A genomic perspective on a transition to symbiosis in ectomycorrhizal *Amanita*

Maryam Chaib De Mares

University of Groningen

Supervisor: Joana Falcão Salles, Ph.D.

External co-supervisors: Jaqueline Hess, Ph.D. and Anne Pringle, Ph.D.

Harvard University

Submitted in partial fulfilment of the requirements for Master's thesis in the Erasmus Mundus Master Programme in Evolutionary Biology (MEME)

Spring Semester 2013

A genomic perspective on a transition to symbiosis in ectomycorrhizal *Amanita*

Chaib De Mares, Maryam[1], Jaqueline Hess[2], Joana Falcao Salles[1], Anne Pringle[2]

m.chaib@student.rug.nl
[1] *CEES Centre for Ecological and Evolutionary Studies, University of Groningen*
[2] *Department of Organismic and Evolutionary Biology, Harvard University*

Abstract

Mycorrhizal fungi form symbioses with plants, providing them with resources such as nitrogen in exchange for carbon. Even though these symbioses are widespread, the genes that have allowed them to originate and diversify remain unknown. The genus *Amanita* provides an excellent model for elucidating genomic changes associated with the evolution of ectomycorrhizal (ECM) symbiosis. Within this group, saprotrophic *Amanita* species form a strongly supported clade basal to a monophyletic ECM clade. We conducted a comparative genomics study between the ECM species *A. muscaria* and the saprotroph *A. thiersii*. Our analysis revealed patterns of loss of plant cell wall degrading enzymes in *A. muscaria* compared to *A. thiersii*, as well as a shift towards small secreted proteins, some of which resemble effector-like receptors. Moreover, there is an increase in the peptidase potential of the ECM fungus evident in the amplification of aspartic proteases and A4 peptidases. Together, these features would allow the ECM fungus to prevent the degradation of its host cell walls, degrade elicitors of plant defenses, and potentially mobilize nutrients from the soil environment while establishing the symbiosis with the plant root. Furthermore, a family of carbohydrate esterases (CEs) is bacterial in origin and expanded in *A. muscaria*. We hypothesize that this transfer is either an additional potential mechanism contributing to the evolution of ECM symbiosis or a form of reversal after major losses in the fungus' ability to metabolize carbon. Our comparative analyses support the view that different genomic changes have facilitated the evolution of ECM symbiosis, and ours is the first report that points to horizontal gene transfer as a potential mechanism contributing to the evolution of mycorrhizal symbioses.

Key words: *Amanita*, carbohydrate-active enzymes, comparative genomics, ectomycorrhizal symbiosis, horizontal gene transfer

Table of contents

List of Figures

List of Tables

Glossary

Carbohydrate-active enzymes (CAZymes): Enzymes involved in the synthesis, metabolism and transport of carbohydrates. These are classified in families in the CAZy database (Cantarel *et al.* 2009; http://www.cazy.org/). Included in this database are the glycoside hydrolase (GH), glycosyl transferase (GT), polysaccharide lyase (PL), carbohydrate esterase (CE) and carbohydrate-binding module (CBM) families.

Motif: Protein sequence motifs are signatures (i.e. short pieces of a protein sequence) of protein families, and can often be used as tools for the prediction of protein function. A conserved motif may be, for example, [L/I]x(9)A. This motif is composed by 11 amino acid positions. The first one, [L/I], means amino acid L or I; the following 9 are any amino acid (denoted by x), and the last position is amino acid A.

PFAM domain: The PFAM database (http://pfam.sanger.ac.uk) contains information about protein domains and families (Punta *et al.* 2012). Each protein family in the database includes its functional annotation (if any) and multiple sequence alignments generated using Hidden Markov Models.

Protein domain: A conserved part of a given protein sequence and structure that can evolve, function, and exist independently of the rest of the protein chain. Each domain forms a compact three-dimensional structure and often can be independently stable and folded. Domains often form functional units, and a given protein can contain multiple domains.

Protein family: A group of proteins that descend from a common ancestor (i.e. are homologous); often nearly synonymous with gene family. Proteins in a family have significant sequence similiarity but not necessarily functional similarity.

Proteome: The protein inventory of an organism. In this particular study, we refer to the proteome as the predicted proteome, which is the complete set of proteins predicted to be encoded in a fully sequenced genome by means of bioinformatics tools.

Secretome: The subset of a proteome consisting of proteins secreted from a cell.

Signal peptide: Secretory N-terminal peptides that target a protein for translocation across the endoplasmic reticulum (ER) membrane in eukaryotes. Signal peptides share several structural features with targeting peptides of chloroplasts, mitochondria, and proteins bound to the outer cellular membrane, but can still be distinguished because they are either enriched in different types of amino acids, or contain a transmembrane domain (in the case of transmembrane proteins).

Small secreted proteins (SSPs): The part of a secretome encompassing all proteins with 300 or fewer amino acids in length.

Transmembrane domain: A domain which usually contains an alpha helix. These helices indicate that proteins are anchored in the membrane, as opposed to being secreted from the cell.

Introduction

Symbioses – associations of different species throughout a significant portion of their life history (Margulis 1991) – are widespread and central to evolution. For instance, the evolution of an endosymbiosis between prokaryotic lineages enabled eukaryotic life (Sagan 1967; Smith and Szathmáry 1999). Mycorrhizas are ubiquitous symbioses formed between fungi and plants (Smith and Read 1997). In exchange for carbon, mycorrhizal fungi provide plants with an array of resources, including nitrogen or phosphorus. A plant grown in the presence of its symbionts shows a consistently higher growth rate, a greater capacity to absorb water and nutrients, and protection from pathogens (Smith and Read 2008). In addition, the establishment of the symbiosis is required for the completion of the fungal life cycle (i.e. formation of fruiting bodies). The improved health and fitness enabled by mycorrhizal symbioses may have been fundamental to the evolution of land colonization by plants (Simon *et al.* 1993).

Traditionally, seven different categories of mycorrhizal symbiosis have been distinguished on the basis of their morphological characteristics and the fungal and plant species involved (Perotto *et al.* 1995; Finlay 2008; Smith and Read 2008). However, it is becoming increasingly clear that the morphology of a fungus can change depending on the plant symbiont (Villareal-Ruiz *et al.* 2004), and that different environmental factors can induce various lifestyles in some fungal species (Talbot *et al.* 2008).

The most ancient and widespread symbiosis involves the arbuscular mycorrhizae (AM), whose fungal partners belong to the phylum Glomeromycota. Molecular sequence data suggest that the first land plants formed associations with these fungi about 460 million years ago (Redecker *et al.* 2000). An AM symbiosis is typified by highly branched fungal structures, called arbuscules, which grow intracellularly and develop inside the lumen of plant cells (Genre and Bonfante 2012). Basidiomycota and Ascomycota fungi, however, usually engage in ectomycorrhizal (ECM) symbioses. These symbioses are characterized by the presence of a fungal mantle around the root surface (Bonfante and Genre 2010), as well as a network of intercellular hyphae penetrating between the epidermal and cortical cells, named the Hartig net (Figure 1).

The dominant components of forest ecosystems, compared to free-living plants, are plant species that engage in ECM symbioses. They grow as diverse communities in which an individual fungus may colonize multiple trees (Kennedy *et al.* 2003). Moreover, a plant can usually establish a mutualistic symbiosis with a broad range of fungal species (Martin *et al.* 2001). In single-species pine stands of approximately 0.1 ha, 15 to 35 species of ECM fungi are typically reported; single soil cores often contain several species (Eberhart *et al.* 1996), and even adjacent root tips are frequently colonized by different fungi (Bruns 1995).

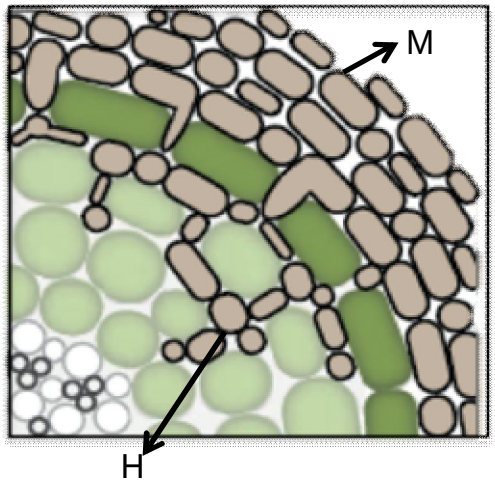


Figure 1. Schematic representation of a transverse cross-section of a mature mycorrhizal root tip. The fungus wraps around the entire root surface, forming a thick, multi-layered mantle (M) constructed from individual hyphae. A number of fungal hyphae also invade between the plant cells of the root, forming a structure called the Hartig net (H). Modified from Plett and Martin (2011).

Several hypotheses have been proposed regarding the origin of ECM symbiosis. One hypothesis posits that ECM symbiosis would be an ancient shared ancestral state followed by numerous losses (Weiss *et al.* 2004). However, it is currently accepted that this symbiosis has evolved repeatedly over the last 130–180 Myr (LePage *et al.* 1997). Hibbett *et al.* (2000) suggested that some lineages would have experienced subsequent reversals to a free-living lifestyle; yet, the most recent evidence points to a history of independent origins with no reversals (Bruns and Shefferson 2004; Matheny *et al.* 2006). It is also a matter of debate whether ECM fungi have a dual lifestyle, capable of living

not only within the plant roots as symbionts, but also able to live in the soil as saprotrophs (free-living organisms that obtain energy from non-living organic matter; see Baldrian 2009 for a review).

Genomics insights into ECM symbiosis

Over the last decade, novel technologies such as whole genome sequencing and high-throughput transcriptomics have enabled substantial advances in the comprehension of cellular and molecular mechanisms shaping fungal genomes. For instance, sequencing approximately 50 genomes of saprotrophic and pathogenic fungi provided an opportunity to elucidate key components determining their various lifestyles (Galagan *et al.* 2005). For example, approximately 30% of predicted genes of *Magnaporthe oryzae* and *Fusarium graminearum* (Dean *et al.* 2005) have no significant homologs in other organisms, and appear to be unique to these two plant pathogenic fungi. The genomics of the AM symbiosis have been reviewed in detail by Bonfante and Requena (2011) and are outside the scope of this study. Regarding ECM symbiosis, the first complete genome sequence of an ECM basidiomycete, *Laccaria bicolor*, was published in 2008 (Martin *et al.* 2008) and the genome of an ECM ascomycete, *Tuber melanosporum*, was completed in 2010 (Martin *et al.* 2010).

Compared with other fungal genomes, the *L. bicolor* genome contains novel and expanded gene families that play a role in signaling and nutrient cycling (e.g. GTPases, proteases and transporters) (Martin *et al.* 2008; Martin and Selosse 2008; Rajashekar *et al.* 2009). It has been hypothesized that these expanded gene families may act as a 'symbiosis toolbox' in EM symbiosis (Martin *et al.* 2008). However, gene family expansion is not a trait shared by all ectomycorrhizal species. For instance, even though the genome of the ascomycete *T. melanosporum* is the largest and most complex sequenced so far, it contains a proportionally low number of multigene families, most of which have only two members and, in total, comprise only 19% of the predicted proteome (Martin *et al.* 2010).

Both genomes show a massive gene loss in their plant cell-wall degradation arsenal (Martin *et al.* 2008, 2010). Nevertheless, a few carbohydrate-cleaving enzymes are induced in symbiotic tissues of *T. melanosporum*, suggesting that this species degrades its host's cell walls during colonization. In contrast, *L. bicolor* appears to be unable to use plant cell wall components (cellulose, pectins and pectates) as a carbon source for its growth. This adaptation may prevent it from degrading its host cells and triggering the plant's defensive reaction. However, because no data are available for closely related free-living species it is impossible to know if features of the *L. bicolor* and *T. melanosporum* genome were driven by a transition to symbiosis, or by some other aspect of their biology.

Loss of plant cell wall degrading enzymes is also found in other ectomycorrhizal basidiomycetes such as *Amanita muscaria* (see below, Wolfe *et al.* 2012a) and *A. bisporigera* (Nagendran *et al.* 2009). Similarly, the genome of the brown-rot fungus *Postia placenta*, which cannot efficiently depolymerize lignin, is composed of a low repertoire of cellulases that act on plant cell walls (Martinez *et al.* 2009). These patterns of losses indicate that in ECM, as well as brown-rot fungi, evolutionary shifts in lifestyle may have been facilitated by changes in the repertoires of carbohydrate degrading enzymes.

Secretome

Plant – fungus symbiotic interactions, either pathogenic or mutualistic, are represented by fungal cells growing within or in close association with plant tissues. Communication through the secretion of proteins and metabolites that are either taken up by the host or detected at the cell surface plays a pivotal role in determining the outcome of the interaction (Bouws *et al.* 2008; Brown *et al.* 2012; Genre and Bonfante 2012). The variety of possible outcomes has resulted in the evolution of a range of mechanisms to acquire nutritients from various habitats, including mutualistic, biotrophic, hemibiotrophic, necrotrophic and non-pathogenic saprophytic lifestyles. In turn, plant hosts have also evolved a range of mechanisms to recognize "non-self" and activate defense responses towards potential pathogens.

Current understanding of the mechanisms underlying plant recognition and defense pose that successful pathogens are able to overcome or suppress non-host plant defenses that are triggered via surface receptor recognition of Pathogen or Microbe associated molecular patterns (PAMPs or MAMPs) (reviewed in Deller *et al.* 2011). These conserved molecules are typically present in, secreted from or on the surface of, the pathogen but absent from the plant and have therefore been selected as "non-self" recognition determinants. Pathogens that are able to overcome this first line of defense do so through the release of effector molecules, which function to inhibit recognition or the subsequent activation of plant defenses triggered by PAMPs. In most, but not all, cases to date, these effector molecules are known to be proteins which are either translocated into the target cell or secreted into the apoplastic leaf spaces (de Wit *et al*. 2009).

Remarkable similarity exists between the mechanisms of colonization used by mutualistic and pathogenic organisms (Hogenhout *et al.* 2009; Sanders 2011; van Ooij 2011). Fungal pathogens of plants are known to release effectors, often small proteins usually encoded by avirulence genes (Kamoun 2007; Stergiopoulos and de Wit 2009). The pathogenic fungus *Ustilago maydis* uses these small secreted effector proteins to subvert plant cell defenses and colonize plant tissues, altering plant signaling (Kämper *et al.* 2006; Mueller *et al.* 2008; Doehlemann 2009). Effector proteins play a part in pathogen attack and act either in the plant's extracellular space or after entering the host plant's cells (Ellis *et al.* 2009).

The *L. bicolor* genome contains 278 distinct small secreted proteins (SSPs, <300 amino acids) of which 69% belong to multigene families. Ten of these SSPs are highly regulated in symbiotic tissue and share structural similarity with fungal pathogenic effector proteins (Martin *et al.* 2008). These up-regulated SSPs are therefore called mycorrhiza-induced small secreted proteins (MiSSPs). The most highly expressed SSP is secreted onto the surface of the fungal hyphae of the intraradicular Hartig net (MiSSP7; Martin *et al.* 2008), and encodes an effector protein (MiSSP7) indispensable for the establishment of mutualism with poplar (*Populus trichocarpa*) roots (Plett *et al.* 2011). MiSSP7 is secreted by the fungus upon receipt of diffusible signals from plant roots, imported into the plant

cell via endocytosis, and targeted to the plant nucleus where it alters the transcriptome of the plant cell.

The AM *Glomus intraradices* secretes a small protein, Sp7, that is taken up by host cells and promotes the establishment of the biotrophic interaction, analogous to the role played by MiSSP7 (Kloppholz *et al.* 2011). In contrast, none of the MiSSPs proteins found in *L. bicolor* ectomycorrhizas are detected among ectomycorrhizal-regulated *T. melanosporum* transcripts (Martin *et al.* 2010). Therefore, understanding the presence, repertoire, mode of evolution and function of secreted and small secreted proteins will continue to provide a mechanistic understanding of the processes underlying plant-fungal interactions.

The Amanita system

As described above, the currently available basidiomycete genomes can only provide information on coarse genomic differences that exist between saptrotrophic and symbiotic fungi (Martin *et al.* 2008, 2010; Nagendran *et al.* 2009). A coherent elucidation of the genomic mechanisms associated with an evolution to ectomycorrhizal symbiosis requires a comparative genomics approach based on more closely related free-living and symbiotic species. The genus *Amanita* (Figure 1) is an excellent model for this approach because it consists of a diverse group of ectomycorrhizal and closely related saprotrophic taxa. Recently a single origin was demonstrated for the ectomycorrhizal symbiosis within this group, where the saprotrophic *Amanita* species form a strongly supported clade basal to a monophyletic clade of ectomycorrhizal species (Wolfe *et al.* 2012a).

Figure 2. A sampling of color and form diversity within the genus *Amanita*. Clockwise, from top left, *Amanita muscaria* subsp. *flavivolvata*, *Amanita frostiana*, *Amanita jacksonii*, an undescribed *Amanita* species, the saprotrophic *Amanita manicata*, and *Amanita phalloides* (from Wolfe *et al.* 2012a).

Over the last decade, the fungal genus *Amanita* has emerged as a model system for understanding the characteristics of the ectomycorrhizal symbiosis (Pringle and Vellinga 2006; Pringle *et al.* 2009; Wolfe *et al.* 2010; Wolfe *et al.* 2012a). Recent genetic and genomic advances in this system will shed light on the genetic pathways needed to establish the ectomycorrhizal association. The genomes of the ectomycorrhizal species *A. muscaria* and the saprotroph *A. thiersii* are currently available at the Joint Genome Institute (JGI, http://genome.jgi.doe.gov/). In addition, four more genomes have been sequenced and are being annotated at the moment: *A. brunnecens* and *A. polypyramis* (ectomycorrhizal), *A. inopinata* (saprotrophic) and the saprotrophic outgroup *Volvariella volvaceae* (Bao *et al.* 2013; J. Hess, *pers. commun.)*.

The evolution of symbiosis in *Amanita* is associated with the loss of two cellulase genes from many EM *Amanita* species, but many saprohtrophic *Amanita* species have copies of one or both of these genes (Wolfe *et al.* 2012a). One is an endoglucanase (*eg1*), that belongs to the carbohydrate-active enzymes (CAZymes) glycoside hydrolase (GH) family 5, and is extracellular (Ding *et al.* 2001); the other is a cellobiohydrolase (*cbhI-I*), that belongs to GH family 7, and is also extracellular (Jia *et al.* 1999). The loss of CBHI

cellobiohydrolases was also observed from the genome sequences of *L. bicolor* and *T. melanosporum* (Martin *et al.* 2008, 2010). These and other plant cell-wall-degrading enzymes are absent from fungal ectosymbiotic lineages, probably because these interfere with the establishment of symbiosis.

In this study we explored the genetic differences among a symbiotic and decomposer species of the genus *Amanita.* Specifically, what novel genes appear after the evolution of symbiosis in *Amanita* species? Based on information both from the *L. bicolor* genome (Martin *et al.* 2008) and previous molecular genetics work of *Amanita* species (Nagendran *et al.* 2009, Wolfe *et al.* 2012), we hypothesized that genes responsible for establishing and maintaining EM symbioses in *Amanita* are unique or expanded in symbiotic *A. muscaria*. We focused on the analysis of secreted and small secreted protein families. Comparative analyses on Whole Genome Sequence (WGS) data presented here suggest that, indeed, ECM symbioses have been facilitated by different genomic changes in the ECM species sequenced to date. In the case of *Amanita*, the ECM genome lacks carbohydrate-active enzymes present in *A. thiersii* and other saprotrophic genomes. However, other gene families are duplicated and retained, suggesting that *A. muscaria* may have retained a weak ability to degrade both, plant cell wall material and elicitors of plant defenses during the establishment of symbiosis.

Materials and Methods

In order to elucidate the patterns of molecular evolution of the secreted and the subset of small secreted proteins (SSPs) in the genus *Amanita*, predicted proteins from the recently completed genome projects (i.e. proteomes) of the EM species *A. muscaria* and the saprotrophic species *A. thiersii* (Hess *et al. submitted*) were extracted from Joint Genome Institute databases (JGI). The *A. muscaria* Koide genome sequence can be accessed at http://genome.jgi.doe.gov/Amamu1/Amamu1.home.html and the *A. thiersii* genome sequence can be accessed at http://genome.jgi.doe.gov/Amamu1/Amamu1.home.html).

Prediction of the secretome

The prediction of the secretome of both species was based on a pipeline kindly provided by Dr. Emmanuelle Morin (Institut National de la Recherche Agronomique, INRA, Champenoux, France). In this pipeline, the subcellular localization of each protein in the proteome is predicted using WoLF PSORT (Horton *et al.* 2000; http://wolfpsort.org/). All proteins predicted to be extracellular are annotated as such. Then, the presence and location of signal peptide cleavage sites in the amino acid sequences is predicted using Signal P v4.1 (Petersen *et al.* 2011; http://www.cbs.dtu.dk/services/SignalP/). Secretory signal peptides are N-terminal peptides that target a protein for translocation across the endoplasmic reticulum (ER) membrane in eukaryotes. They share several structural features with the targeting peptides of chloroplasts and mitochondria, but can still be distinguished because they are enriched in different types of amino acids (von Heijne *et al.* 1989; von Heijne 1990). Proteins with a discrimination score (D-score) ≥ 0.34 were combined into the secretome dataset. This value is used to discriminate signal peptides from non-signal peptides and its value is the standard used within the framework of the mycorrhizal genomics initiative (http://mycor.nancy.inra.fr/IMGC/MycoGenomes/).

The secretome dataset was then scanned for transmembrane α-helices using TMHMM v2.0 (Krogh *et al.* 2001, http://www.cbs.dtu.dk/services/TMHMM/). Proteins containing such helices are known to be bound to the membrane and were therefore removed from the database. The combined results from SignalP (Petersen *et al.* 2011) and WoLF PSORT (Horton *et al.* 2000) were used to run the package Target P (Emanuelsson *et al.*

2007), which enabled us to remove proteins with a signal that sorts to the mitochondria. Proteins with a Target P location = S – secreted were left in the dataset (Target P v1.1; http://www.cbs.dtu.dk/services/TargetP/). These results were used as input for PS-Scan, a package that searches protein sequences for functional amino acid patterns based sequence similarity through the PROSITE database of protein families and domains (Sigrist *et al.* 2002). The length of each protein had been calculated during the prediction of the secretome, and the SSP database for both species was generated by selecting all proteins smaller than 300 amino acids using custom Python scripts (Cock *et al.* 2009).

Functional analyses

We next analyzed the functional annotation available at the JGI genome portal for each genome. We focused on the protein domains predicted using the PFAM database (Punta *et al.* 2012). Additionally, we analyzed the complete proteome for enzymes involved in the synthesis, metabolism and transport of carbohydrates. These are known as carbohydrate-active enzymes (CAZymes) and are classified in families in the CAZy database (Cantarel *et al.* 2009). Included in this database are glycoside hydrolase (GH), glycosyl transferase (GT), polysaccharide lyase (PL), carbohydrate esterase (CE) and carbohydrate-binding module (CBM) families. The CAZyme annotations for *A. muscaria* and *A. thiersii* were kindly provided by Bernard Henrissat (Université de Marseille). For statistical support in both functional analyses, we used chi-square tests to identify specific protein domains or CAZyme families under- or over-represented in either fungal species. P-values were adjusted for multiple testing using Bonferroni correction.

Identification of effector-like motifs among SSPs

We searched for conserved motifs previously identified in candidate effector fungal genes among the SSP subset of both species. We looked for the motifs [LI]xAR and [RK]Cx(2)Cx(12)H of the rice blast fungal pathogen *Magnaporthe oryzae*. These motifs are hypothesized to be involved in protein–protein interactions (Yoshida *et al.* 2009). We also searched for the YxSL[RK] effector candidate, which is enriched at least two-fold in secreted proteins compared to non-secreted proteins in *Pythium ultimum* and four other oomycete genomes (Lévesque *et al.* 2010). Finally, we looked for the [YFW]xC motif,

suggested to be a new class of effectors from haustoria-producing pathogenic fungi (Godfrey *et al.* 2010), and the classical RxLR motif first identified in *Phytophtora* pathogens (Kamoun 2006). All searches used a custom Python script. To assign putative functions to SSPs containing effector-like motifs, we searched for homologs of these protein sequences using HMMER (Finn *et al.* 2011).

Protein family classification

Next, to determine gene families within the secretome of both *A. muscaria* and *A. thiersii*, we used a clustering technique based on pairwise comparisons of full-length protein sequences (Enright *et al.* 2002; Kriventseva 2005; Frech and Chen 2010). BLASTP all-vs-all comparison was performed on the combined FASTA file of the secretome of both species (1099 proteins) with the NCBI BLAST package v2.2.27. We filtered the results to obtain matches with E-value $\leq 10^{-5}$. TRIBE-MCL v.09-308 was used to generate the clusters based on the BLAST output as suggested by the MCL manual (Stijn van Dongen 2000). For the clustering step we used the *mcl* command at varying inflation values, ranging from 1.0 to 3.0 (step size 0.5). All other *mcl* parameters were left as default. Based on the granularity of the clusters formed (not shown), the results produced with an inflation value of 2.0 were used in subsequent analyses. Clusters with two or more proteins were considered putative protein families. Unique families were defined as protein families with members from only one species. The putative functions of proteins in all clusters were annotated by homology searches using the PFAM database (Punta *et al.* 2012).

Phylogenetic analysis of hydrophobins

We used a phylogenetic tree to infer the evolutionary dynamics of the hydrophobin protein family, expanded in the genome of *A. thiersii*. We searched for homologs of the hydrophobin protein family in the genome of *V. volvacea* deposited in NCBI (www.ncbi.nlm.nih.gov; Bao *et al.* 2013). To achieve more sensitivity detecting potential homologs, we used the domain enhanced lookup accelerated BLAST (DELTA-BLAST) tool (Boratyn *et al.* 2012).

Results

A. muscaria *is enriched in SSPs*

In the current study we analyzed the filtered gene catalog of version 1 of the *A. muscaria* and *A. thiersii* genomes (18153 and 10354 unique proteins, respectively) (Table 1). Initially, we predicted all possible secreted proteins (i.e. the secretome) by identifying signal peptides in amino acid sequences. We identified 651 predicted, secreted proteins in *A. muscaria*. 199 of these proteins were subsequently predicted to contain a transmembrane domain (TM). Similarly, we identified 650 secreted proteins in *A. thiersii,* 83 of which had a predicted TM. After exclusion of proteins with predicted TMs, the total secretome of *A. muscaria* comprised 532 proteins (nearly 3 % of the total current predicted protein models), whereas that of *A. thiersii* had 567 proteins (approximately 5% of proteome; Table 1).

Table 1. Summary of secretome in *A. muscaria* and *A. thiersii*.

	Proteome size	Secretome (TM)	TM	Secretome (no TM)	Secretome to proteome (%)	SSP size	SSP to secretome (%)
A. muscaria	18153	651	119	532	2.9	313	58.8
A. thiersii	10354	650	83	567	5.5	246	43.4

The average predicted mature protein lengths for the secretome were significantly lower in *A. muscaria* than in *A. thiersii* ($t = -5.1956$, $p < .001$) (Figure 3a). Similarly, the subset of small secreted proteins (SSPs) is larger in *A. muscaria* than in *A. thiersii* ($t = -4.5179$, $p < .001$), with a total of 313 (SSPs). This is nearly 60% of its predicted secretome. In turn, *A. thiersii* has longer proteins and 246 SSPs correspond to 43% of the predicted secretome (Table 1). Moreover, the saprotrophic *A. thiersii* contains a higher percentage of cysteine (C) residues among the SSP subset (Figure 3b) ($p < .05$).

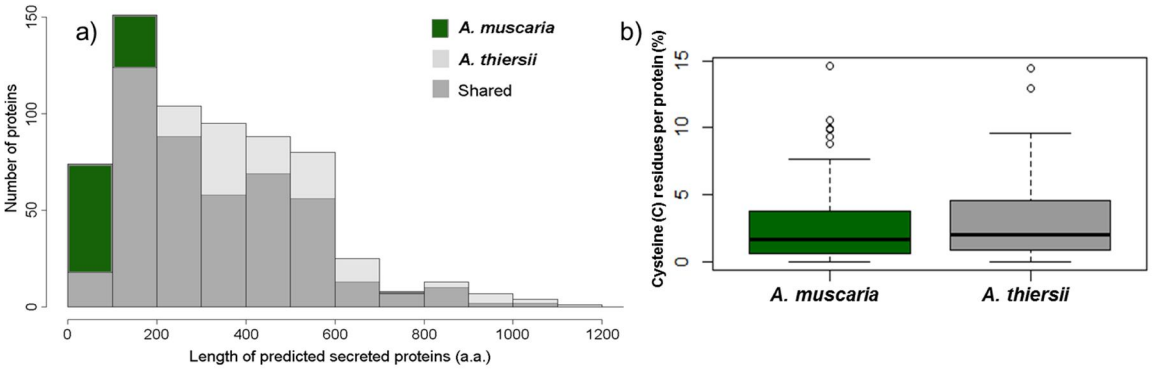


Figure 3. Distribution of a) length of predicted secreted proteins and b) percentage of cysteine (C) residues per small secreted protein (SSP) in *Amanita muscaria* and *A. thiersii*.

Novel functions in predicted secretome of A. muscaria

We next analyzed the functional annotation of each secretome. Only 248 of the *A. muscaria* secretome (47%) showed similarity to documented protein functional domains (http://pfam.sanger.ac.uk/). More strikingly, only 74 (24%) of the SSPs in the same species matched an existing PFAM domain (see Glossary). In *A. thiersii*, 363 (64%) secreted proteins and 104 (42%) SSPs matched known functional domains. The number of uncharacterized secreted and SS protein is large in both species, but significantly larger in the ECM *A. muscaria* compared to the saprotroph *A. thiersii*.

Proteins involved in nutrient acquisition and interaction with the host

Among the secreted proteins for which we could assign a putative function, we observed that both fungi contain numerous genes that code for diverse groups of enzymes involved in oxidation of organic substances. These are oxidoreductases, including FAD binding domains (PF01565), multicopper oxidases (PF07732; Appendix A) and cytochrome P450 mono-oxigenases (PF00067). Moreover, they share a large number of proteases and peptidases. Proteases are key enzymes involved in the extracellular degradation of proteins by fungi. However, the kinds of enriched proteases differ between the two genomes. While the ECM fungus *A. muscaria* codes for a large number of aspartic proteases (PF00026), the saprotroph *A. thiersii* is enriched in metalloproteases (PF02102).

The ECM fungus is also enriched in other genes that may play a role in the nutrition of the fungus such as lipases (class 3, PF01764), which are esterases that can hydrolyze long-chain acyl-triglycerides, glycerol, and free fatty acids at a water/lipid interface. In contrast, it shows a massive reduction in carbohydrate degrading enzymes (data not shown; instead, see section "Distribution of CAZymes in *Amanita*").

The ECM fungus shows a marked enrichment in proteins predicted to have roles in protein-protein interactions, such as phosphoesterases (PF04185). These proteins are potential candidates for communication between the symbiont and host. It is also rich in glycosyl hydrolases (GH) families 16 (PF00722) and 18 (PF00704), involved in chitin metabolism and have been suggested to play a role in modification of the fungal cell wall, presumably either for fungal growth or the formation of the Hartig net during ECM symbiosis.

In our analyses of annotated genes, we highlight proteins that may be related to the saprotrophic lifestyle of *A. thiersii*.The secretome of *A. thiersii* (Appendix B) is highly enriched with proteins involved in carbohydrate metabolism. For example, some domains unique to *A. thiersii* include GH families 61 (PF03443), 43 (PF04616), 28 (PF00245) and 10 (PF00331), which break down cellulose and hemicellulose, major components of plant cell walls. See section "Distribution of CAZymes in *Amanita*" for a description of similar results.

Both secretomes share a group of small (~100 a. a.) cysteine-rich proteins that are expressed only by filamentous fungi and known as hydrophobins (PF01185) (Tagu *et al.* 1996). These SSPs appear expanded in *A. thiersii* (Figures 4 and 5). These proteins may be involved in mediating contact and communication between the fungus and its environment. In contrast, other SSPs are enriched in *A. muscaria* (Figure 4). This is the case of a cysteine-rich domain, common in fungal extracellular membranes (CFEM, Kulkarni *et al.* 2003). Interestingly, CFEM domains are found in some proteins with proposed roles in fungal pathogenesis (Dean *et al.* 2005), and may be involved in fungal

symbiotic interactions (Plett *et al.* 2012). For instance, the expression of a family of SSPs with a CFEM domain was specifically induced when *L. bicolor* was involved in a symbiotic interaction (Martin *et al.* 2008).

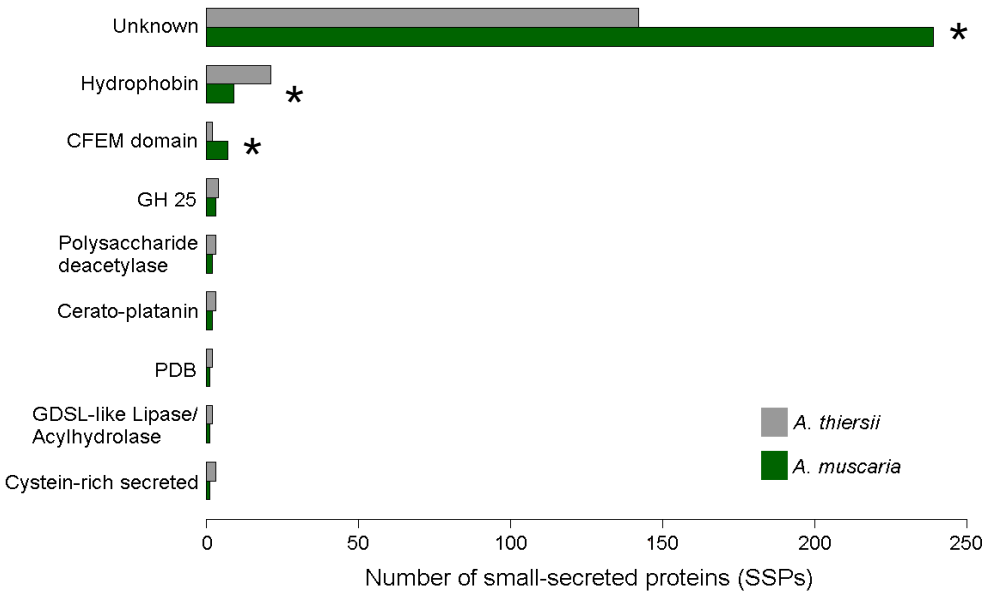


Figure 4. Small secreted proteins with PFAM domains that differ in number between both species shared by *A. muscaria* and *A. thiersii*. (*) Denotes difference is statistically significant after Bonferroni correction ($p < .0001$).

Distribution of CAZymes in Amanita

The query of each predicted *A. muscaria* and *A. thiersii* gene model in the proteome against the CAZy database (http://www.cazy.org) revealed that these fungi present a rich set of CAZymes that are both dissimilar in numbers as well as in distribution (Table 2). It is noteworthy that *A. thiersii* features a more abundant set of GHs both as secreted and non-secreted proteins in comparison to *A. muscaria*. Conversely, it presents a more reduced set of GTs and PLs, although for both species GTs are mostly present in the set of non-secreted proteins. In addition, although CEs are comparable in number between both species, the distribution pattern in *A. thiersii* shows that this class is clearly over-represented in the secretome.

Table 2. Number of proteins in major classes of carbohydrate active enzymes in *Amanita muscaria* and *A. thiersii*. GH, glycoside hydrolases; GT, glycosyl transferases; PL, polysaccharide lyases; CBM, carbohydrate-binding modules; CE, carbohydrate esterases; EXPN, expansins; AA, auxiliary activities.

Location	Species		GH	GT	PL	CBM	CE	EXPN	AA		Total proteins
Non-secreted	***A. muscaria***		**72**	68	1	11	7	6	29		**194**
Secreted			**49**	2	2	6	8	2	16		**85**
Total proteome			**121**	70	3	17	15	8	45		**279**
Non-secreted	***A. thiersii***		**103**	58	4	16	3	1	24		**209**
Secreted			**104**	0	6	9	17	5	20		**161**
Total proteome			**207**	58	10	25	20	6	44		**370**

A. thiersii presents 35 unique CAZy modules (24 GH, 2 PL, 1 AA, 3 CBM and 5 CE) compared to *A. muscaria*, many of which are predicted to act on plant cell wall polysaccharides (PCWP, Table 3). In contrast, only two CAZy modules (GH81 and GH89) are unique to the proteome of *A. muscaria* compared to *A. thiersii*, and they are both involved in fungal cell wall metabolism.

When considering non-secreted proteins only, the number of proteins containing a CAZy module is not significantly different between the two species. However, it does differ significantly for the secretome and SSPs of both species (data not shown). We observed that there is a reduction in the number of proteins in the secretome of *A. muscaria* involved in the degradation of plant cell wall polysaccharides (PCWP, Table 3) as compared to *A. thiersii*. For instance, *A. muscaria* lacks 10 of 12 GH families known to be active on PCWPs, while *A. thiersii* lacks only four. This pattern is similar to that observed in the ECM fungus *L. bicolor*, which lacks nine of those 12 GH families. A comparison of the CAZy modules present in the SSP subsets of both species resulted in a similar pattern of reduction in PCW degrading enzymes in the ECM fungus (Table 4). Here, *A. thiersii* has seven unique CAZy families GH24, GH45, GH61, CE5, CE12, GH12 and GH17, most of which are related to PCW degradation.

A notable difference between the proteomes of *A. muscaria* and *L. bicolor* is the absence in the former of GH family 28 *versus* six in *L. bicolor*. Another ECM fungus, *A. bisporigera*, also lacks these enzymes (Nagendrian *et al.* 2009), and *T. melanosporum* has only two (Martin *et al.* 2010). GH family 28 contains all known fungal polygalacturonases (PGs), a set of well-studied PCW degradative enzymes that hydrolize pectin, induce defense responses and contribute to virulence in some fungal pathogens (Sprockett *et al.* 2011). The species *A. muscaria, A. bisporigera*, *L. bicolor* and *T. melanosporum* all lack cellulases from families GH6 and GH7 (but GH6 cellulases are also absent from *A. thiersii*). *A. muscaria* also lacks proteins in CE family 8 that are present in *L. bicolor* as well as *A. thiersii* and are involved in degradation of pectins in plant cell walls.

The reduction pattern in *A. muscaria* is not observed when comparing numbers of proteins involved in carbohydrate metabolism of fungal cell walls (FCW, Table 3). The enzymes in GH families 18, 20 and 79, as well as CE family 4 are more likely to be involved in modification or degradation of fungal rather than plant cell walls because they contain chitinases and chitin-deacetylases (Cantarel *et al.* 2009), whose substrates are major components of fungal cell walls. Therefore the repertoire of FCW degradative enzymes in both *A. muscaria* and *A. thiersii* may be involved in growth-related remodeling of FCW or saprotrophy on other fungi.

Also contrasting the observed losses in GHs in the ECM fungus, a deeper analysis of the proteins containing the CE1 module in *A. muscaria* revealed that these proteins are expanded and might be of bacterial origin (see Box 1). These proteins were not predicted to be secreted; however they may act on carboxylic ester bonds of xylan and pectin substrates (Kubicek *et al.* 2010). Unlike the CE1 proteins in *A. muscaria*, the CE1 protein present in *A. thiersii* has an attached CBM1 module. The biological functions of proteins with this CBM1 module involve targeting and maintaining the proximity of the enzyme to cellulose (Kubicek *et al.* 2010).

CAZyme comparison to other genomes

When analyzing the *Amanita* genomes in the context of other published genomes of basidiomycete and ascomycete species with different lifestyles, we observed that CAZyme module distribution varies with lifestyle (Table 5; Appendix C) and even among the same lifestyle. For instance, the GH repertoire of the saprotroph *A. thiersii* (207) is comparable in number to that of other related saprotrophs, including *Agaricus bisporus* (291, Morin *et al.* 2012), *Coprinopsis cinerea* (211; Stajich *et al.* 2010) and *Volvariella volvacea* (224; Chen *et al.* 2013). But, the total number of CAZymes in the genome of *C. cinerea* (421) is similar to numbers in pathogenic fungi; for instance, the species' large number of CBM, GT and CE modules is comparable to those in *Magnaporthe grisea* (434; Dean *et al.* 2005) and *Gibberella zeae* (468; Amselem *et al.* 2011). These results evidence the variability found in the CAZyme repertoire among species that share a similar lifestyle.

Our results show that the CAZy repertoires for the ECM fungi *A. muscaria* and *L. bicolor* closely resemble those of the brown rot fungi *Postia placenta* (Martinez *et al.* 2009) and *Serpula lacrymans* (Eastwood *et al.* 2011) (Table 5, Appendix C). These genomes share a reduced spectrum of enzymes targeting PCW polymers compared to phytopathogens and saprotrophs (Appendix C). The genomes of the white rot fungi *Schizophyllum commune* (380, Ohm *et al.* 2010) (380) and *Phanerochaete chrysosporium* (313, Martinez *et al.* 2004) have similar numbers and distribution of CAZy classes (Table 5, Appendix C) and together resemble those of the saprotrophs *V. volvacea, A. thiersii* and *A. bisporus*.

Table 3. Putative carbohydrate active enzymes (CAZymes) active on plant and fungal cell wall polysaccharides in *A. muscaria* (ECM) and *A. thiersii* (saprotroph). Total number of proteins is presented with number of secreted proteins in parentheses. CBM, carbohydrate binding module; CE, carbohydrate esterases; GH, glycoside hydrolases; GT, glycosyl transferases; EXPN, distantly related to plant expansins; PL, polysaccharide lyases.

CAZymes acting on plant cell wall polysaccharides (PCWP)			
Substrate	CAZy module	*A. muscaria*	*A. thiersii*
Cellulose, hemicellulose, xylans	GH61	2(1)	14(11)
	GH43	3(2)	5(3)
	GH45	0	2(2)
	GH7	0	1(1)
	GH10	0	1(1)
	GH51	0	1(1)
	GH74	0	1(1)
	CBM1	0	1(1)
	GH5-CBM1	0	0
	GH6	0	0
	GH11	0	0
	GH67	0	0
Pectins	GH28	0	5(4)
	CE8	0	2(2)
	PL1	0	2(1)
	PL3	0	2(0)
	PL4	0	1(1)
	PL9	0	0
CAZymes acting on fungal cell wall polysaccharides (FCWP)			
Substrate	CAZy module	*A. muscaria*	*A. thiersii*
β-Glucans	GH16	19(7)	20(8)
	GH5	13(4)	20(7)
	GH3	4(1)	9(3)
	GH17	4(0)	8(3)
	GH30	1(1)	2(2)
	GT48	2(0)	2(0)
	CBM43	1(1)	1(1)
	GH72	1(1)	1(1)
	GH9	1(0)	1(0)
	CBM18	0	0
Chitin	GH18	14(8)	20(9)
	GT2	11(0)	11(0)
	CE4	7(5)	7(6)
	GH20	3(2)	3(2)
α-Glucans + other FCWP	GH79	7(2)	6(2)
	GH71	1(0)	3(0)
	GH13 + GT5	1(0)	1(0)
Other	EXPN	8(2)	6(5)

Table 4. The subset of *A. muscaria* (ECM) and *A. thiersii* (saprotroph) genes coding for small secreted proteins (SSPs) involved in carbohydrate metabolism. CBM, carbohydrate binding module; CE, carbohydrate esterases; GH, glycoside hydrolases; EXPN, distantly related to plant expansins; FCWP, fungal cell wall polysaccharides; PCWP, plant cell wall polysaccharides.

Annotation	*A. muscaria*	*A. thiersii*	Substrate	Active on
CBM13	3	0	xylan / mannose	Non-lytic activity
CBM50	1	1	chitin / peptidoglycan	Non-lytic activity
CE12	0	3	rhamnogalacturonan / pectin / xylan	PCWP
CE16	1	0	carbohydrate acetyl esters	PCWP
CE4	2	4	chitin	FCWP
CE5	0	2	xylan / cutin	PCWP
EXPN	2	5	Other FCWP	FCWP
GH12	0	3	glucans / xyloglucans	PCWP
GH16	1	2	glucans / galactans	FCWP
GH17	0	1	1,3-β-D-glucan / 1,3;1,4-β-D-glucan	FCWP
GH24	0	3	peptidoglycan	bacterial polysaccharides
GH25	3	4	peptidoglycan / function in fungi not certain	bacterial polysaccharides
GH45	0	2	cellulose / hemicellulose / xylans	PCWP
GH61	0	8	cellulose / hemicellulose / xylans	PCWP
GH79	1	0	Other FCWP	FCWP

Table 5. Distribution of major classes of carbohydrate active enzymes in selected basidiomycetes and ascomycetes with different lifestyles than *Amanita muscaria* and *A. thiersii*. GH, glycoside hydrolases; GT, glycosyl transferases; PL, polysaccharide lyases; CBM, carbohydrate-binding modules; CE, carbohydrate esterases.

Division	Species	Lifestyle	GH	GT	PL	CBM	CE	TOTAL
Basidiomycete	*Amanita muscaria*	Ectomycorrhizal	121	70	3	17	15	**226**
Basidiomycete	*Laccaria bicolor*	Ectomycorrhizal	163	88	7	26	19	**303**
Basidiomycete	*A. thiersii*	Saprotrophic	207	58	10	25	20	**320**
Basidiomycete	*Agaricus bisporus*	Saprotrophic	172	60	6	17	36	**291**
Basidiomycete	*Volvariella volvacea*	Saprotrophic	224	66	18	21	28	**357**
Basidiomycete	*Coprinopsis cinerea*	Saprotrophic	211	72	13	88	37	**421**
Basidiomycete	*Postia placenta*	Brown rot	157	27	2	12	22	**220**
Basidiomycete	*Serpula lacrymans*	Brown rot	163	71	4	17	19	**274**

Basidiomycete	*Schizophyllum commune*	White rot	225	79	9	28	39	**380**
Basidiomycete	*Phanerochaete chrysosporium*	White rot	180	68	4	45	16	**313**
Basidiomycete	*Cryptococcus neoformans*	Pathogenic	75	68	3	10	9	**165**
Basidiomycete	*Ustilago maydis*	Pathogenic	98	64	1	9	19	**191**
Ascomycete	*Magnaporthe grisea*	Pathogenic	231	92	4	60	47	**434**
Ascomycete	*Gibberella zea*	Pathogenic	243	102	20	61	42	**468**

Effector-like motifs among SSPs

Effector-like motifs are thought, and in a few cases shown, to suppress host defense response, and may be involved in mediating the establishment of symbioses (Yaeno *et al.* 2011; Plett and Martin 2012). We found no evidence of the effector motifs [LI]xAR and [RK]Cx2Cx12H of the rice blast fungal pathogen, *M. oryzae* (Yoshida *et al.* 2009) or the YxSL[RK] motif of *P. ultimum* (Lévesque *et al.* 2010) among the SSPs of *A. muscaria* or *A. thiersii.* However, we found the RxLR motif in 5 SSPs in *A. thiersii* and 6 SSPs in *A.muscaria* (Table 6). No putative functions could be assigned to SSPs containing RxLR motifs in *A. muscaria*, but at least three of these proteins in *A. thiersii* appear to be involved in PCW degradation (Table 6). Interestingly, it was recently shown that binding of an RXLR-like motif of *L. bicolor* to membrane phospholipids of *Populus trichocarpa* roots conferred access to MiSSP7 into the plant cell (Plett and Martin 2012). In addition, we found 92 and 74 proteins containing at least one [YFW]xC motif in *A. thiersii* and *A. muscaria*, respectively (data not shown). Proteins containing this motif have been suggested to be part of a new class of effectors from haustoria-producing pathogenic fungi (Godfrey *et al.* 2010).

Table 6. Small secreted proteins containing a RxLR-motif in *A. muscaria* and *A. thiersii*.

	A. muscaria				***A. thiersii***			
	Protein ID	position	Motif	HMMER	Protein ID	position	Motif	HMMER
1	24045	28	RKLR	hypothetical protein *Fibroporia radiculosa*	62220	284	RGLR	Metalloprotease
2	423730	97	RDLR	hypothetical protein *Fibroporia radiculosa*	49723	176	RDLR	Lysozyme
3	69699	198	RHLR	predicted protein *Laccaria bicolor*	46409	42	RPLR	Copper transporter
4	166224	60	RPLR	No hit	54995	57	RLLR	Hypothetical protein *Coprinopsis cinerea*
5	159751	70	RTLR	No hit	11121	104	RSLR	No significant hit
6	818364	151	RLLR	No hit				

Protein family classification

The combined secretome of *A. muscaria* and *A. thiersii* was clustered according to sequence similarity using an inflation parameter with a value of 2 (see Materials and Methods) and a total of 457 clusters were obtained (Figure 5). Of these, 157 and 116 clusters are composed of only one protein in *A. muscaria* and *A. thiersii*, respectively. These single clusters were removed from subsequent analyses. Moreover, clusters with at least two sequences were considered as putative protein families (Appendix D). These clusters contained 826 proteins, meaning that 75% of the combined secretomes has evolved in protein families. Both secretomes shared 40 protein families. Furthermore, 21 protein families are unique to *A. muscaria*, whereas 43 are unique to *A. thiersii* (Figure 5). Unique families are defined as protein families with members from only one species.

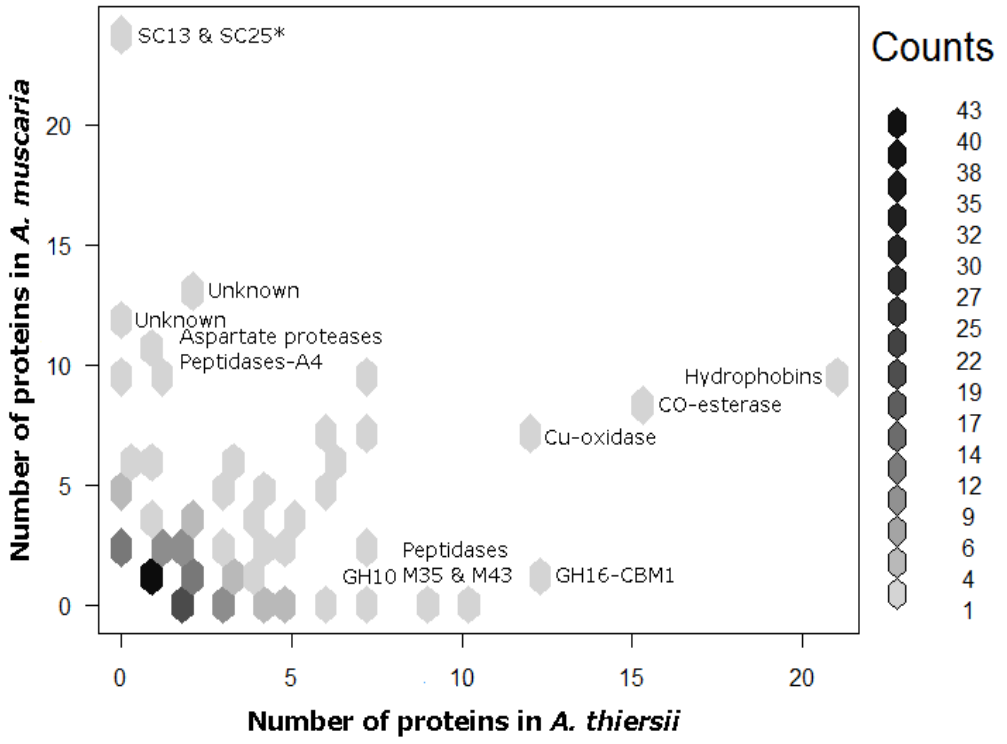


Figure 5. Number of proteins per cluster in *Amanita muscaria* and *A. thiersii*. CO, carboxyl; Cu, multicopper. * ECM-regulated cDNAs described in Nehls *et al.* (1999).

Notable protein families expanded in *A. thiersii* as compared to *A. muscaria* are genes with hydrolytic activity. These include carboxyl-esterases (PF00135), multicopper-

oxidases (PF07731, PF00394, and PF07732) and metallo-endopeptidases such as proteases of the families M35 (PF02102) and M43 (PF05572) (Rawlings *et al.* 2012). Similarly, hydrolytic enzymes specifically involved in degradation of carbohydrates (GH10 and GH16-CBM1) were not only expanded but also unique to the saprotroph's secretome (Figure 5).

Four expanded protein families appear distinctive in *A. muscaria* compared to *A. thiersii* (Figure 5). Two of these families have a potential novel function. The other two include enzymes with catalytic activity such as A4 peptidases (PF01828) and aspartate proteases (PF00026). Strikingly, the largest unique cluster in *A. muscaria* shows homology to SC13 and SC15, two differentially regulated clones in fully developed and functionally active symbiosis between *Picea abies* and *A. muscaria* (Nehls *et al.* 1999).

Protein family amplification in A. thiersii: *Hydrophobins*

Compared to *A. muscaria*, the secretome of *A. thiersii* is expanded in a protein family containing a PFAM domain that corresponds to hydrophobins (PF01185; Tagu *et al.* 1996) (Figure 5). Hydrophobins have been described in various filamentous fungi and are involved in cell morphogenesis, hyphal aggregation, and plant-microbe interactions (Wessels 1997). Recently, it has been suggested that hydrophobins play an important role in the establishment of ectomycorrhizal symbiosis. Plett *et al.* (2012) showed that hydrophobins are expanded in *L. bicolor* and expressed at higher levels when the fungus interacts with less receptive hosts. The authors in that study proposed that hydrophobins would provide a thicker layer around the fungal hyphae to either protect it from plant-based defenses or to hide hyphal surface antigens, which would induce a higher level of defense by the plant. However, that study did not compare the observed patterns of protein evolution and gene expression to those from closely related free-living species, and, therefore, cannot be directly linked to the evolution of ECM symbioses. In *Amanita*, our data shows evidence that hydrophobins are not a hallmark of ECM symbiosis in this clade.

Discussion

Fungi are heterotrophic and thus dependent on organic matter for energy. This dependency has resulted in the evolution of different lifestyles corresponding to the source used for carbon exploitation. Saprotrophic fungi live freely in soils, while biotrophic fungi can be either pathogenic or mutualistic. Ectomycorrhizal (ECM) symbiosis usually function as mutualisms where plants exchange carbon for scarce resources with certain soil fungi. Plant species capable of forming ECM are dominant components of forest and woodland ecosystems over much of the earth's surface. There is evidence that ECM symbiosis has evolved multiple times during evolutionary history (Bruns and Shefferson 2004; Matheny *et al.* 2006), and had a single origin within the genus *Amanita* (Wolfe *et al.* 2012). In this study we conducted a comparative analysis of the species *A. muscaria* (ECM) and *A. thiersii* (free-living) to explore the genetic basis of the transition to symbiosis in this fungal genus, with a particular focus on secreted proteins.

Our comparative analysis revealed that *A. muscaria* is deficient in its capacity to produce extracellular enzymes active on plant cell wall polysaccharides (PCWP). In particular, there is a reduction in cellulases, a consistent pattern among the symbiotic fungi studied so far. Eastwood *et al.* (2011) analyzed losses and expansions in 19 gene families of glycosyl hydrolases (GH), carbohydrate esterases and oxidoreductases across 10 species of fungi, including Agaricomycetes with a range of nutritional modes. They found that ectomycorrhizal fungi had the fewest hydrolytic CAZy genes. In addition, our findings confirm the results obtained by Wolfe et al. (2012a), who documented the loss of two genes in the cellulose degradation pathway used by free-living fungi to obtain carbon: *eg1*, a secreted endoglucanase (Jia *et al.* 1999) and *cbhI-I*, a secreted cellobiohydrolase that belongs to GH family 7 (Ding *et al.* 2001). These results are an indication that the ECM fungus avoids to degrade its host cell walls while in symbiosis with the roots.

Differential patterns of loss and retention of CAZymes involved in degradation of PCWP observed in A. *muscaria* are evidence of distinct genomic changes facilitating the evolution of ECM symbiosis among fungal taxa. For instance, in *A. muscaria* and other

ECM *Amanita* species such as *A. bisporigera* (Nagendrian *et al.* 2009), the loss of polygalacturonases (GH28, Table 4), which degrade pectin, contrasts with the presence and functional significance of endoglucanases (EGs) in the ECM fungi *T. melanosporum* and *L. bicolor* (Martin *et al.* 2008, 2010). These EGs are numerous in the secretome of the saprotroph *A. thiersii* and other saprotrophic and biotrophic fungi. At least in the case of *T. melanosporum*, this species appears to degrade its host cell walls during the formation of symbiosis (Martin *et al.* 2010).

In addition to the pattern of loss of cellulases, we observed a shift in the size of secreted proteins. The ECM fungus is enriched in small proteins, many of which remain uncharacterized (74% of SSPs). A large proportion of these SSPs contained effector-like motifs (Table 6) and all of them have uncharacterized functional domains. It could be hypothesized that these SSPs are novel proteins specific to the establishment of symbiosis between the fungus and host. However, to fully understand the function of pathogenic-like effector motifs in SSPs of both *Amanita* species, it will be necessary to conduct expression and experimental studies assessing their role in mediating PCW decomposition or establishment of ECM symbiosis.

It is matter of current debate whether ECM fungi can act as decomposers. Talbot *et al.* (2008) proposed that ECM fungi might live as facultative saprotrophs, able to degrade and metabolize soil C compounds as an alternative C source when the supplies of photosynthates from the host plant are low. Moreover, some ectomycorrhizal fungi have been demonstrated to utilize proteins as nitrogen and carbon sources (Abuzinadah & Read 1986). However, in the case of *A. muscaria*, it was experimentally demonstrated that this species had no cellulase activity (Wolfe *et al.* 2012) and therefore could not grow on litter; but it is able to grow when protein is provided as the sole nitrogen source. Whether or not ECM species are capable of partially degrading organic matter to acquire nutrients such as nitrogen from plant litter remains an open question.

This ability to degrade organic matter was discovered in the ECM *Paxillus involutus* (Rineau *et al.* 2012). In that study, the authors showed that *P. involutus* partially degraded

polysaccharides and modified the structure of polyphenols through chemical changes that were consistent with a hydroxyl radical attack, involving Fenton chemistry similar to that of brown-rot fungi (Hammel *et al.* 2002). This case may be similar to *A. muscaria*, where there is a wide distribution and amplification of protease gene families in its secretome and suggest that proteolytic activity plays an important role in the acquisition of nutrients (Figure 5; Appendices A, B and C). Both ECM fungi, *P. involutus* and *A.muscaria,* lack transcripts encoding extracellular enzymes needed for metabolizing released carbon after degradation of organic matter. These enzymes, such as GH45 and GH61, are present in brown rot fungi (Kubicek *et al.* 2010). The saprotrophic activity would be then reduced to a radical-based biodegradation system that can efficiently disrupt the organic matter–protein complexes and thereby mobilize the entrapped nutrients. The released carbon must become available for further degradation and assimilation by other organisms, because these ECM fungi have lost the ability to do so as an adaptation to symbiotic growth on host photosynthate.

Proteolytic activity is important not only in the ECM species, as mentioned above, but also in the saprotroph *A. thiersii*. It is important to note, however, that the proteolytic capabilities differ in both species. The saprotroph shows expansion of metallopeptidases, while the ECM fungus is particularly enriched in aspartic proteases (Figure 5; see below). Multicopper oxidases, such as laccases, are common in the genome of *A. thiersii*. These enzymes are involved in various biological processes including lignin degradation, melanin synthesis, and pathogenesis of human and plant hosts (Giardina *et al.* 2010). The genomes of other saprotroph mushrooms such as *A. bisporus* (Morin *et al.* 2012), *V. volvacea* (Chen *et al.* 2013) and *C. cinerea* (Stajich *et al.* 2010) also contain multiple laccase genes (Bao *et al.* 2013) that are hypothesized to play a role in the species' free-living lifestyle.

Aspartic proteases, on the other hand, are common in ECM fungi and *A. muscaria* is no exception. A potential role in suppression of plant defenses has been suggested for these enzymes (Salzer *et al.* 1997). Plant cells excrete a number of hydrolytic enzymes that generate elicitors from invading organisms that could, in turn, activate plant defenses.

Nehls *et al.* (2001) identified two extracellular aspartic proteases in *A. muscaria* and showed that one of them is differentially expressed at the plant/fungus interface with *Populus,* suggesting a role for aspartic proteases in suppressing plant defense mechanisms. In our study we provide evidence that aspartic proteases are expanded in *A. muscaria* as compared to the free-living fungus (Figure 5). Even though the protein identified by Nehls *et al.* (2001) is not part of the expanded family, aspartic proteases may play an important role in the establishment of symbiosis with the host.

In general, *A. muscaria* seems to follow the pattern in ECM species of large-scale loss of CAZymes; but a closer analysis of family 1 of carbohydrate esterases (CE1) revealed an interesting exception. We conducted a deeper molecular and phylogenetic analysis of this family (Box 1) and determined that they have a bacterial origin. Furthermore, they are present in at least 11 ECM *Amanita* species; yet, there is no evidence of these genes being present in the free-living species of this genus. While the function of the CE1 proteins remains undetermined, it is feasible that they may have enabled a shift in lifestyle in *Amanita*. Proteins in the CE1 family are phenolic acid esterases that enable microorganisms to attack and partially degrade plant tissues that contain aromatic compounds (Gupta-Udatha *et al.* 2011). These compounds covalently link and physically mask the potentially fermentable substrates in lignocelluloses, and thus protect them from degradation (Akon 2008).

This would not the first description of an acquisition of CAZymes from bacteria that allowed fungi to establish a habitat within a new environmental niche. García-Vallvé *et al.* (2000) found that the glycosyl hydrolases of rumen fungi were obtained laterally from bacteria, allowing them to degrade cellulose and hemicellulose in the rumen of herbivorous mammals. It would be, however, the first report of an acquisition of a CAZyme family from bacteria that has contributed to the evolution of (ecto) mycorrhizal symbiosis. Nevertheless, it may also be possible that the acquired CE1 proteins have a function unrelated to mycorrhizal formation. Given that *Amanita* ECM species have lost potential to degrade carbon, the horizontal transfer event, followed by an expansion, could be a form of reversal after the major losses of carbohydrate-degrading repertoires.

Conclusion

Prior to this study, available genomic data and observed patterns of molecular evolution in ECM species could not be directly linked to the evolution of ECM symbioses. Our study system allowed us to compare the repertoires of secreted proteins between more closely related free-living and ECM species. This approach allowed us to establish that (a) the genomic changes driving evolution of ECM symbiosis differ between fungal taxa; (b) in *A. muscaria*, patterns of loss of plant cell wall degrading enzymes as well as a shift towards small secreted proteins may have allowed more efficient communication with the plant host during the establishment of symbiosis; and, (c) there is a shift in the metabolic capabilities of the ECM fungus towards the exploitation of nutrients (*i.e.* Nitrogen) in the soil evident in the amplification of aspartic proteases and peptidases. Together, these features would allow the ECM fungus to prevent the degradation of its host cell walls, degrade elicitors of plant defenses, and potentially mobilize nutrients from the soil environment while establishing the symbiosis with the plant root. Finally, we hypothesized that a horizontal transfer event from bacteria is either an additional potential mechanism contributing to the evolution of ECM symbiosis or a form of reversal after major losses in the fungus' ability to metabolize carbon. Future work should include expression analysis to elucidate which SSPs are expressed during the formation of ECM symbiosis, experimental studies to determine the functions of proteins directly involved in the symbiosis and extend the comparison to other species in the genus to address the contractions and expansions of gene families in a phylogenetic context.

References

Abuzinadah, R.A. and D.J. Read. 1986. The role of proteins in the nitrogen nutrition of ectomycorrhizal plants. I. Utilization of peptides and proteins by ectomycorrhizal fungi. *New Phytologist*, 103: 481 – 493.

Akin, D.E. 2008. Plant cell wall aromatics: influence on degradation of biomass. *Biofuels Bioprod. Biorefin.*, 2: 288 – 303.

Amselem, J., C.A. Cuomo, J.A.L. van Kan, M. Viaud, E.P. Benito, *et al*. 2011. Genomic Analysis of the Necrotrophic Fungal Pathogens *Sclerotinia sclerotiorum* and*Botrytis cinerea*. PLoS Genet., 7: e1002230.

Baldrian, P. 2009. Ectomycorrhizal fungi and their enzymes in soils: is there enough evidence for their role as facultative saprotrophs? *Oecologia*, 161: 657 – 660.

Bao, D., M. Gong, H. Zheng, M. Chen, L. Zhang, H. Wang, *et al.* 2013. Sequencing and comparative analysis of the straw mushroom *Volvariella volvacea* genome. *PLoS ONE*, 8: e58294.

Bonfante, P. and A. Genre. 2010. Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat. Commun.*, 1: 48.

Bonfante, P. and N. Requena. 2011. Dating in the dark: how roots respond to fungal signals to establish arbuscular mycorrhizal symbiosis. *Curr. Opin. Plant Biol.*, 14: 451 – 457.

Boratyn, G.M., A.A. Schäffer, R. Agarwala, S.F. Altschul, D.J. Lipman and T.L. Madden. 2012. Domain enhanced lookup time accelerated BLAST. *Biol. Direct*, 7: 1 – 12.

Bouws, H., A. Wattenberg and H. Zorn. 2008. Fungal secretomes – nature's toolbox for white biotechnology. *Appl. Microbiol. Biotechnol.*, 80: 381 – 388.

Brown, N.A., J. Antoniw and K.E. Hammond-Kosack. 2012. The predicted secretome of the plant pathogenic fungus *Fusarium graminearum:* A refined comparative analysis. *PLoS ONE*, 7: e33731.

Bruns, T.D. 1995. Thoughts on the processes that maintain local species diversity of ectomycorrhizal fungi. *Plant and Soil*, 170: 63 – 73.

Bruns, T.D. and R.P. Shefferson. 2004. Evolutionary studies of ectomycorrhizal fungi: recent advances and future directions. *Can. J. Bot.*, 82: 1122 – 1132.

Cantarel, B.L., P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.*, 37: D233 - D238.

Chen, B., F. Gui, B. Xie, Y. Deng, X. Sun, M. Lin, *et al.* 2013. Composition and expression of genes encoding carbohydrate-active enzymes in the straw-degrading mushroom *Volvariella volvacea*. *PLoS ONE,* 8: e58780.

Cock, P.J. T. Antao, J.T. Chang, B.A. Chapman, C.J. Cox, A. Dalke *et al.* 2009. Biophyton: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics,* 25: 1422 – 1423.

Dean, R.A., N.J. Talbot, D.J. Ebbole, M.L. Farman, T.K. Mitchell, M.J. Orbach, *et al.* 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature,* 434: 980 – 986.

Deller, S., K.E. Hammond-Kosack and J.J. Rudd. 2011. The complex interactions between host immunity and non-biotrophic fungal pathogens of wheat leaves. *J. Plant Physiol.*, 168: 63 – 71.

Ding, S.J., W. Ge and J.A. Buswell. 2001. Endoglucanase I from the edible straw mushroom, *Volvariella volvacea* - purification, characterization, cloning and expression. *Eur. J. Biochem.*, 268: 5687 – 5695.

Doehlemann, G. 2009. Pep1, a secreted effector protein of *Ustilago maydis* is required for successful invasion of plant cells. *PLoS Pathog.*, 5: e1000290.

Eastwood, D.C., D. Floudas, M. Binder, A. Majcherczyk, P. Schneider, A. Aerts, *et al.* 2011. The plant cell-wall decomposing machinery underlies the functional diversity of forest fungi. *Science*, 333: 762 – 765.

Eberhart, J. L., D.L. Luoma, and M.P. Amaranthus. 1996. Response of ectomycorrhizal fungi to forest management treatments—a new method for quantifying morphotypes. In: Azcon-Aguilar, C. and J.M. Barea, eds. Mycorrhizas in integrated systems: from genes to plant development. Luxembourg: Office for Official Publications of the European Communities; 96-99.

Ellis, J.G., M. Rafiqi, P. Gan, A. Chakrabarti and P.N. Dodds. 2009. Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Curr. Opin. Plant Biol.*, 12: 399 – 405.

Emanuelsson, O., S. Brunak, G. von Heijne and H. Nielsen. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Prot.,* 2: 953 – 971.

Enright, A.J., S.V. Dongen and C.A. Ouzounis. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, 30: 1575 – 1584.

Finlay, R.D. 2008. Ecological aspects of mycorrhizal symbiosis: with special emphasis on the functional diversity of interactions involving the extraradical mycelium. *J. Exp. Bot.*, 59: 1115 – 1126.

Finn, R.D., J. Clements and S.R. Eddy. 2011. HMMER web server: interactive sequence similarity searching. *Nucl. Acids Res.*, web server issue 39: W29 - W37.

Fitzpatrick, D.A. 2011. Horizontal gene transfer in fungi. *FEMS Microbiol. Lett.*, 329: 1-8.

Frech, C. and N. Chen. 2010. Genome-wide comparative gene family classification. *PLoS ONE,* 5: e13409.

Galagan, J.E., M.R. Henn, L.J. Ma, C.A. Cuomo and B. Birren. 2005. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res.*, 15: 1620 – 1631.

García-Vallvé, S., A. Romeu and J. Palau. 2000. Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Mol. Biol. Evol.*, 17: 352 – 361.

Genre, A. and P. Bonfante. 2012. The interface between plants and mycorrhizal fungi: nutrient exchange, signaling and cell organization. In B. Hook (Ed.), A comprehensive treatise on fungi as experimental systems for basic and applied research (pp.39 – 49). The Mycota IX. Fungal associations. 2nd ed. Heidelberg: Springer.

Giardina P, V. Faraco, C. Pezzella, A. Piscitelli, S. Vanhulle and G. Sannia. 2010. Laccases: a never-ending story. *Cell Mol. Life Sci.*, 67: 369 – 385.

Godfrey, D., H. Böhlenius, C. Pedersen, Z. Zhang, J. Emmersen and H. Thordal-Christensen. 2010. Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics*, 11: 317 – 330.

Gupta-Udatha, D.B.R.K., I. Kouskoumvekaki, L. Olsson and G. Panagiotou. 2011. The interplay of descriptor-based computational analysis with pharmacophore modeling builds the basis for a novel classification scheme for feruloyl esterases. *Biotechnology Advances*, 29: 94 – 110.

Hammel, K.E., A.N. Kapich, K.A. Jensen and R.C. Ryan. 2002. Reactive oxygen species as agents of wood decay by fungi. *Enzyme Microb. Technol.*, 30: 445 – 453.

Hibbet, D.S., L.B. Gilbert and M.J. Donoghue. 2000. Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. *Nature,* 407: 506 – 510.

Hogenhout, S.A., R.A.L. van der Hoorn, R. Terauchi and S. Kamoun. 2009. Emerging concepts in effector biology of plant-associated organisms. Mol Plant Microbe Interact., 22: 115 – 122.

Horton, P., K-J. Park, T. Obayashi, N. Fujita, H. Harada, C.J. Adams-Collier and K. Nakai. 2007. WoLF PSORT: protein localization predictor. *Nucl. Acids Res.*, 35: W585 – 587.

Jia, J., P.S. Dyer, J.A. Buswell and J.F. Peberdy. 1999. Cloning of the *cbhI* and *cbhII* genes involved in cellulose utilisation by the straw mushroom *Volvariella volvacea*. *Mol. Gen. Genet.,* 261: 985 – 993.

Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual Review of Phytopathology,* 44: 41 – 60.

Kamoun, S. 2007. Groovy times: filamentous pathogen effectors revealed. *Curr. Opin. Plant Biol.,* 10: 358 – 365.

Kämper, J., R. Kahmann, M. Bölker, L-J. Ma, T. Brefort, B.J. Saville, *et al.* 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature,* 444: 97 – 101.

Kennedy, P.G., A.D. Izzo and T.D. Bruns. 2003. There is high potential for the formation of common mycorrhizal networks between understory and canopy trees in a mixed evergreen forest. *J. Ecol.*, 91: 1071 – 1080.

Kloppholz, S., H. Kuhn and N. Requena. 2011. A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr. Biol.,* 21: 1204 – 1209.

Kriventseva, E.V. 2005. Classification of proteins by clustering techniques. *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics,* 6: 1 – 7.

Krogh, A., B. Larsson, G. von Heijne and E.L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, 305: 567 – 580.

Kubicek, C.P., V. Seidl and B. Seiboth. Plant Cell Wall and Chitin Degradation. In: The Cellular and Molecular Biology of Filamentous Fungi, eds. Borkovich, K. and D.J. Ebbole, ASM press, Herndon, VA, USA. pp. 396 – 413.

Kulkarni, R. D., H.S. Kelkar and R.A. Dean. 2003. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. *Trends Biochem. Sci.,* 28: 118 – 121.

LePage, B.A., R.S. Currah, R.A. Stockey and G.W. Rothwell. 1997. Fossil ectomycorrhizae from the middle Eocene. *Am. J. Bot.,* 84: 410 – 412.

Lévesque, A.C., H. Brouwer, L. Cano, J.P. Hamilton, C. Holt, E. Huitema, *et al*. 2010. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome biology*, 11: R73 – R95.

Loftus, B.J., E. Fung, P. Roncaglia, D. Rowley, P. Amedeo, *et al*. 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science*, 307: 1321 – 1324.

Margulis, L. 1991. Symbiogenesis and symbionticism. In: Symbiosis as a source of evolutionary innovation: Speciation and morphogenesis. Margulis, L. and R. Fester, eds. pp. 1 – 24. Cambridge: MIT Press.

Martin, F., S. Duplessis, F. Ditengou, H. Lagrange, C. Voiblet and F. Lapeyrie. 2001. Developmental cross talking in the ectomycorrhizal symbiosis: signals and communication genes. *New Phytol.,* 151: 145 – 154.

Martin, F., A. Aerts, D. Ahrén, A. Brun, E.G.J. Danchin, F. Duchaussoy *et al*. 2008. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*, 452: 88-92.

Martin, F. and M.A. Selosse. 2008. Tansley review. The *Laccaria* genome: a symbiont blueprint decoded. *New Phytol.*, 180: 379 – 390.

Martin, F., A. Kohler, C. Murat, R. Ballestrini, P.M. Coutinho, O. Jaillon, *et al*. 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*, 464: 1033 – 1038.

Martinez, D., L.F. Larrondo, N. Putnam, M.D. Sollewijn Gelpke, K. Huang, *et al*. 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat. Biotechnol.*, 22: 695 – 700.

Martinez, D., J. Challacombe, I. Morgenstern, D. Hibbett, M. Schmoll, C.P. Kubicek, P. Ferreira, F.J. Ruiz-Duenas, A.T. Martinez, P. Kersten, *et al.* 2009. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc. Natl. Acad. Sci. USA*, 106: 1954 – 1959.

Matheny, P.B., J.M. Curtis, V. Hofstetter, M.C. Aime, J.M. Moncalvo *et al.* 2006. Major clades of Agaricales: A multilocus phylogeny overview. *Mycologia*, 98: 982 – 995.

Morin, E., A. Kohler, A.R. Baker, M. Foulongne-Oriol, V. Lombard, L.G. Nagy, *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 – 17506.

Mueller, O., R. Kahmann, G. Aguilar, B. Trejo-Aguilar, A. Wu and R.P. de Vries. 2008. The secretome of the maize pathogen *Ustilago maydis*. *Fungal Genet. Biol.* 45, S63 – S70.

Nagendran, S., H.E. Hallen-Adams, J.M. Paper, N. Aslam and J.D. Walton. 2009. Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus *Amanita bisporigera*, based on the secretome of *Trichoderma reesei*. *Fungal genet. Biol.*, 46: 427-435.

Nehls, U., S. Mikolajewski, M. Ecke and R. Hampp. 1999. Identification and expression analysis of two fungal cDNAs regulated by ectomycorrhiza and fruit body formation. *New Phytol.*, 144: 195 – 202.

Nehls, U., A. Bock, W. Einig and R. Hampp. 2001. Excretion of two proteases by the ectomycorrhizal fungus *Amanita muscaria*. *Plant, Cell and Environment*, 24: 741 – 747.

Plett, J.M. and F. Martin. 2011. Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. *Cell*, 27: 14 – 22.

Ohm, R.A., J.F.D. Jong, L.G. Lugones, A. Aerts, E. Kothe, *et al.* 2010. Genome sequence of the model mushroom *Schizophyllum commune*. *Nat. Biotechnol.*, 28: 957 – 963.

Perotto, S., R. Peretto, A. Faccio, A. Schubert, A. Varma and P. Bonfante. 1995. Ericoid mycorrhizal fungi: cellular and molecular bases of their interactions with the host plant. *Can. J. Bot.*, 73: S557 – S568.

Petersen, T.N., S. Brunak, G. von Heijne and H. Nielsen. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8: 785 – 786.

Plett, J.M., M. Kemppainen, S. D. Kale, A. Kohler, V. Legué, A. Brun, B. M. Tyler, A.G. Pardo and F. Martin. 2011. A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr. Biol.*, 21: 1197 – 1203.

Plett, J.M. and F. Martin. 2012. Poplar root exudates contain compounds that induce the expression of MiSSP7 in *Laccaria bicolor*. *Plant Signaling and Behavior*, 7: 12 – 15.

Plett, J.M., J. Gibon, A. Kohler, K. Duffy, P. Hoegger, R. Velagapudi, *et al.* 2012. Phylogenetic, genomic organization and expression analysis of hydrophobin genes in the ectomycorrhizal basidiomycete *Laccaria bicolor*. *Fungal Genet. Biol.*, 49: 199 – 209.

Pringle, A. and E.C. Vellinga. 2006. Last chance to know? Using literature to explore the biogeography and invasion biology of the death cap mushroom *Amanita phalloides*. *Biol. Invasions*, 8: 1131-1144.

Pringle, A., J.D. Bever, M. Gardes, J.L. Parrent, M.C. Rillig and J.N. Klironomos. 2009. Mycorrhizal symbioses and plant invasions. *Annu. Rev. Ecol. Evol. Syst.*, 40: 699-715.

Punta, M., P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, *et al.* 2012. The Pfam protein families database. *Nucl. Acids Res.*, 40: D290 - D301.

Rajashekar, B., A. Kohler, T. Johansson, F. Martin, A. Tunlid and D. Ahrén. 2009. Expansion of signal pathways in the ectomycorrhizal fungus *Laccaria bicolor* – evolution of nucleotide sequences and expression patterns in families of protein kinases and RAS small GTPases. *New. Phytol.*, 183: 365 – 379.

Rawlings, N.D., A.J. Barrett and A. Bateman. 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.*, 40: D343 – D350.

Redecker, D., R. Kodner and L.E. Graham. 2000. Glomalean fungi from the Ordovician. *Science*, 289: 1920 – 1921.

Richards, T.A., G. Leonard, D.M. Soanes and N.J. Talbot. 2011. Gene transfer into the fungi. *Fungal Biol. Rev.*, 25: 98-110.

Rineau, F., D. Roth, F. Shah, M. Smits, T. Johansson, B. Canbäck, *et al.* 2012. The ectomycorrhizal fungus *Paxillus involutus* converts organic matter in plant litter using a

trimmed brown-rot mechanism involving Fenton chemistry. *Environ. Microbiol.*, 14: 1477 – 1487.

Sagan, L. 1967. On the origin of mitosing cells. *J. Theor. Biol.,* 14: 225–193.

Salzer, P., G. Hebe and A. Hager. 1997. Cleavage of chitinous elicitors from the ectomycorrhizal fungus *Hebeloma crustuliniforme* by host chitinases prevents induction of K^+ and Cl^- release, extracellular alkalinization and H_2O_2 synthesis of *Picea abies* cells. *Planta*, 203: 470 – 479.

Sanders, I.R. 2011. Mycorrhizal symbioses: how to be seen as a good fungus. *Curr. Biol.*, 21: R550 – R552.

Sievers, F., A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, *et al.* 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.,* 7: 539 – 545.

Sigrist, C.J.A., L.Cerutti, N. Hulo, A. Gattiker, L. Falquet, M. Pagni, A. Bairoch and P. Bucher. 2002. PROSITE: A documented database using patterns and profiles as motif descriptors. *Brief. Bioinf.,* 3: 265 – 274.

Simon, L., J. Bousquet, R.C. Levesque and M. Lalonde. 1993. Origin and diversification of endomycorrhizal fungi and coincidence with vascular plants. *Nature*, 363: 67-69.

Smith, J.M. and E. Szathmáry. 1999. The Origins of Life. Oxford Univ. Press.

Smith, S.E. and D.J. Read. 1997. Mycorrhizal symbiosis. San Diego, CA, USA: Academic Press.

Smith, S.E. and D.J. Read. 2008. Mycorrhizal symbiosis. New York, NY, USA: Academic Press.

Sprockett, D.D., H. Piontkivska and C.B. Blackwood. 2011. Evolutionary analysis of glycosyl hydrolase family 28 (GH28) suggests lineage-specific expansions in necrotrophic fungal pathogens. *Gene*, 479: 29 – 36.

Stajich, J.E., S.K. Wilke, D. Ahrén, C.H. Au, B.W. Birren, *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 – 11894.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22: 2688 – 2690.

Stergiopoulos, I., and P.J.G.M. de Wit. 2009. Fungal effector proteins. *Annu. Rev. Phytopathol.*, 47: 233 – 263.

Stijn van Dongen. *Performance criteria for graph clustering and Markov cluster experiments*. Technical Report INS-R0012, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, May 2000.

Tagu D., B. Nasse and F. Martin. 1996. Cloning and characterization of hydrophobins-encoding cDNAs from the ectomycorrhizal basidiomycete *Pisolithus tinctorius*. *Gene*, *168*: 93 – 97.

Talbot, J.M., S.D. Allison and K.K. Treseder. 2008. Decomposers in disguise: mycorrhizal fungi as regulators of soil C dynamics in ecosystems under global change. *Functional Ecology,* 22: 955 – 963.

Towler, D.A.; J.I. Gordon, S.P. Adams and L. Glaser. 1988. The biology and enzymology of eukaryotic protein acylation. *Ann. Rev. Biochem.*, 57: 69-97.

Van Ooij, C. 2011. Establishing the roots of a relationship. *Nat. Rev. Microbiol.*, 9: 629.

Villareal-Ruiz, L., I.C. Anderson and I.J. Alexander. 2004. Interaction between an isolate from the *Hymenoscyphus ericae* aggregate and roots of *Pinus* and *Vaccinium*. *New Phytol.*, 164: 183 – 192.

von Heijne, G., J. Steppuhn and S.G. Hermann. 1989. Domain structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem.*, 180: 535 – 545.

von Heijne, G. 1990. The signal peptide. *J. Membr. Biol.*, 115: 195 – 201.

Weiss, M., M.A. Selosse, K.H. Rexer, A. Urban and F. Oberwinkler. 2004. *Sebacinales*: a hitherto overlooked cosm of heterobasidiomycetes with a broad mycorrhizal potential. *Mycol. Res.*, 108: 1003 – 1010.

Wessels, J.G.H. 1997. Hydrophobins: proteins that change the nature of the fungal surface. *Advances in Microbial Physiology* 38: 1-45.

Wolfe, B.E., F. Richard, H.B. Cross and A. Pringle. 2010. Distribution and abundance of the introduced ectomycorrhizal fungus *Amanita phalloides* in North America. *New Phytol.*, 185: 803-816.

Wolfe B.E., R.E. Tulloss and A. Pringle. 2012a. The irreversible loss of a decomposition pathway marks the single origin of an ectomycorrhizal symbiosis. *PloS One*, 7(7): e39597.

Wolfe, B.E., M. Kuo and A. Pringle. 2012b. *Amanita thiersii* is saprotrophic and expanding its range in the United States. *Mycologia*, 104: 22 – 33.

Yaeno, T., H. Li, A. Chaparro-Garcia, S. Schornack, S. Koshiba, S. Watanabe, *et al.* 2011. Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proc. Natl. Acad. Sci. USA*, 108: 14682 – 14687.

Yoshida, K., H. Saitoh, S. Fujisawa, H. Kanzaki, H. Matsumura, K. Yoshida, *et al.* 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell*, 21: 1573 – 1591.

BOX 1. **A potential horizontal gene transfer event in the ectomycorrhizal fungus *Amanita muscaria***

A total of six genes encoding carbohydrate esterases -family 1- (CE1) with an apparent bacterial origin have been found in the genome of *Amanita muscaria* (koide), an ectomycorrhizal (ECM) fungus. CEs belong to the carbohydrate active enzymes (CAZy) in charge of removing acyl groups from compounds (Towler et al. 1988). Family 1 CEs are involved in the biodegradation of small sugars and acids in hemicellulosic side chains (Kubicek *et al.* 2010). For all the potentially horizontally transferred genes, a BLAST search indicates that there is a 39-47% identity to a PHB depolymerase from *Ktedonobacter racemifer*, a bacterium that belongs to the phylum Chlorobacteria and whose genome has been sequenced recently (Eisen 2010, data not published, http://genome.jgi.doe.gov/ktera/ktera.info.html).

In order to prove that these CE1 genes have a bacterial origin (i.e. have been horizontally transferred), a first step was to eliminate the possibility of bacterial contamination in the original culture. It should be noted that: two different initiatives have sequenced the genome of the same strain of *A. muscaria* (koide). Both initiatives used the same starting strain cultured in different laboratories. Moreover, the sequences were obtained from separate DNA extractions and at different sequencing facilities.

The genome of *A. muscaria* (koide) version 1 deposited in JGI shows that all six genes are located on scaffold 57, clustered between positions 64720 and 103161 (http://genome.jgi.doe.gov/cgi-bin/browserLoad/?db=Amamu1&position=scaffold_57:64700-103180) (Table 1 and Figure 1). These genes are also present in the local assembly of the genome (data not shown), and EST and RNAseq gene expression data show that some of the genes are expressed (Figure 1). In addition, five fungal genes are interspersed within the region downstream CE1.1 and upstream CE1.3 (data not shown).

Table 1. Location of the potentially horizontally transferred genes in the genome of *A. muscaria*.

Gene	JGI protein ID	Scaffold	Start	End	Length of encoded protein (a. a.)
CE1.1	166350	57	64720	66009	348
CE1.2	13215	57	84582	85800	351
CE1.3	13216	57	88259	89156	245
CE1.4	166362	57	101624	103161	394
CE1.5	13212	57	77124	77431	105
CE1.6	166373	57	90842	92423	361

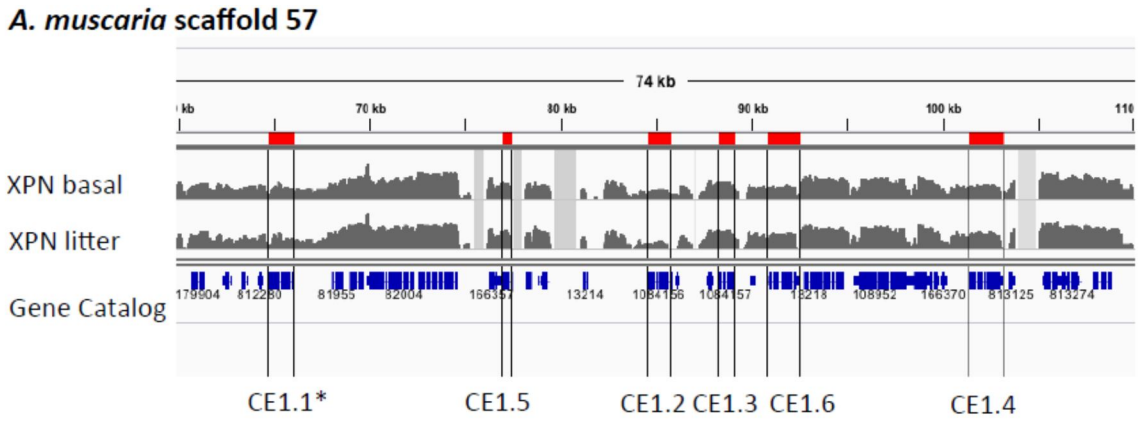


Figure 1. Physical map of scaffold 57 in *A.* muscaria. Gene models are shown in blue. Red boxes correspond to each one of the potentially horizontally transferred genes. In gray: expression data in basal and litter media. * A homolog of the gene CE1.1 was found in the genome sequence of the ECM species *A. brunnescens*.

As a second step to discount bacterial contamination in the culture, we amplified all six CE1 genes in DNA obtained from fresh cultures of the same strain of *A. muscaria* (koide). In addition, the same primers were used to amplify the six CE1 in two *A. muscaria* strains collected at different locations: FP01 and 283. The PCR products of the primer pairs targeting all six CE1 genes were recovered for all three strains and confirmed using Sanger sequencing (Figure 2a). We also searched for these genes in the genomes of other *Amanita* species published in JGI or locally available at Pringle Lab (Figure 2a). A homolog of the gene CE1.1 was found in the local assembly of the genome of the ECM species *A. brunnescens*. In contrast, free-living species of the genus such as

A. inopinata, *A. thiersii* and their sister species *Volvariella volvacea* did not show evidence of carrying these genes. In the case of *A. thiersii* and *V. volvacea*, a non-homologous CE1 was found and it contained a CBM1 (cellulose binding module).

To determine the timing of the HGT, we amplified CE1s from additional species across the phylogeny of *Amanita*. We were able to show that these genes are ubiquitous among ectomycorrhizal species (Figure 2b). Confirmation by Sanger sequencing was possible for most of the species were PCR amplification was successful (darker squares in Figure 2b). It is important to know that empty squares in Figure 2b do not represent confirmed absences of the genes.

Gene trees that are strongly supported and deviate significantly from the species tree are indicative of HGT (Fitzpatrick 2011). Therefore, to provide phylogenetic evidence that a HGT event occurred, the CE1 sequences obtained for all mycorrhizal species of the genus were used to build a gene phylogeny of CE1s in bacteria, basidiomycete and ascomycete fungi (Figure 3). Interestingly, as pointed in Fig. 3, the CE1 copies present in *A. thiersii* and *V. volvacea* are homologous to the so called CE1 clade B (Dimitrios Floudas, *unpublished data*) and are found within the fungal part of the phylogenetic tree. Furthermore, the CE1 clade A (Dimitrios Floudas, *pers. comm*.) of fungi and the clade of ectomycorrhizal *Amanita* share homology with the bacterial taxa included in the phylogeny. However, the bacterial origin of CE1s differs between both clades (Figure 3), suggesting that at least two HGT events have occurred for this family of esterases from bacteria to fungi.

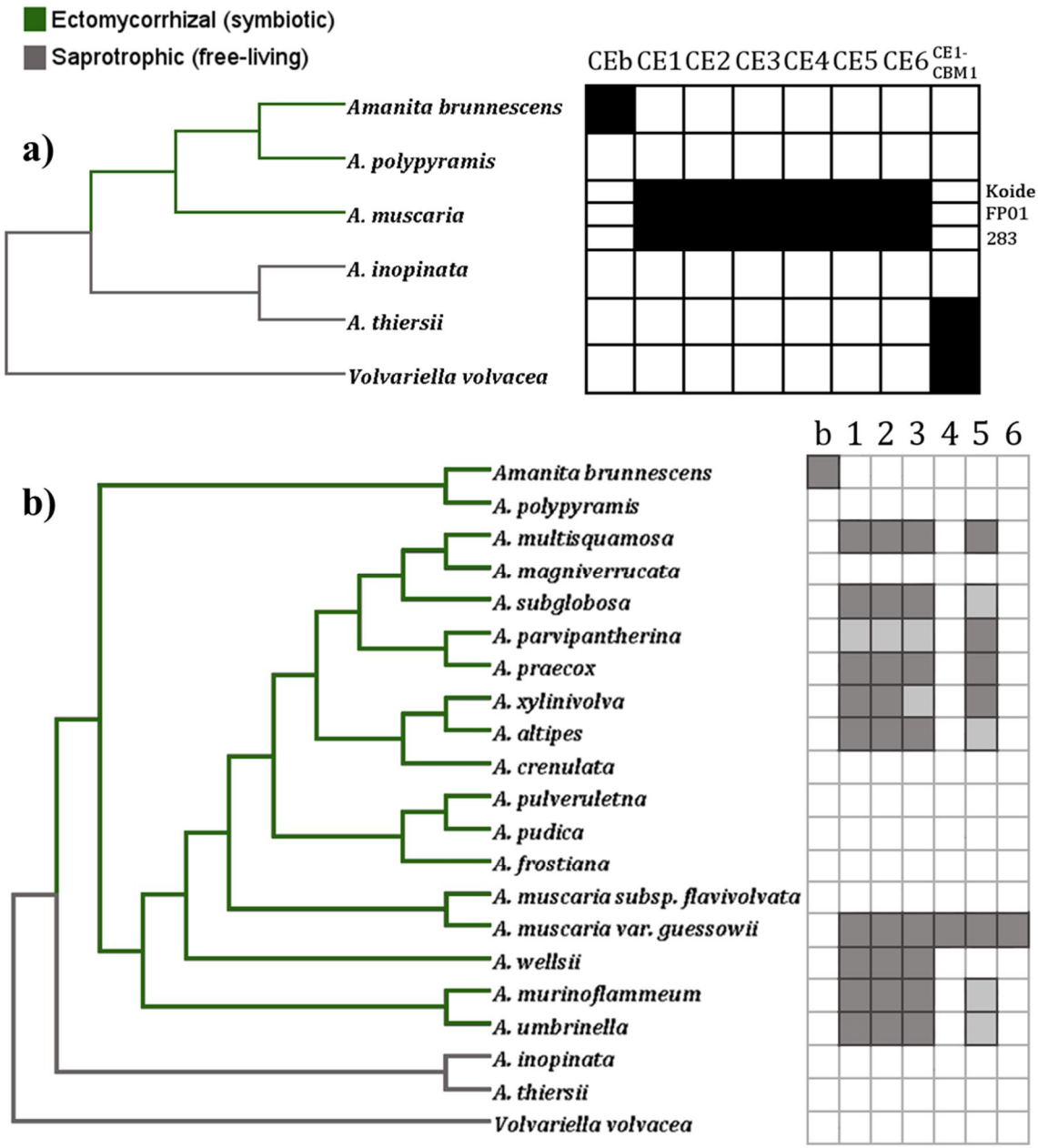
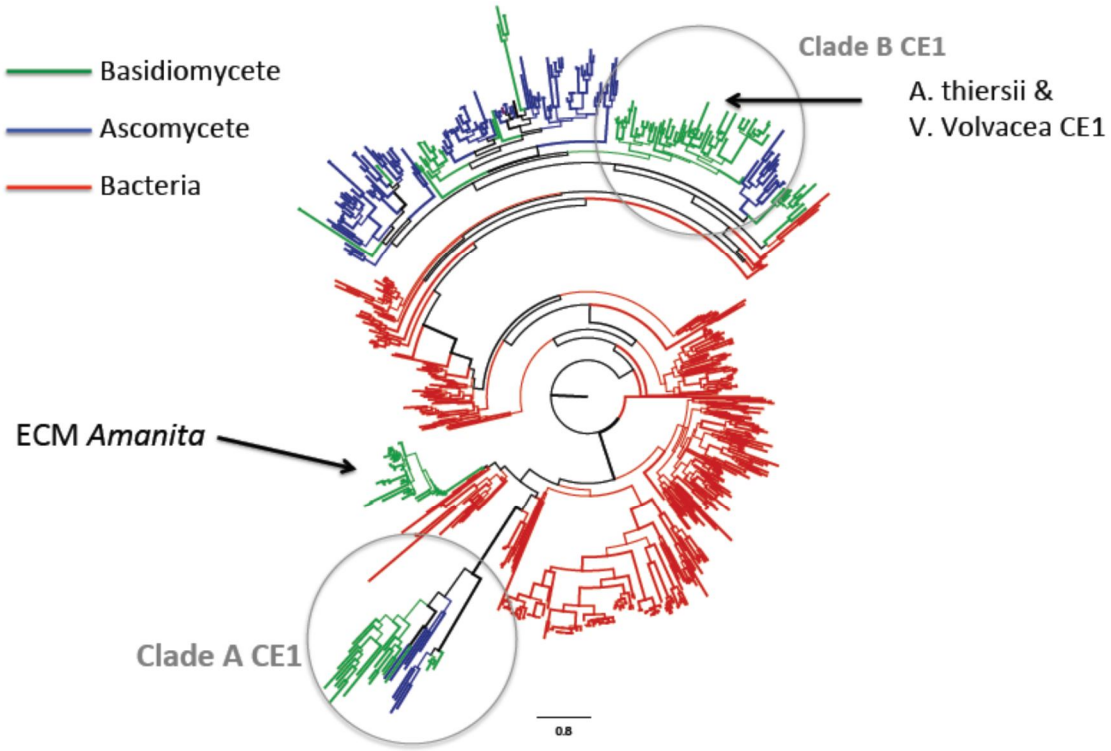


Figure 2. Presence of CE1 genes across the *Amanita* phylogeny and the outgroup *Volvariella volvacea.* Presence of CE1 genes a) based on confirmed genomic sequences and b) specific to the *Amanita* ectomycorrhizal subclade. Dark gray squares correspond to presence of the gene based on Sanger sequencing. Light gray correspond to presence confirmed by PCR amplification only. Empty boxes could not be amplified but do not mean absolute proof of absence of the genes, except for outgroups *A. inopinata*, *A. thiersii* and *V. volvacea*. A four-gene phylogeny built by Wolfe *et al.* (2012) was used as the species tree.

Figure 3. Phylogenetic evidence of horizontal gene transfer of family 1 of carbohydrate esterases in ectomycorrhizal *Amanita*.



Materials and Methods

Screening for CE1 genes across Amanita

Genomic DNA was isolated from *A. muscaria* var. *guessowii* strains Koide BX008, FP01 and PS #283 and *A. brunnescens* (BW HF10C) using a Phenol-Chloroform protocol. For all other strains, genomic DNA was the same as used by Wolfe *et al.* (2012). Primers were designed simultaneously for all seven genes using Geneious version 1.6 created by Biomatters (http://www.geneious.com/).Geneious advanced primer tools allow designing optimal pairs of primers while ensuring these pairs match the sequence of interest only (Table 1). These primer pairs were used to amplify CE1 genes in three different *A. muscaria* var. *guessowii* strains as well as a set of species across the genus *Amanita* (Figures 2a and 2b). When an amplicon of the expected size was detected, it was sequenced using Sanger sequencing.

Table 1. Primers used to amplify CE1 genes across *Amanita.*

Gene Name	Primer Pair	Primer Name	Sequence (5' to 3')
CE1.1	1	CE1.1.i.fw	CCATGGGTGACTCCTGGAAC
		CE1.1.i.rv	CAGCGCTGTACGTATAGCCA
CE1.2	2	CE1.2.i.fw	TGTTTTCGCTGCCATTGGTG
		CE1.2.i.rv	CCAGGAGGCAGCACTATACG
CE1.3	3	CE1.3.i.fw	TCTTTGCATCCTGACGTGCT
		CE1.3.i.rv	GGAGGCAGCGCTGTATGTAT
CE1.4	4	CE1.4.i.fw	CTTTGCCAAACCAGCTGACC
		CE1.4.i.rv	ATCACTCGGGCTACCTGTCT
CE1.5	5	CE1.5.i.fw	TGGAACAACGGCAGAGTTCA
		CE1.5.i.rv	GACAACGACTGGCTTGGGTA
CE1.6	6	CE1.6.i.fw	CGCAGCTATGACCGTCATCT
		CE1.6.i.rv	TCCATTAACCAGTGGCGGAC
CE1.b	1b	A_brun_CE1_fw	ATACAGGTCCTTCCGGCTCT
		A_brun_CE1_rv1	ATTCCATGATGCCACCGTGT

Phylogenetic analysis

Phylogenetic analyses were performed at the a. a. level. CE1 sequences were retrieved from bacteria, basidiomycete and ascomycete fungi based on standard protein-protein BLAST (BLASTP) sequence similarity searches. We also included group A and B CE1 genes previously analyzed by Dimitrios Floudas (*unpublished data*).

Protein multiple sequence alignments were generated using Clustal Omega (Sievers *et al.* 2011). The protein dataset was analyzed by maximum likelihood (ML) using the phylogenetic software RAxML (Stamatakis 2006) with a WAG+ Γ substitution model.

Appendices

Appendix A. Most frequent PFAM domains in *Amanita muscaria*.

PFAM ID	PFAM description	Number of proteins	JGI protein ID
None	Unknown	275	Available on request
PF01565	FAD binding domain	20	15104, 124078, 168234, 168232, 15280, 170711, 86604, 87000, 361258, 82763, 1028575, 86341, 15105, 170300, 108416, 15360, 77085, 13300, 86370, 73925
PF00026	Eukaryotic aspartyl protease	19	1015803, 160710, 374102, 77450, 10037, 23970, 521201, 68030, 194361, 438714, 806948, 26952, 14952, 174642, 27097, 126824, 1014703, 78642, 13161
PF08031	Berberine and berberine like	15	15104, 124078, 168232, 15280, 87000, 82763, 1028575, 86341, 15105, 170300, 108416, 15360, 77085, 86370, 73925
PF01828	Peptidase A4 family	10	13792, 83461, 160347, 82298, 82276, 13366, 13355, 82243, 82209, 13356
PF00704	Glycosyl hydrolases family 18 (chitinase activity)	9	158597, 79620, 7949, 130478, 912317, 16788, 519588, 177245, 174413
PF01185	Fungal hydrophobin	9	13887, 13888, 13078, 805173, 179833, 16404, 13050, 26353, 13892
PF00135	Carboxylesterase family	8	79395, 160722, 660840, 11449, 161330, 163787, 11450, 374064
PF07859	alpha/beta hydrolase fold	8	79395, 160722, 660840, 11449, 161330, 163787, 11450, 374064
PF05730	CFEM domain	7	166314, 160653, 160036, 170951, 811354, 11867, 170950
PF07732	Multicopper oxidase	7	170078, 380044, 366375, 98368, 171605, 10979, 165129
PF00722	Glycosyl hydrolases family 16	7	188094, 157477, 572987, 199495, 420178, 69940, 747950
PF01764	Lipase (class 3)	7	78528, 118868, 158256, 118906, 184433, 762878, 124358
PF00450	Serine carboxypeptidase	7	169205, 198303, 161281, 75756, 98802, 574681, 22923
PF00394	Multicopper oxidase	7	170078, 380044, 366375, 98368, 171605, 10979, 165129
PF07731	Multicopper oxidase	6	380044, 366375, 98368, 171605, 10979, 165129
PF00082	Subtilase family	6	10139, 130625, 23939, 13782, 197494, 68237
PF10342	Ser-Thr-rich GPI-anchored membrane family	6	67439, 6067, 13092, 199551, 162695, 173094
PF01522	Polysaccharide deacetylase	5	171528, 171533, 1042746, 84238, 183520
PF13738	Pyridine nucleotide-disulphide oxidoreductase	5	70326, 176126, 88862, 526943, 84343
PF00067	Cytochrome P450	4	160034, 9529, 122981, 9396
PF00150	Cellulase (glycosyl hydrolase family 5)	4	83540, 7559, 171865, 174786
PF00652	Ricin-type beta-trefoil lectin domain	4	14257, 201047, 283691, 1023467
PF01494	FAD binding domain	4	70326, 27484, 176126, 526943
PF04185	Phosphoesterase family	4	126148, 11192, 168454, 79194

Appendix B. Most frequent PFAM domains in *Amanita thiersii*.

PFAM ID	PFAM description	Number of proteins	JGI protein ID
None	Unknown	204	Available on request
PF01185	Fungal hydrophobin	21	9044, 153465, 8001, 153145, 147541, 51622, 69914, 9327, 147596, 63114, 9328, 63067, 75507, 8987, 148816, 158742, 158744, 9303, 66544, 147588, 153138
PF01565	FAD binding domain	15	71699, 77602, 156140, 8172, 76665, 65924, 55211, 137052, 147027, 147056, 194445, 69208, 7626, 10083, 69520
PF00135	Carboxylesterase family	14	66371, 152568, 59728, 178364, 76935, 50801, 153021, 60268, 154314, 71494, 800, 153008, 3565, 3564
PF00082	Subtilase family	14	139821, 152700, 85772, 6996, 151702, 198228, 7647, 76669, 6424, 50473, 6423, 7712, 187018, 59869
PF07859	alpha/beta hydrolase fold	12	66371, 152568, 59728, 178364, 50801, 153021, 60268, 154314, 71494, 153008, 3565, 3564
PF07732	Multicopper oxidase	12	9610, 43159, 152958, 9259, 5940, 67578, 157107, 152937, 157242, 77109, 6367, 67489
PF03443	Glycosyl hydrolase family 61	12	150098, 148473, 77762, 74874, 75127, 53822, 147046, 198904, 5468, 77096, 64840, 5891
PF07731	Multicopper oxidase	10	9610, 43159, 152958, 9259, 5940, 67578, 152937, 77109, 6367, 67489
PF01764	Lipase (class 3)	10	142006, 197673, 142271, 81145, 137254, 140728, 63172, 193502, 75703, 193495
PF02102	Deuterolysin metalloprotease (M35) family	10	71658, 197192, 43144, 64907, 5929, 76123, 77803, 150837, 152893, 49565
PF00704	Glycosyl hydrolases family 18	10	152190, 141554, 136459, 41705, 2893, 2894, 59398, 153019, 68473, 2688
PF05572	Pregnancy-associated plasma protein-A	10	9181, 62218, 86967, 62220, 67416, 75717, 71908, 51642, 62202, 55272
PF00394	Multicopper oxidase	10	9610, 43159, 152958, 9259, 5940, 67578, 152937, 77109, 6367, 67489
PF02225	PA domain	8	67605, 198228, 76669, 7344, 7712, 85178, 187018, 67596
PF04389	Peptidase family M28	8	67605, 141573, 46806, 142927, 189746, 7344, 56432, 67596
PF00026	Eukaryotic aspartyl protease	8	139688, 73786, 134385, 46374, 45724, 1807, 70417, 138478
PF05922	Peptidase inhibitor I9	7	139821, 152700, 85772, 6996, 6424, 50473, 6423
PF08031	Berberine and berberine like	7	71699, 77602, 156140, 76665, 65924, 137052, 10083
PF00722	Glycosyl hydrolases family 16	7	74678, 76815, 178633, 83659, 72, 75502, 72492
PF00450	Serine carboxypeptidase	7	7364, 69125, 68987, 75331, 152834, 54267, 1338
PF00734	Fungal cellulose binding domain	6	53913, 149157, 134977, 1897, 77096, 135881
PF00144	Beta-lactamase	6	67688, 42916, 2825, 64704, 67690, 2852
PF00150	Cellulase (GH 5)	6	53295, 185172, 65780, 74245, 177045, 145661
PF10342	Ser-Thr-rich GPI-anchored membrane family	6	39973, 4970, 7226, 60326, 187885, 45563
PF01522	Polysaccharide deacetylase	5	45949, 2750, 61932, 5941, 8535
PF06280	Fn3-like domain (DUF1034)	5	198228, 7647, 76669, 7712, 187018
PF01546	Peptidase family M20/M25/M40	5	67605, 141573, 46806, 7344, 56432
PF02128	Fungalysin metallopeptidase	5	9469, 61980, 62981, 182533, 152535

	(M36)		
PF12697	Alpha/beta hydrolase family	4	7364, 68987, 139154, 135563
PF09118	Domain of unknown function (DUF1929)	4	1279, 8122, 82381, 135500
PF12296	Hydrophobic surface binding protein A	4	40345, 151023, 10247, 69743
PF03330	Rare lipoprotein A	4	187285, 5954, 135159, 72811
PF00657	GDSL-like Lipase/Acylhydrolase	4	58713, 147123, 75977, 73100
PF07971	Glycosyl hydrolase family 92	4	149573, 77360, 8125, 82593
PF00141	Peroxidase	4	71171, 65091, 149618, 43976
PF07250	Glyoxal oxidase N-terminus	4	1279, 8122, 82381, 135500
PF01183	Glycosyl hydrolase family 25	4	1476, 49859, 144705, 64911
PF04616	Glycosyl hydrolase family 43	4	67426, 145561, 72917, 270
PF00295	Glycosyl hydrolase family 28	4	73167, 67737, 57538, 77034
PF00331	Glycosyl hydrolase family 10	3	149157, 64703, 46317

Appendix C. Comparative analysis of the number of CAZy families related to plant polysaccharide degradation in *A. muscaria*, *A. thiersii* and other basidiomycetes. Largest number of proteins per class is shown in bold.

Cellulose								
Cazy family	**GH 1**	**GH 3**	**GH 5**	**GH 6**	**GH 7**	**GH 12**	**GH 45**	**Total**
V. volvacea	**3**	**11**	17	0	**14**	2	**0**	**47**
C. cinerea	2	7	**36**	**5**	9	1	0	**60**
A. bisporus	1	6	18	0	1	2	0	**28**
S. commune	**3**	10	16	0	3	1	0	**33**
P. chrysosporium	2	10	15	1	10	2	0	**40**
S. lacrymans	**3**	9	15	0	3	2	0	**32**
P. placenta	2	7	20	0	0	2	0	**31**
A. muscaria	0	4	13	0	0	0	0	**17**
A. thiersii	**3**	9	20	0	2	**3**	**2**	**39**

Xyloglucan												
Cazy family	**GH 2**	**GH 12**	**GH 27**	**GH 29**	**GH 31**	**GH 35**	**GH 36**	**GH 51**	**GH 54**	**GH 74**	**GH 95**	**Total**
V. volvacea	2	2	1	0	6	**4**	0	**3**	0	1	**2**	**21**
C. cinerea	2	1	0	0	3	0	0	1	0	1	0	**8**
A. bisporus	2	2	3	1	6	1	0	1	0	1	1	**18**
S. commune	4	1	0	**2**	4	**4**	0	2	0	1	1	**19**
P. chrysosporium	2	2	2	0	6	3	0	2	0	**4**	1	**22**
S. lacrymans	3	2	2	1	5	3	0	1	0	1	1	**19**
P. placenta	**5**	2	2	0	**7**	3	0	2	0	0	**2**	**23**
A. muscaria	4	0	2	0	5	1	0	0	0	0	1	**13**
A. thiersii	2	**3**	**4**	1	5	**4**	0	1	0	1	1	**22**

Xylan													
Cazy family	**CE 1**	**GH 2**	**GH 3**	**GH 10**	**GH 11**	**GH 27**	**GH 35**	**GH 36**	**GH 43**	**GH 51**	**GH 67**	**GH115**	**Total**
V. volvacea	5	2	**11**	**19**	0	1	**4**	0	8	**3**	0	**3**	**56**
C. cinerea	5	2	7	6	**6**	0	0	0	4	1	0	1	**32**
A. bisporus	2	2	6	2	2	3	1	0	3	1	0	2	**24**
S. commune	**11**	4	10	5	1	0	**4**	0	**12**	2	0	2	**51**
P. chrysosporium	5	2	10	6	1	2	3	0	4	2	0	1	**36**
S. lacrymans	1	3	9	1	0	2	3	0	2	1	0	1	**23**

P. placenta	0	**5**	7	4	0	2	3	0	0	2	0	0	**23**
A. muscaria	**1***	4	4	0	0	2	1	0	3	0	0	0	**15**
A. thiersii	1	2	9	1	0	**4**	**4**	0	5	1	0	**3**	**30**

Galactomannan							
Cazy family	**GH 2**	**GH 5**	**GH 26**	**GH 27**	**GH 35**	**GH 36**	**Total**
V. volvacea	2	17	0	1	**4**	0	**24**
C. cinerea	2	**36**	0	0	0	0	**38**
A. bisporus	2	18	0	3	1	0	**24**
S. commune	4	16	0	0	**4**	0	**24**
P. chrysosporium	2	15	0	2	3	0	**22**
S. lacrymans	3	15	0	2	3	0	**23**
P. placenta	**5**	20	0	2	3	0	**30**
A. muscaria	4	13	0	2	1	0	**20**
A. thiersii	2	20	0	**4**	**4**	0	**30**

Pectin																		
Cazy family	**CE 1**	**CE 8**	**CE 12**	**GH 2**	**GH 3**	**GH 28**	**GH 35**	**GH 43**	**GH 51**	**GH 53**	**GH 54**	**GH 78**	**GH 88**	**GH 93**	**GH 105**	**PL 1**	**PL 4**	**Total**
V. volvacea	5	3	1	2	**11**	3	**4**	8	**3**	**1**	0	0	1	0	2	**11**	2	**57**
C. cinerea	5	0	1	2	7	3	0	4	1	**1**	0	0	1	0	0	1	2	**28**
A. bisporus	2	2	2	2	6	6	1	3	1	**1**	0	**4**	1	0	2	1	1	**35**
S. commune	**11**	2	2	4	10	3	**4**	**12**	2	**1**	0	3	1	0	2	4	**3**	**64**
P. chrysosporium	5	2	0	2	10	4	3	4	2	**1**	0	1	1	0	0	0	0	**35**
S. lacrymans	1	2	0	3	9	7	3	2	1	**1**	0	3	1	0	0	0	0	**33**
P. placenta	0	**4**	0	**5**	7	**8**	3	0	2	0	0	1	1	0	0	0	0	**31**
A. muscaria	**1***	0	0	4	4	0	1	3	0	0	0	0	0	0	0	0	0	**13**
A. thiersii	1	2	3	2	9	5	**4**	6	1	**1**	0	3	0	0	2	1	2	**42**

Starch				
Cazy family	**GH 13**	**GH 15**	**GH 31**	**Total**
V. volvacea	7	**5**	6	**18**
C. cinerea	7	4	3	**14**
A. bisporus	6	2	6	**14**
S. commune	**9**	3	4	**16**
P. chrysosporium	7	3	6	**16**
S. lacrymans	5	3	5	**13**
P. placenta	2	2	**7**	**11**

A. muscaria	5	2	5	**12**
A. thiersii	8	2	5	**15**

β-1,3-1,4-glucan						
Cazy family	**GH 5**	**GH 7**	**GH 12**	**GH 16**	**GH 17**	**Total**
V. volvacea	17	**14**	2	21	3	**57**
C. cinerea	**36**	9	1	26	3	**75**
A. bisporus	18	1	2	21	4	**46**
S. commune	16	3	1	**30**	2	**52**
P. chrysosporium	15	10	2	19	1	**47**
S. lacrymans	15	3	2	17	1	**38**
P. placenta	20	0	2	20	0	**42**
A. muscaria	13	0	0	19	4	**36**
A. thiersii	20	2	**3**	20	**8**	**53**

Appendix D. Gene families in *Amanita*.

Protein family ID	Number proteins *A. muscaria*	Number proteins *A. thiersii*	PFAM Ids	PFAM Description (number of proteins)
1	10	21	PF01185	Hydrophobin(31)
2	24	0		
3	8	15	PF00135	Coesterase(23)
4	7	12	PF07731, PF00394, PF07732	Cu-oxidase2(17), Cu-oxidase(17), Cu-oxidase3(19)
5	9	7	PF00026	Asparteate protease(16)
6	13	2		
7	7	7	PF00450	Serine carboxipeptidase S10(14)
8	1	12	PF03443, PF00734	GH61(13), CBM(1)
9	7	6	PF00722,	GH16(1)
10	11	1	PF01565, PF08031	FAD_binding domain 4(12), Berberine and Berberine-like(10)
11	6	6	PF03981	Ubiquinol-cytochrome C chaperone(1)
12	12	0		
13	10	1	PF00026	Asparteate protease(11)
14	4	6	PF01764	Lipase 3(10)
15	10	0	PF01828	Peptidase A4(9)
16	0	10	PF05572	Peptidase M43(5)
17	6	3		
18	5	4	PF01565, PF08031	FAD_binding domain 4(9), Berberine and Berberine-like(9)
19	4	5	PF01565, PF08031	FAD_binding domain 4(9), Berberine and Berberine-like(5)
20	2	7	PF00082, PF05922	Peptidase S8(9), Subtilisin N(9)
21	0	9	PF02102	Peptidase M35(9)
22	5	4	PF00652	Ricin B lectin(7)
23	2	7		
24	5	3	PF00704	GH 18(8)
25	6	1		
26	2	5	PF03330	Rare lipoprotein A (RlpA)-like double-psi beta-barrel (DPBB, 5)
27	2	5	PF00082, PF02225	Peptidase S8(7), Protease Associated(6)
28	2	5	PF02128, PF07504	Peptidase M36(7), Fungalysin/Thermolysin Propeptide Motif(2)
29	3	4	PF00704, PF02839	GH18(7), CBM5 & 12(4)
30	0	7	PF02012, PF05064, PF01341, PF00331, PF00734	Bacterial neuraminidase(1), Nsp1-like(1), GH6(1), GH10(3), CBM1(5)
31	3	4	PF01183	GH25(7)
32	6	0		
33	0	6		
34	2	4		
35	2	4	PF01522	Polysaccharide deacetylase 1(5)
36	0	6	PF00144	Beta-lactamase(6)

37	3	3		
38	4	1		
39	5	0		
40	3	2	PF00067	p450(5)
41	0	5		
42	0	5		
43	1	4		
44	0	5		
45	0	5	PF01565	FAD binding 4(5)
46	3	2	PF00657	Lipase GDSL(2)
47	3	2	PF01522	Polysaccharide deacetylase 1(5)
48	3	2	PF01764	Lipase 3(3)
49	1	4	PF02065, PF03422	Melibiase (4), CBM6(1)
50	1	4	PF00141	Peroxidase(1)
51	2	3	PF05577,	Peptidase S28(5)
52	2	3	PF07249	Cerato-platanin(5)
53	0	4		
54	2	2		
55	2	2		
56	0	4	PF07142	Repeat of unknown function DUF1388(1)
57	3	1		
58	4	0		
59	4	0		
60	1	3	PF07250, PF09118	Glyoxal oxidase N-terminus(4), DUF1929(4)
61	0	4	PF00295	GH28(4)
62	2	2	PF00149	Metallophosphoesterase(3)
63	2	2	PF00728, PF02838	GH20(4), GH20b(3)
64	1	3	PF02806, PF00686, PF00723, PF00128	Alpha-amylase C(1), CBM20(3), GH15(2), Alpha-amylase(2)
65	0	4		
66	2	2	PF00150	Cellulase(2)
67	1	3	PF01915, PF00933	GH3C(4), GH3(4)
68	2	2	PF04616	GH43(4)
69	2	2		
70	0	4		
71	2	2	PF08531	Alpha-L-rhamnosidase N-terminal domain(2)
72	2	2	PF00082, PF09286	Peptidase S8(2), Pro-kumamolisin activation domain(4)
73	2	2		
74	3	1	PF01494	FAD-binding 3(4)
75	1	3		
76	1	3	PF00188	Cysteine-rich secretory protein & pathogenesis-related 1 proteins (SCP, 4)
77	4	0	PF04185	Phosphoesterase(4)

78	4	0		
79	3	0		
80	0	3	PF04389	Peptidase M28(3)
81	0	3	PF04389, PF02225	Peptidase M28(3), Protease Associated(3)
82	0	3		
83	1	2	PF05730	CFEM(1)
84	1	2	PF00732, PF05199	Glucose-methanol-choline oxidoreductase N(GMCO, 3), GMCO C(3)
85	0	3		
86	1	2	PF01425	Amidase(3)
87	1	2		
88	2	1		
89	1	2	PF03330	DPBB1(1)
90	0	3	PF00959	Phage lysozyme(3)
91	2	1		
92	0	3		
93	1	2	PF07250, PF01822, PF09118	Glyoxal oxid N(2), WSC(3), DUF1929(2)
94	1	2		
95	2	1		
96	0	3	PF07971	GH92(3)
97	2	1	PF00445	Ribonuclease T2(3)
98	1	2	PF00150	Cellulase(2)
99	0	3	PF01670	GH12(3)
100	1	2	PF01161	Phosphatidylethanolamine-binding protein (PBP, 3)
101	1	2		
102	1	2	PF00150	Cellulase(1)
103	2	1	PF01532	GH47(3)
104	1	2	PF01055	GH31(3)
105	2	1	PF00149	Metallophosphoesterase(3)
106	0	3	PF03572	Peptidase S41(2)
107	1	2		
108	2	1	PF02265	S1-P1 nuclease(3)
109	3	0		
110	2	1		
111	2	1		
112	3	0		
113	1	1	PF04389	Peptidase M28(2)
114	1	1	PF04389	Peptidase M28(2)
115	1	1		
116	1	1		
117	1	1		
118	2	0		

119	0	2		
120	0	2		
121	0	2		
122	0	2		
123	0	2		
124	0	2	PF01095	Pectinesterase(2)
125	1	1	PF00646	F-box(1)
126	1	1		
127	0	2	PF00657	Lipase GDSL(2)
128	1	1	PF03009	GDPD(2)
129	0	2	PF00332	GH17(2)
130	0	2		
131	0	2	PF02102	Peptidase M35(2)
132	1	1	PF00775	Dioxygenase C(2)
133	0	2		
134	1	1	PF00190, PF07883	Cupin 1(1), Cupin 2(2)
135	1	1	PF01476	LysM(2)
136	1	1	PF03198, PF07983	GH72(2), X8(2)
137	1	1	PF00722	GH16(2)
138	1	1	PF02055	GH30(2)
139	0	2	PF01083	Cutinase(2)
140	0	2		
141	1	1	PF07971	GH92(2)
142	1	1	PF01532, PF02225	GH47(2), PA(1)
143	0	2	PF01764	Lipase 3(2)
144	1	1	PF08450	SGL(2)
145	0	2	PF05592	GH78(2)
146	1	1		
147	1	1		
148	0	2	PF01975	Survival protein SurE(2)
149	1	1	PF02221	MD-2-related lipid-recognition (ML) domain(2)
150	1	1	PF07992, PF07156	Pyridine nucleotide-disulphide oxidoreductase 2(2), Prenylcysteine lyase(2)
151	1	1		
152	1	1		
153	1	1		
154	0	2		
155	1	1	PF07749, PF00085	Endoplasmic reticulum protein ERp29(2), Thioredoxin(2)
156	1	1		
157	1	1	PF02585	GlcNAc-PI de-N-acetylase(2)
158	1	1	PF07719, PF00515, PF00226	TPR 2(1), TPR 1(2), DnaJ(2)
159	1	1	PF02089	Palm thioest(2)

160	1	1	PF02265	S1-P1nuclease(2)
161	1	1	PF03200	GH63(2)
162	1	1		
163	1	1	PF01464	SLT(2)
164	1	1	PF06824	DUF1237(2)
165	0	2	PF03663	GH76(1)
166	0	2	PF00328	Acid phosphat A(2)
167	1	1		
168	1	1	PF02278	Lyase 8(2)
169	2	0		
170	1	1	PF00160	Pro isomerase(2)
171	0	2	PF00098	Zinc finger-CCHC(1)
172	1	1	PF00025	ADP ribosylation factor(2)
173	1	1	PF00644	Poly(ADP-ribose) polymerase catalytic domain (PARP, 2)
174	2	0	PF00314	Thaumatin(2)
175	1	1		
176	1	1		
177	1	1		
178	1	1	PF08760	DUF1793(2)
179	2	0		
180	2	0		
181	2	0		
182	2	0		
183	2	0		
184	2	0	PF00561	Abhydrolase 1(1)