OXFORD

Resource Article: Genomes Explored

The genome of *Lyophyllum shimeji* provides insight into the initial evolution of ectomycorrhizal fungal genomes

Yuuki Kobayashi[1,*], Tomoko F. Shibata[2], Hideki Hirakawa[3], Tomoaki Nishiyama[4],
Akiyoshi Yamada[5], Mitsuyasu Hasebe[2,6], Shuji Shigenobu[1,6,7], and Masayoshi Kawaguchi[6,8,*]

[1]Laboratory of Evolutionary Genomics, National Institute for Basic Biology, Nishigonaka 38, Myodaiji, Okazaki, Aichi 444-8585, Japan
[2]Division of Evolutionary Biology, National Institute for Basic Biology, Nishigonaka 38, Myodaiji, Okazaki, Aichi 444-8585, Japan
[3]Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
[4]Research Center for Experimental Modeling of Human Disease, Kanazawa University, Kanazawa, Ishikawa 920-0934, Japan
[5]Faculty of Agriculture, Shinshu University, Kamiina, Nagano 399-4598, Japan
[6]Department of Basic Biology, SOKENDAI, Okazaki, Aichi 444-8585, Japan
[7]Trans-omics Facility, National Institute for Basic Biology, Okazaki, Aichi 444-8585, Japan
[8]Division of Symbiotic systems, National Institute for Basic Biology, Okazaki, Aichi 444-8585, Japan

*To whom correspondence should be addressed. Tel.: +81-0564-55-7672, Email: kobayasi@nibb.ac.jp (Y.K.); Tel.: +81-0564-55-7560; E-mail: masayosi@nibb.ac.jp (M.K.)

Abstract

Mycorrhizae are one of the most fundamental symbioses between plants and fungi, with ectomycorrhizae being the most widespread in boreal forest ecosystems. Ectomycorrhizal fungi are hypothesized to have evolved convergently from saprotrophic ancestors in several fungal clades, especially members of the subdivision Agaricomycotina. Studies on fungal genomes have identified several typical characteristics of mycorrhizal fungi, such as genome size expansion and decreases in plant cell-wall degrading enzymes (PCWDEs). However, genomic changes concerning the evolutionary transition to the ectomycorrhizal lifestyle are largely unknown. In this study, we sequenced the genome of *Lyophyllum shimeji*, an ectomycorrhizal fungus that is phylogenetically related to saprotrophic species and retains some saprotroph-like traits. We found that the genome of *Ly. shimeji* strain AT787 lacks both incremental increases in genome size and reduced numbers of PCWDEs. Our findings suggest that the previously reported common genomic traits of mycorrhizal fungi are not essential for the ectomycorrhizal lifestyle, but are a result of abolishing saprotrophic activity. Since *Ly. shimeji* is commercially consumed as an edible mushroom, the newly available genomic information may also impact research designed to enhance the cultivation of this mushroom.

Key words: mycorrhizal fungi, comparative genomics, ectomycorrhizae, edible mushroom

1. Introduction

Mycorrhizal symbiosis is one of the most abundant types of symbioses between fungi and plants, involving more than one hundred fungal genera and 80% of land plant species.[1,2] This type of symbiosis is basically mutualistic in that fungi help plants take up soil-derived nutrients, such as phosphate and nitrate, in exchange for photoassimilates.[3] Mycorrhizae are classified into several types by their morphology and function. Of them, ectomycorrhizae have the most impact on boreal forest biomass because dominant trees in the boreal forest harbour these types of mycorrhizal fungi.[4,5]

Ectomycorrhizal plants appear in several taxa, including the representative families Fagaceae and Pinaceae.[2] These mycorrhizae are proposed to have changed their mycorrhizal style from ancestral arbuscular mycorrhizae and have succeeded in becoming adapted to boreal regions. In contrast, ectomycorrhizal fungi are thought to have evolved from saprotrophic ancestors.[6] Accounting for their taxonomic appearance, the transition from the saprotrophic to

the ectomycorrhizal lifestyle is proposed to have independently evolved dozens of times throughout the history of Basidiomycota and Ascomycota.[1,6]

Besides their ecological importance as forest components, many ectomycorrhizal fungi also impact our lives as foods. Many favoured mushrooms, such as truffles, boletes, and matsutake, are mycorrhizal.[7] However, unlike saprotrophic species, most mycorrhizal mushrooms have difficulty artificially producing basidiocarps, probably because of their host-dependent lifestyle.[7]

Lyophyllum shimeji, an agaricoid fungus known as hon-shimeji or just shimeji in the older literature, has been one of the most favoured mushrooms in Japan for a long time.[8,9] *Ly. shimeji* is known to be an ectomycorrhizal fungus associated with plants in the Fagaceae and Pinaceae families.[10,11] However, unlike many other mycorrhizal fungi, *Ly. shimeji* can metabolize botanical polysaccharides, such as starch.[12] Moreover, the fruiting body of *Ly. shimeji* can be induced on a medium containing barley and sawdust without a host

plant,[13] and it is now commercially produced axenically. Thus, *Ly. shimeji* has physiological similarities to saprotrophic fungi though it has a mycorrhizal lifestyle. Also, *Ly. shimeji* has a systematically close saprotrophic relative, *Ly. decastes*.[14,15] Morphological traits of *Ly. shimeji*, *Ly. decastes* and several other species highly resemble each other. Descriptions of the mycorrhizal ability of these relatives are also confusing. For example, *Ly. decastes* is often described as a mycorrhizal fungus,[16,17] whereas other studies have reported its inability to form mycorrhizae,[11,18] possibly because of improper species identification.[19] Therefore, *Ly. shimeji* and related species include mycorrhizal and non-mycorrhizal fungi within a very close genetic distance. Since ancestral species of Lyophyllaceae are proposed to be saprotrophic,[17] *Ly. shimeji* may be one of the last groups of ectomycorrhizal fungi that has evolved from a saprotrophic ancestor.

Since the genome of *Laccaria bicolor* was sequenced as the first mycorrhizal genome in 2008,[20] many other genomes of mycorrhizal fungi have been sequenced.[21-26] Comparative studies between mycorrhizal and saprotrophic fungi have revealed that mycorrhizal fungi have several characteristic genomic features, such as genome size increments accompanied by the proliferation of transposable elements (TEs) and reductions in genes encoding plant cell-wall degrading enzymes (PCWDEs).[23-26] These convergent genomic traits suggest that such features are adaptive to the mycorrhizal lifestyle. However, to what degree these features are essential for the mycorrhizal lifestyle is largely unknown. In this study, we sequenced the genome of *Ly. shimeji* as a representative of a putative primitive stage of mycorrhizal fungi in order to survey the initial genomic status in the ecological transition. The genome information of *Ly. shimeji* is also expected to be useful for research and cultivation of this praised mushroom.

2. Materials and methods

2.1. Fungal material

The strain of *Ly. shimeji* AT787 was initially isolated from Nagano and maintained at Shinshu University.[27] This strain was confirmed to be mycorrhizal and has been used for research on mycorrhizal formation. Hyphae were cultured on a cellophane membrane placed on MMN (Modified Melin-Norkrans) medium[28] and MMN medium without malt and sugar (hereafter referred to as MMN salt-only medium). After an 1-week incubation at 28°C, hyphae were peeled from the cellophane membrane and used for nucleic acid extraction.

2.2. Extraction of genomic DNA and RNA

Genomic DNA was isolated from hyphae cultured on MMN medium using the CTAB (cetyl trimethylammonium bromide) method[29] with some modifications as described below. Collected hyphae were frozen with liquid nitrogen, ground with a pestle, and dispersed in extraction buffer (100 mM Tris–HCl pH 8.0, 20 mM EDTA, 0.75% sarkosyl, 0.1% PVP, 0.75% CTAB, 0.13 M sorbitol, 0.75 M NaCl, and 0.1 mg ml^{-1} proteinase K). After incubation at 37°C for 30 min, the aqueous phase was centrifuged at 12,000 rpm (rotations per minute) for 10 min at 4°C, and the pellet was discarded. An equal volume of phenol/chloroform (1:1, Vol:Vol) was added and centrifuged (12,000 rpm, 5 min at 4°C). The aqueous phase was collected, and an equal volume of chloroform was added to the sample and

centrifuged (12,000 rpm, 5 min at 4°C). The aqueous phase was collected again, and 1:10 Vol of sodium acetate and 0.7 Vol of isopropanol were added, mixed, and centrifuged (12,000 rpm, 20 min at 4°C). The resulting pellet was washed twice with 70% EtOH and resuspended in TE buffer. Extracted DNA was purified with Genomic-tips (Qiagen, Netherlands). RNAs were extracted from hyphal samples cultured on MMN medium and MMN salt-only medium using an RNeasy Plant Mini Kit (Qiagen, Netherlands) with buffer RLC included in the kit.

2.3. Library construction and sequencing

PacBio libraries were constructed following the manufacturer's protocol (Pacific Bioscience, USA) and sequenced with a PacBio RS II sequencer. Paired-end Illumina genomic libraries with insert sizes of 250 bp and 700 bp were constructed with a TruSeq DNA Sample Preparation v2 Kit (Illumina, USA) and were sequenced with a HiSeq 2500 System (Illumina, USA) for 151 bp from both ends. RNA-seq libraries were constructed with a TruSeq Stranded mRNA Library Prep Kit (Illumina, USA) and sequenced with a HiSeq 1500 System (Illumina, USA) for 126 bp from both ends.

2.4. Genome assembly

The genome size of *Ly. shimeji* was estimated with GenomeScope 2.0[30] using the Illumina reads. PacBio reads were primarily assembled with Falcon v2017.11.02-16.04-py2.7-ucs and Falcon-Unzip v0.4.0.[31] Two allelic draft genomes were integrated into a single haploid-type genome with HaploMerger2.[32] Repetitive elements were predicted using RepeatMasker v4.0.6[33] after modelling the repetitive sequences with RepeatModeler v1.0.8[33] with default settings. The completeness of the draft genome was assessed with BUSCO ver.4.1.4, using the database fungi_odb10.[34]

2.5. Gene prediction and annotation

Genes were predicted with mapped transcripts. Adapters and low-quality sequences were removed from Illumina RNA-seq reads with Cutadapt 2.5.[35] Transcripts were assembled with Trinity v2.11.0.[36] Genes were predicted with BRAKER v2.1.0[37] with the combination of GeneMark-ES v4.57[38] and Augustus v3.3.1.[39] The completeness of the gene set was assessed with BUSCO v4.1.4,[34] using the database fungi_odb10 (*n* = 758). Genes were functionally annotated with eggNOG-mapper v2.[40]

2.6. Reference genome datasets

Reference genomes were downloaded from JGI Mycocosm.[41,42] Assembly genomes of *Agaricus bisporus* var *bisporus* (H97) v2.0,[43] *Amanita muscaria* Koide v1.0,[23] *Amanita thiersii* Skay4041 v1.0,[44] *Aspergillus oryzae* RIB40,[45] *Auricularia subgrabra* v2.0,[46] *Boletus edulis* Prilba v1.0,[25] *Cenococcum geophilum* 1.58 v2.0,[47] *Coprinopsis cinerea*,[48] *Galerina marginata* v1.0,[49] *Hebeloma cylindrosporum* h7 v2.0,[23] *Laccaria bicolor* v2.0,[20] *Lentinula edodes* W1-26 v1.0,[50] *Neurospora crassa* OR74A v2.0,[51] *Pisolithus tinctorius* v1.0,[23] *Pleurotus ostreaus* PC15 v2.0,[49] *Rhizophagus clarus* HR1,[52] *Rhizopus delemer* 99-880,[53] *Saccharomyces cerevisiae* S288C,[54] *Schizosaccharomyces pombe*,[55] *Tricholoma matsutake* 945 v3.0,[25] and *Tuber melanosporum* Mel28 v1.2[21] were used as references. Assembly genome and raw reads of *Ly. shimeji* JCM30591 were downloaded from the NCBI database.[56]

2.7. Comparative analyses

PCWDEs were searched with dbCAN2 web server.[57,58] Only genes that were predicted as CAZymes by two or three of three methods in dbCAN2 were regarded as PCWDEs. Secretory proteins were predicted with SignalP v4.1.[59] Small proteins were selected with Seqkit v0.5.2.[60] Ortholog analyses were conducted with OrthoFinder v2.3.11[61] with default parameters. Three-hundred fifty-one single-copy orthologs were aligned with MAFFT v7.490[62] with default parameters to generate a phylogenetic tree constructed with IQ-tree v2.1.3[63] with 1,000 bootstrapping replicates and 1,000 SH-aLRT (Shimodaira-Hasegawa approximate likelihood ratio test) replicates. 'Q.yeast+F+R6' model was chosen by the ModelFinder in IQ-tree. PCA and correlation efficiency were computed, Welch's *t*-test was executed, and figure drawings were created using R 4.0.3.[64]

## 3. Results and discussion

3.1. Genome sequencing and assembly

We sequenced the entire genome of *Ly. shimeji* AT787 with a short-read sequencer, Illumina HiSeq2500, and a long-read sequencer, PacBio RSII. A total of 13.6 Gbp of Illumina reads and 5.81 Gbp of PacBio reads were obtained. From the k-mer analysis of the Illumina short reads, the genome size was estimated to be approximately 30.5 Mbp with 2.3% heterozygosity (Fig. S1). We assembled a draft genome with Falcon-Unzip, which generates a 'primary contig' as a basal allele and a 'haplotig' as the opposite allele. A total of 36.3 Mbp of primary contigs and 23.7 Mbp of haplotigs were obtained. Since the primary contigs may contain both alleles partially, we merged primary contigs and haplotigs into single alleles, and 31.8 Mbp of the draft genome with an N50 of 1.61 Mbp was generated (Table 1). BUSCO assessment found 96.9% of the conserved single-copy genes were complete genes in the draft genome. From this draft genome, 12,904 genes were predicted (Table 1). The completeness of the BUSCO predicted proteome was estimated to be 97.8%, including 97.0% being single-copy sequences (Table 1). The contiguity of our assembly and BUSCO completeness of the genome are above or comparable to average for representative reference genomes (Table S1).

3.2. Comparison of genome size and content with other fungal genomes

Since mycorrhizal fungi tend to have large genomes with manyTEs, we compared the genome size, gene number, and repeat content, including TEs, of *Ly. shimeji* AT787 with the data registered in JGI Mycocosm (accessed Oct. 2020)[41,42] and the data of another strain of *Ly. shimeji*, JCM30591 (NCBI GenBank GCA_001950515). The plot of genome size and gene number showed that genomes of mycorrhizal fungi are generally larger than those of other fungi while gene numbers are not; however, the genome size of *Ly. shimeji* AT787 is very small among mycorrhizal fungi yet the gene number is relatively similar to other mycorrhizal fungi, while that of *Ly. shimeji* JCM30591 is not (Fig. 1a). The low BUSCO duplication ratio (Table S1) suggest neither are mixture of multiple genomes or polyploid and genome size estimates from raw short reads (57.7 Mb for JCM30591, DRR037069) suggests the genome size of strain JCM30591 and AT787 differ by 1.9-fold. The difference in the genome size implies that JCM30591 and AT787 are different species in the so-called *Lyophyllum decastes* species complex; however, we cannot state which strain is strict '*Ly. shimeji*' because the holotype sequence is unavailable. Since many strains including AT787 have been treated as *Ly. shimeji* in previous studies,[27] we regarded both strains as *Ly. shimeji* in this paper. The repeat content of *Ly. shimeji* AT787 is 6.84% including 6.14% TEs (Fig. 1b). Most of the TEs in the *Ly. shimeji* AT787 genome were Gypsy-type LTR elements (Fig. 1b). No SINEs and LINEs were detected while certain amount of LINEs was found in the *Ly. shimeji* JCM30591 genome (Fig. 1b and c). When compared to representative fungal genomes, the repeat content in *Ly. shimeji* AT787 was smaller than that in most, if not all, ectomycorrhizal genomes (Fig. 1c). These results indicated that neither genome size enlargement nor TE proliferation occurred in the genome of *Ly. shimeji* AT787, unlike typical ectomycorrhizal fungi. Therefore, a large genome size is not an essential trait for the mycorrhizal lifestyle but, rather, a result of mycorrhizal history. Generally, TEs are not adaptive to organisms but provide disadvantages for cell proliferation.[65] Mycorrhizal fungi tend to have a slower cell cycle than saprotrophic species; thus, mycorrhizal fungi may have experienced a weaker selection pressure to eliminate such junk sequences. We hypothesize that such a relaxed selection pressure might be the driving force for large genomes in mycorrhizal species.

Table 1. Basic information about the *Ly. shimeji* AT787 assembled genome and predicted genes

Genome assembly (purged)		Predicted genes	
Total bases	31,834,823 bp	Number of genes	12,904
No. of scaffolds	60	Average of CDS length	1375.9 bp
N50	1,610,882 bp	Average of protein length	457.6 aa
L50	7		
GC%	53.92%		
BUSCO benchmarks of genome assembly		BUSCO benchmarks of predicted genes	
Complete	96.9 %	Complete	97.8%
Complete single copy	96.2 %	Complete single copy	97.0%
Complete duplicated	0.7 %	Complete duplicated	0.8%
Fragmented	0.3 %	Fragmented	1.2%
Missing	2.8 %	Missing	1.0%
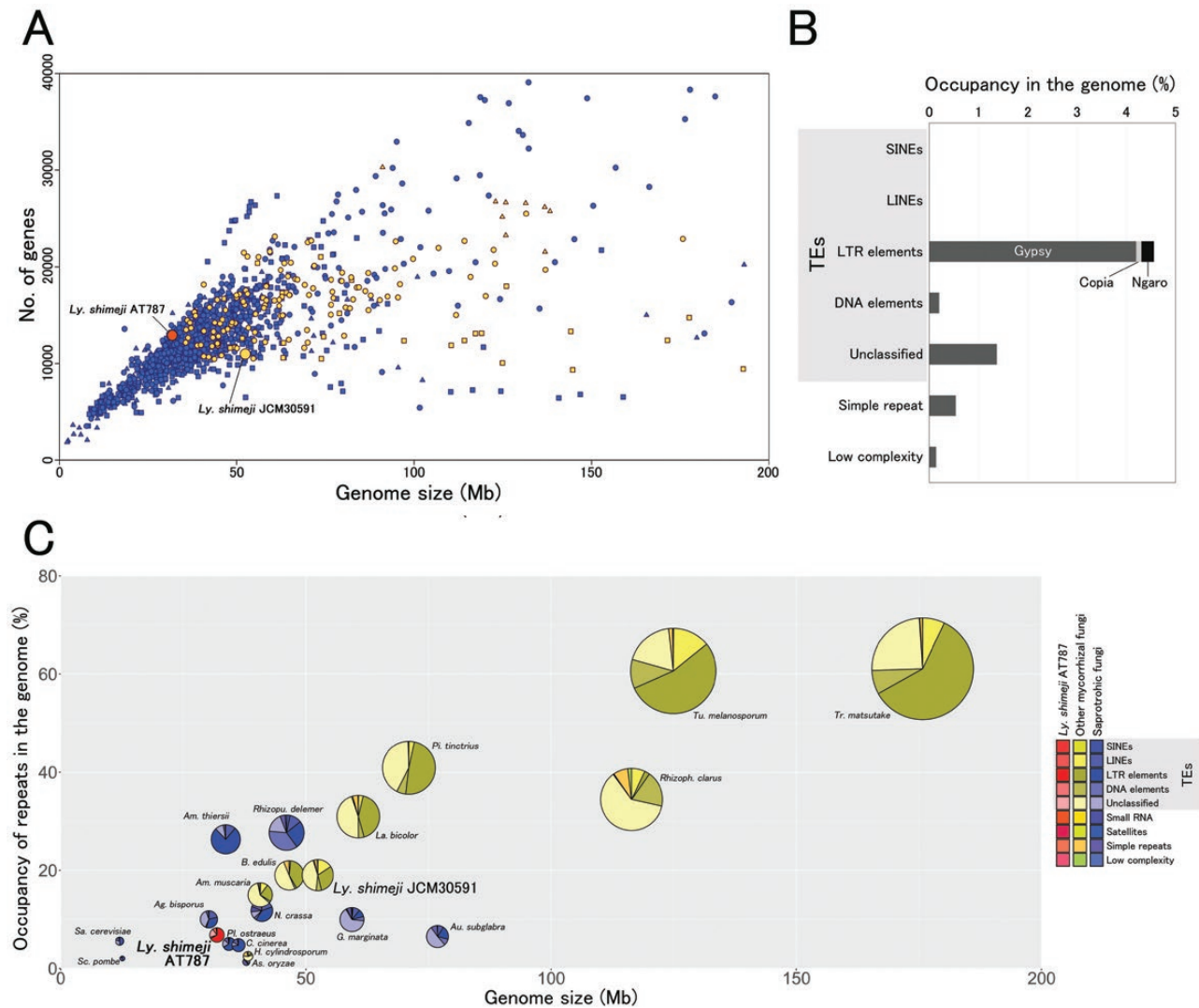
Figure 1. Comparison of genome size and gene number among sequenced fungi Genome size and components of *Ly. shimeji* AT787 compared with other fungi. (A) Genome size and gene numbers of fungi. Red, yellow, and blue markers correspond to *Ly. shimeji* AT787, other mycorrhizal fungi, and non-mycorrhizal fungi, respectively. Circles, squares, and triangles correspond to Basidiomycota, Ascomycota, and other taxa, respectively. (B) Repetitive sequences, including TEs in *Ly. shimeji* AT787. C: Genome size and repeat content of fungi. The size of circles corresponds to the absolute number of repeated sequences. The fungal genomes registered in JGI Mycocosm and the genome of *Ly. shimeji* JCM30591 registered in NCBI were used as references.

3.3. PCWDE gene content

We next compared the repertoire of PCWDE genes in the *Ly. shimeji* AT787 genome compared with other fungal genomes because the number of PCWDEs is known to be lower in mycorrhizal genomes. We found 384 Carbohydrate-Active enZymes (CAZymes), including 178 secretory CAZymes, in *Ly. shimeji* AT787 (Fig. 2a). Genomes of reference species showed that saprotrophic fungi, except yeasts, harbour more CAZymes, especially secretory enzymes, than mycorrhizal fungi (Fig. 2a and b), as reported in previous studies.[23,25] However, the number of CAZymes in *Ly. shimeji* AT787 was much higher than in reference mycorrhizal fungi and even some saprotrophic fungi (Fig. 2) despite its mycorrhizal ecology. Interestingly, the genome of *Ly. shimeji* JCM30591 also harbours high amount of CAZymes while genome size increment was observed like other mycorrhizal species (Fig. 2). Thus, depletion of PCWDEs is not an essential trait of ectomycorrhizal fungi. Therefore, unlike other mycorrhizal

fungi, *Ly. shimeji* may have the capacity to decay plant-derived materials like saprotrophic species, a finding consistent with its ability to live on plant-derived substances. Mycorrhizal fungi have been hypothesized to avoid hurting plant cell walls to prevent the activation of plant defence systems.[47] However, it is more likely that mycorrhizal fungi just abandoned the use of PCWDEs after their ecological change because mycorrhizal fungi can directly derive monosaccharides from their host plants, as also hypothesized by Miyauchi et al.[25]

3.4. Small-secreted proteins

Mycorrhizal fungi harbour many small-secreted proteins (SSPs).[23,66] Since SSPs include mycorrhiza-induced genes, and some of them are essential for the establishment of symbiosis, SSPs are thought to communicate with or modify mycorrhizal host plants.[20,23,67,68] We identified the SSPs from the gene repertoire of *Ly. Shimeji* AT787 and compared
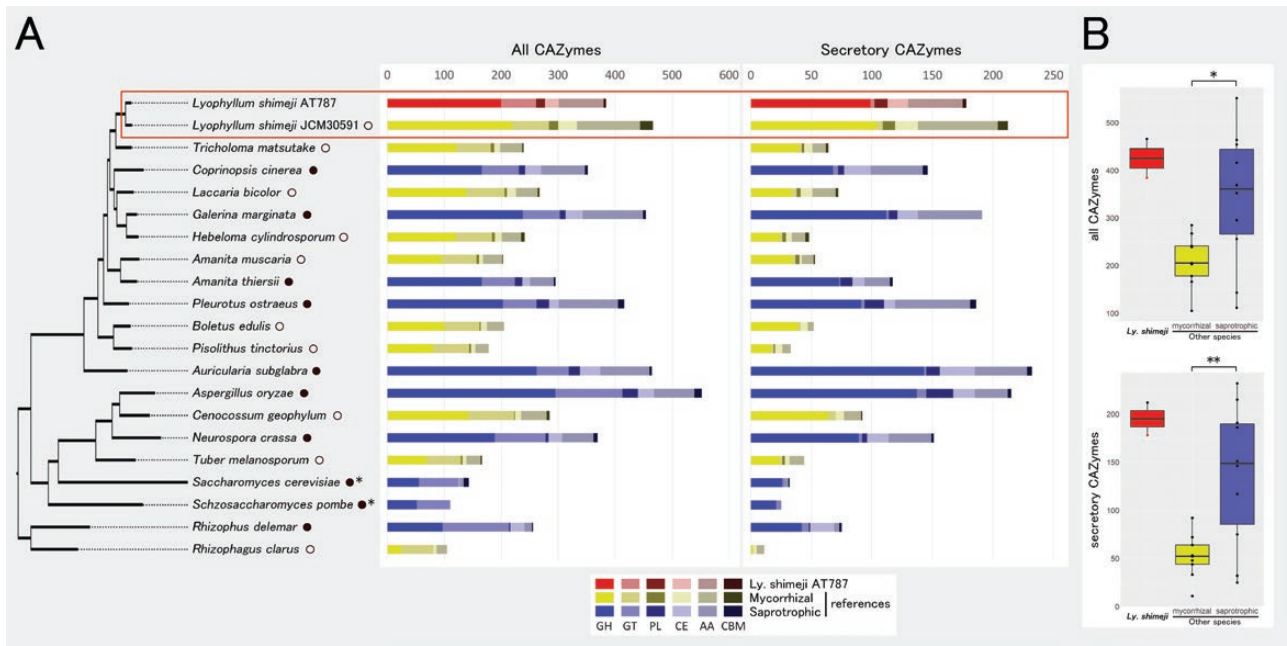
Figure 2. Comparison of PCWDEs with other representative fungi numbers of PCWDEs in the genome of *Ly. shimeji* AT787 in comparison with other representative fungi that have been reported. (A) Numbers of CAZymes found in each genome. The phylogenetic tree was constructed with jointed sequences of 351 single-copy orthologs. GH, GT, PL, CE, AA, and CBM correspond to glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases, auxiliary activities, and cellulose-binding motifs, respectively. Strains of *Ly. shimeji* are bordered in red. Asterisks after the species name indicate yeasts with a small number of total genes. Open and closed circles after the species name indicate mycorrhizal and saprotrophic references, respectively. (B) Comparison of numbers between *Ly. shimeji* strains, other mycorrhizal species, and saprotrophic species. Plots of *Ly. shimeji* AT787 are shown in red. Asterisks indicate significant differences (an asterisk for Welch's *T*-test *P* < 0.05 and two asterisks for Welch's *T*-test *P* < 0.01).

them with those from reference genomes. However, when we selected proteins with predicted secretory signal peptides and sizes of ≤300 aa, there was no apparent difference in gene number between mycorrhizal and saprotrophic fungi even in reference species (Fig. S2). We next checked the functional annotation of predicted SSPs in *Ly. Shimeji* AT787 (Table S2). Some of the predicted SSPs were annotated as hydrophobins, which include proteins having a function in mycorrhizal symbiosis.[69,70] Some annotated SSPs are secretory CAZymes (Table S2). The lack of apparent difference in the total number of SSPs between mycorrhizal and saprotrophic species may be because SSPs include effectors that are considered to be abundant in mycorrhizal fungi and CAZymes that tend to be reduced in mycorrhizal fungi. Approximately 60% of the predicted SSPs were functionally unannotated (Table S2). Since some mycorrhiza-related SSPs are lineage-specific,[23] some of these unannotated SSPs could function in mycorrhizal symbiosis.

3.5. Ortholog analysis

We compared the gene repertoire of *Ly. Shimeji* AT787 with several other fungal genomes. When compared with two mycorrhizal fungi (*La. bicolor* and *H. cylindrosporum*) and two saprotrophic fungi (*Co. cinerea* and *G. marginata*) in the Agaricales, *Ly. shimeji* AT787 and the two saprotrophic species share 126 orthologs that are not shared with the two mycorrhizal species (Fig. 3a, bordered in red). In contrast, *Ly. shimeji* AT787 and the other two mycorrhizal species share only 32 orthologs that are not shared with saprotrophic species (Fig. 3a, bordered in white). Thus, *Ly. shimeji* AT787 shares more orthologs with saprotrophic species than with

other mycorrhizal species. Genes shared by *Ly. shimeji* AT787 and two saprotrophic species include 20 glycoside hydrolases as the most enriched common orthologs (Table S3). This result confirmed the absence of a mycorrhiza-specific reduction in *Ly. shimeji* AT787 CAZymes, found by CAZyme-targeted analysis as described above, by non-targeted orthologous analysis. In contrast, most orthologs shared by only *Ly. Shimeji* AT787 and two mycorrhizal fungi were functionally unannotated and did not show enrichment in specific functions (Table S4).

We next searched for other genes that differ between mycorrhizal and saprotrophic fungi. We conducted PCA analysis of 19 dikaryon fungi, including 9 saprotrophic species, 9 mycorrhizal reference species/strains and *Ly. shimeji* AT787 (Fig. S3). However, there was no evidence of an axis reflecting ecology, whereas PC1 clearly reflected taxonomic differences (Fig. S3).

We also searched ortholog groups (orthogroups) whose member numbers are correlated with ecological types and phylogeny (Fig. S4). Although a higher concordance with phylogeny was observed, the correlation to ecology was limited and only eight orthogroups showed a high correlation (correlation coefficient > 0.6 or < −0.6) with ecological type (Fig. S4). In comparison, 2,778 orthogroups showed a high correlation with phylum (Fig. S4). Among orthogroups showing a high correlation with ecological type, an orthogroup encoding PCWDE, GH11, was absent in all mycorrhizal species, including two *Ly. shimeji* strains (Table S5). No other orthogroup was annotated as having any apparent function related to the mycorrhizal lifestyle. Three genes encoding COX2, COX3, and NADH5 are usually
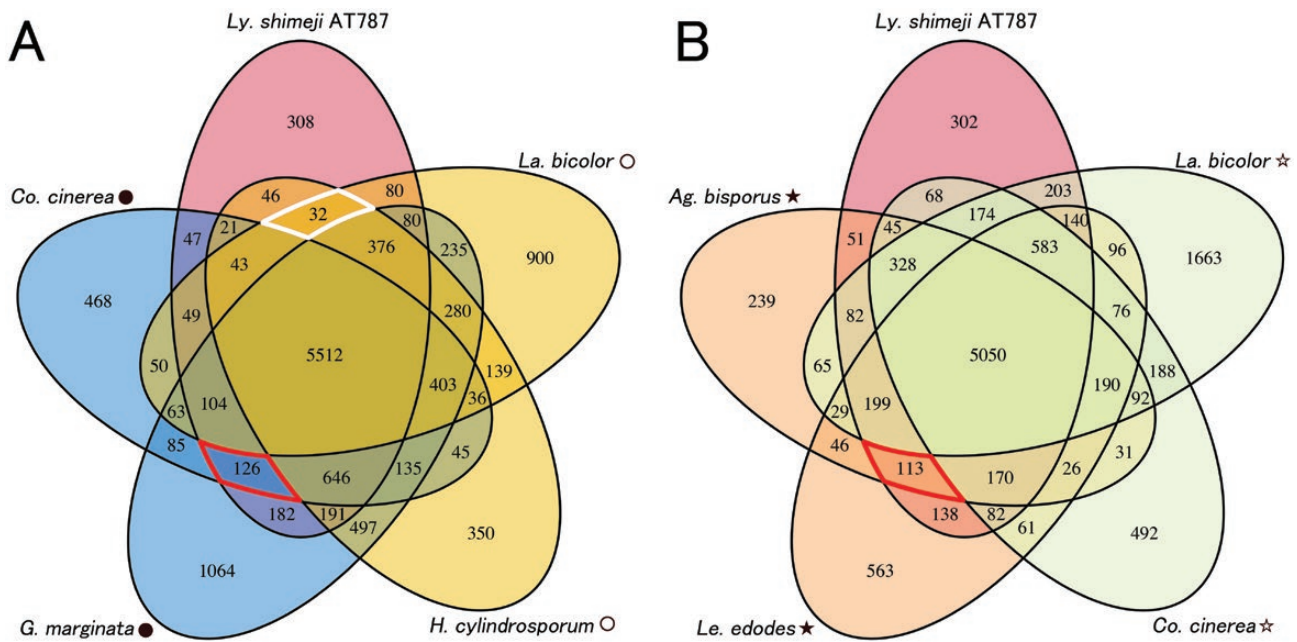
Figure 3. Shared orthologs with other fungi in the Agaricales. Venn diagrams of shared orthologs. (A) Orthologs shared with *Ly. shimeji* AT787, two mycorrhizal reference species (*La. bicolor* and *H. cylindrosporum*), and two saprotrophic reference species (*Co. cinerea* and *G. marginata*). Red, yellow, and blue ellipses correspond to *Ly. shimeji* AT787, other mycorrhizal fungi, and saprotrophic fungi, respectively. Regions with red or white borders correspond to orthologs that *Ly. shimeji* AT787 shares only with saprotrophic or mycorrhizal species, respectively. Open and closed circles after the species name indicate mycorrhizal and saprotrophic references, respectively. (B) Orthologs shared with *Ly. shimeji* AT787, two consumed reference species (*Ag. bisporus* and *Le. edodes*), two non-consumed reference species (*Co. cinerea* and *La. bicolor*). Red, orange, and green colours correspond to *Ly. shimeji* AT787, other consumed fungi, and non-consumed fungi, respectively. The region with a red border corresponds to orthologs that *Ly. shimeji* AT787 shares only with widely consumed species. Open and closed stars after the species name indicate consumed and non-consumed references, respectively.

placed in the mitochondrial genome and could result from having mitochondrial data included in the genome dataset.

3.6. Metabolic processes for compounds related to taste

In addition to its primitive mycorrhizal trait, *Ly. shimeji* is also famous as an edible mushroom. We next surveyed genes involved with taste by comparing *Ly. shimeji* AT787 genes with genes from two of the most consumed mushroom species (*Ag. bisporus* and *Le. edodes*) and two species that are non-toxic but are not generally eaten (*Co. cinerea* and *La. bicolor*). Previous studies showed that major compounds providing the umami taste of mushrooms are glutamate and nucleotide monophosphates, such as inosinic acid and guanylic acid.[71] We surveyed genes in the biosynthetic pathways for these umami molecules, but none of these genes showed a difference in gene numbers between consumed and non-consumed species ([Table S6]). Therefore, the accumulation of umami molecules might be regulated by other processes such as gene expression or degradation. Alternatively, the preference for edible fungi can be caused by other compounds, such as odorant compounds that are well-known for truffles and matsutake mushrooms. We also conducted an orthologous analysis of these fungi and found that 113 genes are shared between *Ly. shimeji* AT787 and widely consumed mushroom species but are not present in the primarily non-consumed species ([Fig. 3b], bordered in red). We found that 14 of these genes were annotated as major facilitator superfamily (or abbreviated as 'MFS') transporters ([Table S7]). This family may be a candidate gene family that has expanded in consumed mushroom species. MFS transporters include sugar transporters.[72]

Some MFS sugar transporters have been mentioned as being involved with strawberry taste via sugar accumulation.[73] Therefore, MFS transporters might affect mushroom taste.

3.7. Concluding remarks

In recent years, accumulating information on the fungal genome has revealed common traits of mycorrhizal genomes such as genome size expansion accompanied by TE proliferation and degradation of the PCWDE repertoire.[23–26] The *Ly. shimeji* AT787 genome, however, did not show either of these genomic traits even though the organism is unquestionably mycorrhizal. Our results suggest that increments of genome size and the depletion of PCWDEs may not have been the initial genomic alterations during mycorrhizal evolution but, rather, a consequence of the mycorrhizal lifestyle.

Ly. shimeji is also a good edible mushroom, especially in Japan.[8,9] Exceptional among mycorrhizal fungi, *Ly. shimeji* can develop basidiocarps in axenic culture and, therefore, can be cultivated commercially.[13,27] Furthermore, research methods, including transformation, have also been developed.[74,75] The genomic information is helpful for agriculture such as the establishment of adequate culture conditions based on metabolic gene repertoire, exploration of genetic markers to select strains with beneficial traits, and application of genetic engineering.[76,77] The genome data reported here will aid future research efforts to improve the commercial uses of *Ly. shimeji*.

Acknowledgements

Conflict of interest

None declared.

Data availability

The sequencing reads, draft genome assembly, and annotations have been deposited in the DDBJ database under BioProject accession PRJDB13731 with some adjustment. The original data including genome assembly, gene prediction, annotation and ortholog tables are available through FigShare (10.6084/m9.figshare.20281272, 10.6084/m9.figshare.21641057).

Supplementary Data

Supplementary data are available at *DNARES* online.

References

1. Tedersoo, L., May, T.W., and Smith, M.E. 2010, Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages, *Mycorrhiza*, **20**, 217–63.
2. Wang, B. and Qiu, Y.L. 2006, Phylogenetic distribution and evolution of mycorrhizas in land plants, *Mycorrhiza*, **16**, 299–363.
3. Smith, S.E. and Read D.J. 2008, *Mycorrhizal Symbiosis*, Academic Press: New York.
4. Brundrett, M.C. 2009, Mycorrhizal associations and other means of nutrition of vascular plants: understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis, *Plant Soil*, **320**, 37–77.
5. Soudzilovskaia, N.A., van Bodegom, P.M., Terrer, C., et al. 2019, Global mycorrhizal plant distribution linked to terrestrial carbon stocks, *Nat. Commun.*, **10**, 5077.
6. Hibett, D.S., Gilbert, L.B., and Donoghue, M.J. 2000, Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes, *Nature*, **28**, 506–8.
7. Zambonelli, A. and Bonito, G.M. 2012, *Edible ectomycorrhizal mushrooms*, Springer: Berlin.
8. Ichioka, T. 1799, *Shin-you-kin-pu (Colored illustration of mushrooms in the southern part of Shinshu area)* (In Japanese), Own publication.
9. Imazeki, R. and Hongo, T. 1987, *Illustrations of mushrooms of Japan*, *vol.* **I** (In Japanese), Osaka: Hoikusha.
10. Hasegawa, M. and Kawai, M. 2000, Evidence of Ectomycorrhiza formation between *Lyophyllum shimeji* and *Quercus phillyraeoides*, *Bull. Nara For. Res. Inst.*, **30**, 21–6.
11. Yamada, A., Ogura, T., and Ohmasa, M. 2001, Cultivation of mushrooms of edible ectomycorrhizal fungi associated with *Pinus densiflora* by in vitro mycorrhizal synthesis, *Mycorrhiza*, **11**, 59–66.
12. Ota, A. 1994, Some cultural characteristics of mycelia of a mycorrhizal fungus, *Lyophyllum shimeji*, *Mycoscience*, **35**, 83–7.
13. Ota, A. 1994, Production of fruit-bodies of a mycorrhizal fungus, *Lyophyllum shimeji*, in pure culture, *Mycoscience*, **35**, 147–51.
14. Larsson, E. and Sundberg, H. 2011, *Lyophyllum shimeji*, a species associated with lichen pine forest in northern Fennoscandia, *Mycoscience*, **52**, 289–95.
15. Bellanger, J.M., Moreau, P.A., Corriol, G., et al. 2015, Plunging hands into the mushroom jar: a phylogenetic framework for Lyophyllaceae (Agaricales, Basidiomycota), *Genetica*, **143**, 169–94.
16. Agerer, R. 1998, *Lyophyllum decastes*. In: Agerer, R. (ed.) *Color atlas of ectomycorrhizae, plate 124*. Enihorn-Verlag: Schwäbisch Gmünd.
17. Hofstetter, V., Redhead, S.A., Kauff, F., et al. 2014, Taxonomic revision and examination of ecological transitions of the Lyophyllaceae (Basidiomycota, Agaricales) based on a multigene phylogeny, *Cryptogam. Mycol.*, **35**, 399–425.
18. Visnovsky, S.B., Cummings, N., Guerin-Laguette, A., et al. 2014, Detection of the edible ectomycorrhizal fungus *Lyophyllum shimeji* colonising seedlings of cultivated conifer species in New Zealand, *Mycorrhiza*, **24**, 453–63.
19. Yamada, A., Masuno, K., and Fukuda, M. 2012, Taxonomic considerations for Japanese Lyophyllum shimeji and its allied species., *Mushroom Sci. Technol.*, **20**, 9–15.
20. Martin, F., Aerts, A., Ahren, D., et al. 2008, The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis, *Nature*, **452**, 88–92.
21. Martin, F., Kohler, A., Murat, C., et al. 2010, Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis, *Nature*, **464**, 1033–8.
22. Tisserant, E., Malbreil, M., Kuo, A., et al. 2013, Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis, *Proc. Natl. Acad. Sci. USA*, **110**, 20117–22.
23. Kohler, A., Kuo, A., Nagy, L.G., et al. 2015, Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists, *Nat. Genet.*, **47**, 410–5.
24. Murat, C., Payen, T., Noel, B., et al. 2018, Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle, *Nat. Ecol. Evol.*, **2**, 1956–65.
25. Miyauchi, S., Kiss, E., Kuo, A., et al. 2020, Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits, *Nat. Commun.*, **11**, 1–17.
26. Looney, B., Miyauchi, S., Morin, E., et al. 2021, Evolutionary transition to the ectomycorrhizal habit in the genomes of a hyperdiverse lineage of mushroom-forming fungi, *New Phytol.*, **233**, 2294–309.
27. Furukawa, H. and Katagiri, K. 2020, Hon-shimeji tou no kinsho saibai gijyutsu nokaihatsu [Development for artificial cultivation for *Lyophyllum shimeji* and related species, in Japanese], *Naganoken Ringyo Sogo Center Kenkyu Hokoku*, **34**, 65–79.
28. Marx, D.H. 1969, The influence of ectotrophic mycorrhizal fungi on the resistance of pine roots to pathogenic infections. I. Antagonism of mycorrhizal fungi to root pathogenic fungi and soil bacteria, *Phytopathology*, **59**, 153–63.
29. Fulton, T.M., Chunwongse, J., and Tanksley, S.D. 1995, Microprep protocol for extraction of DNA from tomato and other herbaceous plants, *Plant Mol. Biol. Rep.*, **13**, 207–9.
30. Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. 2020, GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes, *Nat. Commun.*, **11**, 1432.
31. Chin, C.S., Peluso, P., Sedlazeck, F.J., et al. 2016, Phased diploid genome assembly with single-molecule real-time sequencing, *Nat. Methods*, **13**, 1050.
32. Huang, S., Chen, Z., Huang, G., et al. 2012, HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies, *Genome Res.*, **22**, 1581–8.
33. Smit, A.F.A., Hubley, R., and Green, P. 2013, RepeatMasker Open-4.0. http://www.repeatmasker.org
34. Manni, M., Berkeley, M.R., Seppey, M., and Zdobnov, E.M. 2021, BUSCO: assessing genomic data quality and beyond, *Curr. Protoc.*, **1**, e323.
35. Martin, M. 2011, Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet J.*, **17**, 10–2.
36. Grabherr, M.G., Haas, B.J., Yassour, M., et al. 2011, Full-length transcriptome assembly from RNA-seq data without a reference genome, *Nat. Biotechnol.*, **29**, 644–52.
37. Bruna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodvsky, M. 2021, BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database, *NAR Genom. Bioinform.*, **3**, lqaa108.
38. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. 2005, Gene identification in novel eukaryotic

genomes by self-training algorithm, *Nucleic Acids Res.*, **33**, 6494–506.

39. Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. 2008, Using native and syntenically mapped cDNA alignments to improve de novo gene finding, *Bioinformatics*, **24**, 637–44.

40. Cantalapiedra, C.P., Hernandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. 2021, eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale, *Mol. Biol. Evol.*, **38**, 5825–9.

41. Grigoriev, I.V., Cullen, D., Goodwin, S.B., et al. 2011, Fueling the future with fungal genomics, *Mycology*, **2**, 192–209.

42. Grigoriev, I.V., Nikitin, R., Haridas, S., et al. 2014, MycoCosm portal: gearing up for 1000 fungal genomes, *Nucleic Acids Res.*, **42**, D699–704.

43. Morin, E., Kohler, A., Baker, A.R., et al. 2012, Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche, *Proc. Natl. Acad. Sci. USA*, **109**, 17501–6.

44. Hess, J., Skrede, I., Wolfe, B.E., et al. 2014, Transposable element dynamics among asymbiotic and ectomycorrhizal *Amanita* fungi, *Genome Biol. Evol.*, **6**, 1564–78.

45. Machida, M., Asai, K., Sano, M., et al. 2012, Genome sequencing and analysis of Aspergillus oryzae, *Nature*, **438**, 1157–61.

46. Floudas, D., Binder, M., Riley, R., et al. 2012, The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes, *Science*, **336**, 1715–9.

47. Peter, M., Kohler, A., Ohm, R.A., et al. 2016, Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*, *Nat. Commun.*, **7**, 12662.

48. Stajich, J.E., Wilke, S.K., Ahrén, D., et al. 2010, Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*), *Proc. Natl. Acad. Sci. USA*, **107**, 11889–94.

49. Riley, R., Salamov, A.A., Brown, D.W., et al. 2014, Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi, *Proc. Natl. Acad. Sci. USA*, **111**, 9923–8.

50. Chen, L., Gong, Y., Cai, Y., et al. 2016, Genome sequence of the edible cultivated mushroom *Lentinula edodes* (Shiitake) reveals insights into lignocellulose degradation, *PLoS One*, **11**, e0160336.

51. Galagan, J.E., Calvo, S.E., Borkovich, K.A., et al. 2003, The genome sequence of the filamentous fungus *Neurospora crassa*, *Nature*, **422**, 859–68.

52. Kobayashi, Y., Maeda, T., Yamaguchi, K., et al. 2018, The genome of *Rhizophagus clarus* HR1 reveals a common genetic basis for auxotrophy among arbuscular mycorrhizal fungi, *BMC Genomics*, **19**, 465.

53. Ma, L.J., Ibrahim, A.S., Skory, C., et al. 2009, Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication, *PLoS Genet.*, **5**, e1000549.

54. Goffeau, A., Barrell, B.G., Bussey, H., et al. 1996, Life with 6000 genes, *Science*, **274**, 563–7.

55. Wood, V., Gwilliam, R., Rajandream, M.A., et al. 2002, The genome sequence of Schizosaccharomyces pombe, *Nature*, **415**, 871–80.

56. Sayers, E.W., Bolton, E.E., Brister, J.R., et al. 2022, Database resources of the national center for biotechnology information, *Nucleic Acids Res.*, **7**, D20–D26.

57. Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. 2012, dbCAN: a web resource for automated carbohydrate-active enzyme annotation, *Nucleic Acids Res.*, **40**, W445–451.

58. Zhang, H., Yohe, T., Huang, L., et al. 2018, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation, *Nucleic Acids Res.*, **46**, W95–101.

59. Petersen, T., Brunak, S., von Heijne, G., et al. 2011, SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nat. Methods*, **8**, 785–6.

60. Shen, W., Le, S., Li, Y., and Hu, F. 2016, SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation, *PLoS One*, **11**, e0163962.

61. Emms, D.M. and Kelly, S. 2019, OrthoFinder: phylogenetic orthology inference for comparative genomics, *Genome Biol.*, **20**, 238.

62. Katoh, K. and Standley, D.M. 2013, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Mol. Biol. Evol.*, **30**, 772–80.

63. Minh, B.Q., Schmidt, H.A., Chernomor, O., et al. 2020, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, *Mol. Biol. Evol.*, **37**, 1530–4.

64. R Core Team. 2020, *R: A language and environment for statistical computing.* R Foundation for Statistical Computing: Vienna, Austria. https://www.R-project.org/

65. Jangam, D., Feschotte, C., and Betrán, E. 2017, Transposable element domestication as an adaptation to evolutionary conflicts, *Trends Genet.*, **33**, 817–31.

66. Pereira, M.F., Veneault-Fourrey, C., Vion, P., et al. 2018, Secretome analysis from the ectomycorrhizal Ascomycete *Cenococcum geophilum*, *Front. Microbiol.*, **9**, 141.

67. Plett, J.M., Kemppainen, M., Kale, S.D., et al. 2011, A secreted effector protein of *Laccaria bicolor* is required for symbiosis development, *Curr. Biol.*, **21**, 1197–203.

68. Pellegrin, C., Daguerre, Y., Ruytinx, J., et al. 2019, *Laccaria bicolor* MiSSP8 is a small-secreted protein decisive for the establishment of the ectomycorrhizal symbiosis, *Environ. Microbiol.*, **21**, 3765–79.

69. Tagu, D., Kottke, I., and Martin, F. 1998, Hydrophobins in ectomycorrhizal symbiosis: hypothesis, *Symbiosis*, **25**, 5–18.

70. Casarrubia, S., Daghino, S., Kohler, A., et al. 2018, The hydrophobin-like OmSSP1 may be an effector in the ericoid mycorrhizal symbiosis, *Front. Plant Sci.*, **9**, 546.

71. Mau, J.L. 2005, The umami taste of edible and medicinal mushrooms, *Int. J. Med. Mushrooms*, **7**, 119–25.

72. Pao, S.S., Paulsen, I.T., and Saier, M.H., Jr. 1998, Major facilitator superfamily, *Microbiol. Mol. Biol. Rev.*, **62**, 1–34.

73. Liu, H.T., Ji, Y., Liu, Y., et al. 2020, The sugar transporter system of strawberry: genome-wide identification and expression correlation with fruit soluble sugar-related traits in a *Fragaria × ananassa* germplasm collection, *Hortic. Res.*, **7**, 132.

74. Saito, T., Tanaka, N., and Shinozawa, T. 2001, A transformation system for an ectomycorrhizal basidiomycete, *Lyophyllum shimeji*, *Biosci. Biotechnol. Biochem.*, **65**, 1928–31.

75. Maeta, K., Koike, A., Mukaiyama, H., et al. 2008, Preparation and crossing of mating-capable monokaryons via protoplasting of the dikaryotic mycelia of a mycorrhizal mushroom, *Lyophyllum shimeji*, *J. Wood Sci.*, **54**, 337–40.

76. Fang, J., Zhu, X., Wang, C., and Shangguan, L. 2016, Applications of DNA technologies in agriculture, *Curr. Genomics*, **17**, 379–86.

77. Halewood, M., Lopez-Noriega, I., Ellis, D., et al. 2018, Using genomic sequence information to increase conservation and sustainable use of crop diversity and benefit-sharing, *Biopreserv. Biobank*, **16**, 368–76.