Rapid Divergence of Genome Architectures Following the Origin of an Ectomycorrhizal Symbiosis in the Genus *Amanita*

Jaqueline Hess,*[,1,2] Inger Skrede,[2] Maryam Chaib De Mares,[3] Matthieu Hainaut,[4,5] Bernard Henrissat,[4,5,6] and Anne Pringle[7]

[1]Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

[2]Section for Genetics and Evolutionary Biology, University of Oslo, Oslo, Norway

[3]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

[4]Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, Aix-Marseille University, Marseille, France

[5]INRA, USC1408 AFMB, Marseille, France

[6]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

[7]Departments of Botany and Bacteriology, University of Wisconsin, Madison, Madison, WI

***Corresponding author:** E-mail: jaqueline.hess@univie.ac.at.

Associate editor: Jun Gojobori

Abstract

Fungi are evolutionary shape shifters and adapt quickly to new environments. Ectomycorrhizal (EM) symbioses are mutualistic associations between fungi and plants and have evolved repeatedly and independently across the fungal tree of life, suggesting lineages frequently reconfigure genome content to take advantage of open ecological niches. To date analyses of genomic mechanisms facilitating EM symbioses have involved comparisons of distantly related species, but here, we use the genomes of three EM and two asymbiotic (AS) fungi from the genus *Amanita* as well as an AS outgroup to study genome evolution following a single origin of symbiosis. Our aim was to identify the defining features of EM genomes, but our analyses suggest no clear differentiation of genome size, gene repertoire size, or transposable element content between EM and AS species. Phylogenetic inference of gene gains and losses suggests the transition to symbiosis was dominated by the loss of plant cell wall decomposition genes, a confirmation of previous findings. However, the same dynamic defines the AS species *A. inopinata*, suggesting loss is not strictly associated with origin of symbiosis. Gene expansions in the common ancestor of EM *Amanita* were modest, but lineage specific and large gene family expansions are found in two of the three EM extant species. Even closely related EM genomes appear to share few common features. The genetic toolkit required for symbiosis appears already encoded in the genomes of saprotrophic species, and this dynamic may explain the pervasive, recurrent evolution of ectomycorrhizal associations.

Key words: **gene gain, gene loss, mutualism, evolution of symbiosis, ectosymbiosis, convergent evolution, genome architecture, phylogenomics, oxidative metabolism, gene regulation.**

Article

Introduction

Symbioses, or close physical associations between organisms of two or more interacting species, are ubiquitous. Examples are found across all domains of life and include the gut microbiomes of humans and other vertebrates (Ley et al. 2008), interactions of reef-building corals and photosynthetic algae (Baker 2003) and associations between plants and ants (Bronstein et al. 2006). Despite the ubiquity of symbioses, our knowledge of the evolutionary mechanisms mediating the origin of symbiosis and the potential impacts of symbiosis on genome architectures is limited. Many symbioses have evolved independently and repeatedly across the tree of life, for example, there are multiple origins of associations between nitrogen fixing bacteria and plants (Masson-Boivin et al. 2009). Often, these independently evolved symbioses encompass structurally or functionally similar interactions (Bittleston et al. 2016). But whether convergently evolved symbioses are shaped by similar evolutionary processes remains an open question.

Most insights so far have come from the analysis of prokaryotic endosymbionts of insects. Evolution of symbiosis in convergently evolved bacterial endosymbionts is characterized by an initial period of genome expansion and repetitive element proliferation, followed by extreme genome reduction, gene loss, and the purging of mobile DNA (Moran and Plague 2004; Oakeson et al. 2014; Wernegreen 2015). Bacterial endosymbionts typically also show increased rates of molecular evolution (Moran 1996; Wernegreen et al. 2001). These genome changes are likely mediated by a reduced effective population size (N_e) caused by the strict vertical transmission of the endosymbionts, resulting in repeated bottlenecks and restriction of genetic exchange with other populations (Moran 1996; O'Fallon 2008; Newton and

Bordenstein 2011). Similar processes appear to be driving genome evolution in bacterial endosymbionts of ciliates (Boscaro et al. 2017), but relatively little is known about common evolutionary mechanisms driving other types of symbioses (but see Lutzoni and Pagel 1997; Rubin and Moreau 2016).

The ectomycorrhizal (EM) symbioses of fungi and plants offer a textbook example of convergent evolution and a tractable system to study the impact of symbiosis on eukaryotic genomes. Unlike endosymbiotic bacteria, ectosymbiotic EM fungi interact simultaneously with plants and the external environment; vast portions of the fungal body forage away from plant roots and must cope with the diverse array of antagonists found in soil (Smith and Read 2008). The obligate mutualism has evolved independently and repeatedly at least 11 times among the Agaricales (the gilled mushrooms; Matheny et al. 2006) with multiple, additional origins in other clades of the phyla Ascomycota and Basidiomycota (Tedersoo et al. 2010). In each case, symbiosis evolved from saprotrophy. The hallmark of an EM association is the reciprocal exchange of nutrients—photosynthetically derived sugars from the plant in exchange for nitrogen, phosphorous, or other compounds from the fungus (Smith and Read 2008). Horizontally transmitted fungal symbionts grow around the lateral roots of plants, predominately trees, ensheath root tips and within each root tip project a hyphal network (the Hartig net) into the plant's apoplastic space. Root tips are the site of nutrient exchange. Formation of a functional symbiosis relies on intricate signaling and communication between fungus and plant (Duplessis et al. 2005; Plett et al. 2014, 2015; Tschaplinski et al. 2014; Liao et al. 2016; Doré et al. 2017).

Early genomic analyses of EM fungi focused on two species, *Laccaria bicolor* (Martin et al. 2008) and *Tuber melanosporum* (Martin et al. 2010), but recently a broader sampling of basidiomycete and ascomycete EM fungi derived from independent origins was analyzed (Kohler et al. 2015; Peter et al. 2016). Comparisons of these genomes reveal common features. The first is a trend toward expanded gene repertoires: EM fungi generally house between 15,000 and 23,000 genes (Kohler and Martin 2016), compared with the 10,000–15,000 genes common in wood decaying, saprotrophic, basidiomycetes (Riley et al. 2014). EM species also appear to have lost most enzymes involved in the plant cell wall (PCW) decomposition (Nagendran et al. 2009; Wolfe, Kuo, et al. 2012; Kohler et al. 2015; Martin et al. 2016). In particular, enzymes involved in the breakdown of lignin and crystalline cellulose appear largely absent from EM genomes (Kohler et al. 2015). Genes upregulated in mycorrhizal root tips are typically enriched for effector-like small secreted proteins (SSPs), suggesting SSPs are used to communicate with the plant (Liao et al. 2014, 2016; Kohler et al. 2015; Doré et al. 2017).

Despite global similarities among distantly related EM genomes, between 7% and 38% of symbiosis upregulated genes are lineage-specific orphan genes (Kohler et al. 2015). Within the general pattern of loss of PCW decomposition enzymes there is a more specific dynamic; different EM fungi retain distinct sets of PCW decomposition enzymes, often reflecting the ecology and degradation abilities of the last saprotrophic ancestor (Kohler et al. 2015). Moreover, neither gene repertoire expansion nor the presence of SSPs define an EM genome, as highlighted by the ascomycete *T. melanosporum*: the genome of this species encodes a mere 7,496 genes and houses no mycorrhizae-induced SSPs (Martin et al. 2010).

Current knowledge suggests distantly related EM fungi took distinct evolutionary trajectories during the transition from an asymbiotic (AS) to an EM niche (Kohler et al. 2015), but to date, comparisons have been limited to contrasts between pairs of a derived EM fungus and its closest saprotrophic ancestor, or among EM fungi evolved from distant origins of symbiosis. These approaches identify common features among distantly related EM fungi, but to more precisely establish the evolutionary events defining the origin of EM symbiosis, and to distinguish these from lineage-specific evolutionary changes, comparisons of closely related EM fungi, descended from a single origin of symbiosis, are required.

The Genus *Amanita* as a Model

The genus *Amanita* houses ~500 species, most of which form ectomycorrhizal mutualisms with plants (Bas 1969; Wolfe, Tulloss, et al. 2012). Within *Amanita*, there is a single origin of EM symbiosis, dated to ~50–100 Ma (Wolfe, Tulloss, et al. 2012; Kohler et al. 2015). The *Amanita* species basal to the EM clade, and closely related genera (e.g., *Limacella*, *Pluteus*, and *Volvariella*), are saprotrophs (Wolfe, Kuo, et al. 2012; Wolfe, Tulloss, et al. 2012). Our analyses focus on the EM species *Amanita brunnescens*, *Amanita polypyramis*, and *Amanita muscaria* var. *guessowii* (from this point forward discussed as "*A. muscaria*"), the AS species *Amanita inopinata* and *Amanita thiersii*, and the AS outgroup *Volvariella volvaceae*. The natural history of each species is described more fully by (Hess and Pringle 2014). We have previously shown transposable element (TE) expansions in two out of three of these EM species (but also the asymbiotic *A. thiersii*), suggesting that even within this single genus, distinct evolutionary processes may be influencing the architecture of different genomes (Hess et al. 2014). Symbiotic genomes have also been shaped by at least one horizontal gene transfer event (Chaib De Mares et al. 2015).

We use genome-wide phylogenomic analysis to reconstruct the evolutionary histories of the full complement of protein coding genes across the five *Amanita* and *V. volvacea* genomes. We aimed to i) understand if EM *Amanita* share common genome architectures, and if these bear similarities to genome architectures of other EM fungi and ii) reconstruct the early evolutionary events around the time of the origin of EM symbiosis. Our analyses reveal a dramatic divergence among genome architectures following the origin of symbiosis in the genus. For example, extant EM *Amanita* genomes show a 2-fold difference in numbers of protein coding genes, mirroring the extreme differences found between the EM fungi *L. bicolor* (a basidiomycete) and *T. melanosporum* (an ascomycete), species which diverged over 635 Ma (Kohler et al. 2015). Our data suggest the ancestral EM *Amanita* genome was shaped primarily by gene loss; in particular the loss of PCW decomposition enzymes. Large scale gene amplifications emerge as features of some (but not all) EM lineages.

A

Species	Ecology	Assembly Size	Genome Size (k-mer)	# Scaffolds	N50 (KB)
Amanita brunnescens	EM	35,996,348	262,868,678	5,977	13
Amanita polypyramis	EM	19,668,116	144,950,695	999	64
*Amanita muscaria guess.**	EM	40,699,759	44,318,520	1,101	17
*Amanita thiersii**	AS	33,689,220	48,743,552	1,446	77
Amanita inopinata	AS	20,100,599	29,014,546	896	156
Volvariella volvacea	AS	35,328,295	39,945,087	1,523	84

*sequenced and assembled at JGI

B

Species	Ecology	# Predicted Genes	# Pred. Genes (filtered)	Complete (BUSCO) %	Duplicated (BUSCO) %
Amanita brunnescens	EM	14,532	11,853	97.7 (10.6)	10.8
Amanita polypyramis	EM	7,969	7,683	98.4 (8.4)	8.6
Amanita muscaria guess.	EM	13,854	12,363	97.6 (7.5)	10.6
Amanita thiersii	AS	11,935	8,982	97.4 (10.5)	8.6
Amanita inopinata	AS	7,569	7,303	95.7 (8.9)	8.6
Volvariella volvacea	AS	11,061	10,482	95.3 (7.9)	9.4

C

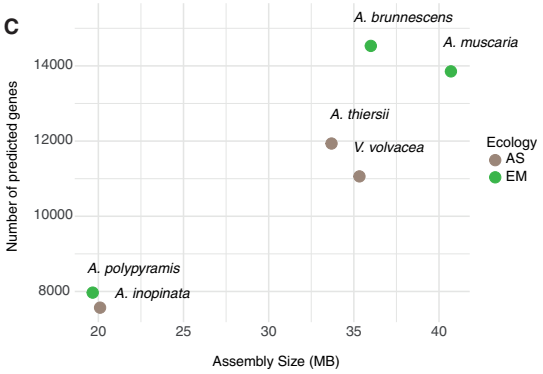


FIG. 1. Genome statistics. In all three panels, EM refers to ectomycorrhizal, AS to asymbiotic. (*A*) Assembly statistics. (*B*) Gene prediction statistics. (*C*) Assembly and annotation statistics according to ecology.

In the aggregate, the data suggest adaptation to the EM niche following the origin of symbiosis was mediated by independent evolutionary trajectories, even within this single genus. Nevertheless, the gene family expansions found in the most recent common ancestor of EM *Amanita* reveal candidate gene families that may have facilitated the multiple, convergent emergences of the EM symbiosis across the fungal tree of life.

Results

The Heterogeneous Genome Architectures of EM *Amanita* Reveal No Single Feature That Distinguishes EM from As Genomes

Comparisons of genome sizes and the numbers of protein coding genes among our *Amanita* genomes suggest no single pattern distinguishes EM from AS species (fig. 1). Finalized genome assemblies range from ~20 to 40 MB, although assemblies are likely to be underestimates of true genome sizes due to collapsed repeats (Alkan et al. 2011; Hess et al. 2014). Genome size estimates based on the k-mer content of unassembled read data also point to a large range in genome sizes and estimated genome sizes of the EM *A. brunnescens* and *A. polypyramis* are considerably larger than the estimate for the EM *A. muscaria* or any AS species, at 263 MB (36 MB assembly size) and 145 MB (19.7 MB assembly size), respectively. The genome of *A. muscaria* (44.3 MB) is not markedly different in size from genomes of the AS species *A. thiersii* and *V. volvacea* (48.7 and 39.9 MB, respectively) although larger than that of *A. inopinata* which we estimated to be ~29 MB in size. Our results are largely explained by our earlier study; large amounts of unassembled TEs are found in the *A. brunnescens* and *A. polypyramis* genomes (Hess et al. 2014). These TEs are reflected in the genome size estimates based on k-mer content, but are not reflected in estimates based on genome assemblies.

A pattern analogous to the genome sizes emerged when comparing gene content among EM and AS genomes. The number of annotated protein-coding genes varies by a factor of two between the different species, ranging from ~7,569 genes in the AS *Amanita inopinata* to up to 14,532 in the EM *A. brunnescens*. In general, smaller assembled genomes house fewer genes (fig. 1C). We found an expanded gene repertoire in *A. brunnescens* and *A. muscaria*, while *A. polypyramis* houses a particularly small number of genes compared with both its EM and AS relatives. *A. inopinata* also appears to have relatively few genes, and particularly when compared with the average Basidiomycete genome size of 10,000–13,000 genes (Riley et al. 2014). When considering only high confidence genes, that is, after removing genes that overlap with annotated TEs (which constitute the majority of filtered genes, between 60% in *V. volvacea* and 99% in *A. thiersii*), or after removing proteins shorter than 200 amino acids with no homology with other known proteins or conserved domains, numbers of annotated genes are slightly reduced. Nevertheless, the dramatic range in gene repertoire sizes among *Amanita* genomes remains, spanning from 7,303 genes in *A. inopinata* to a maximum of 12,363 in the EM *A. muscaria*. No clear signal distinguishes EM from AS genomes.

Because we are interested in the discovery of genes relevant to the EM niche, and because these genes typically include poorly conserved SSPs (Kohler et al. 2015), we decided to retain the full set of predicted genes for downstream analyses, but apply specific filters for individual analyses as described. Note that our gene prediction pipelines resulted in different estimates of the numbers of genes compared with pipelines used by the JGI; we discuss this phenomenon in more detail in the supplementary text (supplementary text, supplementary figs. S1 and S2, Supplementary Material online). Discrepancies do not affect results as we used our pipelines for all genomes, including JGI genomes. We simultaneously assessed the completeness and redundancy of our genome assemblies and gene predictions using BUSCO (Simão et al. 2015), which scores the presence and absence of known single copy orthologs in predicted proteomes. Overall, we recovered between 95% and 98% of BUSCO's 1,438 fungal

genes in our own genomes; the genomes appear to capture most gene space. The proportions of genes only recovered as fragments range from 8.5% to 10.6% and were highest in the *A. brunnescens* and *A. thiersii* genomes. Between 8.6% and 10.8% of BUSCO genes were duplicated in each proteome. Duplicated BUSCOs may suggest redundancy in the genome assembly caused by the sequencing of diploid cultures. However, the *A. thiersii* genome was sequenced from a haploid strain, and we take the 8.6% redundancy estimated for *A. thiersii* as an indicator that factors other than assembly redundancy may cause elevated rates of duplicated BUSCO genes in these species. But because the *A. polypyramis* genome appears very different from other *Amanita* species, we performed additional tests of the quality of its assembly and annotation; in fact, there is no reason to believe either assembly or annotation are problematic (supplementary text, Supplementary Material online).

The Loss of PCW Degrading Enzymes Is Not Strictly Associated with the EM Niche

Current literature identifies the loss of PCW degrading enzymes as a potentially defining feature of EM genomes (Nagendran et al. 2009; Wolfe, Kuo, et al. 2012; Kohler et al. 2015; Martin et al. 2016). To catalogue the depth and breadth of PCW degrading enzyme reductions in EM *Amanita*, we annotated the full complement of carbohydrate active enzymes in genomes using the CAZy database (Lombard et al. 2014). A targeted study of an endoglucanase (GH5_5) and a cellobiohydrolase (GH7) among a broader swath of *Amanita* species (108 species) has already demonstrated that the loss of these two key extracellular cellulases is associated with emergence of EM symbiosis within the genus *Amanita*: they are lacking in all EM *Amanita* sampled to date, while a beta-glucosidase (GH3) is lost in most EM *Amanita* (Wolfe, Tulloss, et al. 2012). Our data extend analyses to the full set of gene families involved in the enzymatic degradation of lignocellulose.

EM *Amanita* appear to have lost the full complement of core hydrolytic CAZymes ("Core CAZymes") acting on cellulose and hemicellulose. Figure 2 lists the PCW degrading enzyme families involved in the breakdown of the plant cell wall among Agaricales (Floudas et al. 2015). With the exception of one GH12 in *A. brunnescens* and the one GH45 in *A. brunnescens* and *A. polypyramis*, as well as a GH9 conserved across all species, core CAZymes are absent. The GH9, GH12 and GH45 are families harboring endoglucanases capable of digesting cellulose into cellooligosaccharides (Rytioja et al. 2014). The near complete loss of core CAZymes is underlined by the complete loss of CBM1 binding modules, which are often attached to key PCW degrading enzymes and mediate the targeting of enzymes to cellulose (Várnai et al. 2014).

By contrast, all EM species retain at least one lytic polysaccharide monooxygenase (LPMOs, AA9), and several laccases and laccase-like enzymes (e.g., AA1). LPMOs are small enzymes that facilitate access to densely packed cellulose substrates by oxidatively modifying crystalline cellulose (Horn et al. 2012), but LPMOs can also act on hemicellulose (Bennati-Granier et al. 2015) as well as chitin (Sabbadin et al. 2018). Laccases are primarily associated with lignolytic activity among wood decay fungi, but are associated with diverse other substrates as well (Kües and Ruhl 2011). Laccases can catalyze the oxidation of phenolic compounds and have been implicated in numerous biological processes beyond wood decay, including the metabolism of plant defense compounds in interactions between fungi and plants and developmental processes (Kües and Ruhl 2011). *Amanita muscaria* appears to encode an expanded set of laccases, even compared with the AS species. A similar amplification of laccases has also been found in the EM fungus *L. bicolor*, and in this fungus a number of laccases are induced in mycorrhizal root tips (Courty et al. 2009).

Genes encoding enzymes mediating the degradation of hemicellulose and pectin ("Accessory CAzymes") are also reduced or completely lost among EM *Amanita*. Many other genera of EM fungi are characterized by widespread reductions in the numbers of PCW degrading enzymes (Wolfe, Tulloss, et al. 2012; Kohler et al. 2015), and our data corroborate these findings. But patterns of losses within *Amanita* imply that broad scale losses are not strictly associated with the EM niche. *Amanita inopinata* closely resembles the EM *Amanita* with respect to its set of PCW degrading enzymes, although this species is not associated with EM hosts, does not form EM structures in nature or in culture, and appears nested within a clade of litter degraders (Wolfe, Tulloss, et al. 2012; Hess and Pringle 2014). While losses and reductions of PCW-degrading enzymes define the EM *Amanita*, an EM *Amanita* cannot be defined by the absence of PCW degrading enzymes.

Phylogenomic Analyses of Gene Families Support Independent, Parallel Amplifications of Gene Content in *A. brunnescens* and *A. muscaria*

To test a hypothesis that observed distributions in gene content (fig. 1) result from either: i) large-scale gene duplication in the EM ancestor, followed by genome erosion in *A. polypyramis* or ii) the independent amplifications of gene families in *A. brunnescens* and *A. muscaria*, we reconstructed the genome-wide evolutionary histories of all genes found in every genome. We sought to understand how gene content has evolved across and after the transition from the AS to EM niche.

To infer homology among genes across the six genomes, we first clustered their predicted protein sequences into orthologous groups, each containing at least two genes (orthogroups, but excluding singletons unique to one species); we now describe general results from clustering into orthogroups. Clustering resulted in a total of 10,386 orthogroups, of which 513 were identified as putative TE clusters. Among the non-TE clusters, 4,852 had members in each of the *Amanita* genomes, 1,781 appear lineage-specific (supplementary fig. S3A, Supplementary Material online). Transcriptomes expressed in EM root tips tend to be enriched for lineage-specific genes (Kohler et al. 2015), and in our analyses, we also found elevated numbers of lineage-specific

Species	*Amanita brunnescens*	*Amanita polypyramis*	*Amanita muscaria*	*Amanita thiersii*	*Amanita inopinata*	*Volvariella volvacea*
Ecology	EM	EM	EM	AS	AS	AS
Oxidate enzymes						
POD (AA2)	0	0	0	0	0	7
MCO (AA1)						
Lac (AA1_1)	6	7	19	14	4	11
Fer (AA1_2)	1	1	1	1	1	0
Lac /Fer (AA1)	0	0	0	3	0	0
CRO (AA5)	0	1	0	0	0	0
CBH (AA3_1)	0	0	0	1	0	1
LPMO (AA9)	1	2	2	16	2	32
Core CAZymes						
GH6	0	0	0	1	0	6
GH7	0	0	0	2	0	15
GH9	1	1	1	1	1	1
GH10	0	0	0	4	0	19
GH12	1	0	0	3	0	2
GH5_5	0	0	0	3	0	1
GH5_7	0	0	0	1	0	2
GH28	0	0	0	7	0	3
GH74	0	0	0	1	0	1
GH44	0	0	0	1	0	1
GH45	1	1	0	3	1	0
GH53	0	0	0	1	0	1
Accessory CAZymes						
GH1	0	0	0	5	0	3
GH2	1	1	4	2	1	2
GH3	2	3	4	8	5	11
GH5_22	0	1	1	2	0	2
GH43	0	0	2	7	0	16
GH51	1	0	0	1	0	3
GH78	0	0	0	6	0	0
GH88	2	1	0	1	1	1
GH105	0	0	0	2	0	2
GH115	0	0	0	3	0	3
GH131	1	0	0	1	0	2
GH145	0	0	0	1	0	1
CE1	0	0	0	1	0	4
CE5	0	0	0	2	0	1
CE8	0	1	0	2	0	3
CE12	0	0	0	3	0	1
CE15	0	0	0	1	0	1
Carbohydrate binding modules						
CBM1	0	0	0	8	0	45
Number of distinct CAZy fams associated with CBM1	0	0	0	7	0	15
Total PCWD enzymes	18	20	34	111	16	159
Proportion of CAZome involved in PCWD	11.83	17.61	24.50	33.89	12.68	40.37

Fig. 2. Carbohydrate active enzymes involved in decay of plant cell wall material. Enzymes were classified according to Floudas et al. (2015), categories include: enzymes involved in oxidative degradation, hydrolytic enzymes degrading the core components of lignocellulose (cellulose, hemicellulose, and pectin) and enzymes degrading side chains ("Accessory CAZymes").

clusters in two of the EM species, *A. brunnescens* and *A. muscaria*, but not in *A. polypyramis*. Genome organization into different size categories of gene families differs significantly among species even when excluding singletons (χ^2 = 1036.1, df = 20, *P* value <2.2e-16; supplementary fig. S3*B*, Supplementary Material online). The genomes of the two species with the smallest gene repertoires, *A. polypyramis and A. inopinata*, are dominated by single-copy genes with ~95% of genes found in single-copy orthogroups in these species. Examination of residuals (significant threshold 0.05, Bonferroni corrected) suggest that *A. brunnescens* and *A. muscaria* house significantly fewer single-copy orthogroups than expected by chance (83% in *A. brunnescens* and 84% in *A. muscaria*) and larger than expected proportions of multigene families in all size classes (*A. muscaria*) and size categories "2" and "3–5" (*A. brunnescens*), suggesting plasticity in gene content is driven by amplifications of individual gene families, and not by polyploidization.

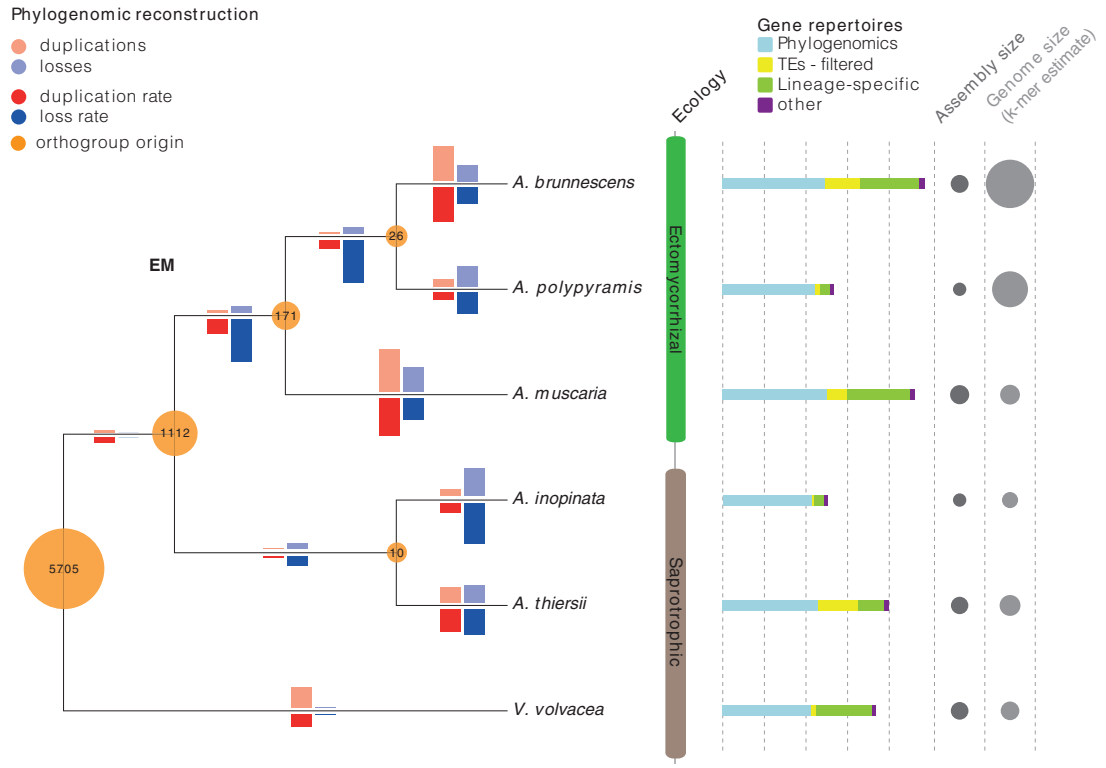


Fig. 3. Genome-wide phylogenomic reconstruction of gene duplications and losses. Bar charts highlight the numbers of duplication (red) and loss (blue) events inferred on each branch, either as total numbers (translucent, above branch) or scaled by the respective length of each branch (opaque, below branch). Circles mark the number of orthogroups rooted at various points of the phylogeny. The bar plot on the right ("Gene repertoires") highlights the proportions of protein coding genes included in phylogenomic analysis, compared with the proportions filtered from the data set for each genome, as well as reasons for filtering. Data on the tree were visualized using Evolview (He et al. 2016).

Phylogenomic inference of gene duplications and losses based on orthogroups with more than three members from at least two different species (7,024 total) suggests independent amplifications of gene repertoires in *A. brunnescens* and *A. muscaria* (light red bars, fig. 3). Only 53 gene duplications map to the MRCA of all EM *Amanita*, while both *A. brunnescens* and *A. muscaria* are characterized by extensive lineage-specific amplifications, with 833 and 1031 inferred duplications, respectively. The gene repertoire of the EM *A. polypyramis* provides a stark contrast. Duplications within *A. polypyramis* are outweighed by inferred losses (176 vs 508 gains and losses, respectively), resulting in an overall reduction in gene space; data emphasize the diversity of evolutionary trajectories taken by even closely related ectomycorrhizal genomes.

Early genomic changes accompanying the transition from the AS to EM niche also appear to be dominated by gene loss, rather than gain. When scaled by branch lengths, we inferred the highest rates of gene loss across the entire phylogeny on the branch marking the MRCA of all EM *Amanita*, as well as in the MRCA of *A. brunnescens* and *A. polypyramis* (dark blue bars, fig. 3). In total, we inferred 156 gene losses in the MRCA of all EM *Amanita*, distributed in 152 orthogroups. Of the 152 orthogroups experiencing gene loss, 120 went entirely extinct at this node. Nevertheless, despite the dynamic of loss, the overall rate of duplication within the MRCA of all EM *Amanita* is still relatively high, especially when compared with other internal branches (dark red bars, fig. 3). Increased rates of change within the MRCA are also suggested by the relatively large number of orthogroups rooted at the MRCA of all EM species: in total, we found 171 orthogroups unique to EM species.

The remarkable pattern of parallel gene space expansion in *A. brunnescens* and *A. muscaria* inspired us to ask whether the pattern is driven by amplification of the same orthogroups. Indeed, net patterns of gene family expansion (the cumulative number of duplications and inferred gene origins compared against gene losses) suggest significant parallel amplification of specific orthogroups in EM lineages (fig. 4A). Overall copy number (CN) increases of orthogroups expanded uniquely on a single branch were generally lower than expected when duplications and losses were placed among gene families at random (1000 permutations). The exception proved to be the branch representing the MRCA of EM *Amanita*, which had larger than expected CN increase specific to it, suggesting a unique signature of amplification on this branch. Larger than expected CN increases were also found among specific orthogroups amplified in parallel on several sets of branches, often involving parallel amplification in the EM ancestor and extant EM lineages, suggesting continued amplification of gene families expanded early after the transition to symbiosis. By contrast, orthogroups expanded in parallel only in *A. brunnescens* and *A. muscaria*, but no other EM branch, did not deviate from the random expectation.

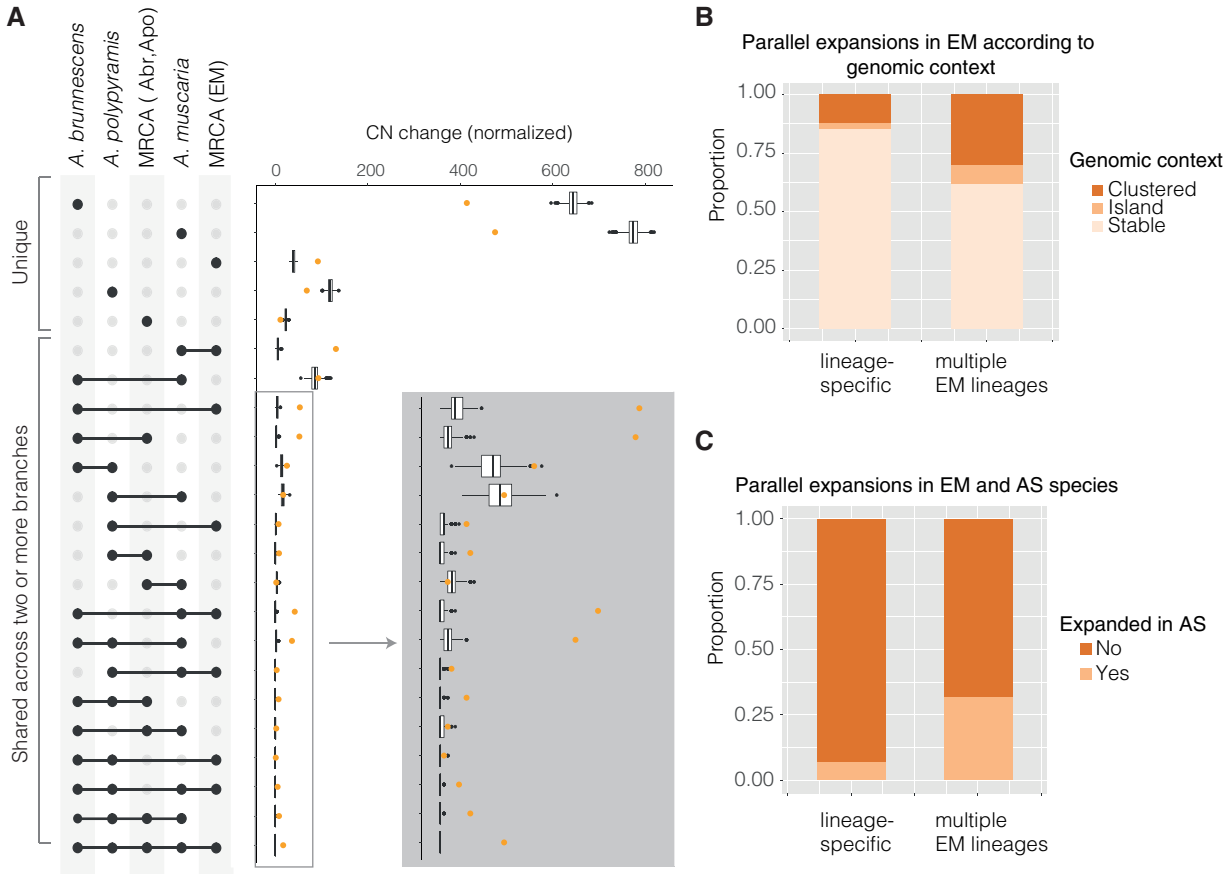
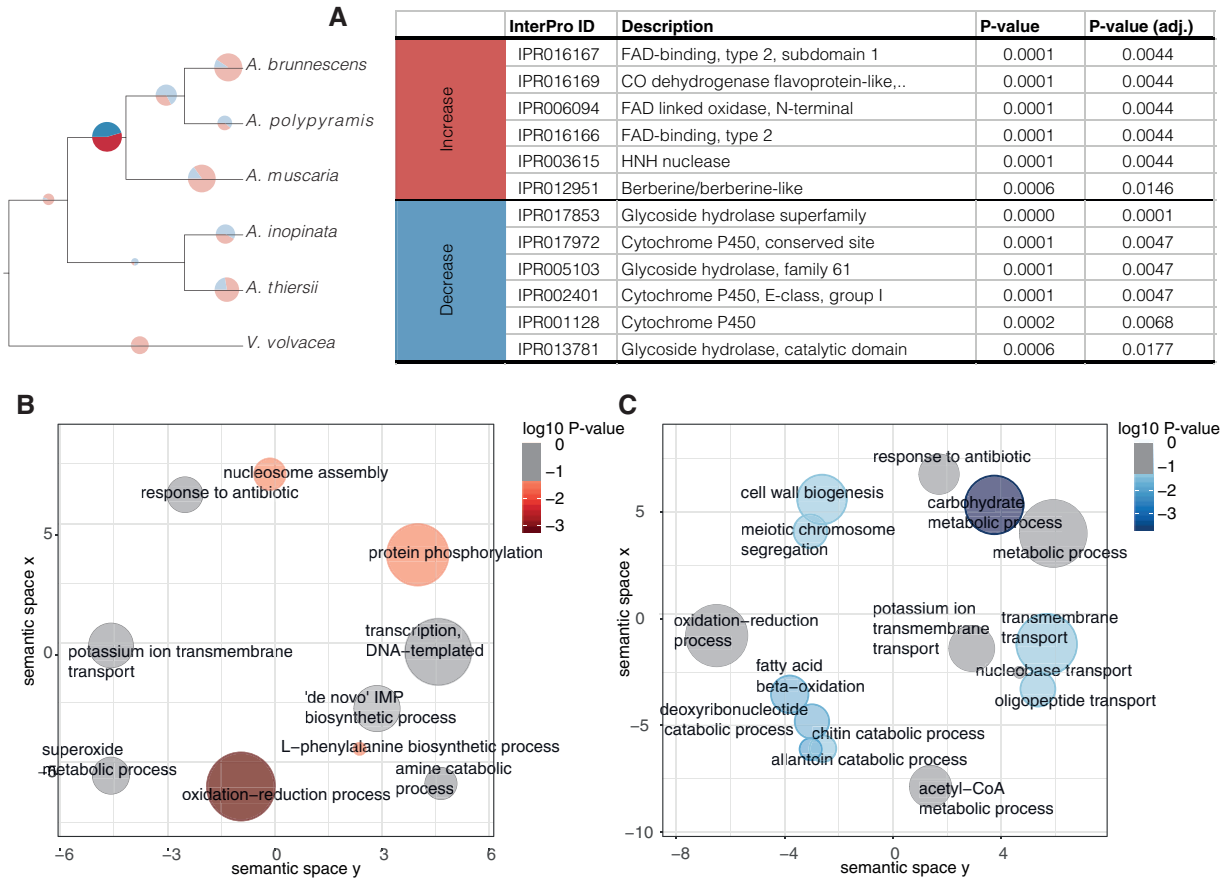


Fig. 4. Expansion profiles of orthogroups amplified in EM lineages. (*A*) Points within "CN change (normalized)" mark the average copy number gain (normalized by the number of branches considered) of orthogroups for which an expansion was inferred on the set of branches indicated by dots and lines on the left side of the plot. Boxplots show the distribution of this metric when the same numbers of duplications and losses are randomly placed across the genome (1000 permutations.) (*B*) Proportional representation of lineage-specific versus convergently amplified orthogroups according to genomic context, and (*C*) according to whether orthogroups also show convergent amplification in AS lineages.

These patterns may reflect differences in the propensity to duplicate among different gene families. To test this hypothesis, we classified lineage-specific and orthogroups showing parallel amplification according to 1) genomic context (fig. 4*B*), and 2) whether amplified orthogroups are also expanded within AS lineages (fig. 4*C*). Genomic context was defined as either "stable," meaning no other genes from the same orthogroup are found within a 10-kb window around the gene, "island," meaning the gene is the only predicted gene on a short scaffold, which may indicate regions embedded in high repeat content, or "clustered," meaning other members of the same orthogroup are found within 10 kb of the focal gene in the respective genome. Genes arranged in clusters on a chromosome may be more prone to duplications and losses, since many mutational mechanisms involve defects in homology-dependent repair machinery or replication slippage (Hastings et al. 2009). Arrangement in clusters also renders genes more susceptible to gene conversion among copies, which can diminish phylogenetic signal and artificially increase similarity among paralogs, causing inferences of a lineage-specific duplication when in fact the paralogs predate speciation (Liao 1999).

Our data confirm these expectations. Orthogroups that were amplified in parallel were more likely to occur in islands or clusters, as compared with orthogroups experiencing lineage-specific amplifications. The analysis suggests parallel amplification may be driven, at least partially, by an increased propensity to duplicate or the inability to accurately infer the evolutionary history of a gene family (fig. 4*B*).

Orthogroups amplified in multiple EM lineages were also more likely to be amplified in AS lineages, compared with orthogroups experiencing lineage-specific amplifications (fig. 4*C*), another line of evidence suggesting parallel expansions among EM lineages may involve dynamics not unique to the EM niche. But in the aggregate, only 44 of the 241 orthogroups that were amplified in more than one EM species are both clustered within genomes, and also expanded in at least one AS species. While the data suggest evolutionary processes not specific to the EM niche may have contributed to parallel amplifications among EM *Amanita*, the majority of such orthogroups appear to be located in stable genomic regions and specifically amplified among EM lineages.



	InterPro ID	Description	P-value	P-value (adj.)
Increase	IPR016167	FAD-binding, type 2, subdomain 1	0.0001	0.0044
	IPR016169	CO dehydrogenase flavoprotein-like,..	0.0001	0.0044
	IPR006094	FAD linked oxidase, N-terminal	0.0001	0.0044
	IPR016166	FAD-binding, type 2	0.0001	0.0044
	IPR003615	HNH nuclease	0.0001	0.0044
	IPR012951	Berberine/berberine-like	0.0006	0.0146
Decrease	IPR017853	Glycoside hydrolase superfamily	0.0000	0.0001
	IPR017972	Cytochrome P450, conserved site	0.0001	0.0047
	IPR005103	Glycoside hydrolase, family 61	0.0001	0.0047
	IPR002401	Cytochrome P450, E-class, group I	0.0001	0.0047
	IPR001128	Cytochrome P450	0.0002	0.0068
	IPR013781	Glycoside hydrolase, catalytic domain	0.0006	0.0177

Fig. 5. Functional enrichment analyses of genes expanded or contracted early in the evolution of an EM niche. (*A*) Complete list of significantly enriched InterPro domains (FDR-corrected *P* value <0.05, Fisher's exact test). Note that enzymes with the InterPro signature IPR005103 "Glycoside hydrolase, family 61" have since been reclassified as LPMOs in family AA9 in the CAZy database. (*B*) Significantly enriched GO terms among expanded genes. (*C*) significantly enriched GO terms among contracted genes. GO term enrichment results include only terms with *P* values <0.1 and are summarized using REVIGO (Supek et al. 2011) for the purpose of visualization. Significant enrichments at a *P* value threshold of 0.05 are colored red and blue (*B* and *C*, respectively), otherwise terms are colored gray.

Functional Signatures of the Ancestral EM *Amanita* Genome

In total, we found 203 orthogroups expanded within the MRCA of EM *Amanita* lineages, and 148 contracted (supplementary table S1, Supplementary Material online). To screen for functional signatures among these gene families, we tested for overrepresentation of functional categories among this gene set using an enrichment analysis of InterPro domains as well as a GO term analysis. We were able to annotate 57% of the expanded clusters with InterPro domains and 44% with GO terms, while 84% and 62% of the contracted clusters were annotated with InterPro domains and GO terms, respectively. Figure 5 plots significantly overrepresented InterPro domains (fig. 5*A*) among expanded and contracted gene families, as well as enriched GO terms in the "Biological Process" category (fig. 5*B* and *C*). To visualize the analysis, GO terms were consolidated into semantically similar terms using REVIGO (Supek et al. 2011). A comprehensive list of enrichment results is provided in supplementary table S2, Supplementary Material online.

Gene families expanded early in the evolution of the EM niche, within the MRCA of EM *Amanita*, are significantly enriched for several domains found in FAD-dependent oxidoreductases, including Berberine Bridge Enzymes (BBEs; IPR006094, IPR016166, IPR016167, IPR016169, IPR012951). These enzymes are involved in a range of different chemical reactions, and in fungi reactions include oxidation of carbohydrates and the biosynthesis of alkaloids (Daniel et al. 2017). The enrichment for oxidoreductases is also reflected in the GO term analysis, where the strongest overrepresentation was found for the term "oxidation–reduction process" (GO: 0055114). Other significantly enriched terms suggest expansion of genes involved in nucleosome assembly (GO: 0006334 "nucleosome assembly") and posttranslational regulation (GO: 0006468, "Protein phosphorylation").

Gene families contracted early in the evolution of the EM niche, within the MRCA of EM *Amanita*, are significantly enriched for genes involved in carbohydrate metabolism, for example, the overrepresentation of InterPro domains IPR017853 "Glycoside hydrolase superfamily" (now obsolete), IPR013781 "Glycoside hydrolase, catalytic domain" and IPR005103 "Glycoside hydrolase, family 61." The latter encompasses a class of enzymes now reclassified as LPMOs in CAZy family AA9 (see above). Losses of carbohydrate metabolism genes are also reflected in the GO term analysis; the most significantly enriched GO term within the contracted gene set

is "carbohydrate metabolic process" (GO: 0005975). Analyses mirror the results from our targeted annotation of carbohydrate active enzymes (fig. 2), highlighting the dynamic of loss of carbohydrate active enzymes. Genes involved in transport also appear to be lost, for example, significantly enriched GO terms include "transmembrane transport" (GO: 0055085) and "oligopeptide transport" (GO: 0006857). We also detected losses of genes involved in specialized metabolism (e.g., Cytochromes P450s: IPR017972, IPR002401, and IPR001128, which predominately occur in the same protein) and a variety of other metabolic processes (e.g., "fatty acid beta oxidation," GO: 0006635; "deoxyribonucleotide catabolic process," GO: 0009264; "allantoin catabolic process," GO: 0000256) as well as "cell wall biogenesis" (GO: 0042546).

Lineage-Specific Changes Reflect Continued Adaptation after the Origin of Symbiosis

The phylogenomic analysis suggests the majority of new genes found in EM species arose within lineages, after the origin of the EM niche (fig. 3). To understand the functional classes of genes either expanded or contracted during the evolution of each EM lineage, we screened the gene families of *A. brunnescens*, *A. polypyramis*, and *A. muscaria* for over-represented InterPro domains and GO terms, as described earlier (fig. 6 and supplementary table S2, Supplementary Material online). Annotation rates varied between 68% and 85% for InterPro domains and 52–74% for GO terms. Among expanded sets of genes, we detected many of the same types of genes as those expanded in the ancestral EM genome, including genes involved in oxidative metabolism, posttranscriptional regulation and signaling (fig. 6*A* and *B*). The GO term "oxidation–reduction process" (GO: 0055114) was strongly enriched in all three species. This dynamic is also reflected in the significant enrichments for various InterPro domains of oxidative enzymes, but especially the enrichment of Cytochrome P450s.

Unlike in the ancestral EM genome, in extant EM species we found strong enrichment for transport-related genes (e.g., "transmembrane transport," GO: 0055085) among expanded orthogroups (a stark contrast to the contraction of these in the MRCA of EM *Amanita*). All three species show strong enrichments for the InterPro domain "Major facilitator superfamily," a transporter (IPR016169), while both *A. muscaria*, but especially *A. brunnescens* show signatures of expansion of amino acid transporters ("amino acid transmembrane transport," GO: 0003333; IPR002293 "Amino acid/polyamine transporter I" and IPR004840 "Amino acid permease"). Thus, our data suggest a pattern of turnover: the transporter repertoire seems to have been reduced in the MRCA of the EM lineages, and then independently expanded in each derived lineage. Besides genes involved in oxidative metabolism and transport, we also detected GO terms pertaining to sugar metabolism ("glycolytic process," GO: 0006096 and "malate metabolic process," GO: 0006108) enriched in *A. muscaria* and *A. brunnescens*, respectively.

Other gene families amplified within extant EM fungi but not in the MRCA of EM *Amanita* emphasize the independent trajectories taken by different species (fig. 6*A*). These include NOD-like receptor gene families, involved in selfnonself recognition and the detection of other organisms (including the InterPro domains IPR003593 "AAA+ ATPase," IPR007111 "NACHT nucleoside triphosphatase," and IPR027417 "P-loop NTPase"; Dyrka et al. 2014). NOD-like receptor gene families were enriched within *A. brunnescens*' expanded gene set in particular. Other large receptor gene families, including NLR-type receptor gene families, clustered within lineage-specific orthogroups and could not be included in phylogenomic analysis (supplementary table S3, Supplementary Material online). These receptors tend to be hypervariable in copy number between species, but also among individuals of the same fungal species (Dyrka et al. 2014). We also characterized the lineage-specific clusters not included in the phylogenomic analysis (supplementary table S3, Supplementary Material online) for functional enrichments. Annotation rates for these clusters were considerably lower than for the clusters included in phylogenomics: 38–42% for InterPro domains and 26–38% for GO terms. Lineage-specific clusters were dominated by TE-related domains but otherwise possessed similar signatures to lineage-specific amplifications within clusters included in the phylogenomic analysis (supplementary table S2, Supplementary Material online).

Gene families contracted within extant EM fungi highlight the continued loss of orthogroups involved in carbohydrate metabolism ("carbohydrate metabolic process" GO: 0005975 and "carbohydrate catabolic process" GO: 0016052). We detected many of the same patterns found in late expanded gene families, for example, a strong enrichment for oxidative enzymes in all three species ("oxidation–reduction process" GO: 0055114), transmembrane transport (GO: 0055085), or basic sugar metabolism ("glycerol ether metabolic process," GO: 0006662), underlining that both gene family expansion and gene family contraction are shaping repertoires of genes with similar functions.

Secretome Size Distributions Show No Conclusive Patterns That Distinguish EM from AS Fungi

Effector-like small secreted proteins (SSPs) can mediate interactions between EM fungi and plants by modulating the plant immune system (Plett et al. 2011, 2014). In fact, ectomycorrhizal fungi often show expanded repertoires of SSPs, compared with asymbiotic relatives (Pellegrin et al. 2015). While the two EM *Amanita* with large genomes, *A. brunnescens* and *A. muscaria*, do appear to possess expanded repertoires of SSPs, once again *A. polypyramis* provides conflicting evidence. Predictions of the numbers of secreted proteins range from 260 in the genome of *A. polypyramis* to 588 in the genome of *V. volvacea*; between 33% (in *A. polypyramis*) and 54% (in *A. brunnescens*) of these are shorter than 300 amino acids, and are defined as SSPs (Dryad Annotation file, supplementary fig. S4*A*, Supplementary Material online). The EM species *A. brunnescens* and *A. muscaria* encode the largest numbers of SSPs, with 282 and 217, respectively (although the AS *A. thiersii* encodes 215), while the EM *A. polypyramis* encodes the smallest number of SSPs, with 86 altogether. Data once again highlight the stark contrast in genome architectures among EM

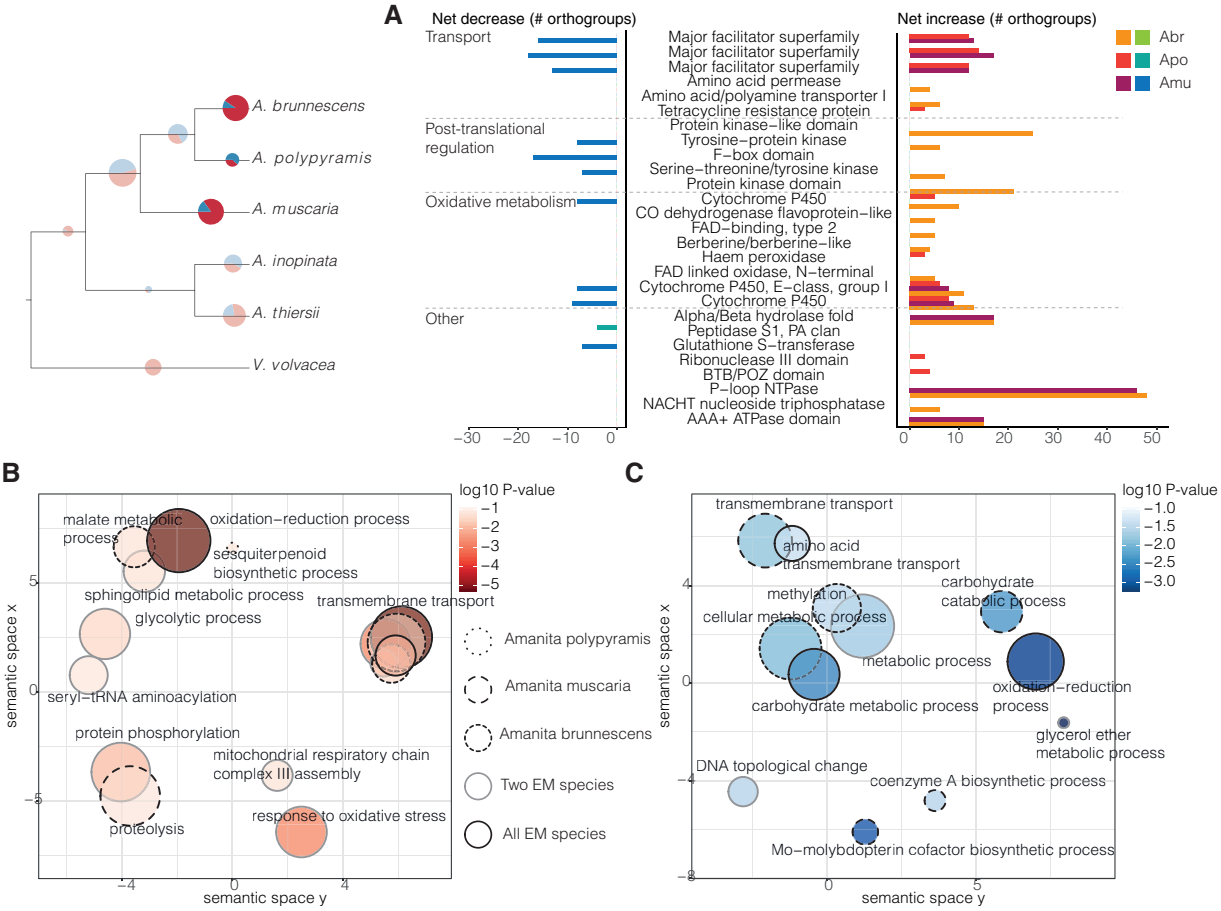


Fig. 6. Functional enrichment among orthogroups amplified or contracted in extant EM lineages. (*A*) InterPro domains. Colors indicate gain (red) and loss (blue), shading indicates lineage-specific enrichment in the respective species, (*B*) GO term enrichments in the biological process category among lineage-specific amplified orthogroups. Circle color indicates *P* value significance, circle outline which species the term was enriched in. Data were summarized using REVIGO (Supek et al. 2011). (*C*) As for (*B*) but with contracted orthogroups.

Amanita. However, the genomes with the largest numbers of predicted SSPs are also the genomes with the most fragmented assemblies (fig. 1A). Fragmented assemblies can lead to fragmented or partial gene models (Denton et al. 2014) and artificially inflate predicted numbers of SSPs.

We also investigated the ages of SSPs, as compared with the ages of the full predicted secretome (supplementary fig. S4*B*, Supplementary Material online). In each of the six species, SSPs were dominated by increased proportions of lineage-specific genes (either amplified on terminal branches, or unclustered as singletons) compared with the complete secretomes, which house higher proportions of genes present in the MRCA of all species analyzed. Data mirror patterns found across larger taxonomic distances (Kohler et al. 2015) and suggests that SSPs tend to be lineage-specific even across shorter evolutionary distances and after a single origin of symbiosis. The importance of lineage-specific SSPs during symbiosis is confirmed by the mapping of *A. muscaria* expression data from (Kohler et al. 2015) to the predicted secretome of *A. muscaria*, and by investigation of the age distribution of these mycorrhiza-induced SSPs (supplementary table S4, Supplementary Material online). Expression data were collected from EM root tips of a different individual of *A. muscaria (var. muscaria)* grown in association with *Populus tremula x tremuloides* (Kohler et al. 2015). All symbiosis-upregulated SSPs in *A. muscaria* were classified as either singletons or recently amplified in *A. muscaria* (with the exception of one gene whose age we dated to the MRCA of the *Amanita*, but that we determined to be a gene prediction artefact). None of the 23% of *A. muscaria* SSPs conserved across the genus *Amanita* were induced during symbiosis, although intriguingly, 4 of the 19 symbiosis-induced *A. muscaria* amplified SSPs have homologs in non-EM species.

Gene Expression Data of *A. muscaria* Root Tips Confirm Significance of Functional Patterns

To validate patterns revealed by our phylogenomic analyses we used the expression data taken from EM root tips (Kohler et al. 2015). A total of 408 genes were significantly upregulated in EM root tips: 183 are genes that arose within the *A. muscaria* lineage, compared with 145 already present at the root of the *Amanita* phylogeny, 50 unclustered singletons, and 7 genes that arose within the MRCA of EM *Amanita* (23 of the upregulated genes were not included in our phylogenomic analysis because of the small size of their corresponding orthogroups; supplementary tables S3 and S5, Supplementary Material online). These data highlight the

Table 1. Genes Active in *A. muscaria* Root Tips Residing in Orthogroups Duplicated Within the MRCA of EM *Amanita*.

GeneID	FC (EM vs FLM)	Orthogroup	Gene Age	Orthogroup Age	Description
Amu1429.1	13.07	ORTHOMCL117	Amu	*Amanita*	BBE oxidoreductase
Amu3629.1	6.26	ORTHOMCL278	Amu	Root	Putative CorA-like transporter
Amu7507.1	3602.15	ORTHOMCL326	Amu	Root	Ferric reductase
Amu10401.1	49.73	ORTHOMCL3379	Amu	EM	PHB depolymerase / esterase[a]
Amu4064.1	12.67	ORTHOMCL343	EM	Root	Potassium transporter
Amu11173.1	11.41	ORTHOMCL41	Amu	EM	Cytochrome P450
Amu4980.1	22.37	ORTHOMCL41	Amu	EM	Cytochrome P450
Amu5003.1	5.28	ORTHOMCL41	Amu	EM	Cytochrome P450
Amu5016.4	10.84	ORTHOMCL41	Amu	EM	Cytochrome P450
Amu4989.1	5.83	ORTHOMCL41	Amu	EM	Cytochrome P450
Amu12396.1	27.05	ORTHOMCL41	Amu	EM	Cytochrome P450
Amu11955.1	5.81	ORTHOMCL517	EM	Root	Cytochrome P450
Amu7518.1	6.45	ORTHOMCL5619	Amu	EM	Acid phosphatase
Amu2358.1	5.17	ORTHOMCL6796	EM	EM	Thioredoxin-like
Amu12324.1	21.11	ORTHOMCL736	Amu	EM	Protein of unknown function
Amu9590.1	7.04	ORTHOMCL7410	EM	EM	Protein of unknown function
Amu13382.1	7.37	ORTHOMCL7623	EM	EM	Protein of unknown function
Amu10418.1	16.86	ORTHOMCL7816	EM	EM	ACC deaminase
Amu2199.2	5.48	ORTHOMCL7837	EM	EM	Probable dipeptyl peptidase

FC, fold change; EM, ectomycorrhizal root tip; FLM, free-living mycelium.
[a]Horizontally transferred (Chaib De Mares et al. 2015).

large contribution of lineage-specific expansions to the symbiotic transcriptome.

While only seven upregulated genes have direct origins within the MRCA of EM *Amanita*, 19 belong to orthogroups that increased in copy number in the EM ancestor and then continued to expand in *A. muscaria* (table 1). Upregulated genes include two transporter genes, as well as ten genes involved in oxidative metabolism: a BBE-like gene, a highly induced ferric reductase, seven Cytochrome P450s, and a Thioredoxin-like gene. Our functional enrichment analyses appear to have identified genes directly relevant to the functioning of EM symbiosis in *A. muscaria*. Moreover, one upregulated gene is a PHB depolymerase esterase ([a], table 1). This gene belongs to an orthogroup with origins in a horizontal gene transfer event, likely from a soil bacterium into the EM ancestor (Chaib De Mares et al. 2015). In addition to gene duplications and losses, horizontal gene transfer may also have facilitated adaptations to the EM niche.

Discussion

Genome Architectures of EM *Amanita* Suggest Rapid Divergence following an Origin of Symbiosis

The basidiomycete *Laccaria bicolor* (Martin et al. 2008) and ascomycete *Tuber melanosporum* (Martin et al. 2010) are separated by 635 Ma of evolution (Kohler et al. 2015), and yet each is an ectomycorrhizal fungus, the result of convergent evolution to a symbiotic niche. Our analyses suggest genome architectures as divergent as those of *L. bicolor* and *T. melanosporum* can also evolve within tens of millions of years, and following a common origin of symbiosis. While in general the genomes of ectomycorrhizal fungi appear larger than genomes of asymbiotic fungi, as measured by both genome size and the number of encoded proteins (Floudas et al. 2012; Kohler and Martin 2016), to date analyses have focused on a heterogeneous array of species. By contrast the genomes of closely related *Amanita* reveal no clear patterns distinguishing the genome architectures of EM species from asymbiotic *Amanita*, or the outgroup *V. volvacea*. Instead, the three EM species show distinct genome architectures: *Amanita brunnescens* and *A. muscaria* both harbor expanded gene repertoires, similar to a typical EM genome (Kohler and Martin 2016) while the gene repertoire of *A. polypyramis* appears to be greatly reduced, and similar to the gene repertoire of the asymbiotic *A. inopinata*. The data mirror an earlier analysis of transposable element dynamics among the *Amanita* (Hess et al. 2014); in both cases, ecology emerges as a poor predictor of genome architecture.

Our genome-wide phylogenomic analysis of protein-coding genes suggests considerable reconfiguration of the ancestral symbiont genome. Absolute gene family expansion was relatively modest within the MRCA of EM *Amanita*, during the transition from saprotrophy to symbiosis, compared with expansions on terminal branches, after the origin of symbiosis. Data suggest the additional set of genes required for symbiosis may have been minimal. Nevertheless, some of the highest rates of both duplications and losses are found at this node of the phylogeny (fig. 3). Moreover, data support a scenario of parallel gene space expansion in *A. brunnescens* and *A. muscaria* (fig. 3). Patterns of orthogroup amplification (fig. 4A) suggest both a tendency for continued amplification of families originally duplicated within the MRCA of EM *Amanita*, as well as other independent but convergent amplifications of gene families on many EM branches. Lineage-specific parallel amplification and loss of gene families, particularly of families arranged in tandem arrays along chromosomes, appears to be a widespread phenomenon and has been associated with adaptation to new niches in other organisms, for example, in fruit flies (Zhong et al. 2013), trypanosomatids (Drini et al. 2016), plants (Cannon et al. 2004),

and snakes (Aird et al. 2017). Amplified gene families are likely to be functionally relevant and emerge as candidates implicated in the adaptation of genomes to the EM niche.

Loss May Be the Predominant Mechanism Enabling a Transition to an EM Niche

While gene family expansion dynamics are clearly relevant to the evolution of EM symbiosis, gene loss emerges as perhaps the more critical defining mechanism of an EM origin. Gene loss is particularly concentrated within the ancestral symbiont genome of EM *Amanita*. An estimated 79% of the orthogroups contracted within the MRCA of EM *Amanita* appear purged from the genome completely, suggesting the early loss of a wide variety of conserved genes was a key to the initial ability to form mycorrhizal associations. Gene families contracted within the MRCA of EM *Amanita* are strongly enriched for genes involved in the degradation of plant cell walls (figs. 2 and 5), but also include twelve orthogroups harboring transporters (fig. 5 and supplementary table S1, Supplementary Material online). We detected the near complete loss of hydrolytic CAZymes acting on cellulose (GH6, GH7, and GH5_5), hemicellulose and pectin as well as broad reductions in the numbers of accessory CAZymes. This pattern of loss, already considered a hallmark of diverse EM fungi, may shield the fungus from detection by the plant immune system (Martin et al. 2008; Kohler et al. 2015), and enable a fungus to live stably within a root, without decomposing the plant.

However, the lack of PCW decomposing enzymes in the AS species *A. inopinata* leads us to speculate whether the loss of key PCW decomposing enzymes may in fact predate the ectomycorrhizal symbiosis. The fungus *A. inopinata* is not ectomycorrhizal: it is often collected near plants that do not form ectomycorrhizal symbioses (e.g., the genera *Taxus* or *Chaemaecyparis* in Europe, or *Chaemaecyparis* or *Cupressus* in New Zealand; Fraiture and Giangregorio 2013), it cannot form ectomycorrhizal symbioses in the laboratory, and a search for root tips in a natural population was unsuccessful (Wolfe, Tulloss, et al. 2012). There is no evidence the fungus is a pathogen. For the moment, *A. inopinata*'s niche remains unknown, but the loss of PCW decomposing enzymes within its genome provides a tantalizing connection to recent data suggesting saprotrophic fungi are capable of facultative biotrophy (Smith et al. 2017; Selosse et al. 2018). We hypothesize *A. inopinata* is in fact a facultative or potentially obligate biotroph, growing near plant roots and using root exudates as carbon sources but not providing any resources to plants, and so not forming any ectomycorrhizal structures. If we are correct, *A. inopinata* and potentially other species within its clade may represent an emerging, second origin of symbiosis.

While the MRCA of EM *Amanita* experienced a contraction of transmembrane transporter families, extensive, lineage-specific expansions in derived EM lineages suggest the function of EM *Amanita* in symbiosis may vary. Transmembrane transporters serve as the interface between fungus and plant, and also enable a fungus to assimilate and secrete diverse compounds from the environment (Garcia et al. 2016). Differences among lineages suggest different species may move substances across the mycorrhizal interface differently, or vary in the assimilation of nitrogen or other nutrients from soil.

Functional Signatures of Gene Amplifications in EM Amanita

While gene loss dominates the evolutionary dynamic within the MRCA of EM *Amanita* (fig. 3), various gene families are also duplicated. While amplifications within the ancestral symbiont genome reflect sweeping functional changes in gene regulation and oxidative metabolism, convergent amplifications within *A. brunnescens* and *A. muscaria* reflect more specific tinkerings of pathways implicated in transport, sugar metabolism, and terpenoid metabolism, as well as the continued evolution of signaling pathways and oxidative metabolism (Kohler and Martin 2016).

Expanded gene families in the ancestral symbiont genome are strongly enriched for regulatory genes, including protein kinases, and expansions of the protein kinase repertoire may have facilitated new signaling or developmental processes involved in interactions with plants. Protein kinases are involved in posttranscriptional regulation of a variety of cellular processes including general metabolism, development, the cell cycle, and the perception and integration of biotic and abiotic signals. Protein kinases typically belong to large protein families with hundreds of members (Kosti et al. 2010), and are also among the gene families greatly expanded in the *L. bicolor* genome (Martin et al. 2008) as well as in the genome of the arbuscular mycorrhizal fungus *Rhizophagus intraradices* (Tisserant et al. 2013).

However, the most prominent expansions in the *Amanita*, both in the ancestral symbiont genome but also within the lineage-specific expansions of *A. brunnescens* and *A. muscaria*, involve oxidative enzymes, including laccases, Cytochrome P450s and Berberine Bridge Enzymes (BBEs). Expansions of oxidative enzymes may facilitate transitions from AS to EM ecology (fig. 6). In fact, a reshaping of oxidative metabolism (together with Fenton chemistry) underpins other ecological transitions in fungi, for example, the evolution of brown rot decomposers from ancestral white rots (Eastwood et al. 2011; Floudas et al. 2012; Balasundaram et al. 2018). Accumulating evidence suggests EM fungi are able to decompose soil organic matter (SOM), probably in order to liberate nitrogen embedded in recalcitrant agglomerates of lignin fragments, polyphenols, polysaccharides, lipids, peptides, and minerals (Rineau et al. 2012; Shah et al. 2016). Laccases appear as potential players in the degradation of SOM by EM fungi and the retention of both classes of enzymes by EM *Amanita* suggests that these species employ mechanisms similar to other EM fungi to obtain nitrogen (fig. 2). Laccases also play a role in developmental processes, including fruit body and rhizomorph development (Courty et al. 2009; Sipos et al. 2017) and expansions may have contributed to changes facilitating the emergence of mycorrhizal structures.

In addition to being involved in wood degradation, laccases and cytochrome P450s detoxify xenobiotics, and are implicated in the biosynthesis of defense compounds.

These enzymes may play a more direct role in interactions with EM plants (Črešnar and Petrič 2011; Kües and Ruhl 2011). And in fact, seven Cytochrome P450s belonging to gene families expanded within the MRCA of EM *Amanita* are among the genes upregulated in *A. muscaria* root tips (table 1). Metabolomic analyses of compatible and incompatible pairings of *L. bicolor* with either *Populus trichocarpa* (compatible) or *Populus deltoides* (incompatible) demonstrated a defining difference between interactions is the concentration of phenolic defense compounds. During compatible interactions, at colonization, defense compounds are degraded by the benzoate degradation pathway of *L. bicolor* (Tschaplinski et al. 2014). Furthermore, the interactions between cytochrome P450s, FAD-dependent oxidoreductases and plant defense compounds may mediate host specificity; a metatranscriptomic study of different *Suillus* species paired with compatible and incompatible hosts (Liao et al. 2016) demonstrated these genes as upregulated, but only in particular fungal–plant combinations.

Berberine Bridge Enzymes (BBEs) belonging to three different gene families are an additional class of oxidative enzymes enriched within the MRCA of EM *Amanita*, and within *A. brunnescens* and *A. muscaria* where they may help to mediate interaction with the host plant. BBEs were first described as involved in the biosynthesis of alkaloids by California poppy (Hauschild et al. 1998). Members of this gene family have since been identified in other plants, bacteria, and fungi where they play a role in alkaloid biosynthesis, oxidation of mono- and polysaccharides and synthesis of antibiotics among a diverse set of functions (Daniel et al. 2015, 2017). The oomycete plant pathogen *Phytophtora infestans* secretes BBEs *in planta*, suggesting a possible role for BBEs in the pathogen's interaction with the plant, either through modification of plant defensive compounds or by oxidation of breakdown products of the PCW that would otherwise signal the presence of the pathogen to the plant immune system (Raffaele et al. 2010; Daniel et al. 2017). In tomato and tobacco, the pathogen *Septoria lycopersici* appears to modify a plant defense compound using a BBE-type enzyme, the product of which then acts as a signaling molecule and dials down the plant defense response (Bouarab et al. 2002). At least two of the gene families encoding BBEs expanded in the MRCA of EM *Amanita* are dominated by secreted proteins (Dryad Annotation file) and one is significantly upregulated in mycorrhizal root tips (table 1), suggesting a potential role for *Amanita* BBEs in the interaction with its host.

Genome Evolution among Asymbiotic and Ectomycorrhizal Amanita and Implications for Convergent Evolution of the Symbiosis

Whether and to what extent gene repertoire expansions or contractions during the transition to symbiosis were driven by natural selection or are the result of neutral processes remains unclear. The diverse genome architectures encountered among EM *Amanita* hint at the influence of nonadaptive processes. In particular, the effects of changing population sizes on genome architecture can be profound and may obscure the effects of selection (Lynch and Conery 2003). For example, the reductive and repeat-rich genome architecture of *T. melanosporum* is linked to its demographic history, and a small effective population size caused by a retreat to glacial refugia during the last ice age (Murat et al. 2004; Martin et al. 2010). The streamlining of the *A. polypyramis* genome may be the result of either demography (as in *Tuber*) or selection for a compact and efficient genome (Wolf and Koonin 2013). Levels of heterozygosity estimated during the genome assembly process (supplementary table S6, Supplementary Material online) appear extremely low (approximately one polymorphism per 48,000 bases), suggesting *A. polypyramis* has a small effective population size and possibly shares a comparable demographic history with *T. melanosporum*.

Similarly, whether the potential expansions of functional space caused by gene duplications confer selective advantages, or result from fixation through genetic drift, is unknown (Lynch and Conery 2003; Kelkar and Ochman 2012). The distinction between these two processes is less critical for duplications that arose in the MRCA of all EM *Amanita* since duplicate genes have been maintained in extant EM genomes for tens of million years and are therefore likely under selective constraint (irrespective of the original mechanism driving fixation). By contrast, understanding the significance of lineage-specific expansions in this context is more challenging, especially for large, dynamic gene families, since we do not know the precise age of gene duplicates, or whether they are variable among individuals of a species. Studies of snake venom gene cluster evolution show a complex interplay of adaptive evolution and neutral processes in the expansion and maintenance of these rapidly evolving gene families (Aird et al. 2017). A focused analysis of individual gene families will be required to determine the proportions of pseudogenes and gene prediction artefacts in *Amanita* genomes. Furthermore, functional evidence for neo- or subfunctionalization of retained duplicates, for example, through selection screens or expression would be useful. Nevertheless, we are confident that the patterns uncovered by our analyses are relevant to the evolution of EM symbiosis based on i) pervasive parallel gene space expansion, in particular of families that were already amplified in the EM ancestor and ii) the overlap of our phylogenomic data with data of genes upregulated in *A. muscaria* root tips. Lineage-specific duplicates are overrepresented among genes upregulated in EM root tips (supplementary fig. S5 and table S5, Supplementary Material online).

Patterns of evolution within the genus highlight the malleability of fungal genomes and the diversity of genome architectures resulting in EM symbiosis (Martin et al. 2010; Kohler et al. 2015; Kohler and Martin 2016). Unlike the shared evolutionary paths to endosymbiosis taken by independently evolved bacterial endosymbionts (Wernegreen 2015; Boscaro et al. 2017), it seems that no single feature defines an EM genome, even following a single origin of symbiosis. While gene loss is clearly involved in the transition to symbiosis, a critical, open question raised by these data (and by the data of *A. inopinata* in particular) is whether gene loss was a

prerequisite to evolving the symbiotic interaction, or a process restricted within the MRCA of EM *Amanita*. We also discovered expansions of quickly evolving and large families of regulatory genes and different families of oxidoreductases at the core of the early evolution of the EM symbiosis in *Amanita*, and gene expression studies in other EM genera suggest a more widespread relevance of these to EM symbiosis (see above; Tschaplinski et al. 2014; Liao et al. 2016). The ubiquitous distribution of these enzymes across many different fungal niches, including saprotrophic and pathogenic as well as EM fungi, highlights their evolvability and broad functional diversity, and suggests these are gene families that can contract or expand to enable adaptation to different niches (Syed et al. 2014; Qhanya et al. 2015; Daniel et al. 2017). Plants protect their tissues using phenylpropanoids and these compounds present chemical challenges to fungi interacting with live and dead plants alike (Gluck-Thaler and Slot 2018). The cytochrome P450 family expanded and upregulated in EM root tips (ORTHOMCL41) most closely matches a CYP512 from the brown rot fungus *Serpula lacrymans*. CYP512 is enriched among wood-decaying Basidiomycetes and displays affinity for a wide range of substrates (Syed et al. 2014). Available data suggest both a relevance to the decay of lignocellulose, and that EM *Amanita* have successfully managed to co-opt this class of enzymes, already present in the genomes of their saprotrophic ancestors, for navigating interactions with hosts. In the aggregate, available data suggest that the chemical challenges a fungus meets when interacting with a dead plant may not be too dissimilar from those of interacting with a live plant and that a rudimentary capability to form symbioses is already encoded in saprotrophic genomes.

Conclusions

Our data support an emerging hypothesis that the transition from saprotroph to symbiont does not require extensive sets of new genes, rather, the origin of symbiosis appears to stem primarily from the loss of key gene families, including carbohydrate active enzymes and transporter genes. However, one AS *Amanita* also has lost PCW degrading enzymes, suggesting loss is not strictly associated with the evolution of symbiosis and may in fact predate the evolution of symbiosis. Extensive lineage-specific amplifications of symbiosis relevant families suggest extant EM *Amanita* may vary in fundamental aspects of the symbiosis, including host compatibility or the transport of molecules to and from the host and surrounding soils. We hypothesize that in the absence of a derived repertoire of effector genes evolved to specifically target the host immune system, a broad and variable repertoire of oxidative enzymes capable of mitigating the wave of defense chemicals produced by a plant may have enabled the ancestral EM *Amanita* fungus to begin its intimate association with its host. The ubiquitous distribution of these enzymes among fungal species with diverse ecologies and their arrangement in gene clusters may underpin the apparently intrinsic propensity for fungi to evolve ectomycorrhizal symbioses.

Materials and Methods

Fungal Strains, Sequencing, and Assembly

Sources and cultures of sequenced fungi, DNA extraction and genome sequencing protocols, as well as initial genome assembly procedures are described in detail by (Hess et al. 2014). In order to reduce redundancy in the genome sequences due to heterozygosity inherent in the first round of assemblies, we used AllpathsLG v. 46452 (Gnerre et al. 2011) to reassemble the *Amanita brunnescens*, *Amanita polypyramis*, *Amanita inopinata*, and *Volvariella volvacea* genomes. The *Amanita muscaria* var. *guessowii* and *Amanita thiersii* genomes assembled at the Department of Energy, Joint Genome Institute (JGI) were used as is, since AllpathsLG was used for assembly at the JGI. Because AllpathsLG requires custom sequencing libraries, we simulated a 150-bp insert size overlapping paired-end library and a 3-kb insert size mate pair library for each genome using wgsim (https://github.com/lh3/wgsim; last accessed september 27, 2018). We used the parameters: -e 0.0 -1 100 -2 100 -r 0.0 -R 0.0 -X 0.0 with either -d 180 -s 20 or -d 3000 -s 500 to generate 100-bp paired end libraries of the overlapping or mate pair kind, respectively. The original quality filtered sequencing libraries were then combined with the simulated reads and used for de novo assembly with the parameters PLOIDY = 2 and HAPLOIDIFY=True. The optimal number of simulated reads was determined by maximizing assembly contiguity, using iterative rounds of assembly for each genome, testing library sizes of 1 million, 5 million, 10 million, and 25 million reads for the overlapping library and 0.25 million, 0.5 million, 1 million, and 5 million reads for the mate pair library. The numbers of simulated reads and coverage cutoffs used for the respective assemblies are listed in supplementary table S6, Supplementary Material online.

Transcriptome Sequencing

To aid genome annotation, we sequenced transcriptomes of each species grown on two different conditions: a pure culture medium as well with the addition of sterile grass litter to encourage expression of genes potentially involved in degradation of plant cell wall material (see supplementary methods, Supplementary Material online, for the full media recipe). Cultures were grown on liquid media for 14 days, harvested, and stored in RNAlater (Albion) at −20°C until further processing. On the day of extraction, frozen samples were thawed at room temperature for 3 h. Tissue was removed from the thawed solution using sterile forceps and ground to a fine powder using mortar and pestle under liquid nitrogen. RNA was extracted using the Qiagen RNeasy Maxi kit (Qiagen, Hilden Germany) according to the manufacturer's protocol. cDNA libraries for *A. thiersii* and *A. inopinata* were prepared and sequenced at the Duke Sequencing Core Facilities, using the Roche rapid library preparation method and a 454 GS-FLX instrument with Titanium chemistry. *Amanita muscaria* cDNAlibraries were prepared using the TruSeq RNA Library Prep kit (Illumina), pooled, and sequenced on an Illumina HiSeq 2000 instrument at the Harvard Biopolymers Facility. Because initial trials of cDNA generation from *A. brunnescens* and *A. polypyramis* were unsuccessful using the TruSeq

protocol, libraries from these species were prepared at CoFactor Genomics using the NuGEN Ovation RNA-seq system (NuGEN Technologies, San Carlos, California), and pooled for sequencing on an Illumina HiSeq 2000 instrument. For *Volvariella volvacea*, a 454 transcriptome data set of a previously sequenced strain V23 (Bao et al. 2013), was downloaded from NCBI (accession number SRR619832).

Genome Annotation

Genomes were annotated using the JamG pipeline (Papanicolaou 2013). A detailed description of all elements of the pipeline and the parameters used are elaborated in the supplementary methods, Supplementary Material online. Briefly, RNA-seq data from species for which Illumina data are available were aligned to the respective genomes using GMAP (Wu and Watanabe 2005) and used for combined de novo and genome-guided assembly with Trinity 2.0.6 (Grabherr et al. 2011). The 454 transcriptomes were assembled either with Newbler v. 2.5 (Roche 454, Inc.) or MIRA 4.0.2 (Chevreux et al. 2004), in the case of the later assembly of *V. volvacea* data. Assembled transcriptomes were used to generate initial draft gene models with PASA (Haas et al. 2008), which in turn were used to train de novo gene predictors. We chose a combination of gene predictors that work well for gene dense fungal genomes, namely CodingQuarry v1.1 (Testa et al. 2015), Augustus v3.0.2 (Stanke et al. 2008), Genemark-ES v2.3e (Ter-Hovhannisyan et al. 2008), and SNAP (Korf 2004). Predicted gene models together with transcript and homologous protein alignments were passed to EvidenceModeler followed by a final round of PASA curation to tidy up putative gene models and generate a comprehensive set of predictions (Haas et al. 2008). Completeness and redundancy of our genome assemblies and gene annotations using BUSCO v.1.1 (Simão et al. 2015) by running the software in "gene set" mode with the fungal-specific set of single-copy orthologs.

Functional Annotation

Conserved domains in protein coding sequences were annotated using InterProScan v5.4 (Jones et al. 2014) with the options -iprlookup -goterms -pa. Assignment to conserved orthologous groups was computed using eggNOG mapper in one2one ortholog mode with the fungal database fuNOG (Huerta-Cepas et al. 2016). Secreted proteins were predicted using a pipeline consisting of SignalP v4.1 (Petersen et al. 2011), TargetP 2.0 (Emanuelsson et al. 2007), TMHMM 2.0 (Krogh et al. 2001), PS_scan (Hulo et al. 2006), and WolfPSort v0.2 (Horton et al. 2007) as implemented in (Kohler et al. 2015). The search and family assignment of all CAZymes were performed as previously described (Knapp et al. 2018) using the CAZy database (www.cazy.org) annotation pipeline (Cantarel et al. 2009; Lombard et al. 2014). This pipeline examines all classes of CAZymes, namely Glycoside Hydrolases (GH), Polysaccharide Lyases (PL), Carbohydrate Esterases (CE), Glycosyltransferases (GT), and Carbohydrate-Binding Modules (CBM), augmented by Auxiliary Activities (AA), which break down cell wall components oxidatively (Levasseur et al. 2013). Briefly, each protein sequence encoded by the genome was subjected to a BLAST (Altschul et al. 1997) search against a sequence library that contains the sequences of the domains listed in the CAZy database, in parallel to a HMMer (Eddy 2009) search performed against HMM profiles corresponding to these families and subfamilies thereof. All positive hits were manually examined for final validation.

Although our gene prediction pipeline includes a repeat masking step, we expect a small number of gene models to derive from transposable elements (TEs). To identify such putative TE-derived genes, we ran RepeatMasker 4.0.5 (Smit et al. 2015) with the TE libraries constructed in (Hess et al. 2014) and scored the overlap percentage with predicted protein-coding genes using BEDtools intersect (Quinlan and Hall 2010). Predicted genes with >20% overlap with an annotated TE were designated as putative TEs.

Gene Family Reconstruction and Analysis

We chose the longest predicted protein at each locus to assemble the proteomes of all species analyzed. The predicted proteomes were combined and clustered into gene families using FastORTHO (http://enews.patricbrc.org/fastortho/; last accessed September 27, 2018) with BLASTP e-value cutoff $1e^{-5}$, percent match cutoff 60, percent identity cutoff 30, and an inflation value of 3. Prior to phylogenomic reconstruction, we filtered the resulting clusters to exclude TE families in order to avoid introducing bias in the estimation of background birth and death rates, which we expect to be different between TEs and protein-coding genes, and between different families of TEs. We applied the following criteria for removal of a cluster: 1) the cluster contains >50% of sequences flagged as TEs, or 2) the cluster contains between 15% and 50% of flagged TEs and either contains members that encode TE-associated PFAM domains, or has representatives from less than three species indicating erratic distribution, or both. Physical clustering of genes within the same orthogroup along the genome was determined using a custom script available on Github (https://github.com/JackyHess/Amanita_comparative_genomics; last accessed September 27, 2018).

Multiple sequence alignments of each orthogroup were inferred using the phylogeny-aware aligner PRANK (Löytynoja and Goldman 2005) with five rounds of iterative refinement of alignment and guide tree estimated with RAxML (Stamatakis 2014), implemented in the tool Canopy (http://wasabiapp.org/software/canopy/; last accessed September 27, 2018). We used TrimAl v1.2 to remove poorly aligned regions from the alignments with the option –automated1 (Capella-Gutiérrez et al. 2009). Sequences retaining <20 residues postfiltering were removed from the alignments. Focusing on nontrivial alignments, that is, those of at least four sequences and with more than two different species, we then reconstructed individual gene trees using the following pipeline: first, we determined the best-fitting model of protein evolution using ProtTest 3.4.1 (Darriba et al. 2011) which was then used to infer a Maximum Likelihood tree (ML) with RAxML v8.0.26 using the default search algorithm (-f d) (Stamatakis 2014). Since gene tree

reconstruction is an error-prone process, we used TreeFix v.1.1.10 (Wu et al. 2013) for species-tree aware error correction of the ML tree. TreeFix was run using default settings apart from the options –nquickiter = 100 and –niter = 1000 and the best-fit model determined by ProtTest. Finally, the error-corrected gene tree was reconciled with the species tree using DLCpar (Wu et al. 2014), which implements an incomplete lineage sorting-aware reconciliation algorithm. Wherever possible, we ran DLCpar using a full, unbounded search (–max_dups=−1 –max_losses=−1), resorting to either bounded heuristics –max_dups = 4, –max_losses = 4), or finally a hill-climbing search with 20 prescreen iterations and 1000 search iterations if gene trees were too complex for the full search. The resulting locus trees were mapped onto the species tree to infer numbers of duplications and losses, gene copy number at internal branches and net copy number change using Notung v.2.8 (Stolzer et al. 2012) in phylogenomics mode (–phylogenomics).

In order to test significance of parallel gene space expansion among EM lineages, we implemented a sampling protocol to generate background distributions of expected copy number increase for orthogroups amplified on specific sets of branches (fig. 4A). To achieve this, we randomly placed the numbers of duplications and losses inferred for each branch among the orthogroups present in the ancestor of the branch. Each orthogroup was represented by the number of members in the ancestor. In order to reflect the increasing and decreasing mutational target size for each orthogroup with each duplication and loss drawn, the number of orthogroup members were adjusted dynamically. The code has been made available on Github (https://github.com/JackyHess/Amanita_comparative_genomics; last accessed September 27, 2018)

Functional Enrichment Analysis

Enrichment of protein domains among gene sets of interested were calculated using a Fisher's exact test with Benjamini–Hochberg correction for multiple testing. GO term enrichments were calculated using topGO with the weight01 algorithm and a *P* value threshold of 0.05 (Alexa et al. 2006). "EM early" enrichment analyses were run at the level of orthogroups, by assigning all unique annotations found in genes belonging to that orthogroup. This set included orthogroups with inferred expansions in the MRCA of EM *Amanita*, as well as orthogroups originating at this node. "EM late" enrichment analyses include conserved gene families with inferred expansions and lineage-specific clusters not included in the phylogenomic analysis.

Data Availability

RNA-seq data generated for this study were deposited at the Sequence Read Archive under accession SRP157961. Updated genome assemblies for all non-JGI genomes were deposited in NCBI under BioProject accession PRJNA485427. Genome annotations, orthogroup alignments, trees inferred for the phylogenomic analysis, and a database flat file containing functional annotations, orthogroup membership, genomic context, gene ages, and secretome predictions for all predicted proteomes were deposited in Dryad under accession doi:10.5061/dryad.g63c748.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by the National Science Foundation award number 1021606 to A.P. and Research Council of Norway project 221840 to I.S. Computational work was performed on the Abel Cluster, owned by the University of Oslo and Uninett/Sigma2, and operated by the Department for Research Computing at USIT, the University of Oslo IT-department. http://www.hpc.uio.no/. We would like to thank Benjamin Wolfe for storage and maintenance of RNAseq samples, as well as Laszlo Nagy and two anonymous reviewers for their helpful comments on this manuscript.

References

Aird SD, Arora J, Barua A, Qiu L, Terada K, Mikheyev AS. 2017. Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol Evol.* 9(10):2640–2649.

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22(13):1600–1607.

Alkan C, Sajjadian S, Eichler EE. 2011. Limitations of next-generation genome sequence assembly. *Nat Methods* 8(1):61–65.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389–3402.

Baker AC. 2003. Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of symbiodinium. *Annu Rev Ecol Evol Syst.* 34(1):661–689.

Balasundaram SV, Hess J, Durling MB, Moody SC, Thorbek L, Progida C, LaButti K, Aerts A, Barry K, Grigoriev IV, et al. 2018. The fungus that came in from the cold: dry rot's pre-adapted ability to invade buildings. *ISME J.* 19:535.

Bao D, Gong M, Zheng H, Chen M, Zhang L, Wang H, Jiang J, Wu L, Zhu Y, Zhu G, et al. 2013. Sequencing and comparative analysis of the straw mushroom (*Volvariella volvacea*) genome. *PLoS One* 8(3):e58294.

Bas C. 1969. Morphology and subdivision of *Amanita* and a monograph of its section Lepidella. *Persoonia* 5:285–573.

Bennati-Granier C, Garajova S, Champion C, Grisel S, Haon M, Zhou S, Fanuel M, Ropartz D, Rogniaux H, Gimbert I, et al. 2015. Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*. *Biotechnol Biofuels* 8:90.

Bittleston LS, Pierce NE, Ellison AM, Pringle A. 2016. Convergence in multispecies interactions. *Trends Ecol Evol.* 31(4):269–280.

Boscaro V, Kolisko M, Felletti M, Vannini C, Lynn DH, Keeling PJ. 2017. Parallel genome reduction in symbionts descended from closely related free-living bacteria. *Nat Ecol Evol.* 1(8):1160–1167.

Bouarab K, Melton R, Peart J, Baulcombe D, Osbourn A. 2002. A saponin-detoxifying enzyme mediates suppression of plant defences. *Nature* 418(6900):889–892.

Bronstein JL, Alarcón R, Geber M. 2006. The evolution of plant-insect mutualisms. *New Phytol.* 172(3):412–428.

Cannon SB, Mitra A, Baumgarten A, Young ND, May G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC Plant Biol*. 4:10.

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

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.

Chaib De Mares M, Hess J, Floudas D, Lipzen A, Choi C, Kennedy M, Grigoriev IV, Pringle A. 2015. Horizontal transfer of carbohydrate metabolism genes into ectomycorrhizal *Amanita*. *New Phytol*. 205(4):1552–1564.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res*. 14(6):1147–1159.

Courty PE, Hoegger PJ, Kilaru S, Kohler A, Buée M, Garbaye J, Martin F, Kües U. 2009. Phylogenetic analysis, genomic organization, and expression analysis of multi-copper oxidases in the ectomycorrhizal basidiomycete *Laccaria bicolor*. *New Phytol*. 182(3):736–750.

Črešnar B, Petrič Š. 2011. Cytochrome P450 enzymes in the fungal kingdom. *Biochim Biophys Acta* 1814(1):29–35.

Daniel B, Konrad B, Toplak M, Lahham M, Messenlehner J, Winkler A, Macheroux P. 2017. The family of berberine bridge enzyme-like enzymes: a treasure-trove of oxidative reactions. *Arch Biochem Biophys*. 632:88–103.

Daniel B, Pavkov-Keller T, Steiner B, Dordic A, Gutmann A, Nidetzky B, Sensen CW, van der Graaff E, Wallner S, Gruber K, et al. 2015. Oxidation of monolignols by members of the berberine bridge enzyme family suggests a role in plant cell wall metabolism. *J Biol Chem*. 290(30):18770–18781.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.

Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, Hahn MW. 2014. Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Comput Biol*. 10(12):e1003998.

Doré J, Kohler A, Dubost A, Hundley H, Singan V, Peng Y, Kuo A, Grigoriev IV, Martin F, Marmeisse R, et al. 2017. The ectomycorrhizal basidiomycete *Hebeloma cylindrosporum* undergoes early waves of transcriptional reprogramming prior to symbiotic structures differentiation. *Environ Microbiol*. 19:1338–1354.

Drini S, Criscuolo A, Lechat P, Imamura H, Skalický T, Rachidi N, Lukeš J, Dujardin J-C, Späth GF. 2016. Species- and strain-specific adaptation of the hsp70 super family in pathogenic trypanosomatids. *Genome Biol Evol*. 8(6):1980–1995.

Duplessis S, Courty P-E, Tagu D, Martin F. 2005. Transcript patterns associated with ectomycorrhiza development in *Eucalyptus globulus* and *Pisolithus microcarpus*. *New Phytol*. 165(2):599–611.

Dyrka W, Lamacchia M, Durrens P, Kobe B, Daskalov A, Paoletti M, Sherman DJ, Saupe SJ. 2014. Diversity and variability of NOD-like receptors in fungi. *Genome Biol Evol*. 6(12):3137–3158.

Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M, et al. 2011. The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* 333(6043):762–765.

Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Inform*. 23(1):205–211.

Emanuelsson O, Brunak S, Heijne von G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*. 2(4):953–971.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otillar R, Spatafora JW, Yadav JS, et al. 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* 336(6089):1715–1719.

Floudas D, Held BW, Riley R, Nagy LG, Koehler G, Ransdell AS, Younus H, Chow J, Chiniquy J, Lipzen A, et al. 2015. Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. *Fungal Genet Biol*. 76:78–92.

Fraiture A, Giangregorio MD. 2013. *Amanita inopinata*, its ecology and expansion in Europe. *Cryptogamie, Mycologie* 34(3):211–222.

Garcia K, Doidy J, Zimmermann SD, Wipf D, Courty P-E. 2016. Take a trip through the plant and fungal transportome of Mycorrhiza. *Trends Plant Sci*. 21(11):937–950.

Gluck-Thaler E, Slot JC. 2018. Specialized plant biochemistry drives gene clustering in fungi. *ISME J*. 2:510.

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A*. 108(4):1513–1518.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 29(7):644–652.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol*. 9(1):R7.

Hastings PJ, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy number. *Nat Rev Genet*. 10(8):551–564.

Hauschild K, Pauli HH, Kutchan TM. 1998. Isolation and analysis of a gene bbe1 encoding the berberine bridge enzyme from the California poppy *Eschscholzia californica*. *Plant Mol Biol*. 36(3):473–478.

He Z, Zhang H, Gao S, Lercher MJ, Chen W-H, Hu S. 2016. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res*. 44(W1):W236–W241.

Hess J, Pringle A. 2014. The natural histories of species and their genomes. In: Fungi. Francis M. Martin, editor. Vol. 70. Advances in Botanical Research. Elsevier. pp. 235–257.

Hess J, Skrede I, Wolfe BE, LaButti K, Ohm RA, Grigoriev IV, Pringle A. 2014. Transposable element dynamics among asymbiotic and ectomycorrhizal Amanita fungi. *Genome Biol Evol*. 6(7):1564–1578.

Horn S, Vaaje-Kolstad G, Westereng B, Eijsink VG. 2012. Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels* 5(1):45.

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*. 35(Web Server):W585–W587.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*. 44(D1):D286–D293.

Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJA. 2006. The PROSITE database. *Nucleic Acids Res*. 34(Database issue):D227–D230.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.

Kelkar YD, Ochman H. 2012. Causes and consequences of genome expansion in fungi. *Genome Biol Evol*. 4(1):13–23.

Knapp DG, Németh JB, Barry K, Hainaut M, Henrissat B, Johnson J, Kuo A, Lim JHP, Lipzen A, Nolan M, et al. 2018. Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. *Sci Rep*. 8(1):6321.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, et al. 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet*. 47(4):410–415.

Kohler A, Martin F. 2016. The evolution of the mycorrhizal lifestyles – a genomic perspective. In: Molecular Mycorrhizal Symbiosis. Hoboken (NJ): John Wiley & Sons, Inc. pp. 87–106.

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59.

Kosti I, Mandel-Gutfreund Y, Glaser F, Horwitz BA. 2010. Comparative analysis of fungal protein kinases and associated domains. *BMC Genomics* 11:133.

Krogh A, Larsson B, Heijne von G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 305(3):567–580.

Kües U, Ruhl M. 2011. Multiple multi-copper oxidase gene families in Basidiomycetes – what for? *Curr Genomics* 12(2):72–94.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* 6(1):41.

Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol*. 6(10):776–788.

Liao D. 1999. Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet*. 64(1):24–30.

Liao HL, Chen Y, Bruns TD, Peay KG, Taylor JW, Branco S, Talbot JM, Vilgalys R. 2014. Metatranscriptomic analysis of ectomycorrhizal roots reveals genes associated with Piloderma-Pinussymbiosis: improved methodologies for assessing gene expression in situ. *Environ Microbiol*. 16(12):3730–3742.

Liao H-L, Chen Y, Vilgalys R. 2016. Metatranscriptomic study of common and host-specific patterns of gene expression between pines and their symbiotic ectomycorrhizal fungi in the genus *Suillus*. *PLoS Genet*. 12:e1006348.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42(Database issue):D490–D495.

Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A*. 102(30):10557–10562.

Lutzoni F, Pagel M. 1997. Accelerated evolution as a consequence of transitions to mutualism. *Proc Natl Acad Sci U S A*. 94(21):11422–11427.

Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302(5649):1401–1404.

Martin F, Aerts A, Ahrén D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, et al. 2008. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452(7183):88–92.

Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, et al. 2010. P|[eacute]|rigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464(7291):1033–1038.

Martin F, Kohler A, Murat C, Veneault-Fourrey C, Hibbett DS. 2016. Unearthing the roots of ectomycorrhizal symbioses. *Nat Rev Microbiol*. 14(12):760–773.

Masson-Boivin C, Giraud E, Perret X, Batut J. 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol*. 17(10):458–466.

Matheny PB, Curtis JM, Hofstetter V, Aime MC, Moncalvo J-M, Ge Z-W, Slot JC, Ammirati JF, Baroni TJ, Bougher NL, et al. 2006. Major clades of Agaricales: a multilocus phylogenetic overview. *Mycologia* 98(6):982–995.

Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev*. 14(6):627–633.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A*. 93(7):2873–2878.

Murat C, Díez J, Luis P, Delaruelle C, Dupré C, Chevalier G, Bonfante P, Martin F. 2004. Polymorphism at the ribosomal DNA ITS and its relation to postglacial re-colonization routes of the Perigord truffle *Tuber melanosporum*. *New Phytol.* 164(2):401–411.

Nagendran S, Hallen-Adams HE, Paper JM, Aslam N, Walton JD. 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(5):427–435.

Newton ILG, Bordenstein SR. 2011. Correlations between bacterial ecology and mobile DNA. *Curr Microbiol*. 62(1):198–208.

Oakeson KF, Gil R, Clayton AL, Dunn DM, Niederhausern von AC, Hamil C, Aoyagi A, Duval B, Baca A, Silva FJ, et al. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol Evol.* 6(1):76–93.

O'Fallon B. 2008. Population structure, levels of selection, and the evolution of intracellular symbionts. *Evolution* 62(2):361–373.

Papanicolaou A. 2013. Just annotate my genome. Available from: http://jamg.sourceforge.net/; last accessed September 27, 2018.

Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. 2015. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Front Microbiol.* 6:1278.

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

Petersen TN, Brunak S, Heijne von G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786.

Plett JM, Daguerre Y, Wittulsky S, Vayssières A, Deveau A, Melton SJ, Kohler A, Morrell-Falvey JL, Brun A, Veneault-Fourrey C, et al. 2014. Effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proc Natl Acad Sci U S A*. 111(22):8299–8304.

Plett JM, Kemppainen M, Kale SD, Kohler A, Legué V, Brun A, Tyler BM, Pardo AG, Martin F. 2011. A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr Biol*. 21(14):1197–1203.

Plett JM, Tisserant E, Brun A, Morin E, Grigoriev IV, Kuo A, Martin F, Kohler A. 2015. The mutualist *Laccaria bicolor* expresses a core gene regulon during the colonization of diverse host plants and a variable regulon to counteract host-specific defenses. *Mol Plant Microbe Interact*. 28(3):261–273.

Qhanya LB, Matowane G, Chen W, Sun Y, Letsimo EM, Parvez M, Yu J-H, Mashele SS, Syed K. 2015. Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in basidiomycete biotrophic plant pathogens. *PLoS One* 10(11):e0142100.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

Raffaele S, Win J, Cano LM, Kamoun S. 2010. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC Genomics* 11:637.

Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, Levasseur A, Lombard V, Morin E, Otillar R, et al. 2014. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proc Natl Acad Sci U S A*. 111(27):9923–9928.

Rineau F, Roth D, Shah F, Smits M, Johansson T, Canbäck B, Olsen PB, Persson P, Grell MN, Lindquist E, 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(6):1477–1487.

Rubin BER, Moreau CS. 2016. Comparative genomics reveals convergent rates of evolution in ant-plant mutualisms. *Nat Commun.* 7:12679.

Rytioja J, Hildén K, Yuzon J, Hatakka A, de Vries RP, Mäkelä MR. 2014. Plant-polysaccharide-degrading enzymes from Basidiomycetes. *Microbiol Mol Biol Rev*. 78(4):614–649.

Sabbadin F, Hemsworth GR, Ciano L, Henrissat B, Dupree P, Tryfona T, Marques RDS, Sweeney ST, Besser K, Elias L, et al. 2018. An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. *Nat Commun.* 9(1):756.

Selosse M-A, Schneider-Maunoury L, Martos F. 2018. Time to re-think fungal ecology? Fungal ecological niches are often prejudged. *New Phytol*. 217(3):968–972.

Shah F, Nicolás C, Bentzer J, Ellström M, Smits M, Rineau F, Canbäck B, Floudas D, Carleer R, Lackner G, et al. 2016. Ectomycorrhizal fungi decompose soil organic matter using oxidative mechanisms adapted from saprotrophic ancestors. *New Phytol*. 209(4):1705–1719.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation

completeness with single-copy orthologs. *Bioinformatics* 31(19): 3210–3212.
Sipos G, Prasanna AN, Walter MC, O'Connor E, Bálint B, Krizsán K, Kiss B, Hess J, Varga T, Slot J, et al. 2017. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. *Nat Ecol Evol.* (12):515.
Smit A, Hubley R, Green P. 2015. *RepeatMasker Open-4.0.* Available from: http://www.repeatmasker.org; last accessed September 27, 2018.
Smith GR, Finlay RD, Stenlid J, Vasaitis R, Menkis A. 2017. Growing evidence for facultative biotrophy in saprotrophic fungi: data from microcosm tests with 201 species of wood-decay basidiomycetes. *New Phytol*. 215(2):747–755.
Smith SE, Read DJ. 2008. Mycorrhizal symbiosis. 3rd ed. Cambridge (MA): Academic Press.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9): 1312–1313.
Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24(5):637–644.
Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28(18):i409–i415.
Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6(7): e21800.
Syed K, Shale K, Pagadala NS, Tuszynski J. 2014. Systematic identification and evolutionary analysis of catalytically versatile cytochrome p450 monooxygenase families enriched in model basidiomycete fungi. *PLoS One* 9(1):e86683.
Tedersoo L, May TW, Smith ME. 2010. Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza* 20(4):217–263.
Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. 18(12): 1979–1990.
Testa AC, Hane JK, Ellwood SR, Oliver RP. 2015. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* 16:170.
Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, Frei dit Frey N, Gianinazzi-Pearson V, et al. 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc Natl Acad Sci U S A*. 110(50):20117–20122.
Tschaplinski TJ, Plett JM, Engle NL, Deveau A, Cushman KC, Martin MZ, Doktycz MJ, Tuskan GA, Brun A, Kohler A, et al. 2014. *Populus trichocarpa* and *Populus deltoides* exhibit different metabolomic responses to colonization by the symbiotic fungus Laccaria bicolor. *Mol Plant Microbe Interact*. 27(6):546–556.
Várnai A, Mäkelä MR, Djajadi DT, Rahikainen J, Hatakka A, Viikari L. 2014. Carbohydrate-binding modules of fungal cellulases: occurrence in nature, function, and relevance in industrial biomass conversion. *Adv Appl Microbiol*. 88:103–165.
Wernegreen JJ, Richardson AO, Moran NA. 2001. Parallel acceleration of evolutionary rates in symbiont genes underlying host nutrition. *Mol Phylogenet Evol*. 19(3):479–485.
Wernegreen JJ. 2015. Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann N Y Acad Sci*. 1360:16–35.
Wolf YI, Koonin EV. 2013. Genome reduction as the dominant mode of evolution. *Bioessays* 35(9):829–837.
Wolfe BE, Kuo M, Pringle A. 2012. *Amanita thiersii* is a saprotrophic fungus expanding its range in the United States. *Mycologia* 104(1):22–33.
Wolfe BE, Tulloss RE, Pringle A. 2012. The irreversible loss of a decomposition pathway marks the single origin of an ectomycorrhizal symbiosis. *PLoS One* 7(7):e39597.
Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21(9):1859–1875.
Wu Y-C, Rasmussen MD, Bansal MS, Kellis M. 2013. TreeFix: statistically informed gene tree error correction using species trees. *Syst Biol*. 62(1):110–120.
Wu Y-C, Rasmussen MD, Bansal MS, Kellis M. 2014. Most parsimonious reconciliation in the presence of gene duplication, loss, and deep coalescence using labeled coalescent trees. *Genome Res*. 24(3):475–486.
Zhong Y, Jia Y, Gao Y, Tian D, Yang S, Zhang X. 2013. Functional requirements driving the gene duplication in 12 Drosophila species. *BMC Genomics* 14:555.