

High rate of gene family evolution in proximity to the origin of ectomycorrhizal symbiosis in Inocybaceae

Faheema Kalsoom Khan¹ , Marisol Sánchez-García² , Hanna Johannesson^{1,3,4}  and Martin Ryberg¹ 

¹Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, 752 36, Sweden; ²Department of Forest Mycology and Plant Pathology, Uppsala Biocentre, Swedish University of Agricultural Sciences, Uppsala, SE-75005, Sweden; ³Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, 106 91, Sweden;

⁴The Royal Swedish Academy of Sciences, Stockholm, 114 18, Sweden

Summary

Authors for correspondence:

Faheema Kalsoom Khan

Email: faheema.kalsoom.khan@ebc.uu.se

Martin Ryberg

Email: martin.ryberg@ebc.uu.se

Received: 3 May 2024

Accepted: 9 July 2024

New Phytologist (2024) **244**: 219–234

doi: 10.1111/nph.20007

Key words: CAZymes, comparative genomics, fungi, Inocybaceae, mycorrhiza, phylogenomics, saprotrophs.

- The genomes of ectomycorrhizal (ECM) fungi have a reduced number of genes encoding Carbohydrate-Active EnZymes (CAZymes), expansions in transposable elements (TEs) and small secreted proteins (SSPs) compared with saprotrophs. Fewer genes for specific peptidases and lipases in ECM fungi are also reported. It is unclear whether these changes occur at the shift to the ECM habit or are more gradual throughout the evolution of ECM lineages.
- We generated a genomic dataset of 20 species in the ECM lineage Inocybaceae and compared them with six saprotrophic species.
- Inocybaceae genomes have fewer CAZymes, peptidases, lipases, secondary metabolite clusters and SSPs and higher TE content than their saprotrophic relatives. There was an increase in the rate of gene family evolution along the branch with the transition to the ECM lifestyle. This branch had very high rate of evolution in CAZymes and had the largest number of contractions. Other significant changes along this branch included expansions in transporters, transposons-related genes and communication genes such as fungal kinases.
- There is a high concentration of changes in proximity to the transition to the ECM lifestyle, which correspond to the identified key changes for the gain of this lifestyle.

Introduction

Fungi serve important roles for nutrient recycling such as degrading complex organic matter and facilitating nutrient uptake by plants. The latter function is performed in a mutualistic relation called mycorrhiza (Lebreton *et al.*, 2021). Ectomycorrhiza (ECM) is the dominant mycorrhizal type among stand-forming trees in terrestrial ecosystems, and *c.* 60% of all trees on Earth depend on this symbiosis (Steidinger *et al.*, 2019). ECM fungi have evolved independently from saprotrophic ancestors in multiple lineages through convergent evolution (Brundrett & Teder-soo, 2018; Sheikh *et al.*, 2022).

Carbohydrate-Active enZymes (CAZymes), peptidases and lipases are integral to fungal metabolism, enabling them to utilize efficiently a wide range of substrates for energy, growth and facilitate adaptation to various habitats (El-Gendi *et al.*, 2021). Their diverse functions support fungal ecological roles in nutrient cycling, decomposition and environmental interactions (Tláškal *et al.*, 2021; Corbu *et al.*, 2023). These enzymes are also major components of the fungal secretome, playing a key role in organic matter decomposition and symbiosis development (Alfaro *et al.*, 2014; Venkatesh *et al.*, 2021). Other components of the secretome include small secreted proteins (SSPs) and secondary metabolites (SMs), which mediate fungal interactions with their hosts and environment (Martin *et al.*, 2008; Pellegrin *et al.*, 2015; Lofgren *et al.*, 2021).

Comparative genomic studies have revealed characteristic genomic features common in several ECM lineages, such as reductions in the number of genes encoding CAZymes, especially lignocellulolytic enzymes such as cellulases and hydrolases (Kohler *et al.*, 2015; Miyauchi *et al.*, 2020). However, individual ECM fungal genomes have also retained unique sets of CAZymes, for example endoglucanases, pectinases and oxidoreductases/laccases, suggesting diverse capabilities of ECM lineages to degrade plant materials from litter and soil (Kohler *et al.*, 2015). Lower gene content of specific peptidases (serine peptidases) and lipases (versatile lipases) in the genomes of ECM fungi as compared to saprotrophs has also been reported (Barriuso & Martínez, 2017; Muszewska *et al.*, 2017). Other ECM-related changes include an increase in transposable elements (TEs) and expansions of symbiosis-induced SSPs (Kohler *et al.*, 2015; Peter *et al.*, 2016; Hess *et al.*, 2018). However, it is yet unclear, how general these patterns of evolution are for the ECM lineages. A few studies, based on single lineages of ECM fungi, have indicated that some of the ECM-related genomic changes, for example contractions in CAZymes and expansions in TEs, have occurred before the transition to ECM in the saprotrophic ancestors as a pre-adaptation for this mutualistic lifestyle (Hess *et al.*, 2018; Looney *et al.*, 2022). Despite the recent progress in the ECM fungal genomics, it is not well-known whether the ECM-linked genomic changes occur at the transition to the

ECM habit or are part of a trend before and/or after the transition.

To identify when the changes have occurred in relation to the transition to ECM, it is important to choose a fungal group with a long evolutionary history as ECM since changes may accumulate slowly and gradually. Additionally, including a well-supported saprotrophic sister outgroup is important to pinpoint the branch in the phylogeny with the transition to the ECM lifestyle. Similarly, a diverse lineage is needed to be able to break up long branches and avoid inference of branch-specific changes as part of a general trend. Hence, we chose Inocybaceae lineage for our investigation.

Inocybaceae is a diverse lineage of ECM fungi, with over 1050 estimated species (Matheny *et al.*, 2006; Matheny & Kudzma, 2019). It has an estimated origin around the Cretaceous–Paleogene boundary (Matheny *et al.*, 2009; Ryberg & Matheny, 2012; Sheikh *et al.*, 2022), and the relationship with the sister taxon (Crepidotaceae) is well-known (Matheny *et al.*, 2009). The Crepidotaceae species are predominantly wood saprotrophs, cosmopolitan in distribution and are relatively easy to culture (Singer, 1986; Aime, 1999). However, Inocybaceae is notoriously difficult to culture from both sporocarp tissues and single spores (Fries, 1982), and to circumvent this issue for genomic analyses, we used DNA extracted directly from sporocarp samples in this study. This also decreased the processing time per sample. To further facilitate handling of samples, we preserved the specimens by desiccation instead of freezing in liquid nitrogen like other genomic studies of mushroom-forming fungi (Chang *et al.*, 2019; Looney *et al.*, 2022; Wu *et al.*, 2022). While RNA extraction is not possible from the dried sporocarps, they are easier to transport and store, which simplifies the overall fieldwork and makes it less critical to do the laboratory work immediately after the fieldwork.

To resolve what evolutionary processes acted and what evolutionary events occurred before, at and after the origin of the ECM symbiosis, we performed comparative genomic analyses of 20 species of Inocybaceae and six saprotrophic outgroup species (two of those from Crepidotaceae). We present the first extensive genomic study of Inocybaceae and introduce a method that enables genomic data extraction from dried sporocarps. Our analyses identified the rapidly evolving ECM-related gene families in relation to the transition to the ECM in Inocybaceae. Because of their major importance in fungal metabolism and direct connection to nutritional mode, we focused on CAZymes, peptidases and lipases for gene family evolution analyses and revealed changes in evolutionary rates in these gene categories before, in close proximity to, and/or after the transition to the ECM lifestyle. In addition, we compared the gene content of major components of the secretome between ECM and saprotrophic species.

Materials and Methods

Fungal material, DNA extraction and sequencing

We generated sequence data for 19 Inocybaceae genomes so that each of the four genera present in the north temperate region are

represented by at least two genomes. As the *Inocybe sensu stricto* is the most species-rich genus, most genomes (13) were generated for this genus. Additionally, we sequenced two genomes from the Inocybaceae's sister lineage Crepidotaceae. All fungal material used for genome sequencing was sampled fresh and thereafter dried using a food dehydrator for later DNA extraction, except *Crepidotus cessati* (Rabenh.) Sacc., which was obtained as dried material from the Herbarium GB in Sweden. *Mallocybe terrigena* (Fr.) Vizzini *et al.* sequenced in Bahram *et al.* (2018) was also added to the ingroup, while four other genomes, that is *Agaricus bisporus* (J.E. Lange) Imbach, *Panaeolus papilionaceus* (Bull.) Quél., *Gymnopilus junonius* (Fr.) P.D. Orton and *Hypholoma sublateritium* (Fr.) Quél., were retrieved from the JGI MycoCosm (Grigoriev *et al.*, 2014) database (<https://mycocosm.jgi.doe.gov>) and added as distant outgroups. This resulted in a dataset of 26 genomes (Supporting Information Table S1).

DNA was extracted from the lamella (spore-bearing structure; 20–50 mg) to capture high amount of total DNA using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, Irvine, CA, USA) and following the manufacturer's instructions. Tissue was taken from only one sporocarp per species to avoid intraspecific genetic variation, except for *C. calolepis*, where two sporocarps from the same collection were used to obtain sufficient amount of DNA for genome sequencing. DNA libraries for each sample were prepared using PCR-free TrueSeq from the total DNA, and all samples together were sequenced on one lane of Illumina HiSeq X (PE 150 bp insert size) by the Sci-Life Laboratory in Uppsala, Sweden.

Quality check, contaminants filtering and assembly

Raw genomic reads of the sequenced samples were quality-checked using FASTQC 0.11.8 (Andrews, 2010) and adapters trimmed using TRIMMOMATIC 0.36 (Bolger *et al.*, 2014). To extract reads of the target genome, the median coverage and GC content of the Kmers were used to identify reads sourcing from contaminants such as bacteria using Kmer spectra in the Kmer Analysis Toolkit (KAT) 2.4.2 (Mapleson *et al.*, 2017) with default settings. The reads outside the expected GC and the reads with low coverage were filtered out using the Kat filter function in the KAT, and the remaining reads were assembled using the haplotype-aware assembler SPADes 2.4.2 (Prijbelski *et al.*, 2020) in the metaspades mode. Bacterial contaminations in the assemblies were further assessed using the blobplots function and removed using the workflow A in BLOBTOOLS 1.1.1 (Laetsch & Blaxter, 2017).

Contamination from other fungi was identified first, based on low coverage of contigs and then confirmed by BLAST 2.10.1+ (Altschul *et al.*, 1990) searches of the entire assemblies against GenBank's nucleotide (nt) database. The mitochondrial contigs were detected by Blast searches of the assemblies against the mitochondrial genome of *Laccaria bicolor* (Maire) P.D. Orton and were confirmed by the relatively low GC content of the contigs. The reads, comprising the contigs from contaminants or mitochondria, were removed from the total reads. This final set of filtered reads was assembled, and the assemblies were improved by

polishing with Pilon 1.22 (Walker *et al.*, 2014) in one round. Contaminant removal was verified by BLASTing the re-assembled assemblies against a custom-made database containing ribosomal large subunit sequences from common contaminants and expected target species. Completeness of the assemblies were assessed by BUSCO 4.1.4 (Simão *et al.*, 2015) using the dataset basidiomycota_odb10. To evaluate the potential loss of genes from the target genome due to the filtration process, the BUSCO scores of the assemblies before and after the filtration were compared.

Annotation

Gene predictions and annotation of all assemblies, including the downloaded outgroups, were performed with the FUNANNOTATE pipeline 1.8.4 (Palmer & Stajich, 2019). Transcriptomes of four closely related fungal species (*C. versutus* as *C. cesatii* in the database, *Hebeloma cylindrosporum* Romagn., *Cortinarius austrovenetus* Cleland and *L. bicolor*), retrieved from the MycoCosm (JGI) database, were used to facilitate protein prediction. Specific protein-coding genes such as CAZymes were annotated by the DBCAN2 pipeline 2.0.11 (Yin *et al.*, 2012), which predicts the different classes (Glycoside Hydrolases, Carbohydrate-Binding Modules, Polysaccharide Lyases, Auxiliary Activities, Carbohydrate Esterases, Glycosyl Transferases) of CAZymes. Secretomes were annotated using the pipeline described by Pellegrin *et al.* (2015). Secondary metabolite gene clusters (SMCs) and peptidases were annotated using annotate function in the funannotate pipeline with antiSMASH and MEROPS databases, respectively. Lipases were annotated by BLASTing the predicted protein sequences of all samples against the Lipase Engineering (LED) database (<http://www.led.uni-stuttgart.de>) at $E^{-\text{value}} = 10^{-5}$ cut-off and classified into three major classes of lipases, that is GX, GGGX and Y according to the LED database.

Phylogenomic analysis

The orthologous genes, both single-copy and orthogroups, in the 26 genomes were identified using phylogeny-based orthology inference approach implemented in ORTHOFINDER 2.5.2 (Emms & Kelly, 2019), with DIAMOND for sequence similarity search and local alignment (Buchfink *et al.*, 2015) and DendroBLAST algorithm for gene tree inference (Kelly & Maini, 2013). Each of the resulting single-copy orthologous gene sets were aligned using MAFFT 7.407 (Katoh & Standley, 2013) with -auto option, and the ambiguous/poorly aligned regions were removed using -automated1 option in TRIMAL 1.4.1 (Capella-Gutiérrez *et al.*, 2009) and were then concatenated into a supermatrix. A Maximum Likelihood (ML) phylogeny was constructed using IQ-TREE 2.0 (Nguyen *et al.*, 2015) with -testmerge option and 1000 rapid bootstrap iterations. ML gene trees were also obtained using IQ-Tree with -mfp option. Branch support values for each gene tree were calculated in IQ-Tree using SH-aLRT (Guindon *et al.*, 2010) with 1000 replicates. The ML gene trees were used to infer a species tree using the coalescent-based method implemented in ASTRAL III 5.7.7 (Mirarab & Warnow, 2015).

To further estimate the support for our obtained topology, we counted the total number of individual gene trees, which supported the monophyly of genera within Inocybaceae, as well as the relationships between these genera proposed by Matheny *et al.* (2020) and the relationships between the genera as derived from our concatenated and Astral analyses. In cases where discrepancies were found between our analyses and Matheny *et al.* (2020), we tested differences in the tree length and support values of the gene trees, which either supported or did not support the monophyly of the group in question. Using a randomization test, we compared the mean tree length and support values for the relationship supported by the fewest trees to 1000 random draws of an equivalent number of gene trees from the total pool of relevant gene trees (i.e. those containing the taxa needed to infer the relationship).

The estimation of divergence times was done on the ML phylogeny with r8s 1.81 (Sanderson, 2003), using a penalized likelihood method. Cross-validation was performed to identify the best smoothing parameter value, and the tree was calibrated by fixing the age of the Inocybaceae crown group to 89.9 million years ago (Sheikh *et al.*, 2022). Even though there is uncertainty associated with this age estimate, we did not include confidence intervals in the analysis. These intervals would also be approximate, and there is no intrinsic way of incorporating the uncertainty into the downstream analyses.

Comparative genomic analysis

The number of total and secreted CAZymes, peptidases, lipases, SSPs and SMCs was counted and compared between the genomes. In addition, gene family evolution was estimated using CAFE 4.0 (Han *et al.*, 2013) and with branch-sets defined to identify changes in relation to the transition to ECM (Fig. 1). The branch between the stem and crown node of Inocybaceae is referred to as the branch with the transition to the ECM. The branches from the root of the tree leading up to the stem node of Inocybaceae are referred to as pre-ECM branches. The branches within the Inocybaceae clade are referred to as the ECM branches. The analyses were conducted in two sets. The first set was performed with all identified orthogroups, in order to investigate the overall evolution of gene families. To identify rapidly evolving orthogroups at each branch-set, the orthogroups were sorted according to the *P*-value of the changes occurring along each branch-set. In addition, to test whether there has been any rate change associated with transition to ECM, the branches of the tree were divided into three categories, as above, the transition branch to ECM, and ECM branches, but the third category with all non-ECM branches (including the pre-ECM branches; Fig. 1). The ML of the gene family evolution was estimated in CAFE under a model with the same rate for all branches (one-rate model), all combinations of the same rate between two of the rate categories and different rate for the third (two-rate model), and different rate for all branch-sets (three-rate model). For the three-rate model, ML estimation was repeated multiple times to ensure convergence on a single optimum. The one-rate model, the two-rate model with the lowest likelihood score and

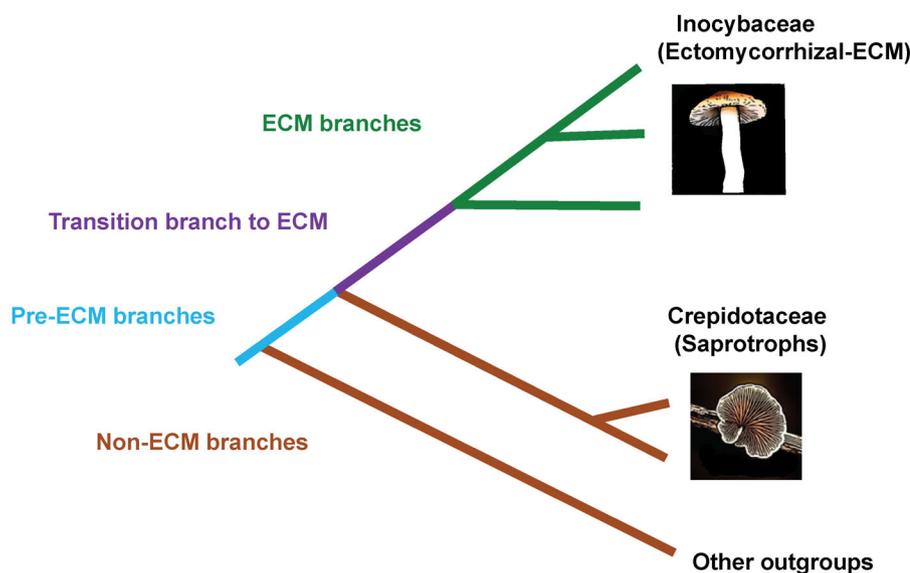


Fig. 1 Representation of the branch-sets used in the phylogeny of Inocybaceae, Crepidotaceae and other outgroups. Non-ectomycorrhizal (ECM) branches represent the ancestral and saprotrophic outgroup branches and are colored in brown or blue. Pre-ECM branches represent the branches from the root of the tree up to the stem node of the ECM and are colored in blue. Transition branch to ECM is represented by purple color. The ECM in-group branches are colored in green.

the three-rate model were compared with a likelihood-ratio test (LRT) using the ‘scipy.stats’ module from the SciPy library in Python (Virtanen *et al.*, 2020).

In the second set of the comparative analyses, we analyzed the CAZyme, peptidase and lipase gene families. The one-rate, two-rate and three-rate models (as mentioned in the previous section) were compared. Furthermore, we assessed probability of the observed difference in ML score between the models and the fit of the three models to our dataset through simulations conducted in CAFE. This was done by generating 1000 simulated datasets for each of these gene categories, based on the one-rate, two-rate and three-rate models. Thereafter, the three models were optimized on each of the simulated datasets in all gene categories.

Results

Phylogeny of Inocybaceae

A total of 339 single-copy orthologues were identified and a concatenated alignment of 183 927 amino acid positions was compiled. The ML tree based on the concatenated alignment and the Astral tree yielded identical topologies (Figs 2a, S1). All the branches in the tree had 100% bootstrap support based on the concatenated alignment, and local posterior probability (lpp) values were 1.00 in the Astral analysis except for the branch, clustering *I. asterospora* and *I. oblectabilis*, which had a lpp support of 0.98. Quartet supports from the Astral analyses were, as expected lower, showing the conflict between gene trees. The monophyly of Inocybaceae had a quartet support of 89% (Fig. 2a). The relationship of Inocybaceae and Crepidotaceae had 87% quartet support. All the genera in Inocybaceae had strong quartet support (> 93%; Fig. 2a). However, the placement of the *Mallocybe* branch within the Inocybaceae got a quartet support of only 53%, and many branches within *Inocybe* had lower support.

The analysis of individual gene trees also showed that the vast majority support the genera in Inocybaceae as monophyletic. *Inocybe* was the genus where the least number of gene trees indicated that it is monophyletic (88%). Our phylogeny indicates that the *Mallocybe* is sister to the rest of *Inocybaceae* instead of sister to *Inosperma* as inferred by Matheny *et al.* (2020). Only 68 (20%) of the gene trees in our analysis indicate monophyly of *Mallocybe* with *Inosperma* while the rest (266 trees; 80%) indicate nonmonophyly. The gene trees, supporting the sister-relationship of the two genera, had no significant difference in tree length or clade support compared to the overall population of gene trees.

Genomic features

The filtration of contaminants, such as bacteria and other fungi such as Ascomycetes, resulted in the removal of 11–55% of the total reads depending on the sample (Table 1). We noticed that the assemblies of some of the heavily contaminated samples, with the presence of both bacterial and fungal contaminants, had high Busco scores (up to 97%) initially with high count of duplicates (up to 15%). These scores dropped significantly after filtration of the contaminants from the assemblies, resulting in improved duplicate counts of *c.* 1% (Table S2). However, heavily contaminated samples lost more Busco genes in the process than samples with low contamination (Table S2).

The genome assemblies from the Inocybaceae species were in the range of 27–51 MB and the two Crepidotaceae species had assembly sizes of 48 and 54 MB. *Inocybe petiginosa* had the smallest assembly size (27 MB) while *Pseudosperma bulbosissimum* (51 MB) from the ECM fungi and *C. cesatii* (58 MB) from the saprotrophs had the largest assembly sizes (Table 1). The median Busco completeness was 94.5%, with a median of only 2% duplicated genes (Table 1). The total number of predicted protein-coding genes ranged from 8615 in *I. petiginosa* to 17 563 in *C. calolepis*. Smaller assemblies were noticed to have smaller

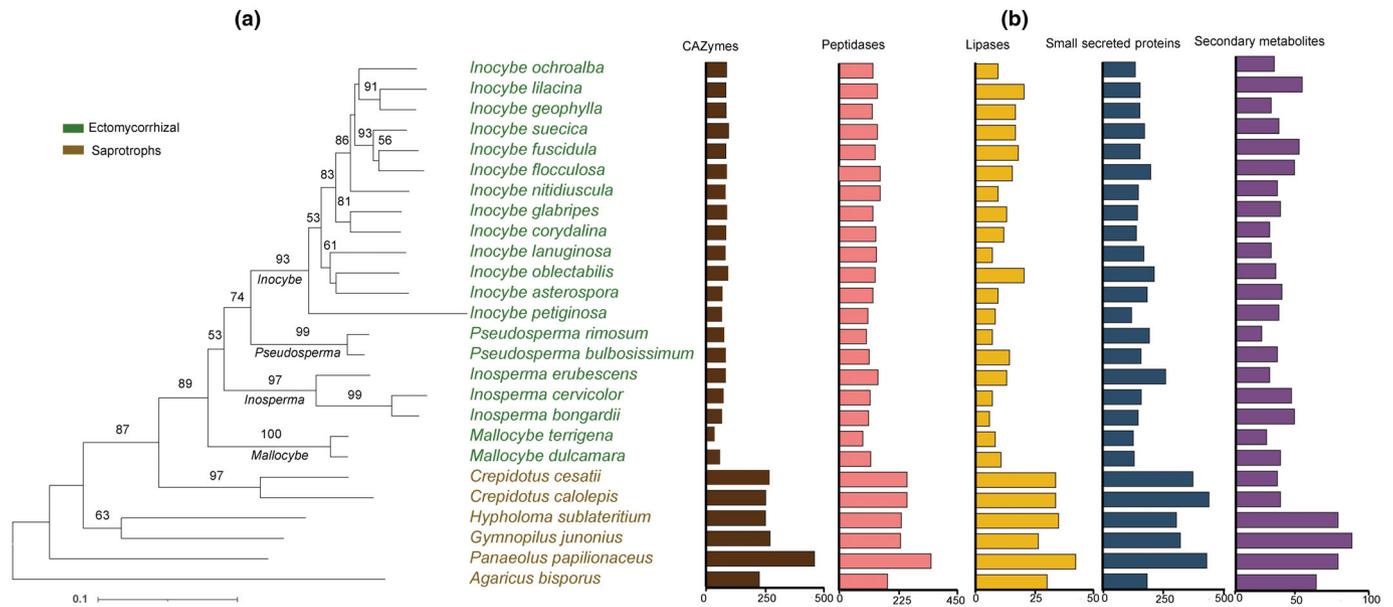


Fig. 2 The Maximum Likelihood (ML) Phylogeny and genome features of Inocybaceae and outgroup species. (a) The ML phylogenomic tree based on 339 single-copy orthologues, in which all nodes got 100% bootstrap support in the ML analysis. The quartet score support, received from the Astral analysis, > 50 is displayed above the branches. The taxa are color coded according to the lifestyles: green for the ECM and brown for saprotrophs (b) Genomic features: Number of genes encoding secreted Carbohydrate-Active enZymes; peptidases; lipases; small secreted proteins; secondary metabolite gene clusters. The scale bar represents average number of nucleotide substitutions per site.

number of total protein-coding genes (Brunner Munzer test, $P < 0.01$). The percentage of masked TEs in the genomes of Inocybaceae were higher (7–21%) than the Crepidotaceae species (3–4%; Table 1). The total secretome was smaller in the ECM species as compared to all the saprotrophic outgroups (Table S3), and there are fewer genes in each gene category of CAZymes, peptidases, lipases, SSPs and SMCs in the genomes of the ECM species than the saprotrophs (Fig. 2b), though the number of SMCs in Crepidotaceae is similar to Inocybaceae species compared with distant outgroups. The comparison using the total gene repertoire in each of the gene categories above showed the same pattern (Table S3). The number of genes encoding SSPs were in the range of 155–259 in the genomes of the ECM species and 338–496 in all saprotrophic species (Table S3). The SSPs constitute 39 to 53% of the total secretome in our ECM species and 37 to 53% in all saprotrophic species.

The total number of orthogroups in the whole dataset, with both the ECM and all outgroup species included, were 17 428. The number of orthogroups in the Inocybaceae species were 13 689, which comprised 2269 (17%) core orthogroups (present in all Inocybaceae species), 9762 (71%) accessory orthogroups (present in at least two species) and 1658 (12%) species-specific orthogroups.

Rapidly evolving gene families

The first part of our CAFE analysis, based on orthogroups in the whole dataset, highlighted rapidly evolving orthogroups at each branch-set, that is pre-ECM branches, transition branch to ECM and the ECM branches (Tables 2, S4). We found 137 significantly large contractions or expansions on the three branch-sets (Fig. S2). The pre-ECM branches showed contractions in gene families with

Pfam domains associated with genes responsible for lignocellulose degradation, such as gloxal oxidase, as well as expansions in gene families related to transposons including retrotransposon gag protein and associated with protein interactions, such as F-box genes. The transition branch to the ECM showed the largest number of significant changes, characterized by expansions in gene families related to transport proteins (e.g. coatamer), repeat-related proteins (e.g. hAT transposon superfamily and tetratricopartite repeats), interaction-related proteins (e.g. BTZ/POZ) and signaling proteins (e.g. pheromone alpha receptor). Subsequently, the ECM branches after the transition were dominated by the expansions in various types of repeat-related proteins (e.g. KDZ transposases and WD repeat-containing domain), polysaccharide deacetylases, fungal protein kinases, NACHT repeat domains and other gene families as the most significant changes.

The three-rate model with different rates for all branch-sets had a slightly but significantly ($P < 0.001$) higher ML (−134 233) (Table S5) than the two-rate model with same rate for non-ECM branches and the transition branch (−134 246). This was also in line with the parameter estimates for the three-rate model, which estimated the highest rate for the ECM branches (0.0046), and similar rates for the non-ECM branches and the transition branch (0.0014, respectively, 0.0019). This is consistent with a rate shift along the transition branch and higher rate of gene family evolution in the ECM state than non-ECM state (Table S5).

Large number of CAZyme gene family contractions along the transition branch to ECM

Inocybaceae genomes have fewer CAZymes (Fig. 2b), both in total genome and the secretome. In addition, they do not have

Table 1 Summary statistics of the 26 assemblies used in the study. The 22 newly assembled genomes are in bold font.

Species	Ecology	Read pairs before filtering (BP)	Read pairs after filtering (BP)	Assembly (MB)	Scaffolds (no.)	N50	BUSCO ^a (Comp. Single + Comp. Duplicates + Fragmented)	Masked TEs ^b (%)	No. of protein-coding genes
<i>Inocybe ochroalba</i>	Ectomycorrhizal	22M	16M	33	4929	12 258	96% (92 + 2 + 2)	7	10 833
<i>Inocybe lilacina</i>	Ectomycorrhizal	24M	11M	34	4634	12 881	81% (77 + 1 + 3)	11	10 857
<i>Inocybe geophylla</i>	Ectomycorrhizal	21M	16M	34	4462	15 594	94% (91 + 1 + 2)	18	10 954
<i>Inocybe suecica</i>	Ectomycorrhizal	30M	15M	44	7765	8410	95% (91 + 2 + 2)	10	13 083
<i>Inocybe fuscidula</i>	Ectomycorrhizal	26M	20M	40	5701	12 535	96% (93 + 1 + 2)	11	11 793
<i>Inocybe flocculosa</i>	Ectomycorrhizal	21M	17M	48	8860	7645	95% (91 + 1 + 3)	11	14 660
<i>Inocybe nitidiuscula</i>	Ectomycorrhizal	20M	15M	49	9473	7142	93% (89 + 1 + 3)	13	14 166
<i>Inocybe glabripes</i>	Ectomycorrhizal	28M	22M	38	5906	11 075	94% (91 + 1 + 2)	10	11 676
<i>Inocybe corydalina</i>	Ectomycorrhizal	26M	22M	42	7022	9548	95% (92 + 1 + 2)	15	11 560
<i>Inocybe lanuginosa</i>	Ectomycorrhizal	24M	17M	35	6426	7907	92% (84 + 4 + 4)	10	11 528
<i>Inocybe oblectabilis</i>	Ectomycorrhizal	25M	20M	42	6786	10 461	94% (91 + 1 + 2)	15	12 638
<i>Inocybe asterospora</i>	Ectomycorrhizal	38M	24M	40	6897	8956	94% (88 + 3 + 3)	16	11 420
<i>Inocybe petiginosa</i>	Ectomycorrhizal	24M	13M	27	2678	29 137	96% (93 + 1 + 2)	8	8563
<i>Pseudosperma rimosum</i>	Ectomycorrhizal	26M	17M	46	10 002	5324	95% (90 + 1 + 4)	15	13 490
<i>Pseudosperma bulbosissimum</i>	Ectomycorrhizal	22M	18M	51	10 617	5625	96% (92 + 1 + 3)	21	12 839
<i>Inosperma erubescens</i>	Ectomycorrhizal	25M	16M	40	5975	13 546	97% (94 + 1 + 2)	14	11 484
<i>Inosperma cervicolor</i>	Ectomycorrhizal	27M	18M	36	6655	7162	96% (93 + 1 + 2)	20	9694
<i>Inosperma bongardii</i>	Ectomycorrhizal	24M	18M	39	7915	5055	94% (91 + 1 + 2)	19	10 258
<i>Mallocybe terrigena</i>	Ectomycorrhizal	15M	9M	28	5613	6302	89% (85 + 0 + 4)	3	9277
<i>Mallocybe dulcamara</i>	Ectomycorrhizal	31M	27M	41	8392	5452	94% (91 + 1 + 2)	13	11 557
<i>Crepidotus cesatii</i>	Saprotroph	19M	17M	48	6835	12 622	95% (89 + 3 + 3)	4	16 045
<i>Crepidotus calolepis</i>	Saprotroph	21M	15M	54	8272	10 283	85% (80 + 1 + 4)	3	17 453
<i>Hypholoma sublateritium</i>	Saprotroph	–	–	48	700	298 747	99% (97 + 1 + 1)	7	13 194
<i>Gymnopilus junonius</i>	Saprotroph	–	–	59	1161	152 469	95% (82 + 11 + 2)	15	16 391
<i>Panaeolus papilionaceus</i>	Saprotroph	–	–	51	78	1820 100	98% (97 + 1 + 0)	14	14 065
<i>Agaricus bisporus</i>	Saprotroph	–	–	30	29	2334 609	98% (96 + 1 + 1)	12	9184

^aBUSCO (Benchmarking Universal Single Copy Orthologues), which indicates genome completeness; Comp (Complete).
^bTEs (Transposable Elements).

Table 2 The top 10 orthogroups with highly significant changes, that is lowest *P*-values ($P \ll 0.001$) along the three branch-sets used in the analyses, that is pre-ECM (pre-Ectomycorrhizal) branches, transition branch to ECM, and ECM branches.

Orthogroups	<i>P</i> -value	No. of significant changes	Pfam	Description	Function/putative function
Pre-ECM branches					
OG0000056	4.54E−04	(+4*)	PF03732	Retrotransposon gag protein	Transposition (Casacuberta <i>et al.</i> , 2007)
OG0000118	2.59E−03	(+3*)	PF07999	Retrotransposon hot spot protein	Regulation of gene expression (Bernardo <i>et al.</i> , 2020)
OG0000139	2.59E−03	(+3*)	NA	Hypothetical Protein	–
OG0000305	3.11E−03	(−4*)	PF09118	Glyoxal Oxidase	Lignin degradation (Wohlschlager <i>et al.</i> , 2021)
OG0000006	7.73E−03	(+4*)	PF01926	50S Ribosome Binding GTPase	Regulation of cellular processes (Chakraborty <i>et al.</i> , 2022)
OG0000055	7.86E−03	(+3*)	PF12937	F-box protein	Protein–protein interactions (Kipreos & Pagano, 2000)
Transition branch to ECM					
OG0000121	7.37E−04	(+3*)	NA	Hypothetical Protein	–
OG0000124	7.37E−04	(+3*)	PF04053	Coatomer transport protein	Transport (Thompson & Brown, 2012)
OG0000149	7.37E−04	(+3*)	NA	Hypothetical Protein	–
OG0000150	7.37E−04	(+3*)	PF05699	hAT transposon superfamily	Transposition (Rubin <i>et al.</i> , 2001)
OG0000153	7.37E−04	(+3*)	PF09261	Alpha mannosidase	Cleavage of mannose oligosaccharides (Van Petegem <i>et al.</i> , 2001)
OG0000189	7.37E−04	(+3*)	PF00651	BTB/POZ protein domain	Protein–protein interactions (Wen <i>et al.</i> , 2000)
OG0000193	7.37E−04	(+3*)	PF02076	Pheromone A receptor	Signaling (Turrà <i>et al.</i> , 2015)
OG0000202	7.37E−04	(+3*)	NA	Hypothetical Protein	–
OG0000206	7.37E−04	(+3*)	PF13374	Tetratricopeptide repeats	Protein–protein interactions (Blatch & Lässle, 1999)
OG0000214	7.37E−04	(+3*)	PF13191	AAA ATPase domain	Regulation of cellular processes (G. Zhang <i>et al.</i> , 2021)
ECM branches					
OG0000021	1.57E−06	(+5*)	NA	Hypothetical Protein	–
OG0000018	2.60E−06	(−5*)	NA	Hypothetical Protein	–
OG0000174	3.60E−06	(+4*)	PF01522	Polysaccharide Deacetylase	Fungal cell wall modifications (Xu <i>et al.</i> , 2020)
OG0000137	8.57E−06	(+4*)	PF18803	KDZ transposases	Transposition (Iyer <i>et al.</i> , 2014)
OG0000372	9.73E−06	(+3*)	PF05729	NACHT Domain	Regulation of cellular processes (MacDonald <i>et al.</i> , 2013)
OG0000840	9.73E−06	(+3*)	NA	Hypothetical Protein	–
OG0000169	3.74E−05	(+3*)	PF17667	Serine/Threonine Fungal Protein Kinase	Signaling (Nagarajan <i>et al.</i> , 2022)
OG0000379	3.74E−05	(+3*)	PF10294	Lysine Methyltransferase	Regulation of gene expression (Lukinović <i>et al.</i> , 2020)
OG0000482	3.74E−05	(+3*)	NA	Hypothetical Protein	–
OG0000025	5.84E−05	(+4*)	PF00400	WD containing repeat domain	Regulation of cellular processes (van Nocker & Ludwig, 2003)

The plus sign represents expansions, minus sign represents contractions, asterisk and *P*-values represent significant changes in the gene families were generated using CAFE analysis. NA, not applicable.

higher number of, and sometimes completely missing, genes in most of the different classes of CAZymes compared with all the saprotrophic species (Fig. 3a). The genes encoding enzymes acting on crystalline cellulose such as cellobiohydrolases (GH6 and GH7), involved in the oxidation of lignocellulose (e.g. AA8, AA12, AA16 and CE16), acting on hemicellulose (e.g. CE1, CE2 and GH11), pectin (e.g. CE12 and PL4) and lignin (e.g. CE15) are totally absent in the genomes of Inocybaceae (Table 3) as compared to all the saprotrophic relatives. None of the Inocybaceae genomes have the GH32 invertase gene, coding for the enzyme, which helps in the conversion of sucrose into glucose and fructose (Table S6).

The Inocybaceae genomes have retained a wide range of gene copy numbers encoding laccases and laccase-like enzymes, for

example AA1 (1–5) and AA3 (4–11); xylanases, for example GH10 (0–7); polygalacturonases, for example GH28 (0–2); mannanases, for example GH92 (0–2); and lytic polysaccharide oxygenases, for example AA9 (0–3). Cellulose binding modules such as CBM1 and CBM2 are also present in lower copy numbers in Inocybaceae genomes (1–5 and 0–4, respectively). Inocybaceae species have similar gene copy numbers of chitinases, for example GH18 (5–12) and xylogucosyltransferases, for example GH16 (7–11) as all the saprotrophic species (9–17 and 11–18, respectively) (Table S6). Most Inocybaceae and Crepidotaceae species have only one (range 0–2) copy of class II lignin degrading PODs (AA2), as compared to the other included taxa, which have at least two but usually many more of these genes (Table S6).

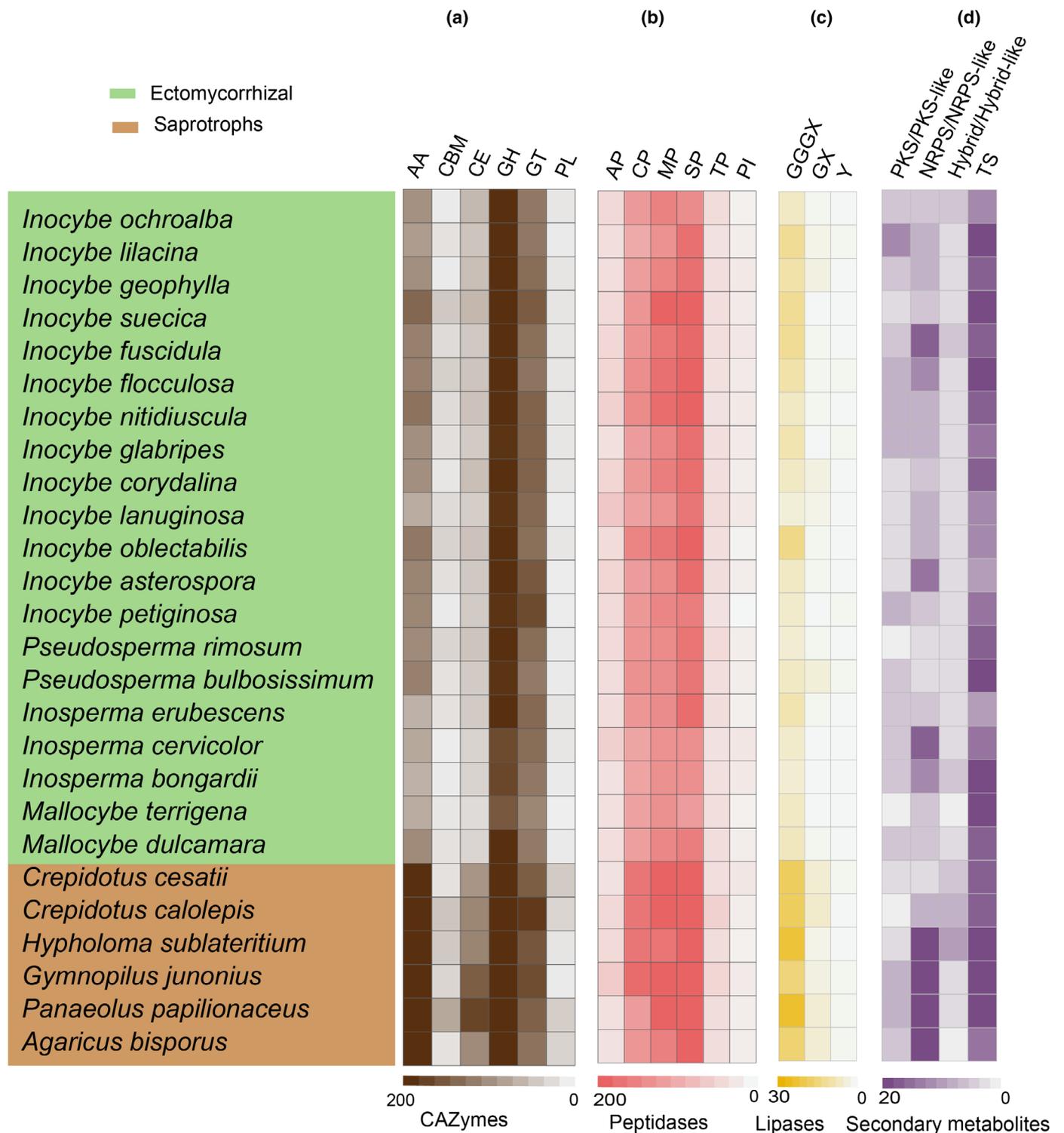


Fig. 3 Heat map of the gene copy numbers in different gene categories in Inocybeaceae and outgroup species. The taxa are color coded according to their lifestyles: green for the ECM and brown for saprotrophs. (a) CAZymes, the gene copy numbers of different classes of Carbohydrate-Active Enzymes, that is Auxiliary Activities (AA), Carbohydrate-Binding Modules (CBM), Carbohydrate Esterases (CE), Glycoside Hydrolases (GH), Glycosyltransferases (GT), Polysaccharide Lyases (PL) (b) Peptidases, the gene copy number of different classes of peptidases, that is Aspartic Peptidases (AP), Cysteine Peptidases (CP), Metallopeptidases (MP), Serine Peptidases (SP), Threonine Peptidases (TP), Protein Inhibitors (PI) (c) Lipases, the gene copy number of different classes of lipases, that is GGGX, GGX and Y (d) Secondary Metabolites, the number of gene clusters in selected classes of secondary metabolites, that is Polyketide Synthases PKS/PKS-like, non-ribosomal peptide synthetases NRPS/NRPS-like and terpene synthases.

Table 3 Distribution of selected Carbohydrate-Active Enzymes (CAZymes) in the analyzed 26 genomes.

CAZyme gene families																	
Species	AA1	AA3	AA8	AA9	AA12	CBM1	CBM2	CE1	CE12	CE15	CE16	GH6	GH7	GH10	GH28	GH92	PL4
<i>Inocybe ochroalba</i>	2	8	0	1	0	4	1	0	0	0	0	0	0	0	0	1	0
<i>Inocybe lilacina</i>	2	9	0	1	0	3	3	0	0	0	0	0	0	1	1	1	0
<i>Inocybe geophylla</i>	2	7	0	1	0	2	1	0	0	0	0	0	0	0	1	0	0
<i>Inocybe seucica</i>	3	13	0	1	0	3	2	0	0	0	0	0	0	0	1	1	0
<i>Inocybe fuscidula</i>	3	11	0	1	0	3	1	0	0	0	0	0	0	0	0	1	0
<i>Inocybe flocculosa</i>	3	7	0	2	0	4	2	0	0	0	0	0	0	0	0	1	0
<i>Inocybe nitidiscula</i>	4	9	0	2	0	2	4	0	0	0	0	0	0	1	0	1	0
<i>Inocybe glabripes</i>	4	6	0	1	0	4	4	0	0	0	0	0	0	1	2	1	0
<i>Inocybe corydalina</i>	3	7	0	3	0	4	3	0	0	0	0	0	0	1	0	1	0
<i>Inocybe lanuginosa</i>	4	5	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0
<i>Inocybe oblectabilis</i>	3	10	0	2	0	5	2	0	0	0	0	0	0	1	1	0	0
<i>Inocybe asterospora</i>	5	7	0	2	0	5	2	0	0	0	0	0	0	0	2	2	0
<i>Inocybe petiginosa</i>	5	10	0	2	0	1	0	0	0	0	0	0	0	1	1	1	0
<i>Pseudosperma rimosum</i>	4	9	0	0	0	2	1	0	0	0	0	0	0	0	0	1	0
<i>Pseudosperma bulbosissimum</i>	4	9	0	2	0	2	4	0	0	0	0	0	0	1	0	1	0
<i>Inosperma erubescens</i>	2	5	0	2	0	3	2	0	0	0	0	0	0	1	1	0	0
<i>Inosperma cervicolor</i>	1	4	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0
<i>Inosperma bongardii</i>	1	4	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0
<i>Mallocybe terrigena</i>	3	5	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0
<i>Mallocybe dulcamara</i>	4	5	0	2	0	4	0	0	0	0	0	0	0	0	0	1	0
<i>Crepidotus cessati</i>	11	18	1	10	0	30	8	1	3	1	4	4	3	4	5	9	1
<i>Crepidotus calolepsis</i>	14	26	6	8	3	35	10	3	1	1	2	5	3	3	1	6	1
<i>Hypholoma sublateritium</i>	13	26	1	6	1	35	8	2	1	2	3	1	4	6	2	4	1
<i>Gymnopilus junonius</i>	9	30	1	19	1	52	9	1	0	4	4	4	6	7	6	2	0
<i>Panaeolous papilionaceus</i>	9	22	3	23	3	98	24	5	4	4	5	2	11	16	2	4	2
<i>Agaricus bisporus</i>	11	34	1	7	1	23	4	1	2	1	1	1	1	1	3	5	1

Ecmycorrhizal and saprotrophic species are shaded differently with lighter shade for the ECM and darker shade for the saprotrophs. AA, Auxiliary Activities; CBM, Carbohydrate-Binding Module; CE, Carbohydrate Esterases; GH, Glycoside Hydrolases; PL, Polysaccharide Lyases.

The CAFE analysis showed contractions in the largest number of CAZyme gene families (76) along the transition branch to ECM, compared with the other branches of the tree (median 13; Fig. 4a). However, the median number of gene families that contracted were 22 along the branches before the transition and a median of 10 contractions after the transition to the ECM in Inocybaceae. On the other hand, there were expansions only

in four gene families at the transition branch and a median of six expansions in the other branches of the tree (Fig. S3a). The comparison of models with different rates for different branch-sets showed that the three-rate model fit the data best with the highest rate for the transition branch to ECM and the lowest rate for the outgroup branches ($P = 0.001$ compared with both one- and two-rate models; Fig. 5; Table S5). However, the simulations

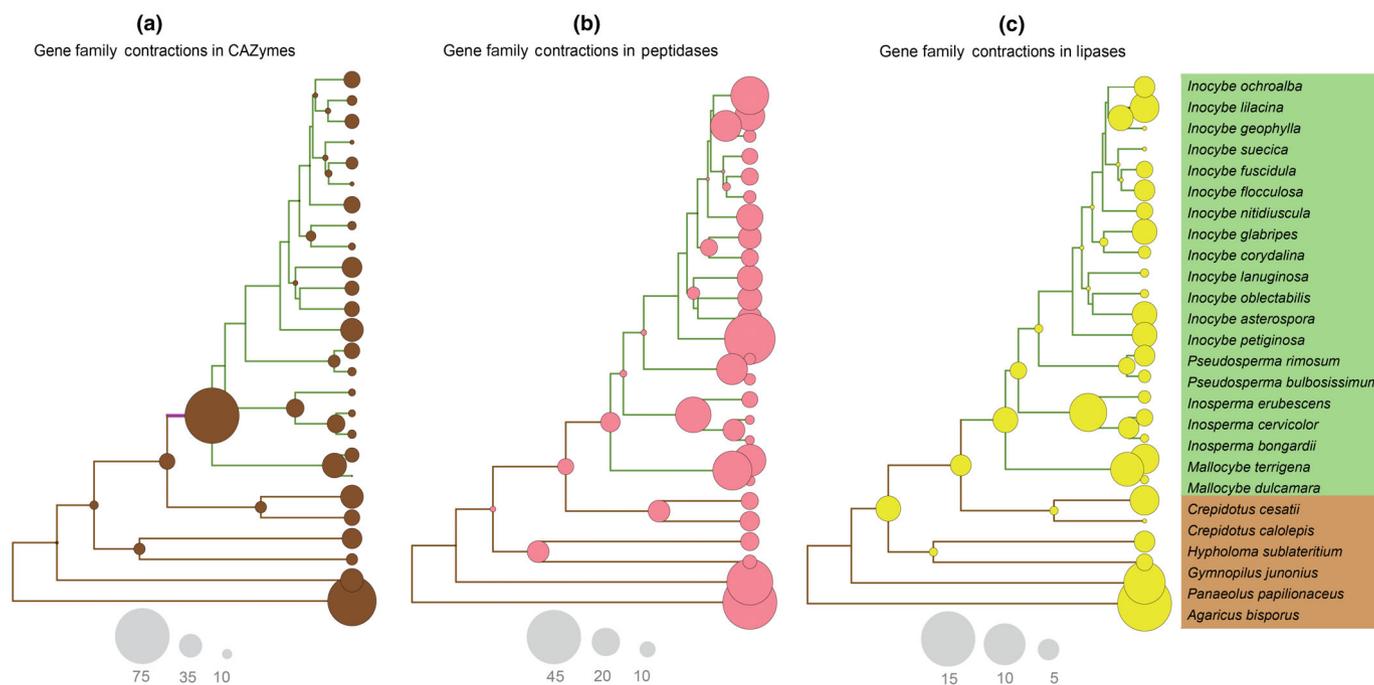


Fig. 4 Gene family contractions in Carbohydrate-Active enZymes (CAZymes) (a), peptidases (b) and lipases (c) in Inocybaceae and outgroup species. The size of the bubbles is proportional to the number of gene family contractions in those branches and summarized at the nodes. The color of the bubbles represents the gene categories. The taxa in green shades are the ectomycorrhizal and brown are saprotrophs. All trees are annotated on the basis of CAFE analysis using best-fitting model on the respective datasets of CAZymes, peptidases and lipases, that is three-rate, two-rate and two-rate model. Branches are colored based on the best-fitting model, with branches with the same rate having the same color.

showed that the observed data were not likely to be generated by any of the tested models ($P = 0.001$; Fig. 5a).

Gradual decrease in peptidase gene family sizes

The Inocybaceae genomes have fewer genes coding for total and secreted peptidases as compared to all saprotrophic species (Fig. 2b; Table S3), except *A. bisporus*, which has similar content of total peptidases as some Inocybaceae species. The total number of peptidases in ECM species range from 228 in *M. terrigena* to 311 in *I. oblectabilis*. *Crepidotus cesatii* and *Crepidotus calolepis* have 409 and 430 peptidases, respectively. The secreted peptidases are in the range of 33–58 in species with ECM ecology and in the range of 94–127 in saprotrophs. Serine peptidases such as S33 and S09X were specifically found to be in lower copy numbers in genomes of ECM species as compared to saprotrophs (Fig. 3b; Table S7).

The CAFE analysis revealed that the contractions in peptidases have been gradual, with only a slow increase in the number of gene family contractions from 14 on the branch leading to the MRCA of Inocybaceae and Crepidotaceae to 17 on the transition branch to ECM in Inocybaceae. This is unlike the large number of contractions (76) in CAZymes along the same branch of the phylogeny (Fig. 4b). There is no noticeable change in the number of peptidase gene family expansions (Fig. S3b). The comparison of models with different rates for different branch-sets showed that the two-rate model, with different rates for ECM branches

and non-ECM branches including the transition branch to ECM, is the simplest model that cannot be rejected ($P = 0.001$ vs one-rate model; $P = 0.897$ vs three-rate model; Fig. 5; Table S5). On the basis of simulations, we can reject the one-rate model ($P = 0.001$) to explain the observed data, but not by the two-rate ($P = 0.085$) or three-rate models ($P = 0.103$; Fig. 5b).

Lower suite of lipase genes in the ECM fungi

While both the ECM and saprotrophic fungi possess a limited set of lipase genes compared with other gene categories, the saprotrophic species have a higher number of both the total and secreted lipases than the ECM species (Table S3; Fig. 2b). The ECM species *I. bongardii* has the lowest number with five genes for secreted lipases, whereas the saprotrophic species *P. papilionaceus* has the highest number with 35 genes in this study (Table S8). ECM species have lower number of genes from the GGGX and GX classes compared with saprotrophs (Fig. 3c). Within the GGGX class, ECM species have consistently lower number of *Candida rugosa*-like lipases (AbH03), and in the GX class, they have lower number of filamentous fungi lipases (AbH23). There is no clear pattern in the other families in these classes (Table S9).

The CAFE analysis for lipases did not reveal any clear discernible pattern of gene family contractions and expansions associated with the transition to ECM (Figs 4c, S3c). The comparison of models with different rates for different branch-sets showed that

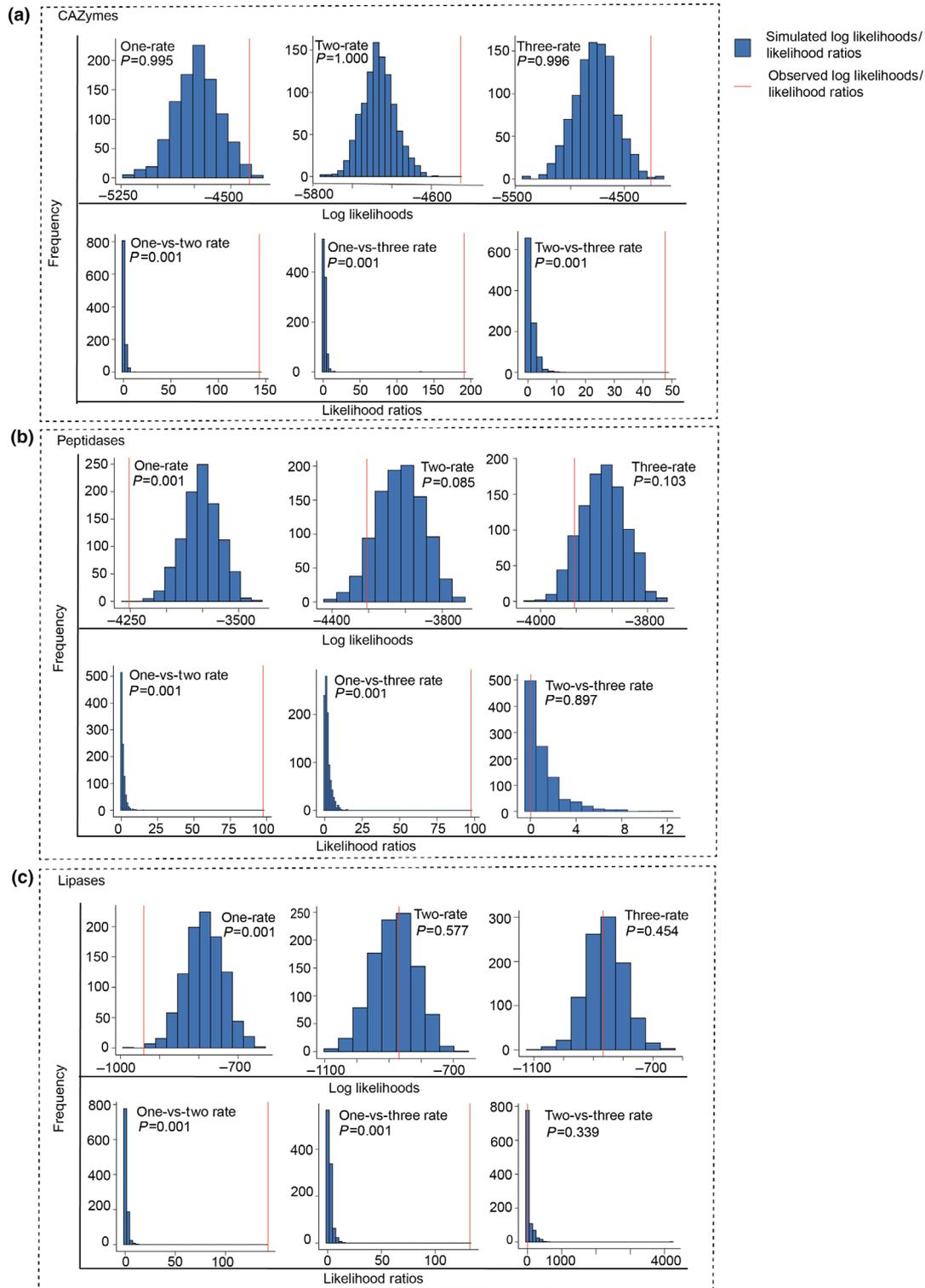


Fig. 5 Simulation results obtained from the CAFE analysis for three datasets (a) Carbohydrate-Active enzymes (CAZymes), (b) peptidases and (c) lipases. The first row of each dataset shows the frequency distribution of the log likelihood values and the second row shows the likelihood ratios of the different model comparisons. The first histogram of each dataset shows the log likelihood values of data simulated under the one-rate model, followed by the two-rate and the three-rate model. The first histogram in the second row of each dataset shows the comparison of one-rate vs two-rate models, followed by one-rate vs three-rate models and two-rate vs three-rate models. P -values are the probability of observing higher Maximum Likelihood (ML) estimates, respectively, higher difference in estimated ML scores.

the two-rate model with higher rate (0.013) for the transition branch and the ECM branches, and lower rate (0.005) for the non-ECM branches was the best-fitting model among the two-rate models and was significantly better than the one-rate model ($P = 0.001$). This two-rate model could not be confidently rejected compared with the three-rate model (that implies the highest rate (0.031) at the transition branch), but the P -value was still low ($P = 0.091$ based on LRT; but 0.339 based on simulations), which may be due to a lack of power to discern that level of difference (Table S5). On the basis of simulations, we can reject the one-rate model ($P = 0.001$) to explain the observed data, but not by the two-rate ($P = 0.577$) or three-rate models ($P = 0.454$, Fig. 5c).

Similar number of SMCs in ECM and closely related saprotrophs

The number of SMCs is lower in Inocybaceae species as compared to the distant saprotrophic outgroups (Fig. 2b; Table S3). Specifically, there are fewer non-ribosomal peptide synthases (NRPS)/NRPS-like gene clusters in most Inocybaceae species, typically ranging from two to three gene clusters, with the exception of *I. asterospora*, *I. cervicolor* and *I. fuscidula*, which have six to seven NRPS-like SMCs. The closely related Crepidotaceae species similarly have one to three of these gene clusters, but the more distant saprotrophic outgroups typically have between eight to 10 gene clusters (Fig. 3d).

All sampled species, except *I. lilacina*, have zero to three polyketide synthases (PKS)/PKS-like gene clusters. *Inocybe lilacina* stands out with the highest count of five PKS/PKS-like gene clusters when compared to all other genomes included in this study. The terpene synthases (TS) gene clusters in Inocybaceae species range from four to 11. The number of TS gene clusters in Crepidotaceae, like other SMCs, show minimal differences from that of Inocybaceae species, while most of the distantly related outgroups possess a noticeably larger number of TS gene clusters.

Discussion

The use of sporocarp tissues for whole-genome sequencing has facilitated sampling of many obligate biotrophic fungal groups, which are otherwise uncultivable, for comparative genomic studies (Bahram *et al.*, 2018; Chang *et al.*, 2019; Looney *et al.*, 2022; Wu *et al.*, 2022). However, sporocarps can harbor diverse communities of bacteria and fungicolous fungi (Maurice *et al.*, 2021), leading to additional genomes in the sequencing data (Bahram *et al.*, 2018; Chang *et al.*, 2019). We noticed that the presence of fungal contaminants in the assembly could inflate the overall Busco scores with an increased number of duplicates. We therefore ensured that our fungal assemblies were contaminant-free by careful and thorough filtering of nontarget sequence reads and were able to recover near complete assemblies for our target species with a median Busco completeness of 94.5%, and only 0–4% of gene duplicates (which is within the range/lower than other studies) (Hess *et al.*, 2018; Wu *et al.*, 2022).

Our phylogenomic analyses confirm that Inocybaceae is monophyletic and the sister lineage of Crepidotaceae. The main genera within Inocybaceae are well-supported, and the topology obtained here is largely in agreement with previous inferences (Matheny *et al.*, 2009; Ryberg *et al.*, 2010). However, the position of *Mallocybe* in our tree disagrees with recent topologies (e.g. Matheny *et al.*, 2020), emerging as a sister clade to the rest of Inocybaceae, as in Matheny *et al.* (2006), rather than as the sister clade to *Inosperma*. We found no systematic bias in terms of tree length or support values between the gene trees showing *Inosperma* and *Mallocybe* as monophyletic vs nonmonophyletic, suggesting that our results are not due to biases associated with fast-evolving genes or poor resolution, and strengthening our inference that *Mallocybe* is sister to the rest of Inocybaceae.

The assembly size and total number of protein-coding genes of Inocybaceae species are small compared with most other ECM species (Kohler *et al.*, 2015). However, it is within the expected average fungal genome size of 30–50 MB (Mohanta & Bae, 2015), except for a few assemblies. Both the average assembly size and total gene content of these species are similar to those of *H. cylindrosporum*, which is a closely related species and that has the smallest ECM genome reported to date (Doré *et al.*, 2015). Interestingly, *I. petiginosa* with the smallest genome size in Inocybaceae and with few predicted genes and SSPs has a narrow ECM host range, primarily associating with *Fagus* and occasionally *Quercus*. Specialist fungal pathogens tend to exhibit smaller genomes and gene content, shedding genes not needed in their constrained lifestyle (de Man *et al.*, 2016; Schuelke *et al.*, 2017), and the same process may be active in the lineage of *I. petiginosa*. We found that the Busco gene loss in the filtration process, mainly affected two heavily contaminated samples (Table S2). The remaining samples showed minimal changes in Busco genes after the filtration. Hence, filtration appears to have a small effect on protein-coding areas of the assemblies but could affect repeat regions with deviating GC-content and coverage. The overall smaller assemblies could therefore be due to multiple repeat regions being collapsed owing to the use of short-read sequencing or the strict filtering scheme employed in this study. However, our results are congruent with other studies showing that small genomes in general carry a small number of genes (Hess *et al.*, 2018).

The most significant changes in the transition branch to ECM include expansions in genes encoding coatomer transporters, pheromone alpha receptors, tetratricopeptides and BTB/POZ domain proteins (Table 2). All these genes are either directly related to symbioses with plants or are abundant in the genomes of plant symbiotic fungi. The coatomer transporters play an important role in the accurate transport of proteins and proper functioning of the secretory pathway (Thompson & Brown, 2012) and have been found to be highly expressed in symbiotic roots (Song *et al.*, 2015). The pheromone alpha signaling receptors, as shown in fungal pathogens, are used to receive chemical stimuli from host plants (Turrà *et al.*, 2015). Tetratricopeptide and BTB/POZ domain proteins, involved in the regulation of gene expression, have been found to be substantially abundant in arbuscular mycorrhizal fungi (Miyachi *et al.*, 2020). These gene families may add to the list of promising

candidate genes important for the transition to the ECM and should be studied in other ECM lineages for further evolutionary insight.

Among the highly significant changes in the ECM branches after the transition to the symbiotic lifestyle were expansions in different types of repeat-related proteins (e.g. KDZ transposases and WD40 repeats), NACHT domain, polysaccharide deacetylases and communication genes such as serine/threonine protein kinases (Table 2). WD40 repeats, involved in protein–protein interactions, are also expanded in the ECM *Cenococcum geophyllum* (Peter *et al.*, 2016). KDZ transposases have been linked to lineage-specific gene family expansions in *L. bicolor* (Iyer *et al.*, 2014). The overall increase in gene family evolution within the ECM branches could also be a consequence of the heightened transposon activities. Fungal polysaccharide deacetylases evade plant immunity by converting chitin, a conserved component in the fungal cell wall and prime target of the plant defense system, into chitosan (Xu *et al.*, 2020; Rizzi *et al.*, 2021). This modification prevents detection of the fungal cell wall by the plant surveillance system while preserving its integrity (Xu *et al.*, 2020). This may be a strategy used by some ECM fungi to infiltrate, develop and grow within plant tissues, possibly reducing the need for SSPs (Ryberg *et al.*, 2022), which could then partly explain the low set of SSP genes in the Inocybaceae genomes. Serine/threonine and tyrosine kinases have previously been found to be highly expressed in ECM fungi as compared to free-living saprotrophs and suggested to play a role in signaling pathways during early ECM fungal–plant interactions (Martin *et al.*, 2001). The continued evolution of these signaling-related kinase proteins after the transition to ECM may be a consequence of continued adaptation to new hosts.

Our results show that Inocybaceae share the common CAZyme signatures with other ECM species, with a complete absence of certain CAZymes targeting cellulose, hemicellulose, pectin and lignin, for example GH6, GH7 and GH32, but also retention of some sets of plant cell wall degrading enzymes (a subset of CAZymes), for example AA3, AA9 and GH28 (Kohler *et al.*, 2015; Miyauchi *et al.*, 2020; Lebreton *et al.*, 2021). However, CBM1 is absent in most other ECM species, but like *C. geophyllum* and *L. bicolor*, Inocybaceae species have retained a few copies of these genes (Kohler *et al.*, 2015; Peter *et al.*, 2016; Miyauchi *et al.*, 2020). Inocybaceae species may utilize these mycorrhiza-induced CAZyme genes for cell-wall remodeling during the formation of symbiotic structures such as hartig net and mantle (Zhang *et al.*, 2018).

Serine peptidase families such as S33 (prolyl aminopeptidase) and S09X (glutamyl endopeptidase) are suggested to be a marker of saprotrophy (Muszewska *et al.*, 2017) and are particularly reduced in Inocybaceae genomes. Given that some fungal serine peptidases trigger host plant immune responses (Figueiredo *et al.*, 2018), the lower repertoire of these saprotroph-specific peptidases in the ECM fungi might be a way to evade host immune reactions (Pellegrin *et al.*, 2015).

Lipases are reported to be under-represented in the fungal secretome (Pellegrin *et al.*, 2015). The lower gene counts of filamentous fungal lipases (abH23) and versatile lipases (abH03) in Inocybaceae is similar to previous observations in the ECM as compared to

saprotrophs (Pellegrin *et al.*, 2015; Barriuso & Martínez, 2017). However, we observed a smaller suite of the total and secreted lipases in the ECM compared with saprotrophs, which contrasts with previous findings (Pellegrin *et al.*, 2015; Miyauchi *et al.*, 2020; Wu *et al.*, 2022), but this could just reflect the generally smaller number of genes in Inocybaceae. The higher content of versatile lipases in saprotrophs is attributed to their broader substrate range, unlike the more specific substrates utilized by ECM fungi (Barriuso & Martínez, 2017). Furthermore, it has been proposed that numerous fungal lipases along with CAZymes have a potential role in eliciting the host plant immune responses (Piombo *et al.*, 2023).

Like other biotrophic fungi, Inocybaceae have fewer SMCs compared to saprotrophs (Spanu, 2012; Pellegrin *et al.*, 2015; Peter *et al.*, 2016; Looney *et al.*, 2022). As observed previously, Inocybaceae sister taxon Crepidotaceae have a similar smaller number of SMCs, indicating that it may be a common pre-adaptation to lose SMCs (Miyauchi *et al.*, 2020; Looney *et al.*, 2022). Inocybaceae species particularly possess fewer NRPS/NRPS-like SMCs, and certain well-characterized NRPS genes are implicated to directly activate plant defense-related genes (Vinale *et al.*, 2012). There are also speculations on the possible role of NRPS/NRPS-like secondary metabolites in host specificity in the ECM fungi (Lofgren *et al.*, 2021). Though most of the Inocybaceae species show some level of host specificity, the three species with highest number of NRPS/NRPS-like genes do not stand out as particularly host-specific compared to others. More knowledge about the specificity level in these species is needed to draw definite conclusions.

The orthogroup based analyses indicate that there is a general shift in gene family evolution along the transition branch to ECM, this could be driven at least partially by increased transposon activity. The higher rate of evolution in ECM branches seem to mainly lead to losses of genes coding for peptidases and CAZymes. For CAZymes, the rate is additionally much higher at the transition branch to ECM, and most of these changes are also contractions. The rapid decrease in this branch is probably driven by selection, but it is unclear whether it reflects a shift toward a new equilibrium driven by the cost of maintaining unnecessary genes and/or loss of genes that could trigger plant defenses; or it is mainly drift from a state with many genes, initially maintained by selection, to a subsequent drop to a lower level, which is then sustained by selection (Miyauchi *et al.*, 2020; Lebreton *et al.*, 2021). However, as our simulations show that none of the tested models fit the CAZymes dataset well, further models should be tested and we caution against routine use of the default model in CAFE assuming that it is an adequate fit. Our results indicate that the rate of evolution is elevated for lipases within the ECM branches as well, and possibly an even higher rate at the transition branch to the ECM. Nevertheless, the current dataset lacks sufficient statistical power to confidently discern a rate change of that magnitude at a single branch.

Conclusion

Inocybaceae genomes show the common molecular signatures of ECM basidiomycetes. The rate of overall gene family evolution

increases in close proximity to the transition to the ECM habit. The evolution of CAZymes, possibly also lipases and some specific gene families is also higher around the actual transition. Taken together, it appears that in addition to the widespread losses of CAZymes, the loss or otherwise tight control of saprotrophic specific peptidases, lipases and secondary metabolite genes, all of which could have possible roles in triggering the host plant immune responses, may be key for the transition to the ECM lifestyle.

Acknowledgements

This study was supported by the grant no. 2016-04216 from the Swedish Research Council (VR) to MR. We acknowledge the Science for Life Laboratory and UPPMAX for the support and services in the provision of computational infrastructure and sequencing of the samples. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. We thank Nahid Heidari, at the program of Systematic Biology in Uppsala University, for performing the molecular lab work and Ellen Larson, at the Gothenburg University Herbarium GB, for the timely provision of *Crepidotus cesatii* sample. We would also like to thank the Swedish Mycological Association for organizing Mycology Week, 2017 as we obtained most of the mushroom samples during this week.

Competing interests

None declared.

Author contributions

MR conceived the idea and MR and FKK designed the project. FKK performed the data analyses, that is quality control of the raw data, assembly and annotation of the genomes, phylogenetic and comparative genomic analyses. MR, HJ and MS-G supervised the project and helped in the interpretation of the results. FKK wrote the first draft of the manuscript and revised it with inputs from MR, HJ and MS-G. All authors read and approved the final version of the manuscript.

ORCID

Hanna Johannesson  <https://orcid.org/0000-0001-6359-9856>
Faheema Kalsoom Khan  <https://orcid.org/0000-0002-4891-953X>

Martin Ryberg  <https://orcid.org/0000-0002-6795-4349>

Marisol Sánchez-García  <https://orcid.org/0000-0002-0635-6281>

Data availability

The complete genomic data generated in this study are deposited at the National Center for Biotechnology Information Sequence

Read Archive and the genome assemblies for all the newly generated samples are in the NCBI GenBank under the BioProject PRJNA1048439. The specific accession nos. of these data can be found in Table S1. All other data supporting the findings of this study are available within the article and its [Supporting Information](#) files or at doi: [10.5281/zenodo.12723518](https://doi.org/10.5281/zenodo.12723518) or are available from the corresponding authors upon request.

References

- Aime M. 1999. *Generic concepts in the Crepidotaceae as inferred from nuclear large subunit, ribosomal DNA sequences, morphology, and basidiospore dormancy patterns*. MS Thesis. Blacksburg, VA, USA: Virginia Polytechnic Institute and State University.
- Alfaro M, Oguiza JA, Ramírez L, Pisabarro AG. 2014. Comparative analysis of secretomes in basidiomycete fungi. *Journal of Proteomics* 102: 28–43.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- Andrews S. 2010. *FASTQC: a quality control tool for high throughput sequence data (Online), v.0.11.8*. [WWW document] URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> [accessed 6 December 2019].
- Bahram M, Vanderpool D, Pent M, Hiltunen M, Ryberg M. 2018. The genome and microbiome of a dikaryotic fungus (*Inocybe terrigena*, Inocybaceae) revealed by metagenomics. *Environmental Microbiology Reports* 10: 155–166.
- Barriuso J, Martínez MJ. 2017. Evolutionary history of versatile-lipases from Agaricales through reconstruction of ancestral structures. *BMC Genomics* 18: 12.
- Bernardo WP, Souza RT, Costa-Martins AG, Ferreira ER, Mortara RA, Teixeira MMG, Ramirez JL, Da Silveira JF. 2020. Genomic organization and generation of genetic variability in the RHS (Retrotransposon Hot Spot) protein multigene family in *Trypanosoma cruzi*. *Genes (Basel)* 11: 1085.
- Blatch GL, Lässle M. 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays* 21: 932–939.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Brundrett MC, Tedersoo L. 2018. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist* 220: 1108–1115.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12: 59–60.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. TRIMAL: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.
- Casacuberta E, Marín FA, Pardue ML. 2007. Intracellular targeting of telomeric retrotransposon Gag proteins of distantly related *Drosophila* species. *Proceedings of the National Academy of Sciences, USA* 104: 8391–8396.
- Chakraborty A, Halder S, Kishore P, Saha D, Saha S, Sikder K, Basu A. 2022. The structure-function analysis of Obg-like GTPase proteins along the evolutionary tree from bacteria to humans. *Genes to Cells* 27: 469–481.
- Chang Y, Desiró A, Na H, Sandor L, Lipzen A, Clum A, Barry K, Grigoriev IV, Martin FM, Stajich JE *et al.* 2019. Phylogenomics of Endogonaceae and evolution of mycorrhizas within Mucoromycota. *New Phytologist* 222: 511–525.
- Corbu VM, Gheorghe-Barbu I, Dumbravă A, Vrâncianu CO, Şesan TE. 2023. Current insights in fungal importance—a comprehensive review. *Microorganisms* 11: 1384.
- Doré J, Perraud M, Dieryckx C, Kohler A, Morin E, Henrissat B, Lindquist E, Zimmermann SD, Girard V, Kuo A *et al.* 2015. Comparative genomics, proteomics and transcriptomics give new insight into the exoproteome of the basidiomycete *Hebeloma cylindrosporium* and its involvement in ectomycorrhizal symbiosis. *New Phytologist* 208: 1169–1187.
- El-Gendi H, Saleh AK, Badierah R, Redwan EM, El-Maradny YA, El-Fakharany EM. 2021. A comprehensive insight into fungal enzymes: structure, classification, and their role in mankind's challenges. *Journal of Fungi* 8: 23.
- Emms DM, Kelly S. 2019. ORTHOFINDER: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20: 238.

- Figueiredo J, Sousa Silva M, Figueiredo A. 2018. Subtilisin-like proteases in plant defence: the past, the present and beyond. *Molecular Plant Pathology* 19: 1017–1028.
- Fries N. 1982. Effects of plant roots and growing mycelia on basidiospore germination in mycorrhizaforming fungi. In: Laursen G, Ammirati J, eds. *Arctic and alpine mycology*. Seattle, WA, USA: University of Washington Press, 493–508.
- 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: D699–D704.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PHYML 3.0. *Systematic Biology* 59: 307–321.
- Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution* 30: 1987–1997.
- 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.
- Iyer LM, Zhang D, De Souza RF, Pukkila PJ, Rao A, Aravind L. 2014. Lineage-specific expansions of *TET1/JBP* genes and a new class of DNA transposons shape fungal genomic and epigenetic landscapes. *Proceedings of the National Academy of Sciences, USA* 111: 1676–1683.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kelly S, Maini PK. 2013. DENDROBLAST: approximate phylogenetic trees in the absence of multiple sequence alignments. *PLoS ONE* 8: e58537.
- Kipres ET, Pagano M. 2000. The F-box protein family. *Genome Biology* 1: 3002.
- 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.
- Laetsch DR, Blaxter ML. 2017. BLOBTOOLS: interrogation of genome assemblies. *F1000Research* 6: 12232.
- Lebreton A, Zeng Q, Miyauchi S, Kohler A, Dai Y-C, Martin FM. 2021. Evolution of the mode of nutrition in symbiotic and saprotrophic fungi in forest ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 52: 385–404.
- Lofgren LA, Nguyen NH, Vilgalys R, Ruytinx J, Liao HL, Branco S, Kuo A, Labutti K, Lipzen A, Andreopoulos W *et al.* 2021. Comparative genomics reveals dynamic genome evolution in host specialist ectomycorrhizal fungi. *New Phytologist* 230: 774–792.
- Looney B, Miyauchi S, Morin E, Drula E, Courty PE, Kohler A, Kuo A, Labutti K, Pangilinan J, Lipzen A *et al.* 2022. Evolutionary transition to the ectomycorrhizal habit in the genomes of a hyperdiverse lineage of mushroom-forming fungi. *New Phytologist* 233: 2294–2309.
- Lukinović V, Casanova AG, Roth GS, Chuffart F, Reynoird N. 2020. Lysine Methyltransferases signaling: histones are just the tip of the iceberg. *Current Protein and Peptide Science* 21: 655–674.
- Macdonald JA, Wijekoon CP, Liao KC, Muruve DA. 2013. Biochemical and structural aspects of the ATP-binding domain in inflammasome-forming human NLRP proteins. *IUBMB Life* 65: 851–862.
- de Man TJ, Stajich JE, Kubicek CP, Teiling C, Chenthamara K, Atanasova L, Druzhinina IS, Levenkova N, Birnbaum SS, Barribeau SM *et al.* 2016. Small genome of the fungus *Escovopsis weberi*, a specialized disease agent of ant agriculture. *Proceedings of the National Academy of Sciences, USA* 113: 3567–3572.
- Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* 33: 574–576.
- Martin F, Aerts A, Ahrén D, Brun A, Danchin EG, 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, Duplessis S, Ditengou F, Lagrange H, Voilet C, Lapeyrie F. 2001. Developmental cross talking in the ectomycorrhizal symbiosis: signals and communication genes. *New Phytologist* 151: 145–154.
- Matheny PB, Aime MC, Bougher NL, Buyck B, Desjardin DE, Horak E, Kropp BR, Lodge DJ, Soyong K, Trappe JM *et al.* 2009. Out of the palaeotropics? Historical biogeography and diversification of the cosmopolitan ectomycorrhizal mushroom family Inocybaceae. *Journal of Biogeography* 36: 577–592.
- Matheny PB, Curtis JM, Hofstetter V, Aime MC, Moncalvo JM, Ge ZW, Slot JC, Ammirati JF, Baroni TJ, Bougher NL *et al.* 2006. Major clades of Agaricales: a multilocus phylogenetic overview. *Mycologia* 98: 982–995.
- Matheny PB, Hobbs AM, Esteve-Raventós F. 2020. Genera of Inocybaceae: new skin for the old ceremony. *Mycologia* 112: 83–120.
- Matheny PB, Kudzma LV. 2019. New species of *Inocybe* (Inocybaceae) from eastern North America. *Journal of the Torrey Botanical Society* 146: 213–235.
- Maurice S, Arnault G, Nordén J, Botnen SS, Miettinen O, Kausserud H. 2021. Fungal sporocarps house diverse and host-specific communities of fungicolous fungi. *The ISME Journal* 15: 1445–1457.
- Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31: i44–i52.
- 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.
- Mohanta TK, Bae H. 2015. The diversity of fungal genome. *Biological Procedures Online* 17: 8.
- Muszevska A, Stepniewska-Dziubinska MM, Steczkiewicz K, Pawlowska J, Dziedzic A, Ginalski K. 2017. Fungal lifestyle reflected in serine protease repertoire. *Scientific Reports* 7: 9147.
- Nagarajan SN, Lenoir C, Grangeasse C. 2022. Recent advances in bacterial signaling by serine/threonine protein kinases. *Trends in Microbiology* 30: 553–566.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.
- Palmer J, Stajich J. 2019. Nextgenusfs/funannotate: funannotate, v.1.5.3. *Zenodo*. doi: 10.5281/zenodo.2604804.
- 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: 1278.
- 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.
- Piombo E, Guaschino M, Jensen DF, Karlsson M, Dubey M. 2023. Insights into the ecological generalist lifestyle of Clonostachys fungi through analysis of their predicted secretomes. *Frontiers in Microbiology* 14: 1112673.
- Prijbelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. *Current Protocols in Bioinformatics* 70: e102.
- Rizzi YS, Happel P, Lenz S, Urs MJ, Bonin M, Cord-Landwehr S, Singh R, Moerschbacher BM, Kahmann R. 2021. Chitosan and chitin deacetylase activity are necessary for development and virulence of *Ustilago maydis*. *MBio* 12: e03419.
- Rubin E, Lithwick G, Levy AA. 2001. Structure and evolution of the hAT transposon superfamily. *Genetics* 158: 949–957.
- Ryberg M, Khan FK, Sanchez-Garcia M. 2022. On the evolution of ectomycorrhizal fungi. *Mycosphere* 13: 1–12.
- Ryberg M, Larsson E, Jacobsson S. 2010. An evolutionary perspective on morphological and ecological characters in the mushroom family Inocybaceae (Agaricomycotina, Fungi). *Molecular Phylogenetics and Evolution* 55: 431–442.
- Ryberg M, Matheny PB. 2012. Asynchronous origins of ectomycorrhizal clades of Agaricales. *Proceedings of the Royal Society B: Biological Sciences* 279: 2003–2011.
- Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19: 301–302.

- Schuelke TA, Wu G, Westbrook A, Woeste K, Plachetzki DC, Broders K, MacManes MD. 2017. Comparative genomics of pathogenic and nonpathogenic beetle-vectored fungi in the genus *Geosmithia*. *Genome Biology and Evolution* 9: 3312–3327.
- Sheikh S, Khan FK, Bahram M, Ryberg M. 2022. Impact of model assumptions on the inference of the evolution of ectomycorrhizal symbiosis in fungi. *Scientific Reports* 12: 22043.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.
- Singer R. 1986. *The Agaricales in modern taxonomy*. Königstein, Germany: Sven Koeltz Scientific Books.
- Song F, Qi D, Liu X, Kong X, Gao Y, Zhou Z, Wu Q. 2015. Proteomic analysis of symbiotic proteins of *Glomus mosseae* and *Amorpha fruticosa*. *Scientific Reports* 5: 18031.
- Spanu PD. 2012. The genomics of obligate (and nonobligate) biotrophs. *Annual Review of Phytopathology* 50: 91–109.
- Steidinger BS, Crowther TW, Liang J, Van Nuland ME, Werner GDA, Reich PB, Nabuurs GJ, 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.
- Thompson JA, Brown JC. 2012. Role of coatamer protein I in virus replication. *Journal of Virology and Antiviral Research* 1: 102.
- Tlálaski V, Brabčová V, Větrovský T, Jomura M, López-Mondéjar R, Oliveira Monteiro LM, Saraiva JP, Human ZR, Cajthaml T, Nunes Da Rocha U *et al.* 2021. Complementary roles of wood-inhabiting fungi and bacteria facilitate deadwood decomposition. *mSystems* 6: 1078.
- Turrà D, El Ghalid M, Rossi F, Di Pietro A. 2015. Fungal pathogen uses sex pheromone receptor for chemotropic sensing of host plant signals. *Nature* 527: 521–524.
- Van Nocker S, Ludwig P. 2003. The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC Genomics* 4: 50.
- Van Petegem F, Contreras H, Contreras R, Van Beeumen J. 2001. Trichoderma reesei alpha-1,2-mannosidase: structural basis for the cleavage of four consecutive mannose residues. *Journal of Molecular Biology* 312: 157–165.
- Venkatesh N, Koss MJ, Greco C, Nickles G, Wiemann P, Keller NP. 2021. Secreted secondary metabolites reduce bacterial wilt severity of tomato in bacterial-fungal co-infections. *Microorganisms* 9: 2123.
- Vinale F, Sivasithamparam K, Ghisalberti EL, Ruocco M, Wood S, Lorito M. 2012. Trichoderma secondary metabolites that affect plant metabolism. *Natural Product Communications* 7: 1545–1550.
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J *et al.* 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17: 261–272.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9: e112963.
- Wen Y, Nguyen D, Li Y, Lai ZC. 2000. The N-terminal BTB/POZ domain and C-terminal sequences are essential for Tramtrack69 to specify cell fate in the developing *Drosophila* eye. *Genetics* 156: 195–203.
- Wohlschlager L, Kracher D, Scheiblbrandner S, Csarman F, Ludwig R. 2021. Spectroelectrochemical investigation of the glyoxal oxidase activation mechanism. *Bioelectrochemistry* 141: 107845.
- Wu G, Miyauchi S, Morin E, Kuo A, Drula E, Varga T, Kohler A, Feng B, Cao Y, Lipzen A *et al.* 2022. Evolutionary innovations through gain and loss of genes in the ectomycorrhizal Boletales. *New Phytologist* 233: 1383–1400.
- Xu Q, Wang J, Zhao J, Xu J, Sun S, Zhang H, Wu J, Tang C, Kang Z, Wang X. 2020. A polysaccharide deacetylase from *Puccinia striiformis* f. sp. *tritici* is an important pathogenicity gene that suppresses plant immunity. *Plant Biotechnology Journal* 18: 1830–1842.
- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. DBCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 40: W445–W451.
- Zhang F, Anasztzisz GE, Labourel A, Champion C, Haon M, Kempainen M, Commu C, Deveau A, Pardo A, Veneault-Fourrey C *et al.* 2018. The ectomycorrhizal basidiomycete *Laccaria bicolor* releases a secreted β -1,4 endoglucanase that plays a key role in symbiosis development. *New Phytologist* 220: 1309–1321.
- Zhang G, Li S, Cheng KW, Chou TF. 2021. AAA ATPases as therapeutic targets: structure, functions, and small-molecule inhibitors. *European Journal of Medicinal Chemistry* 219: 113446.

Supporting Information

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

Fig. S1 Coalescent-based tree generated by Astral for 26 species used in the study.

Fig. S2 Significantly evolving orthogroups in the 26 species used in the study.

Fig. S3 Gene family expansions in Carbohydrate-Active enZymes, peptidases and lipases in Inocybaceae and saprotrophic species.

Table S1 Details of the samples used in this study.

Table S2 Busco scores of the assemblies before and after the filtration.

Table S3 Genomic features of Inocybaceae and all outgroup species.

Table S4 Full list of rapidly evolving orthogroups at the different branch-sets used in the study.

Table S5 Model comparisons with different rates for different branch-sets in the CAFE analysis.

Table S6 CAZyme gene families.

Table S7 Peptidase gene families.

Table S8 Lipase gene families.

Table S9 Number of secreted lipases in different lipase classes in all samples used in this study.

Please note: Wiley is 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.