

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/158800> since 2016-08-05T09:45:08Z

Published version:

DOI:10.1038/ng.3223

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



UNIVERSITÀ DEGLI STUDI DI TORINO

This is an author version of the contribution published on:

Questa è la versione dell'autore dell'opera:

Nature Genetics, Volume 47, Issue 4, 2015, pp 410-415

DOI:10.1038/ng.3223

The definitive version is available at:

La versione definitiva è disponibile alla URL:

<http://www.nature.com/ng/journal/v47/n4/full/ng.3223.html>

Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists

Annegret Kohler^{1,2,25}, Alan Kuo^{3,25}, Laszlo G Nagy^{4,5,25}, Emmanuelle Morin^{1,2,25}, Kerrie W Barry³, Francois Buscot^{6,7}, Björn Canbäck⁸, Cindy Choi³, Nicolas Cichocki^{1,2}, Alicia Clum³, Jan Colpaert⁹, Alex Copeland³, Mauricio D Costa¹⁰, Jeanne Doré¹¹, Dimitrios Floudas⁴, Gilles Gay¹¹, Mariangela Girlanda¹², Bernard Henrissat^{13–15}, Sylvie Herrmann^{6,7}, Jaqueline Hess¹⁶, Nils Högberg¹⁷, Tomas Johansson⁸, Hassine-Radhouane Khouja¹², Kurt LaButti³, Urs Lahrmann¹⁸, Anthony Levasseur¹³, Erika A Lindquist³, Anna Lipzen³, Roland Marmeisse¹¹, Elena Martino^{1,12}, Claude Murat^{1,2}, Chew Y Ngan³, Uwe Nehls¹⁹, Jonathan M Plett^{1,2}, Anne Pringle²⁰, Robin A Ohm³, Silvia Perotto¹², Martina Peter²¹, Robert Riley³, Francois Rineau^{9,24}, Joske Ruytinx⁹, Asaf Salamov³, Firoz Shah⁸, Hui Sun³, Mika Tarkka^{6,7}, Andrew Tritt³, Claire Veneault-Fourrey^{1,2}, Alga Zuccaro^{18,22}, Mycorrhizal Genomics Initiative Consortium²³, Anders Tunlid⁸, Igor V Grigoriev^{3,26}, David S Hibbett^{4,26} & Francis Martin^{1,2,26}

To elucidate the genetic bases of mycorrhizal lifestyle evolution, we sequenced new fungal genomes, including 13 ectomycorrhizal (ECM), orchid (ORM) and ericoid (ERM) species, and five saprotrophs, which we analyzed along with other fungal genomes. Ectomycorrhizal fungi have a reduced complement of genes encoding plant cell wall-degrading enzymes (PCWDEs), as compared to their ancestral wood decayers. Nevertheless, they have retained a unique array of PCWDEs, thus suggesting that they possess diverse abilities to decompose lignocellulose. Similar functional categories of nonorthologous genes are induced in symbiosis. Of induced genes, 7–38% are orphan genes, including genes that encode secreted effector-like proteins. Convergent evolution of the mycorrhizal habit in fungi occurred via the repeated evolution of a 'symbiosis toolkit', with reduced numbers of PCWDEs and lineage-specific suites of mycorrhiza-induced genes.

Main

Fungi are often described as either saprotrophs, which degrade complex organic substrates, or biotrophs, which obtain carbon compounds from living hosts. Among the latter, ECM fungi provide crucial ecological services in interacting with forest trees. They are portrayed as mutualists trading host photoassimilates for nutrients and having limited capacity to decompose soil lignocellulose^{1, 2, 3}, as a result of their reduced repertoire of PCWDEs^{4, 5, 6}. However, recent studies are challenging this view^{7, 8, 9, 10}. An improved understanding of the ability of ECM fungi to decompose lignocellulose is needed to resolve mechanisms of nutrient cycling in forests. The ECM lifestyle in *Laccaria bicolor* is associated with the expression of new mycorrhiza-induced small secreted proteins (MiSSPs) that are required for establishment of symbiosis^{11, 12}. Mycorrhizal symbioses have arisen repeatedly during fungal evolution and include not only ECM associations but also those with ERM and ORM mycorrhizae¹³. It is not known whether these symbioses share the genomic features found in *L. bicolor*⁴ and *Tuber melanosporum*⁵. Here we assess whether there are general evolutionary and functional properties of mycorrhizal Basidiomycota and Ascomycota, using comparative analysis of 18 newly sequenced genomes ([Supplementary Tables 1 and 2](#)), including 13 from ECM, ERM and ORM fungal symbionts.

Genomes of mycorrhizal fungi range in size from approximately 38 to 125 Mb, and their predicted gene contents range from approximately 15,000 to 23,000 genes (Supplementary Tables 3 and 4). We combined the 18 new genomes with 31 others and performed phylogenomic analyses (Supplementary Note, Supplementary Tables 1 and 5–10 and Supplementary Figs. 1–23). An organismal phylogeny resolved six to eight independent lineages of ECM and two lineages of ORM symbionts in Agaricomycetes as well as two mycorrhizal lineages in Ascomycota (Fig. 1). Molecular-clock analyses have suggested that the Pezizomycotina and Agaricomycetes are approximately 400 and 300 million years old, respectively. In contrast, the Pinaceae (which may be the oldest ECM hosts), rosids and orchids are of Jurassic or Cretaceous origin (~100 million years ago), according to fossil and molecular-clock evidence; this is consistent with the view that ECM and other mycorrhizal associations have evolved repeatedly across Dikarya¹⁴.

Figure 1: Evolution of mycorrhizal symbiosis inferred from 49 fungal genomes.

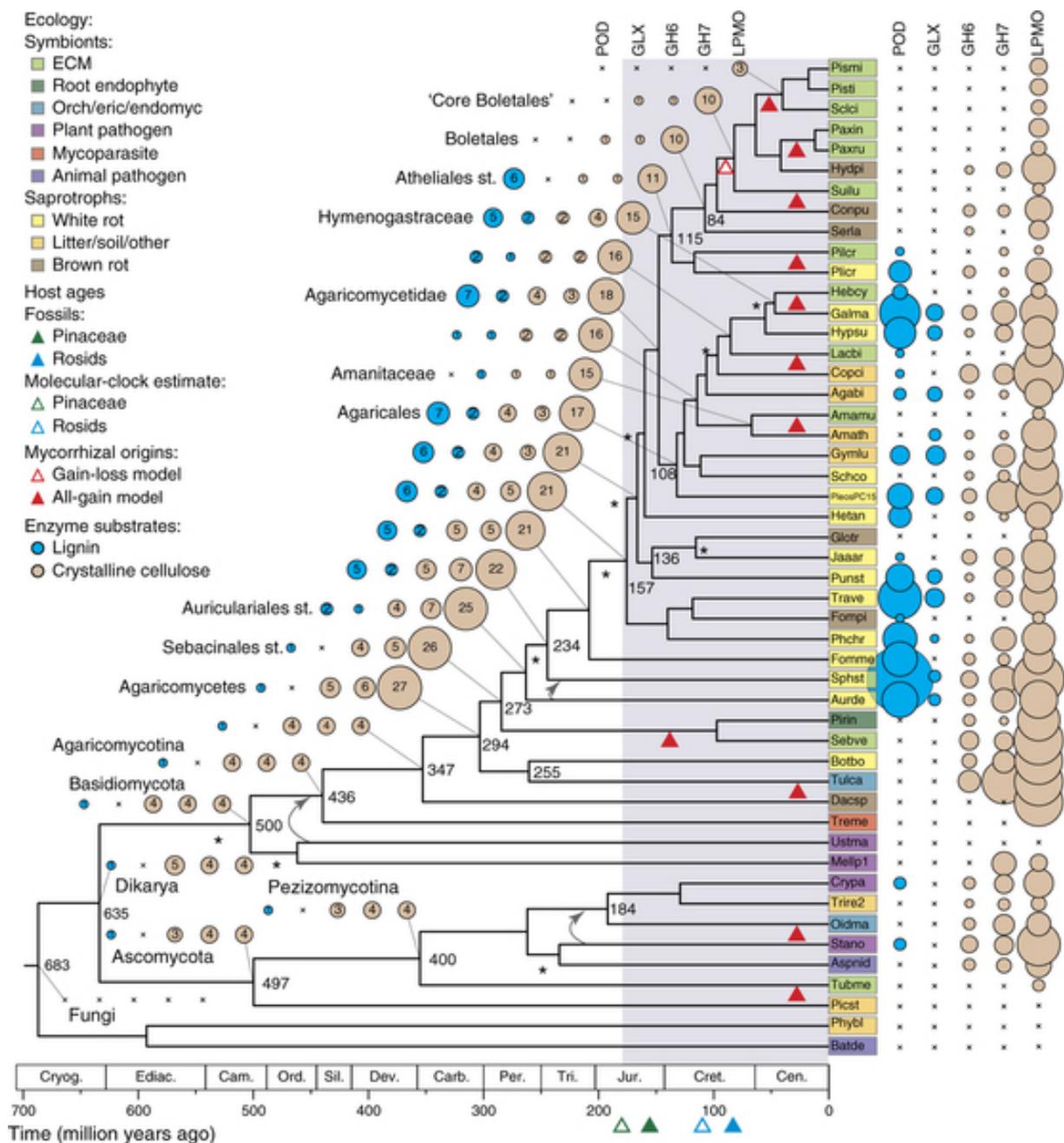


Figure 1: The tree is a chronogram estimated with r8s on the basis of a maximum-likelihood phylogeny inferred with RAxML. Nodes receiving less than maximal support in all analyses are indicated with asterisks. Curved arrows indicate alternate placements for Ustilaginomycotina and Auriculariales ([Supplementary Fig. 2](#)). Mean ages (ma) are indicated adjacent to selected nodes. Circles indicate observed (right of tree) and reconstructed (left) copy numbers for selected genes encoding enzymes involved in decay of lignin (POD and GLX; blue circles) or crystalline cellulose (GH6, GH7 and LPMO; beige circles). Absence of gene copies is indicated with 'x'. Areas of circles are proportional to gene copy numbers. (Copy numbers are indicated for internal nodes; gene counts in terminal taxa are shown in [Supplementary Table 9](#).) Selected clades are labeled at internal nodes; 'st.' indicates the stem node for a taxon. Shading of terminal taxon names indicates nutritional modes (as shown in the key). Solid red triangles, estimated origins of ECM or ERM and ORM mycorrhizal symbioses; unfilled red triangle, alternate reconstruction with a single origin of ECM in Boletales and at least one reversal to saprotrophy; colored triangles below the geological timescale, ages of major ECM hosts based on fossils (solid triangles) and molecular-clock estimates (unfilled triangles); light-gray shading, temporal period when the origins of ECM are most plausible. Orch/eric/endomyc, orchid, ericoid or endomycorrhizae; litter/soil/other, litter, soil or other saprotroph; Cryog., Cryogenian; Ediac., Ediacaran; Cam., Cambrian; Ord., Ordovician; Sil., Silurian; Dev., Devonian; Carb., Carboniferous; Per., Permian; Tri., Triassic; Jur., Jurassic; Cret., Cretaceous; Cen., Cenozoic. Abbreviations for taxon names are defined in [Supplementary Note](#).

The first three sequenced mycorrhizal genomes, i.e., the ECM *L. bicolor* and *T. melanosporum*^{4, 5} and the arbuscular mycorrhizal fungus *Rhizophagus irregularis*¹⁵, demonstrated substantial losses in PCWDE genes. To assess the diversity of lignocellulose-decay capabilities of other mycorrhizal lineages, we cataloged 28 gene families and subfamilies encoding oxidoreductases and carbohydrate-active enzymes (CAZymes) associated with plant cell wall degradation ([Fig. 2](#), [Supplementary Note](#), [Supplementary Tables 7, 8, 9, 10](#) and [Supplementary Fig. 3](#)). Among Agaricomycetes, saprotrophs (including white-rot decayers, which can degrade lignin as well as cellulose), *Botryobasidium botryosum*, *Jaapia argillacea* and soil and litter saprotrophs possess on average 133 gene copies across all PCWDE families, whereas brown-rot species (which lack the ability to degrade lignin) possess 81 gene copies, and ECM species maintain 62 gene copies, on average. Thus, the evolution of ECM parallels that of brown-rot lineages, with both guilds having lost much of the plesiomorphic enzymatic apparatus of white rots ([Supplementary Note](#)).

Figure 2: Summary of lignocellulose-degrading oxidoreductases, carbohydrate-active enzymes (CAZymes) and cellulose-binding modules in 49 fungal genomes.

| Phylum | Subphylum | Species | Ecology 1 | Ecology 2 | Oxidoreductases | CAZymes | CBM1 | |
|-----------------|------------------------------------|-----------------------------------|---------------------------------------|-----------------|---------------------------------|---------|------|----|
| Basidiomycota | Agaricomycotina | <i>Amanita muscaria</i> | Biotroph | Ectomycorrhizal | 29 | 28 | 0 | |
| | | <i>Hebeloma cylindrosporium</i> | | | 22 | 39 | 2 | |
| | | <i>Laccaria bicolor</i> | | | 34 | 47 | 1 | |
| | | <i>Paxillus involutus</i> | | | 23 | 39 | 0 | |
| | | <i>Paxillus rubicundulus</i> | | | 24 | 35 | 0 | |
| | | <i>Piloderma croceum</i> | | | 41 | 50 | 0 | |
| | | <i>Pisolithus microcarpus</i> | | | 17 | 29 | 0 | |
| | | <i>Pisolithus tinctorius</i> | | | 22 | 26 | 0 | |
| | | <i>Scleroderma citrinum</i> | | | 24 | 25 | 0 | |
| | | <i>Suillus luteus</i> | | | 29 | 40 | 0 | |
| | | <i>Tremella mesenterica</i> | | | 5 | 14 | 0 | |
| | | <i>Sebacina vermifera</i> | | | 21 | 121 | 50 | |
| | | <i>Tulasnella calospora</i> | | | 9 | 178 | 110 | |
| | | <i>Piniformospora indica</i> | | | 15 | 103 | 64 | |
| | | <i>Auricularia delicata</i> | | | 60 | 140 | 42 | |
| | | <i>Botryobasidium botryosum</i> | | | 16 | 97 | 29 | |
| | | <i>Fomitiporia mediterranea</i> | | | 40 | 81 | 10 | |
| | | <i>Galerina marginata</i> | | | 77 | 128 | 52 | |
| | | <i>Heterobasidion annosum</i> | | | 35 | 65 | 17 | |
| | <i>Hypholoma sublateritium</i> | 52 | 83 | 28 | | | | |
| | <i>Jaapia argillacea</i> | 17 | 102 | 24 | | | | |
| | <i>Phanerochaete chrysosporium</i> | 31 | 75 | 28 | | | | |
| | <i>Pleurotus ostreatus</i> | 43 | 111 | 31 | | | | |
| | <i>Plicaturopsis crispa</i> | 21 | 70 | 9 | | | | |
| | <i>Punctularia strigosozonata</i> | 46 | 93 | 26 | | | | |
| | <i>Schizophyllum commune</i> | 8 | 103 | 5 | | | | |
| | <i>Sphaerobolus stellatus</i> | (284) | (195) | 0 | | | | |
| | <i>Trametes versicolor</i> | 50 | 94 | 23 | | | | |
| | <i>Coniophora puteana</i> | 16 | 80 | 3 | | | | |
| | <i>Dacryopinax</i> sp. | 14 | 50 | 1 | | | | |
| | <i>Fomitopsis pinicola</i> | 16 | 65 | 0 | | | | |
| | <i>Gloeophyllum trabeum</i> | 13 | 61 | 1 | | | | |
| | <i>Hydnomerulius pinastri</i> | 21 | 86 | 16 | | | | |
| | <i>Serpula lacrymans</i> | 12 | 52 | 8 | | | | |
| | <i>Agaricus bisporus</i> | 47 | 65 | 13 | | | | |
| | <i>Amanita thiersii</i> | 27 | 71 | 10 | | | | |
| | <i>Coprinopsis cinerea</i> | 41 | 96 | 44 | | | | |
| | <i>Gymnopus luxurians</i> | 66 | 118 | 32 | | | | |
| | Ustilaginomycotina | <i>Ustilago maydis</i> | 6 | 22 | 0 | | | |
| | Pucciniomycotina | <i>Melampsora larici-populina</i> | 31 | 71 | 0 | | | |
| | Ascomycota | Pezizomycotina | <i>Tuber melanosporum</i> | Biotroph | Ectomycorrhizal | 5 | 19 | 3 |
| | | | <i>Oidiodendron maius</i> | | Ericoid symbiont | 31 | 119 | 35 |
| | | | <i>Cryphonectria parasitica</i> | | Plant pathogen | 18 | 102 | 12 |
| | | | <i>Stagonospora nodorum</i> | | Plant pathogen | 16 | 104 | 10 |
| | | | <i>Aspergillus nidulans</i> | | Animal pathogen | 9 | 83 | 6 |
| | | | <i>Trichoderma reesei</i> | | Soil, litter or other Saprotoph | 10 | 39 | 14 |
| | | Saccharomycotina | <i>Pichia stipitis</i> | Saprotoph | Soil, litter or other Saprotoph | 2 | 10 | 0 |
| NA | | Mucoromycotina | <i>Phycomyces blakesleeana</i> | NA | Soil, litter or other Saprotoph | 2 | 18 | 1 |
| Chytridiomycota | | NA | <i>Batrachochytrium dendrobatidis</i> | Biotroph | Animal pathogen | 3 | 5 | 0 |

Figure 2: Relative abundance of genes is represented by a color scale, from the minimum (blue) to maximum (red) number of copies per species (excluding *Sphaerobolus stellatus*; values in parentheses). Individual gene families are shown in [Supplementary Table 7](#). NA, not applicable.

To reconstruct the evolution of decay capabilities, we performed gene tree–species tree reconciliation for 16 gene families encoding PCWDEs ([Supplementary Note](#) and [Supplementary Figs. 4–23](#)). Here we focus on enzymes acting on lignin and crystalline cellulose, including class II peroxidases (PODs), glyoxal oxidase (GLX; an accessory enzyme that generates H₂O₂ as a substrate for PODs), GH6 and GH7 (cellobiohydrolases) and lytic polysaccharide monooxygenases (LPMOs) ([Supplementary Figs. 4–23](#)). Expansion of LPMO genes, from 4 copies in the ancestor of Basidiomycota to 27 copies in the ancestor of Agaricomycetes, preceded the initial duplication of ligninolytic POD genes and the origin of the GLX gene, which are localized to the lineage leading to the ancestor of Auriculariales ([Fig. 1](#)). From that point, multiple copies of genes encoding LPMOs, PODs and GLXs, as well as GH6 and GH7, are reconstructed in the 'backbone' nodes of the Agaricomycete phylogeny ([Fig. 1](#)), thus suggesting that white rot was a conserved trait in the early evolution of mushroom-forming fungi.

Although all ECM species have reduced complements of PCWDEs ([Figs. 1](#) and [2](#), and [Supplementary Figs. 3](#) and [12–23](#)), they have arisen from functionally diverse saprotrophic precursors and have retained distinct suites of PCWDEs. For example, within the Agaricales, the ECM *Amanita muscaria* is nested within a group of soil and litter saprotrophs, including *Amanita thiersii*, which lacks PODs, whereas another ECM species, *Hebeloma cylindrosporum*, is nested within a group of white-rot decayers that have multiple POD, GLX, GH6, GH7 and LPMO enzymes ([Fig. 1](#) and [Supplementary Figs. 12–16](#)). *A. muscaria* has lost genes encoding GLXs, GH6 and GH7, and it has a reduced complement of LPMO genes; both may be consequences of a shift to ECM, but the loss of POD genes in *A. muscaria* preceded (and therefore could not have been caused by) the evolution of the ECM habit. In contrast, *H. cylindrosporum* has three POD genes encoding manganese peroxidases ([Supplementary Fig. 4](#)), thus suggesting that it may possess ligninolytic capabilities similar to those in *Cortinarius glaucopus*¹⁰.

The most densely sampled clade of ECM species in our analysis, the Boletales, contains six ECM species that are nested within a paraphyletic assemblage of brown-rot wood decayers lacking PODs and GLXs ([Fig. 1](#)). The ECM Boletales have further lost all copies of GH6 and GH7 genes and have no more than five copies of LPMO genes, thus suggesting that they have limited capacity to degrade lignocellulose ([Fig. 1](#) and [Supplementary Figs. 22](#) and [23](#)). Gene tree–species tree reconciliations of LPMOs and other decay-related CAZymes suggest that parallel losses of brown-rot saprotrophy have occurred in the three ECM lineages in Boletales (Boletineae, Suillineae and Sclerodermatineae; [Supplementary Figs. 15–23](#)), although the ability to oxidize organic matter remains in *Paxillus involutus*⁹.

Cantharellales and Sebaciniales are early-diverging lineages of Agaricomycetes that include the ORM symbionts *Sebacina vermifera* and *Tulasnella calospora*. In contrast to ECM taxa in Boletales and Agaricales, symbiotic members of Cantharellales and Sebaciniales have a robust apparatus for degradation of crystalline cellulose, particularly *T. calospora*, which has 7 GH6, 27 GH7 and 33 LPMO genes (even more than its putatively saprotrophic sister taxon, *B. botryosum*) ([Fig. 1](#) and [Supplementary Figs. 15](#) and [16](#)). Moreover, proteins with a cellulose-binding domain (CBM1) are rarely detected in ECM, whereas they are abundant in the ORM and ERM symbionts ([Fig. 2](#) and [Supplementary Table 7](#)). Unlike typical ECM fungi, ORM fungi transfer carbohydrates to their hosts in the juvenile state of orchids, and they exploit nonliving organic substrates to feed their host¹⁶. Similarly, the ERM *Oidiodendron maius* (Ascomycota) maintains multiple copies of GH6, GH7 and LPMO genes ([Fig. 1](#) and [Supplementary Table 7](#)), thus explaining its saprotrophic ability in *Sphagnum* peat¹⁷. Reconciliation analyses suggest that the divergence of Cantharellales and Sebaciniales occurred before the diversification of PODs in Agaricomycetes and the origin of white rot ([Fig. 1](#)). The well-developed capacity to attack crystalline cellulose in these lineages and in the Ascomycota may reflect a primitive mode of symbiotic life substantially relying on the saprotrophic ability to decay nonwoody substrates with modest lignin content.

In *L. bicolor*, symbiosis requires lineage-specific genes encoding small secreted effector proteins that control host-plant development and immunity^{4, 11, 12}. By sequencing RNA from free-living mycelia and mycorrhizal roots, we identified genes regulated by symbiosis development in the ECM *A. muscaria*, *H. cylindrosporum*, *P. involutus*, *Piloderma croceum* and *Suillus luteus*, the ERM *O. maius* and the ORM *S. vermifera* and *T. calospora*. Of the expressed genes, 1.3–6.0% are upregulated during symbiosis and 0.2–4.9% are downregulated by the interaction (fold change >5, false discovery rate–corrected $P < 0.05$; [Supplementary Note](#) and [Supplementary Fig. 24](#)). Though genes belonging to similar

functional gene ontology (GO) categories (signaling, information storage and processing, and metabolism) were upregulated in these interactions ([Supplementary Fig. 25](#); downregulated genes in [Supplementary Fig. 26](#)), each species expressed a distinct set of genes thought to be involved in redox reactions, nutrient transport and metabolism. A large set of symbiosis-upregulated genes have orthologs in brown- and white-rot fungi (cluster III, [Fig. 3](#)), thus suggesting that they are not unique to mycorrhizal symbionts and tend to be associated with essential core metabolic pathways. These conserved mycorrhiza-induced genes might provide clues to how ancestral gene complements have been adapted to the ECM lifestyle. For example, the few retained PCWDEs acting on pectin (GH28, GH88 and CE8), hemicellulose (GH30) and cellulose (GH5_5 and LPMOs) are expressed in ECM root tips likely to modify the plant cell wall during colonization of the host root apoplastic space ([Supplementary Fig. 27](#)). In contrast, ORM and ERM symbionts expressed a full complement of PCWDEs in symbiosis ([Supplementary Figs. 27 and 28](#)), thus suggesting that they are used to penetrate host cells.

Figure 3: Presence and sequence similarity of symbiosis-upregulated genes from *L. bicolor* in 55 genomes of saprotrophic (white rots, brown rots, soil and litter decayers), mycorrhizal (ECM, ERM and ORM), pathogenic and endophytic fungi.

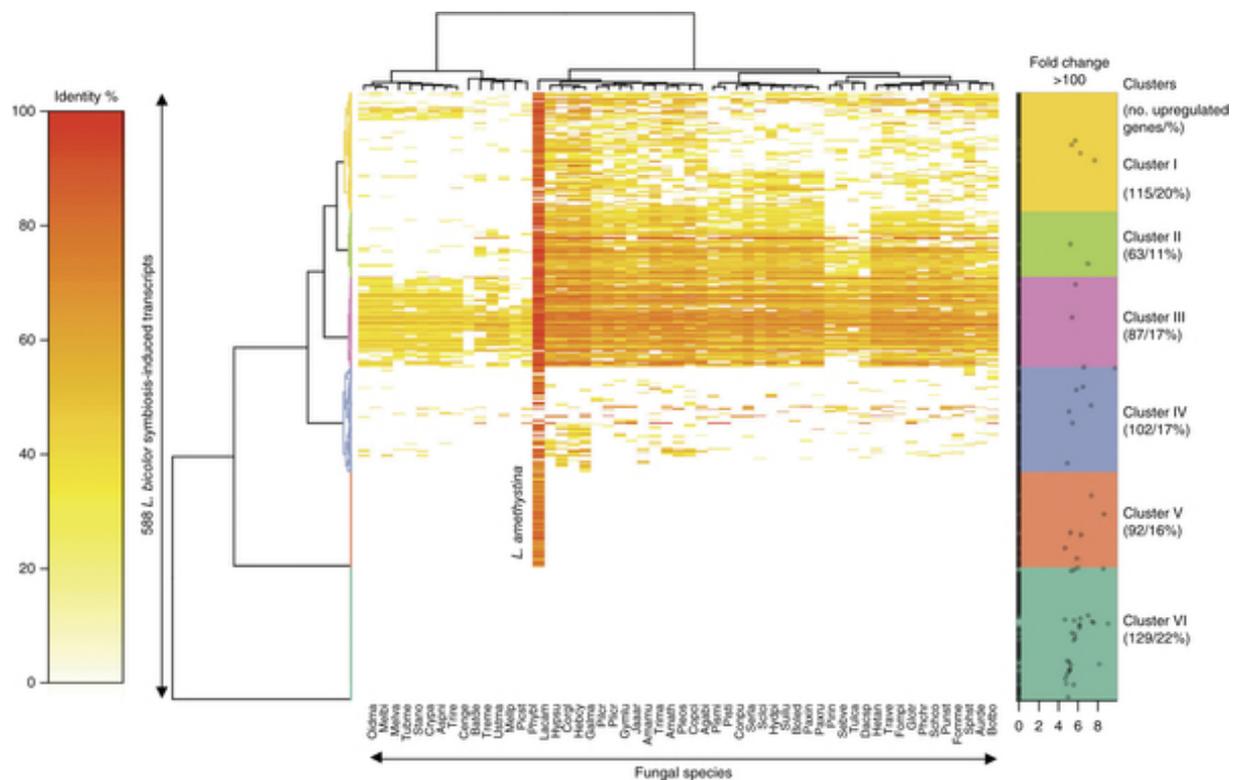


Figure 3: The heat map depicts a double-hierarchical clustering of 588 symbiosis-upregulated *L. bicolor* genes (rows, fold change >5, false discovery rate-corrected $P < 0.05$; [Supplementary Data Set 1](#)) based on their percentage sequence identity (color scale at left) with their orthologs (if any) in selected fungal species (columns). Right, color-coded gene clusters (clusters I to VI), with dots representing differentially expressed genes upregulated >100-fold; for genes with regulation less than 100-fold, the fold-change value is set to 1 (majority of the genes). The x axis is a logarithmic scale of the fold change in gene expression (natural algorithm with Euler's number). Genes of cluster VI are *L. bicolor*-specific (orphan) genes, whereas genes of cluster V have sequence similarity only with predicted proteins of the sister species *L. amethystina*. Double-hierarchical-clustering heat maps of symbiosis-upregulated genes for the ECM *A. muscaria*, *H. cylindrosporum*, *P. involutus*, *P. croceum*,

S. luteus, the orchid mycorrhizal *T. calospora*, the ericoid mycorrhizal *Oidiodendron maius*, the endophytic *S. vermifera* and the brown-rot *S. lacrymans* are shown in [Supplementary Figure 25](#), whereas double-hierarchical-clustering heat maps of symbiosis-downregulated genes are shown in [Supplementary Figure 26](#). Species abbreviations are in [Supplementary Table 1](#).

Many (7–38%) of the symbiosis-induced genes are restricted to a single ECM species (clusters V and VI, [Fig. 3](#); [Supplementary Fig. 25](#)), even in the densely sampled Boletales (23% in *S. luteus*; [Supplementary Fig. 25](#)). Only one-third of the *Laccaria* symbiosis-induced orphan genes have homologs in both *L. bicolor* and *Laccaria amethystina* ([Fig. 3](#)). The latter diverged from the *L. bicolor* lineage ~20 million years ago¹⁸, thus indicating that even after the evolution of the ECM habit species within a genus continued to develop a specific symbiosis protein 'toolkit' and to diverge from each other. Lineage-specific orphan genes may represent either ancestral genes that have diverged so far that their similarities to other sequences have been obscured, a phenomenon frequently observed for effector genes¹⁹, or genes formed *de novo* from previously noncoding sequences²⁰. Orphan fungal genes are over-represented among up- and downregulated genes found in mycorrhizal roots ([Figs. 3 and 4](#), and [Supplementary Figs. 25 and 26](#)). To assess whether previously uncharacterized but highly induced genes may represent candidates for secreted effector proteins (CSEPs), we assessed which and how many of these genes encode proteins that possess properties (for example, predicted signal peptide, size <300 amino acid residues and no sequence similarity) common to known secreted effectors, such as mycorrhiza-induced small secreted protein of 7 kDa (MiSSP7)^{11, 12}. Of the induced genes, 8–28% encode proteins with a predicted signal peptide, a significantly higher fraction than the 3–7% of noninduced and non-symbiosis-related genes that encode a protein with a signal peptide ($P = 2.04 \times 10^{-6}$ to $<2.20 \times 10^{-16}$, Fisher's exact test; [Fig. 4](#) and [Supplementary Fig. 25](#)). Except for *S. vermifera*, there is a significantly higher proportion of genes encoding secreted proteins of <300 amino acid residues (4–11% compared to 1–2% in the noninduced set; $P = 0.00333$ to $<2.20 \times 10^{-16}$, Fisher's exact test; [Fig. 4](#) and [Supplementary Fig. 25](#)).

Figure 4: Distribution of symbiosis-upregulated genes of clusters I to VI into functional categories, including MiSSPs and nonsecreted orphan genes (no KOG).

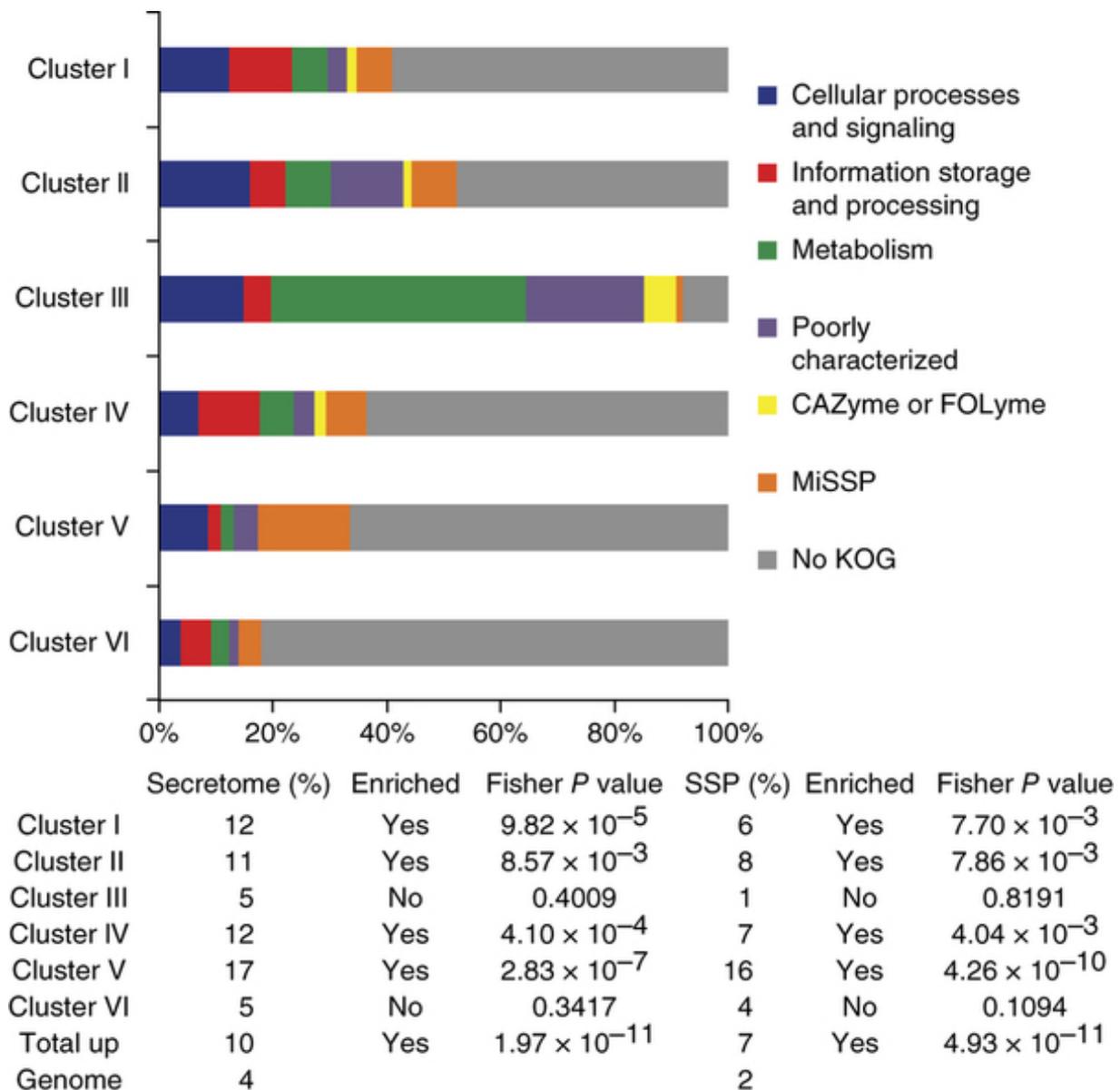


Figure 4: MiSSPs represent 16% of the symbiosis-upregulated transcripts in cluster V, a significant enrichment compared to their percentage in the total gene repertoire (i.e., 2%). CAZyme, carbohydrate-active enzymes; FOLyme, fungal oxidative lignin-degrading enzymes.

Serpula lacrymans is a member of a paraphyletic clade of brown-rot lineages from which the ECM Boletales are derived (Fig. 1). *S. lacrymans* is able to form a loose hyphal mantle around pine roots, and this may provide clues to the antecedent condition of ECM²¹. Transcript profiling of the root-associated mycelium showed a significant proportion of root-induced transcripts, but we detected no small secreted proteins, whereas ECM Boletales expressed a large set of MiSSPs (Supplementary Fig. 25). Although proteins known to interact with host immunity and development, such as MiSSP7, are not the only functional class to be over-represented among lineage-specific short secreted proteins, the prevalence of CSEPs among symbiosis-induced genes suggests that at least some of these genes encode new symbiosis-related effectors.

Mycorrhizal symbioses evolved from ecologically diverse decayer precursors and radiated in parallel, following the origins of their host-plant lineages ([Fig. 1](#)). Polyphyletic evolution of the ECM lifestyle is marked not only by convergent losses of different components of the ancestral saprotrophic apparatus but also by rapid genetic turnover in symbiosis-induced genes ([Fig. 3](#)), some of which may reflect lineage-specific functional innovations, such as MiSSPs. In contrast, ERM and ORM fungi retained an extensive decay apparatus that is probably exploited indirectly by the plant for carbohydrate supply, thus explaining their known saprotrophic ability¹⁷. The available genome sequences of mycorrhizal fungi will represent foundational information for understanding symbiosis development and functioning. These resources will facilitate field studies aiming to predict responses of mycorrhizal communities to environmental shifts, such as altered forest-management practices and climate change.

Methods

Sequencing, assembly and annotation.

The 18 new genomes were sequenced with a combination of the Sanger, 454, Illumina and PacBio sequencing platforms ([Supplementary Note](#) and [Supplementary Tables 2](#) and [3](#)). All genomes were assembled in a platform-dependent manner and annotated with the JGI Annotation Pipeline²³, which combines several gene prediction and annotation methods with transcriptomics data and integrates the annotated genomes into MycoCosm²², a web-based fungal resource for comparative analysis ([Supplementary Note](#)).

Transcript profiling.

Gene expression in mycorrhizal root tips and free-living mycelium from the ECM *A. muscaria*, *H. cylindrosporum*, *P. involutus*, *P. croceum* and *S. luteus*, the ERM *O. maius* and the ORM *S. vermifera* and *T. calospora* was assessed with RNA-seq ([Supplementary Note](#) and [Supplementary Table 11](#)).

Protein sequence clustering.

Predicted protein sequences were clustered with the Markov Cluster Algorithm (MCL) program²⁴, with an inflation parameter of 2.0 ([Supplementary Note](#)).

Phylogeny.

From protein sequence clustering, we identified single-copy gene families (1,809 clusters) that contained genes for >15 species ([Supplementary Note](#) and [Supplementary Fig. 1](#)). We found 617 such clusters, 5 of which contained suspected nonorthologous genes and were excluded. The remaining 611 gene families were retained, and ambiguously aligned sites were removed by 3 different settings altogether (with GBlocks and PRANK posterior probabilities). The single-gene alignments were then concatenated to result in 3 data sets with 19,567 sites and 149 loci (PRANK, 1.0 exclusion threshold), 114,814 sites and 542 loci (PRANK, 0.95 exclusion threshold) and 34,323 sites in 259 loci (MAFFT + GBlocks) ([Supplementary Fig. 2](#)). We then inferred maximum-likelihood and Bayesian trees by using RAxML 7.2.8 and PhyloBayes 3.3, respectively. The inferred trees were largely congruent across data sets and analyses. The inferred maximum-likelihood tree (data set PR0.95; [Supplementary Fig. 2](#)) was used for subsequent analyses.

Accession codes.

Genome assemblies and annotations for the organisms used in this study are available via the JGI fungal genome portal MycoCosm²² ([Supplementary Tables 2](#) and [3](#)). In addition, the newly sequenced genome assemblies and annotations have been deposited in GenBank under the following accession codes and BioProjects, respectively: *A. muscaria* Koide BX008, [JMDV00000000](#) and [PRJNA207684](#); *Gymnopus luxurians* FD-317 M1, [JJNP00000000](#) and [PRJNA68535](#); *H. cylindrosporum* h7, [JMDQ00000000](#) and [PRJNA207849](#); *Hydnomerulius pinastri* MD 312, [JMSK00000000](#) and [PRJNA207871](#); *Hypholoma sublateritium* FD-334 SS-4, [JMSJ00000000](#) and [PRJNA70685](#); *L. amethystina* LaAM-08-1, [JMSL00000000](#) and [PRJNA196025](#); *O. maius* Zn, [JMDP00000000](#) and [PRJNA74727](#); *P. involutus* ATCC 200175, [JOMD00000000](#) and [PRJNA60449](#); *Paxillus rubicundulus* Ve08.2h10, [JMDR00000000](#) and [PRJNA243391](#); *P. croceum* F 1598, [JMDN00000000](#) and [PRJNA61203](#); *Pisolithus microcarpus* 441, [JMDM00000000](#) and [PRJNA60815](#); *Pisolithus tinctorius* Marx 270, [JMDO00000000](#) and [PRJNA207840](#); *Plicaturopsis crispa* FD-325 SS-3, [JOMB00000000](#) and [PRJNA207847](#); *Scleroderma citrinum* Foug A, [JMDU00000000](#) and [PRJNA207859](#); *S. vermifera* MAFF 305830, [JMDS00000000](#) and [PRJNA207844](#); *Sphaerobolus stellatus* SS14, [JOMA00000000](#) and [PRJNA207858](#); *S. luteus* UH-Slu-Lm8-n1, [JMSM00000000](#) and [PRJNA242126](#); *T. calospora* MUT 4182, [JMDT00000000](#) and [PRJNA207843](#)). Electronic files of the phylogenomic and phylogenetic analyses have been deposited in the Dryad Repository under [doi:10.5061/dryad.f2g0f](https://doi.org/10.5061/dryad.f2g0f). The complete gene expression data sets have been submitted to Gene Expression Omnibus (superseries [GSE63947](#)). For *L. bicolor* and *S. lacrymans*, microarray data^{4, 21} were reanalyzed ([GSE9784](#), [GSE27839](#) and [GSE63929](#)).

References

1. Högberg, M.N. & Högberg, P. Extramatrical ectomycorrhizal mycelium contributes one-third of microbial biomass and produces, together with associated roots, half the dissolved organic carbon in a forest soil. *New Phytol.* 154, 791–795 (2002).
2. Read, D.J., Leake, J.R. & Perez-Moreno, J. Mycorrhizal fungi as drivers of ecosystem processes in heathland and boreal forest biomes. *Can. J. Bot.* 82, 1243–1263 (2004).
3. Averill, C., Turner, B.L. & Finzi, A.C. Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. *Nature* 505, 543–545 (2014).
4. Martin, F. *et al.* Symbiosis insights from the genome of the mycorrhizal basidiomycete *Laccaria bicolor*. *Nature* 452, 88–92 (2008).
5. Martin, F. *et al.* Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464, 1033–1038 (2010).
6. Veneault-Fourrey, C. *et al.* Genomic and transcriptomic analysis of *Laccaria bicolor* CAZome reveals insights into polysaccharides remodelling during symbiosis establishment. *Fungal Biol. Gen.* 72, 168–181 (2014).
7. Clemmensen, K.E. *et al.* Roots and associated fungi drive long-term carbon sequestration in boreal forest. *Science* 339, 1615–1618 (2013).
8. Treseder, K.K., Torn, M.S. & Masiello, C.A. An ecosystem-scale radiocarbon tracer to test use of litter carbon by ectomycorrhizal fungi. *Soil Biol. Biochem.* 38, 1077–1082 (2006).
9. Rineau, F. *et al.* Carbon availability triggers the decomposition of plant litter and assimilation of nitrogen by an ectomycorrhizal fungus. *ISME J.* 7, 2010–2022 (2013).
10. Bödeker, I.T.M. *et al.* Ectomycorrhizal *Cortinarius* species participate in enzymatic oxidation of humus in northern forest ecosystems. *New Phytol.* 203, 245–256 (2014).

11. Plett, J.M. *et al.* A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr. Biol.* 21, 1197–1203 (2011).
12. Plett, J.M. *et al.* The effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the *Populus* JAZ6 protein and represses JA-responsive genes. *Proc. Natl. Acad. Sci. USA* 111, 8299–8304 (2014).
13. Smith, S.E. & Read, J.R. *Mycorrhizal Symbiosis* (Academic Press, 2008).
14. Tedersoo, L., May, T.W. & Smith, M.E. Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza* 20, 217–263 (2010).
15. Tisserant, E. *et al.* The genome of an arbuscular mycorrhizal fungus provides insights into the oldest plant symbiosis. *Proc. Natl. Acad. Sci. USA* 110, 20117–20122 (2013).
16. Dearnaley, J.D.W., Martos, F. & Selosse, M.-A. in *Fungal Associations* 2nd edn., Vol. 9 (ed. Hock, B.) Ch. 12, 207–230 (Springer, 2012).
17. Rice, A.V. & Currah, R.S. in *Microbial Root Endophytes* (eds. Schulz, B., Boyle, C. & Sieber, T.) Ch. 13, 227–246 (Springer, 2006).
18. *Soc.* 279, 2003–2011 (2012).
19. Stergiopoulos, I. & de Wit, P. Fungal effector proteins. *Annu. Rev. Phytopathol.* 47, 233–263 (2009).
20. Carvunis, A.R. *et al.* Proto-genes and *de novo* gene birth. *Nature* 487, 370–374 (2012).
21. Eastwood, D.C. *et al.* The plant cell-wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* 333, 762–765 (2011).
22. Grigoriev, I.V. *et al.* MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42, D699–D704 (2014).
23. Grigoriev, I.V. *et al.* The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res.* 40, D26 (2012).
24. Enright, A.J., Van Dongen, S. & Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575 (2002).

Acknowledgments

This material is based on work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported under contract no. DE-AC02-05CH11231. This work was also supported by the Laboratory of Excellence ARBRE (ANR-11-LABX-0002-01), the Genomic Science Program (Plant-Microbe Interactions project) funded by the US Department of Energy, Office of Science, Biological and Environmental Research (contract DE-AC05-00OR22725), the Lorraine Region Council (to F.M.), US National Science Foundation grants DEB-1208719 and DEB-0933081 (both to D.S.H. and DEB-1021606 (to A.P.)), the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (DFG FTZ 118) and the German Science Foundation (DFG, grant BU941/20-1) (to F.B.) and the Swedish Research Council (to A. Tunlid). This work was partly supported by the Laboratory of Excellence TULIP (ANR-10-LABX-41 and ANR-11-IDEX-0002-02). F.M. would like to acknowledge M.A. Selosse and B. Lindahl for helpful discussions.

Author information

Author footnotes

- 1. Present address: Institut de Systématique, Évolution, Biodiversité (ISYEB), UMR 7205, CNRS, Muséum National d'Histoire Naturelle, Université Pierre et Marie Curie, Ecole Pratique des Hautes Etudes, Sorbonne Universités, Paris, France.**
 - Francois Rineau
- 2. These authors contributed equally to this work.**
 - Annegret Kohler,
 - Alan Kuo,
 - Laszlo G Nagy &
 - Emmanuelle Morin
- 3. These authors jointly supervised this work.**
 - Igor V Grigoriev,
 - David S Hibbett &
 - Francis Martin

Affiliations

- 1. Institut National de la Recherche Agronomique (INRA), Laboratory of Excellence Advanced Research on the Biology of Tree and Forest Ecosystems (ARBRE), UMR 1136, Champenoux, France.**
 - Annegret Kohler,
 - Emmanuelle Morin,
 - Nicolas Cichocki,
 - Elena Martino,
 - Claude Murat,
 - Jonathan M Plett,
 - Claire Veneault-Fourrey &
 - Francis Martin
- 2. University of Lorraine, Laboratory of Excellence ARBRE, UMR 1136, Champenoux, France.**
 - Annegret Kohler,
 - Emmanuelle Morin,
 - Nicolas Cichocki,
 - Claude Murat,
 - Jonathan M Plett,
 - Claire Veneault-Fourrey &
 - Francis Martin
- 3. US Department of Energy Joint Genome Institute (JGI), Walnut Creek, California, USA.**
 - Alan Kuo,
 - Kerrie W Barry,
 - Cindy Choi,
 - Alicia Clum,
 - Alex Copeland,
 - Kurt LaButti,
 - Erika A Lindquist,
 - Anna Lipzen,
 - Chew Y Ngan,
 - Robin A Ohm,
 - Robert Riley,
 - Asaf Salamov,
 - Hui Sun,
 - Andrew Tritt &
 - Igor V Grigoriev
- 4. Department of Biology, Clark University, Worcester, Massachusetts, USA.**
 - Laszlo G Nagy,
 - Dimitrios Floudas &
 - David S Hibbett

5. **Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary.**
 - Laszlo G Nagy
6. **Department of Soil Ecology, Helmholtz Centre for Environmental Research–Helmholtz Zentrum fuer Umweltforschung, Halle, Germany.**
 - Francois Buscot,
 - Sylvie Herrmann &
 - Mika Tarkka
7. **German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany.**
 - Francois Buscot,
 - Sylvie Herrmann &
 - Mika Tarkka
8. **Department of Biology, Microbial Ecology Group, Lund University, Lund, Sweden.**
 - Björn Canbäck,
 - Tomas Johansson,
 - Firoz Shah &
 - Anders Tunlid
9. **Centre for Environmental Sciences, Hasselt University, Diepenbeek, Belgium.**
 - Jan Colpaert,
 - Francois Rineau &
 - Joske Ruytinx
10. **Departamento de Microbiologia, Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Universidade Federal de Viçosa, Viçosa, Brazil.**
 - Mauricio D Costa
11. **UMR CNRS 5557, Unité Sous Contrat INRA 1364, Université de Lyon, Université Lyon 1, Villeurbanne, France.**
 - Jeanne Doré,
 - Gilles Gay &
 - Roland Marmeisse
12. **Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università degli Studi di Torino, Torino, Italy.**
 - Mariangela Girlanda,
 - Hassine-Radhouane Khouja,
 - Elena Martino &
 - Silvia Perotto
13. **CNRS, UMR 7257, Aix-Marseille Université, Marseille, France.**
 - Bernard Henrissat &
 - Anthony Levasseur
14. **Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, Marseille, France.**
 - Bernard Henrissat
15. **Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.**
 - Bernard Henrissat
16. **Department of Biosciences, University of Oslo, Oslo, Norway.**
 - Jaqueline Hess
17. **Department of Forest Mycology and Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden.**
 - Nils Högberg
18. **Department of Organismic Interactions, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany.**
 - Urs Lahrman &
 - Alga Zuccaro
19. **Department of Ecology, Biology/Chemistry, Botany, University of Bremen, Bremen, Germany.**
 - Uwe Nehls
20. **Harvard Forest, Harvard University, Petersham, Massachusetts, USA.**
 - Anne Pringle
21. **Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Forest Dynamics, Birmensdorf, Switzerland.**
 - Martina Peter

22. University of Cologne, Botanical Institute, Cluster of Excellence on Plant Sciences (CEPLAS), Cologne, Germany.

- Alga Zuccaro

Consortia

- 1. Mycorrhizal Genomics Initiative Consortium**
2. A full list of members is provided in the [Supplementary Note](#).

Contributions

F.M. conceived and guided the entire Mycorrhizal Genomics Initiative. I.V.G. directed all genome-sequencing and analysis efforts at JGI. D.S.H. directed the phylogenomic analyses. A. Tunlid coordinated the *P. involutus* genome project. F.M., D.S.H., I.V.G. and A. Kohler wrote the paper with contributions and input from all project members. A. Kohler performed overall management of biological materials and DNA and RNA preparations at INRA. K.W.B. performed overall project management at JGI. N.C., M.D.C., J.D., D.F., T.J., U.L., E. Martino, M.P., J.R. and M.T. extracted genomic DNA for sequencing. J.D., N.H., M.G., M.T., H.-R.K., S.H., U.L., E. Martino, U.N., J.M.P., T.J. and F.R. provided RNA for transcript profiling. A. Kohler, E. Morin, F.M., C.M., C.V.-F., G.G., S.H., S.P., M.G., A.Z., F.B., J.C., R.M. and A.P. analyzed genomic and transcriptomic data. C.C. and C.Y.N. performed genome sequencing. A. Clum, A. Copeland, K.L., A. Tritt and H.S. assembled the genomes. A. Kuo, R.A.O., R.R. and A.S. annotated and analyzed the genomes. B.C. and F.S. analyzed the *P. involutus* genome. E.A.L. and A. Lipzen performed transcriptome sequencing and analysis at JGI. L.G.N., D.F. and D.S.H. performed organismal phylogeny and analyzed evolution of decay-related gene families. A. Kuo and E. Morin performed protein clustering analysis. B.H. and A. Levasseur annotated the CAZyme and fungal oxidative lignin enzyme (FOLyme) families. J.H. analyzed the repeated elements in *A. muscaria*.

Competing financial interests

The authors declare no competing financial interests.

Corresponding authors

Correspondence to:

- Francis Martin or
- David S Hibbett