain implicated in the recognition of lipids, including pathogen-related products (Table S8). The *L. quietus* genome possesses 552 genes coding for SSPs, almost twice as high as other ECM Russulaceae.

Secondary metabolite diversity

In all, 409 biosynthetic gene clusters (BGCs) involved in secondary metabolism were detected in Russulales genomes with the addition of the recently sequenced fungi (Fig. S7). Species relatedness drove differences in the number of secondary metabolite modules (41%; $P < 0.05$, PERMANOVA; Fig. 1c). Members of Russulales outside of Russulaceae possess more nonribosomal peptide-synthase (NRPS)-like, combined T1PKS and NRPS-like, and higher total number of BGCs than Russulaceae species, whereas Russulaceae members possess more siderophore BGCs. Four NRPS BGCs share homology among some or most of Russulaceae, which represent two different metabolic pathways (Fig. S8a). A diverse group of up to eight BGCs for terpene synthase are conserved among at least some members of Russulaceae, with two BGCs also represented in Auriscalpiaceae (Fig. S8b). Two BGCs representing a single polyketide synthase pathway with a ferric reductase transporter domain, as well as another group of two BGCs representing a single pathway with an aspartyl protease domain are present in most members of Russulaceae, including the saprotrophic *G. convolvens* (Fig. S8c,d).

Gene evolution rate in Russulaceae

Across the total gene repertoire, ECM Russulaceae and saprotrophic Russulales experienced comparable numbers of gene duplications and losses with a slightly lower overall rate of gene gain (Figs 3, S9). The overall rate of gene evolution (gain, loss and duplication) was accelerated in the ancestors of both ECM Russulaceae (internode 8–9) and saprotrophic Russulaceae (internode 9–12). Species-specific gene evolutionary rates were higher for *L. quietus* and *S. hirsutum* across the total gene repertoire and only for ECM Russulaceae members across the secreted gene repertoire. Gene evolution rate varied across the secreted gene repertoire, with a higher rate of gene loss (0.08), gene duplication rates at about half of the loss rate (0.04), and gene gain rates that were 10-fold less (0.004). An accelerated rate of gene

loss occurred in the total gene repertoire of the ancestor of Russulaceae (node 8). However, gene loss was the greatest in species-specific lineages, indicating a high evolutionary rate of secreted gene repertoire modification. Gene loss rates are over twice as high as the total gene repertoire rates (0.18) and gene gain rates are three orders of magnitude lower (8×10^{-6}).

Impact of TEs on the genome landscape

Transposable elements comprise higher proportions of the genome (%) in ECM Russulaceae than in nonECM Russulales, ranging from 29% to 67% of genome assemblies (Fig. S10). *Gypsy*, *Copia* long terminal repeat (LTR) retrotransposons, hAT families and other unclassified repeats are among the most abundant TEs in genomes from ECM Russulaceae. hAT repeats are involved in RNA processing and are unique to ECM Russulaceae in the Russulales dataset (Hammani *et al.*, 2012). Notably, *Penelope* nonLTR retrotransposons are only present in *L. quietus*. The larger size of genomes of ECM Russulales species is mainly a result of their higher content in repeated elements ($P < 0.05$, PERMANOVA; Figs 1c, S3c). DNA transposons contributed *c.* 42% of variation in the genome size ($P < 0.05$, PERMANOVA; Fig. S3b). We found a significant correlation between the genome size and DNA transposons, EnSpm/CACTA, hAT and Mariner/Tc1 repeats (Spearman coefficient = 0.53, $P < 0.0001$; Fig. S3c).

The distribution of TEs and genes across the 10 largest scaffolds of Russulales genomes were visualized using ‘Hanabi’ plots, showing TEs forming ‘transposon nests’, or dense aggregations of TEs made up of multiple families and unclassified repeats (Fig. 4). To examine the potential association of gene structure and TE content, a permutation test was performed to compare distances between TEs and coding genes of 18 Russulales (Table S9). The ECM Russulaceae genomes show statistically significant associations between SSPs and TE-rich areas, with SSPs being more closely linked to TEs than in nonECM Russulales (Fig. 5; $P < 0.01$). Nonsecreted genes were statistically more distant from TEs in ECM Russulaceae than in nonECM Russulales (Fig. 5; $P < 0.05$). The exception to these patterns was *V. minispora*, a known plant pathogen. Secreted CAZyme families showed little or no association with TEs; however, associations between TE and lipases/proteases were detected in some ECM species (Figs 5, S11). We selected genes and TEs within a defined distance for further examinations. We defined a maximal TE–gene distance as 4.5 kb (4 kbp for gene size + 500 bp to the nearest TEs) as the majority of genes are no > 4 kb in size (Fig. S12). Small secreted proteins were more strongly associated with TE pockets in the ECM group compared with nonECM species, with most of the ECM SSP genes located within a range of 500 bp to the nearest TE (Fig. S13). Genes associated with TEs contained recurring PFAM domains, such as WSC domain, cupredoxin or fungal hydrophobin for SSPs. However, the majority of SSPs have unidentified functions (82.4%; data not shown).

Gene synteny in Russulales

Gene synteny for five *Russula* species and *G. convolvens* was compared across the dataset (Figs S14, S15–S17; Table S10).

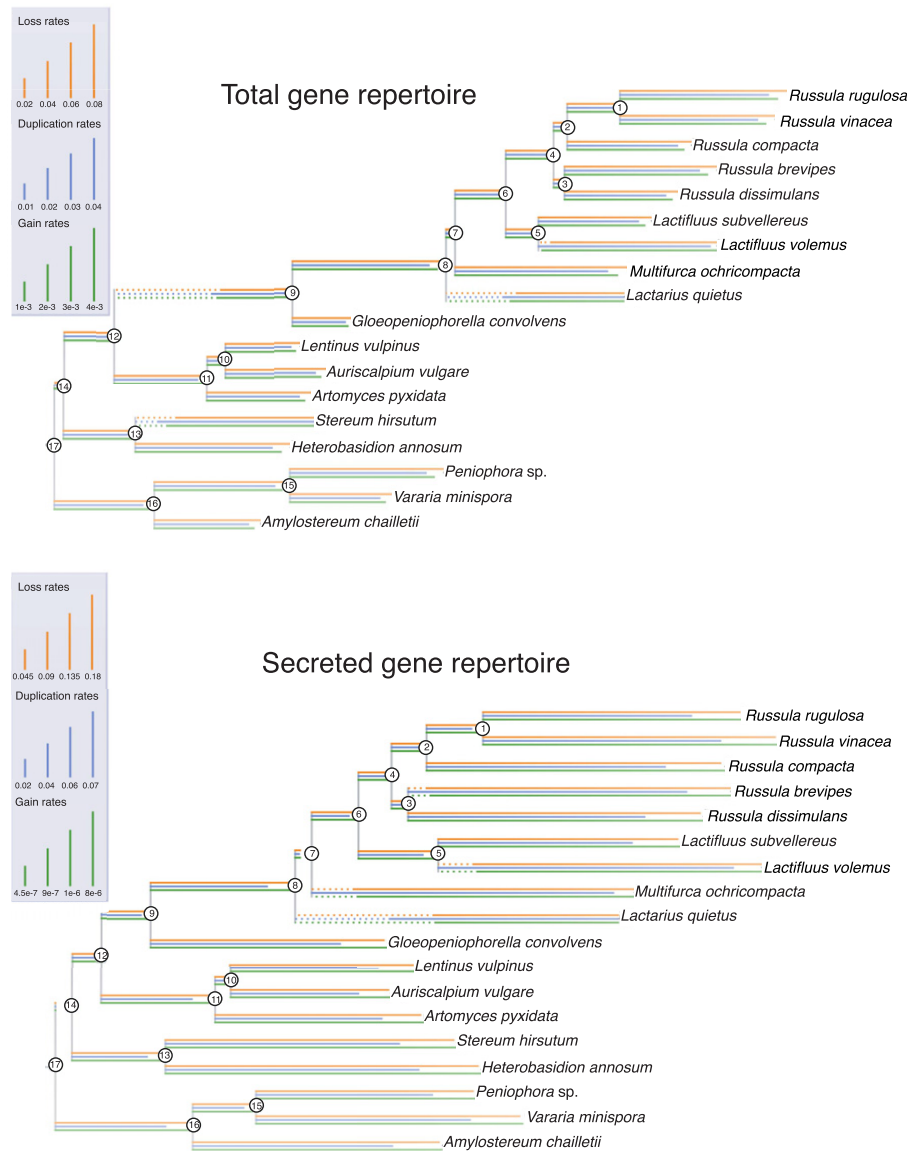


Fig. 3 Evolutionary rate COUNT analysis of Russulales genomes. Top: rates of gene loss, duplication and gain for the total gene repertoire along branches. The Poisson distribution of the birth-and-death model is given for the root node. Bottom: rates of gene gain, loss and duplication for the secreted gene repertoire along branches, with dotted lines representing the rate length exceeding the total line. Ancestral nodes are numbered.

Russulaceae share the highest degree of synteny with each other even though the size of scaffolds for the 10 largest scaffolds is variable within Russulales genomes (Figs S18, S19). Syntenic regions are disrupted by clusters of TEs (Fig. 4). The frequency of TE insertions suggests that TEs accumulated further in TE-rich regions during the course of evolution, forming ‘transposon nests’ that cause transposon breakpoints (Hess & Pringle, 2014). The single largest syntenic block among five *Russula* species was used as a landmark to investigate gene order and mesosynteny (Figs S20, S21).

The concerted loss of CAZymes has been observed in conserved portions of Russulaceae genomes by comparing the largest syntenic region. This comparison reveals contrasting retention of the genes of interest (e.g. AA3_2, AA2, AA9 and GH74) between the saprotrophic *G. convolvens* and closely related ECM species that have lost these traits, such as *R. vinacea* or *R. brevipes* (Fig. S13). A conserved syntenic cluster of secreted CAZymes was detected as a core secretory capacity for Russulaceae, which includes

glycosyltransferases and carbohydrate esterases. Less conserved syntenic clusters are widespread for a larger array of secreted proteins that includes glycoside hydrolases, polysaccharide lyases, expansin-like proteins, carbohydrate binding modules and glycosyltransferases. Transposable element nests are infrequent across the syntenic region and do not show a correlation with secreted genes, indicating that association between repeated elements and secreted genes are nonsyntenic owing to the activity of TEs.

Discussion

Evolutionary transition to the ECM lifestyle

Genomes of ECM fungi in the Russulaceae family present many of the established hallmarks of the transition to ECM symbiosis, including an expansion in genome size as a result of the accumulation of repeated elements and a contraction in gene families involved in the enzymatic breakdown of plant organic matter

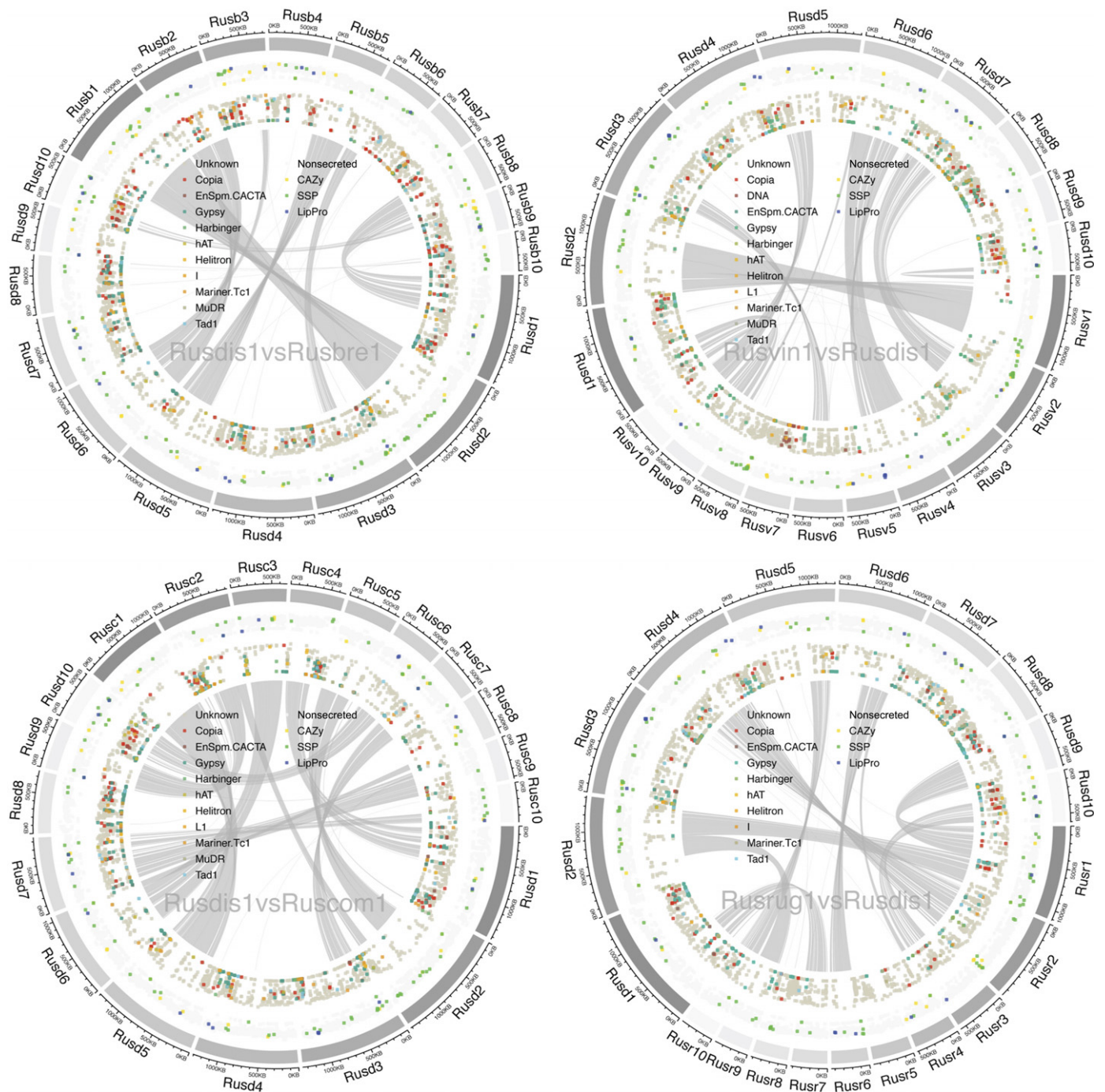


Fig. 4 Macrosynteny comparison with five *Russula* species (Hanabi (fireworks) plot shows pairwise syntentic comparison of scaffold 1–10): *R. dissimulans* (Rusdis1), *R. brevipes* (Rusbre1), *R. vinacea* (Rusvin1), *R. compacta* (Ruscom1) and *R. rugulosa* (Rusrug1). Outer circle, the size of scaffold 1–10; first inner circle, genes located in the scaffolds (genes coding for CAZymes, small secreted proteins (SSPs), lipases and proteases are highlighted (see the legend for details)); second inner circle, transposable element (TE) families and unknown repeats in the scaffolds (see the legend for details); vertical axis of each inner circle, mean distance of neighboring genes/TEs. Short distances between the genes/TEs result in dots towards the center of the Hanabi plot, whereas long distances result in dots towards the outer circle. Grey lines represent shared syntentic regions.

(Kohler *et al.*, 2015; Martin *et al.*, 2016; Hess *et al.*, 2018; Miyauchi *et al.*, 2020). The contraction of the PCWDE gene repertoire coincides with the evolution of the ECM habit after the split with the closest extant saprotrophic species, *G. convolvens*. Expansions in TE content in the ancestor of ECM Russulaceae may have facilitated adaptive shifts, such as loss of

PCWDEs, expansion of proteases/lipases, and diversification of SSPs involved in mycorrhiza development. This remodeling is inferred to have coincided with the K–Pg boundary extinction event, suggesting that the shift may have been driven by a drastic shift in plant community composition as a result of mass extinction (Nichols & Johnson, 2008). Ectomycorrhizal plant hosts at

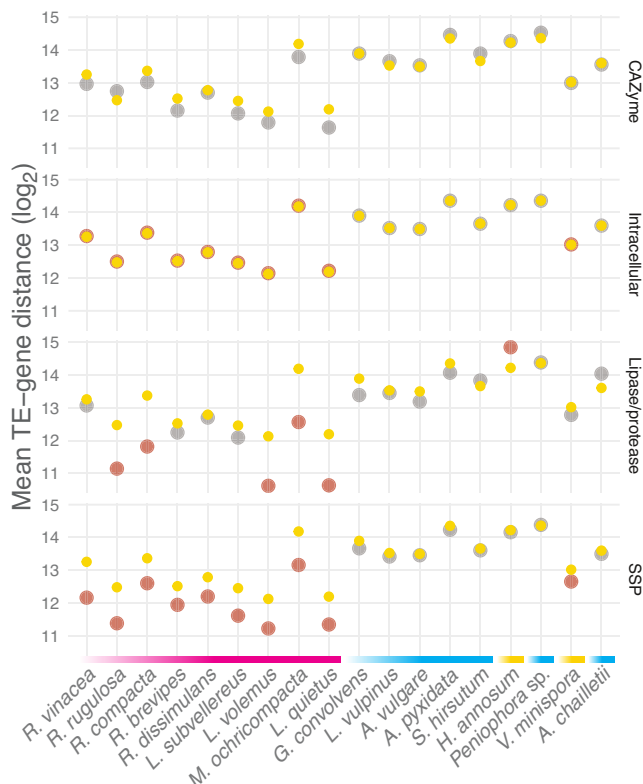


Fig. 5 Observed and permuted transposable element (TE)–gene distances. Yellow, mean TE–gene distances averaged from the 5000 reshuffled models; red, mean TE–gene distances observed in the genomes with statistical significance ($P < 0.01$); gray, mean TE–gene distances observed in the genomes. Distances (base) are \log_2 -transformed (Table S15). Color bars above the species names indicate fungal ecology. Red, blue and yellow refer to ectomycorrhizal, saprotroph and saprotroph/pathogen, respectively.

this time began a shift towards highly variable root evolution within plant families, and diversification of many angiosperm host species occurred later, concurrent with the early diversification of lineages within ECM Russulaceae (Looney *et al.*, 2016; Brundrett & Tedersoo, 2018).

Despite over 65 Myr of divergence time between lineages of Russulaceae, species share a high degree of conserved gene similarity and synteny. Compared with the nonECM Auriscalpiaceae, which has equivalent time of diversification, divergent lineages within Russulaceae maintain at least five times more syntenic links. A higher degree of gene order conservation is unexpected in ECM fungi as the development of TE pockets is hypothesized to break up synteny (Hess & Pringle, 2014); however, this conserved gene order may be important for maintaining their conserved niche and lifestyle. While a large proportion of SSPs are species-specific, a conserved SSP was detected for Russulaceae coding for a phosphatidylglycerol/phosphatidylinositol transfer protein. A homolog of this protein was significantly accumulated in cork oak roots developing ectomycorrhizas with *Pisolithus tinctorius* (Sebastiania *et al.*, 2017). This particular SSP may function as an effector to control the membranes of host plant cells for efficient nutrition exchange and hints at a latent mechanism for ECM symbiosis. This amount of conservation is counter to the

paradigm of effectors arising through convergent evolution and is quite uncommon (Kohler *et al.*, 2015).

We detected a reduction in the repertoire of gene clusters involved in secondary metabolism among the Russulaceae, which has not been widely investigated in other ECM groups (Lofgren *et al.*, 2021). This may constitute another hallmark feature of the evolution of the ECM habit or may be specific to Russulaceae. This was particularly pronounced in NRPS-like SMCs, which have diverse functions but are most known to produce mycotoxins and antibiotics (Bushley & Turgeon, 2010). We hypothesize that his reduction in NRPS-like SMCs may correspond to a reduction in defensive compound diversity in ECM as a result of cooption of plant host defenses and subsequent release on selective pressure. Russulaceae saw an expansion in siderophore SMCs containing conserved N-terminal iron uptake chelate (IucC) domains, which have been implicated in pathogenesis in *Rhizopus* and may be important for iron sequestration (Carroll & Moore, 2018). Based on gene cluster similarity, Russulaceae possesses a conserved set of terpene-related SMCs that may be involved in the production of diverse lactarane sesquiterpenes that have frequently been characterized in Russulaceae (Clericuzio *et al.*, 2012).

Across the total gene repertoire, a heightened rate of gene evolution comprising a high rate of gene loss and duplication was detected as preceding the evolution of the ECM habit. Gene loss was demonstrated in the POD gene family of the common ancestor of Amanitaceae, another group containing extant saprotrophic species as a sister group to a diverse ECM lineage (Hess *et al.*, 2018). Protracted rates of gene gain, loss and duplications across the ancestral total gene repertoire of Russulaceae suggest that priming of the ECM habit may be tied to changes in the mode of nutrition (i.e. polysaccharide metabolism) instead of signaling pathways controlling biotic interactions (e.g. effector-like SSPs). These pre-existing traits may have emerged more frequently in facultative saprotrophs loosely interacting with tree roots, with the ability to switch nutritional modes being latent across a wide diversity of fungi (Smith *et al.*, 2017). The saprotrophic *G. convolvens* is known to widely colonize well-decayed logs using a suite of oxidases, which may necessitate the ability to circumvent plant root defenses within the wood for substrate occupation (Nakasone, 1990). Some species of ECM Russulaceae form mushrooms on well-decayed logs and are thought to associate with roots within the wood and may utilize the POD ligninases retained from white rot ancestry to occupy this niche (Mäkipää *et al.*, 2017). Other elements of the secreted gene repertoire, such as PCWDEs, FCWDEs and proteases, that saw modifications concurrent with the evolution of the ECM habit are potentially more essential to the ECM lifestyle for Russulaceae than host recognition pathways, effectors to circumvent host defense or competitive interactions with other rhizospheric fungi.

Functional specialization within Russulaceae

The peculiar expansions of secreted protease and chitinase families may indicate a possible specialization of Russulaceae fungi to target nonplant-derived organic sources of nitrogen and phosphorus, such as fungal and bacterial necromass. Aspartyl

proteases have been implicated in working in conjunction with hydroxyl radicals to access organic nitrogen from necromass, and these genes have been detected as upregulated in soil organic matter in the presence of the ECM species *Paxillus involutus* (Shah *et al.*, 2016; Beeck *et al.*, 2018). The most extreme gene expansion of 283 genes in a single orthologous cluster of subtilases with pro-kumamolisin activation domains was observed for *L. quietus*. Proteases encoded by these genes are involved in pathogenicity in fungal parasites of animals and other fungi (Muszewska *et al.*, 2011). *Russula gracillima* and *Russula exalbicans* have been shown to parasitize *Lactarius mycorrhizas* (Beenken, 2004), but parasitism in Russulaceae is undoubtedly rare. An alternate and more likely explanation is that this pathway has been coopted for plant–host interaction.

While the prevailing trend of a loss of PCWDEs in ECM lineages is evident in ECM Russulaceae, the opposite trend of functional specialization within secreted CAZymes can be seen in different lineages of Russulaceae. Marked expansions in shared and unique homologous clusters of secreted enzymes as well as different gene families indicates selection for specialized decomposition capabilities. Ectomycorrhizal decomposition, such as ‘litter bleaching,’ has been proposed to contribute significantly to soil carbon turnover in forested ecosystems with phylogeny significantly predicting enzymatic activity (Talbot *et al.*, 2008, 2015; Bödeker *et al.*, 2014; Zak *et al.*, 2019). The ability of an ECM symbiont to scavenge nitrogen, phosphorus and trace elements is mediated through the activity of these enzymes in soil organic matter, and the differential contribution of individual species can be detected by the plant host and used as a criterion to mediate and select for its own mycorrhizal community (Hortal *et al.*, 2017). Traits that have been highlighted as potential drivers of diversification in ECM fungi have primarily looked at morphological traits of sexual reproduction and general ecological strategies, but adaptive functional specialization within ECM lineages is understudied and may be the key to understanding diversification and host dynamics within these diverse groups (Looney *et al.*, 2018; Sánchez-García *et al.*, 2020). The extent to which differential ECM decomposition ability within other ECM lineages is present and how these traits are partitioned within an ECM community should be explored further.

Transposable elements driving gene innovations and ECM regulation

Accumulation of repeated elements in the genomes of ECM Russulaceae and the close physical proximity of TE clusters and SSP genes suggest that TEs may promote gene innovation (e.g. duplication in SSPs and proteases) in ECM Russulaceae fungi. When TE insertions occur near host genes, expression is potentially altered as a result of the silencing of the TE through methylation mechanisms or TE activity on host *cis*-regulatory elements (Chuong *et al.*, 2017). The patterns of localized TEs and their associated SSPs and proteases seem to be species-specific. Such localized TEs may have contributed to unique transcriptional regulation and gene expression. Our

findings are consistent with the view that accumulation of TEs in particular genomic regions have affected certain genes that trigger morphological and physiological changes that are key to the ECM symbiosis (Chuong *et al.*, 2017; Sultana *et al.*, 2017).

Conclusions

In some lineages, such as Russulaceae and Amanitaceae to some extent, genetic traits typifying the evolution of ECM fungi (e.g. loss of PCWDE orthologs, expansion of TE content, reduction of SMCs) are already observed in the genomes of closely related saprotrophic species, and these pre-existing traits may explain the pervasive, recurrent evolution of ECM associations. While the evolution of the ECM habit releases selection on genes required to access plant carbon in the soil, these genes can be coopted for functional specialization in the fungus’s ability to access nutrients, colonize the apoplastic space of the host roots, and/or gain a competitive advantage during community assembly. This specialization may be tightly linked to coevolutionary host interactions, mediated by a heightened adaptability of ECM fungi through a heightened rate of gene expansion and turnover through TE association. Genomic sampling within Russulaceae for this study targeted single representatives of a highly diverse group, so additional sequencing of targeted groups will help to test hypotheses of functional specialization and its relationship to diversification.

Acknowledgements

This research was supported by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research as part of the Plant Microbe Interfaces Scientific Focus Area, at the Oak Ridge National Laboratory. The Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. The work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. This was done within the framework of the Community Sequencing Program #1974 ‘1KFG: Deep Sequencing of Ecologically-relevant Dikarya’ and the Community Sequencing Program #305 ‘Mycorrhizal Genomics Initiative’. Research in the Martin laboratory is also funded by the Laboratory of Excellence Advanced Research on the Biology of Tree and Forest Ecosystems (ARBRE; grant ANR-11-LABX-0002-01), the Region Lorraine Research Council and the European Commission (European Regional Development Fund). BL would like to thank the Laboratory of Excellence ARBRE for a visiting researcher grant. Thanks are also due to Daniel Lindner at the USFS Northern Research Station and Otto Miettinen at the University of Helsinki for taxonomic expertise and sample support. We would like to thank Prof. Joseph Spatafora and Rytas Vilgalys for their support for this project, as well as the managing

editor and anonymous reviewers for constructive comments. The authors declare there are no competing interests.


Author contributions

BL, JL, PBM and FMM conceived the project. FMM coordinates the Mycorrhizal Genomics Initiative. IVG and KWB coordinated the genome sequencing, assembly and annotation at JGI. SM, BL, AKohler and EM performed comparative genome analyses. ED and BH provided updated CAZy profiles. BL, DH, LGN and PEC extracted DNA and RNA. AL, KL, AKuo, MN, WA, GH, JJ, RR, JP and AT carried out DNA sequencing, RNA sequencing, genome assembly or gene prediction at JGI. BL, SM and FMM wrote and edited the manuscript. All co-authors read and edited the manuscript. BL and SM are joint first authors, and JL and FMM are joint senior authors.

Code availability

Visual omics SHINGOTOOLS including PRINGO, TINGO, SynGO and VINGO are available at GitHub: <https://github.com/ShingoMiyachi>.

ORCID

Kerrie W. Barry  <https://orcid.org/0000-0002-8999-6785>
 Pierre Emmanuel Courty  <https://orcid.org/0000-0003-2789-7818>
 Elodie Drula  <https://orcid.org/0000-0002-9168-5214>
 Igor V. Grigoriev  <https://orcid.org/0000-0002-3136-8903>
 Bernard Henrissat  <https://orcid.org/0000-0002-3434-8588>
 David Hibbett  <https://orcid.org/0000-0002-9145-3165>
 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>
 Jesse Labbé  <https://orcid.org/0000-0003-0368-2054>
 Brian Looney  <https://orcid.org/0000-0001-5342-9909>
 Francis M. Martin  <https://orcid.org/0000-0002-4737-3715>
 Shingo Miyachi  <https://orcid.org/0000-0002-0620-5547>
 László G. Nagy  <https://orcid.org/0000-0002-4102-8566>
 Robert Riley  <https://orcid.org/0000-0003-0224-0975>

Data availability

The genome assemblies used for this study are available on the JGI fungal genome portal, MYCOCOSM (<https://mycocosm.jgi.doe.gov>). The details of JGI genomes used for the study, including DDBJ/ENA/GenBank accession numbers, are shown in Supporting Information Table S1. The latest CAZyme annotations are available upon request from CAZy team, Aix-Marseille University, France.

References

Alfaro ME, Faircloth BC, Harrington RC, Sorenson L, Friedman M, Thacker CE, Oliveros CH, Černý D, Near TJ. 2018. Explosive diversification of

marine fishes at the Cretaceous-Palaeogene boundary. *Nature Ecology & Evolution* 2: 688–696.

- Beeck MOD, Troein C, Peterson C, Persson P, Tunlid A. 2018. Fenton reaction facilitates organic nitrogen acquisition by an ectomycorrhizal fungus. *New Phytologist* 218: 335–343.
- Beenken L. 2004. *Die Gattung Russula: Untersuchungen zu ihrer Systematik anhand von Ektomykorrhizen*. Munich, Germany: Ludwig-Maximilians-Universität Press.
- 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 4 October 2021].
- Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. ANTI-SMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* 47: W81–W87.
- Bödeker ITM, Clemmensen KE, De Boer W, Martin F, Olson Å, Lindahl DB. 2014. Ectomycorrhizal *Cortinarius* species participate in enzymatic oxidation of humus in northern forest ecosystems. *New Phytologist* 203: 245–256.
- Brundrett MC, Tedersoo L. 2018. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist* 220: 1108–1115.
- Bushley KE, Turgeon BG. 2010. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evolutionary Biology* 10: 1–26.
- Carroll CS, Moore MM. 2018. Ironing out siderophore biosynthesis: a review of non-ribosomal peptide synthetase (NRPS)-independent siderophore synthetases. *Critical Reviews in Biochemistry and Molecular Biology* 53: 356–381.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17: 540–552.
- 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.
- Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics* 18: 71–86.
- Clericuzio M, Cassino C, Corana F, Vidari G. 2012. Terpenoids from *Russula lepida* and *R. amarissima* (Basidiomycota, Russulaceae). *Phytochemistry* 84: 154–159.
- Csurös M. 2018. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26: 1910–1912.
- DeBie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22: 1269–1271.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUTI and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.
- Gel B, Diez-Villanueva A, 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. KARYOPLOTER: an R/BIOCONDUCTOR package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33: 3088–3090.
- Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S *et al.* 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences, USA* 108: 1513–1518.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, Riley R, Salamov A, Zhao X, Korzeniewski F *et al.* 2014. MYCOCOSM portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* 42: 699–704.
- Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. *circIzize* implements and enhances circular visualization in R. *Bioinformatics* 30: 2811–2812.
- Hage H, Miyachi 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.

- Hammani K, Cook WB, Barkan A. 2012. RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *Proceedings of the National Academy of Sciences, USA* 109: 5651–5656.
- Hess J, Pringle A. 2014. The natural histories of species and their genomes: asymptotic and ectomycorrhizal *Amanita* fungi. *Advances in Botanical Research* 70: 235–257.
- Hess J, Skrede I, Chaib De Mares M, 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.
- Hibbett DS. 2006. A phylogenetic overview of the Agaricomycotina. *Mycologia* 98: 917–925.
- Hibbett DS, Grimaldi D, Donoghue MJ. 1997. Fossil mushrooms from Miocene and Cretaceous ambers and the evolution of Homobasidiomycetes. *American Journal of Botany* 84: 981–991.
- Hortal S, Plett KL, Plett JM, Cresswell T, Johansen M, Pendall E, Anderson IC. 2017. Role of plant–fungal nutrient trading and host control in determining the competitive success of ectomycorrhizal fungi. *The ISME Journal* 11: 2666–2676.
- Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software v.7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- 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.
- Koide RT, Fernandez C, Malcolm G. 2014. Determining place and process: functional traits of ectomycorrhizal fungi that affect both community structure and ecosystem function. *New Phytologist* 201: 433–439.
- Krassilov VA, Makulbekov NM. 2003. The first finding of Gasteromycetes in the Cretaceous of Mongolia. *Paleontological Journal* 37: 439–442.
- Krizsán K, Almási É, Merényi Z, Sahu N, Virágh M, Kószó T, Mondo S, Kiss B, Bálint B, Kües U *et al.* 2019. Transcriptomic atlas of mushroom development reveals conserved genes behind complex multicellularity in fungi. *Proceedings of the National Academy of Sciences, USA* 116: 7409–7418.
- Kuo A, Bushnell B, Grigoriev IV. 2014. Fungal genomics: sequencing and annotation. In: Martin F, ed. *Fungi. Advances in botanical research*. Cambridge, UK: Elsevier Academic Press, 1–52.
- Labbé J, Murat C, Morin E, Tuskan GA, Le Tacon F, Martin F. 2012. Characterization of transposable elements in the ectomycorrhizal fungus *Laccaria bicolor*. *PLoS ONE* 7: e40197.
- Larsson E, Larsson KH. 2003. Phylogenetic relationships of russuloid basidiomycetes with emphasis on aphyllophorean taxa. *Mycologia* 95: 1037–1065.
- Lilleskov EA, Hobbie EA, Fahey TJ. 2002. Ectomycorrhizal fungal taxa differing in response to nitrogen deposition also differ in pure culture organic nitrogen use and natural abundance of nitrogen isotopes. *New Phytologist* 154: 219–231.
- 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.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42: D490–D495.
- 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 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.
- López-Giráldez F, Townsend JP. 2011. PHYDESIGN: an online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology* 11: 1–4.
- Lutzoni F, Nowak MD, Alfaro ME, Reeb V, Miadlikowska J, Krug M, Arnold AE, Lewis LA, Swofford DL, Hibbett D *et al.* 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications* 9: 1–11.
- Mäkipää R, Rajala T, Schigel D, Rinne KT, Pennanen T, Abrego N, Ovaskainen O. 2017. Interactions between soil- and dead wood-inhabiting fungal communities during the decay of Norway spruce logs. *The ISME Journal* 11: 1964–1974.
- Martin F, Aerts A, Ahrén D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V *et al.* 2008. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452: 88–92.
- Martin F, Kohler A, Murat C, Veneault-Fourrey C, Hibbett DS. 2016. Unearthing the roots of ectomycorrhizal symbioses. *Nature Reviews Microbiology* 14: 760–773.
- Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, Sherlock G, Snyder M, Wang Z. 2010. RNNOTATOR: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics* 11: 663.
- 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: 1–17.
- Morin E, Miyauchi S, San Clemente H, Chen EC, Pelin A, de la Providencia I, Ndikumana S, Beaudet D, Hainaut M, Drula E *et al.* 2019. Comparative genomics of *Rhizopogon irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. *New Phytologist* 222: 1584–1598.
- Muszewska A, Taylor JW, Szczesny P, Grynberg M. 2011. Independent subtilase expansions in fungi associated with animals. *Molecular Biology and Evolution* 28: 3395–3404.
- Nakasono KK. 1990. *Cultural studies and identification of wood-inhabiting Corticiaceae and selected Hymenomycetes from North America*. St Paul, MN, USA: APS Press.
- Navarro-Muñoz JC, Selem-Mojica N, Mallowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S *et al.* 2020. A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology* 16: 60–68.
- Nichols DJ, Johnson KR. 2008. *Plants and the KT Boundary*. Boston, MA, USA: Cambridge University Press.
- 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: 1–15.
- Pellitier PT, Zak DR. 2018. Ectomycorrhizal fungi and the enzymatic liberation of nitrogen from soil organic matter: why evolutionary history matters. *New Phytologist* 217: 68–73.
- Plett JM, Yin H, Mewalal R, Hu R, Li T, Ranjan P, Jawdy S, De Paoli HC, Butler G, Burch-Smith TM *et al.* 2017. *Populus trichocarpa* encodes small, effector-like secreted proteins that are highly induced during mutualistic symbiosis. *Scientific Reports* 7: 1–13.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarisation in Bayesian phylogenetics using TRACER 1.7. *Systematic Biology* 67: 901–904.
- Sánchez-García M, Matheny PB. 2017. Is the switch to an ectomycorrhizal state an evolutionary key innovation in mushroom-forming fungi? A case study in the Tricholomatineae (Agaricales). *Evolution* 71: 51–65.
- Sánchez-García M, Ryberg M, Khan FK, 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.
- Schloerke B, Cook D, Larmarange J, Briatte F, Marbach M, Thoen E, Elberg A, Crowley J. 2021. *GGALLY: extension to 'ggplot2'*. R package v.2.1.1. [WWW document] URL <https://CRAN.R-project.org/package=GGally> [accessed 3 September 2021].
- Sebastiana M, Martins J, Figueiredo A, Monteiro F, Sardans J, Peñuelas J, Silva A, Roepstorff P, Pais MS, Coelho AV. 2017. Oak protein profile alterations upon root colonization by an ectomycorrhizal fungus. *Mycorrhiza* 27: 109–128.
- Shah F, Nicolás C, Bentzer J, Ellström M, Smits M, Rineau F, Canbäck B, Floudas D, Carleer R, Lackner G *et al.* 2016. Ectomycorrhizal fungi decompose soil organic matter using oxidative mechanisms adapted from saprotrophic ancestors. *New Phytologist* 209: 1705–1719.

- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. Genome analysis BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.
- Smith GR, Finlay RD, Stenlid J, Vasaitis R, Menkis A. 2017. Growing evidence for facultative biotrophy in saprotrophic fungi: data from microcosm tests with 201 species of wood-decay basidiomycetes. *New Phytologist* 215: 747–755.
- Smith SY, Currah RS, Stockey RA. 2004. Cretaceous and eocene poroid hymenophores from Vancouver Island, British Columbia. *Mycologia* 96: 180–186.
- Smith SE, Read DJ. 2010. *Mycorrhizal symbiosis*. New York, NY, USA: Academic Press.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
- 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.
- Sultana T, Zamborlini A, Cristofari G, Lesage P. 2017. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nature Reviews Genetics* 18: 292–308.
- Talbot JM, Allison SD, Treseder KK. 2008. Decomposers in disguise: mycorrhizal fungi as regulators of soil C dynamics in ecosystems under global change. *Functional Ecology* 22: 955–963.
- Talbot JM, Martin F, Kohler A, Henrissat PKG. 2015. Functional guild classification predicts the enzymatic role of fungi in litter and soil biogeochemistry. *Soil Biology and Biochemistry* 88: 441–456.
- 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.
- Tedersoo L, Smith ME. 2017. Ectomycorrhizal fungal lineages: detection of four new groups and notes on consistent recognition of ectomycorrhizal taxa in high-throughput sequencing studies. In: Tedersoo L, eds. *Biogeography of mycorrhizal symbiosis. Ecological studies (analysis and synthesis), vol. 230*. Cham, Switzerland: Springer.
- Veneault-Fourrey C, Commun C, Kohler A, Morin E, Balestrini R, Plett J, Danchin E, Coutinho P, Wiebenga AD, de Vries RP *et al.* 2014. Genomic and transcriptomic analysis of *Laccaria bicolor* CAZome reveals insights into polysaccharides remodelling during symbiosis establishment. *Fungal Genetics and Biology* 72: 168–181.
- Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R *et al.* 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research* 42: D581–591.
- Weiß M, Waller F, Zuccaro A, Selosse MA. 2016. Sebaciales – one thousand and one interactions with land plants. *New Phytologist* 211: 20–40.
- Wright ES. 2015. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics* 16: 1–14.
- Zak DR, Pellitier PT, Argiroff WA, Castillo B, James TY, Nave LE, Averill C, Beidler KV, Bhatnagar J, Blesh J *et al.* 2019. Exploring the role of ectomycorrhizal fungi in soil carbon dynamics. *New Phytologist* 223: 33–39.
- Zerbino DR, Birney E. 2008. VELVET: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.
- Zhang Z, Yuan Y, Liu Q, Yin H. 2019. Plant nitrogen acquisition from inorganic and organic sources via root and mycelia pathways in ectomycorrhizal alpine forests. *Soil Biology and Biochemistry* 136: 107517.

Supporting Information

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

Fig. S1 Phylogenomic reconstruction of Agaricomycetes and Russulales under maximum likelihood criterion.

Fig. S2 Total gene repertoire of 18 Russulales.

Fig. S3 Genomic features driving the evolution of Russulales.

Fig. S4 Presence and absence of CAZyme domains.

Fig. S5 Presence and abundance of CAZyme groups coded in genomes.

Fig. S6 Phylogenetic reconstruction of the class-II peroxidase (POD) gene family from Russulales that codes for lignin peroxidase.

Fig. S7 Predicted secondary metabolite biosynthesis gene clusters.

Fig. S8 Secondary metabolite cluster homology analysis for different classes of SMCs.

Fig. S9 Evolutionary rate Dollo parsimony COUNT analysis of Russulales.

Fig. S10 Genome coverage of TEs identified.

Fig. S11 Distribution of TE–CAZyme coding gene distance.

Fig. S12 Density of the size of genes coding for various protein categories.

Fig. S13 Density of TE–gene distance of the selected genes.

Fig. S14 Genomic locations of genes for CAZymes in syntenic regions.

Fig. S15 Number of CAZyme-coding genes in synteny with *G. convolvans*.

Fig. S16 Number of genes coding for nonsecreted proteins in synteny with *G. convolvans*.

Fig. S17 Number of TEs in synteny with *G. convolvans*.

Fig. S18 Genome size and scaffold number.

Fig. S19 Overall synteny comparison of 18 Russulales.

Fig. S20 Mesosynteny Hanabi plots with five *Russula*.

Fig. S21 Number of genes in Mesosynteny.

Methods S1 DNA extraction.

Methods S2 RNA extraction.

Methods S3 Quality assessment.

Methods S4 Genome sequencing and assembly.

Methods S5 Transcriptome sequencing and assembly.

Methods S6 Genome annotation.

Methods S7 References.

Table S1 Features of Russulales genomes.

Table S2 JGI taxon set for molecular dating analysis.

Table S3 Count of predicted secreted gene repertoire.

Table S4 Orthologous gene clusters contracted in the ancestry of Russulales.

Table S5 CAFE orthogroup gene counts.

Table S6 Orthologue annotation from CAFE.

Table S7 Orthologous gene clusters expanded in the ancestry of Russulales.

Table S8 Small secreted proteins in synteny.

Table S9 Genome-wide TE–gene distance for species of Russulales.

Table S10 Genomic feature count in microsynteny.

Table S11 Genomic features used for PERMANOVA.

Table S12 PERMANOVA results.

Table S13 Pairwise PERMANOVA for statistical significance in fungal ecology.

Table S14 Number of cell wall-degrading enzymes.

Table S15 Permuted TE–gene mean distances.

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.