Title
The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis

Authors
Martin, F.
Aerts, A.
Ahren, D.
et al.

Publication Date
2008-03-06

The Genome of *Laccaria bicolor* Provides Insights into Mycorrhizal Symbiosis

March 2008

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

Nature manuscript 2007-08-08601B

The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis

F Martin[1*], A Aerts[2+], D Ahrén[3+], A Brun[1+], F Duchaussoy[1+], J Gibon[1+], A Kohler[1+], E Lindquist[2+], V Pereda[1+], A Salamov[2+], HJ Shapiro[2+], J Wuyts[1,4+], D Blaudez[1], M Buée[1], P Brokstein[2], B Canbäck[3], D Cohen[1], PE Courty[1], PM Coutinho[5], EGJ Danchin[5], C Delaruelle[1], JC Detter[2], A Deveau[1], S DiFazio[6], S Duplessis[1], L Fraissinet-Tachet[8], E Lucic[1], P Frey-Klett[1], C Fourrey[1], I Feussner[7], G Gay[8], J Grimwood[9], PJ Hoegger[10], P Jain[11], S Kilaru[10], J Labbé[1], YC Lin[4], V Legué[1], F Le Tacon[1], R Marmeisse[8], D Melayah[8], B Montanini[1], M Muratet[11], U Nehls[12], H Niculita-Hirzel[13], MP Oudot-Le Secq[1], M Peter[1,14], H Quesneville[15], B Rajashekar[3], M Reich[1,10], N Rouhier[1], J Schmutz[9], T Yin[16], M Chalot[1++], B Henrissat[5++], U Kües[10++], S Lucas[2++], Y Van de Peer[4++], GK Podila[11++], A Polle[10++], PJ Pukkila[17++], PM Richardson[2++], P Rouzé[4,18++], IR Sanders[13++], JE Stajich[19++], A Tunlid[3++], G Tuskan[16++] & IV Grigoriev[2++]

(1) UMR 1136, INRA-Nancy Université, Interactions Arbres/Microorganismes, 54280 Champenoux, France. (2) US DOE Joint Genome Institute, Walnut Creek, CA 94598. (3) Microbial Ecology, Lund University, Sweden. (4) Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, Ghent, Belgium. (5) Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS-Universités Aix-Marseille I & II, Marseille, France. (6) Department of Biology, West Virginia University, Morgantown, WV 26506 USA. (7) Department for Plant Biochemistry, Georg-August-Universität Göttingen, Göttingen, Germany. (8) Université Lyon 1, UMR CNRS - USC INRA d'Ecologie Microbienne, Villeurbanne, France. (9) Stanford Human Genome Center, Stanford/JGI, USA. (10) Institute of Forest Botany, Georg-August-Universität, Göttingen, Germany. (11) Department of Biological Sciences, University of Alabama, Huntsville. (12) Eberhard-Karls-Universität, Physiologische Oekologie der Pflanzen, Tübingen, Germany. (13) Dept. of Ecology & Evolution, University of Lausanne, Lausanne, Switzerland. (14) Swiss Federal Research Institute WSL, Birmensdorf, Switzerland. (15) Unité de Recherches en Génomique-Info, 91034 Évry Cedex. (16) Environmental Science Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831. (17) Department of Biology, The University of North Carolina, Chapel Hill, NC 27599-3280. (18) Laboratoire Associé de l'INRA, Ghent University, B-9052 Gent, Belgium. (19) Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720-3102.

* to whom correspondence should be addressed: E-mail: fmartin@nancy.inra.fr
+ These authors contributed equally to this work as second authors.
++ These authors contributed equally to this work as senior authors.

Mycorrhizal symbioses -- the union of roots and soil fungi -- are universal in terrestrial ecosystems and may have been fundamental to land colonization by plants[1,2]. Boreal, temperate, and montane forests all depend upon ectomycorrhizae[1]. Identification of the primary factors that regulate symbiotic development and metabolic activity will therefore open the door to understanding the role of

ectomycorrhizae in plant development and physiology, allowing the full ecological significance of this symbiosis to be explored. Here, we report the genome sequence of the ectomycorrhizal basidiomycete *Laccaria bicolor* (Fig. 1) and highlight gene sets involved in rhizosphere colonization and symbiosis. This 65-million-base genome assembly contains ~ 20,000 predicted protein-encoding genes and a very large number of transposons and repeated sequences. We detected unexpected genomic features most notably a battery of effector-type small secreted proteins (SSP) with unknown function, several of which are only expressed in symbiotic tissues. The most highly expressed SSP accumulates in the proliferating hyphae colonizing the host root. The ectomycorrhizae-specific SSP likely play a decisive role in the establishment of the symbiosis. The unexpected observation that the genome of *L. bicolor* lacks carbohydrate-active enzymes involved in degradation of plant cell walls, but maintains the ability to degrade non-plant cell walls, reveals the dual saprotrophic and biotrophic lifestyle of the mycorrhizal fungus which enables it to grow within both soil and living plant roots. The predicted gene inventory of the *L. bicolor* genome, therefore, points to previously unknown mechanisms of symbiosis operating in biotrophic mycorrhizal fungi. The availability of this genome provides an unparalleled opportunity to develop a deeper understanding of the processes by which symbionts interact with plants within their ecosystem in order to perform vital functions in the carbon and nitrogen cycles that are fundamental to sustainable plant productivity.

The 65 million base pairs genome of *Laccaria bicolor* (Maire) P.D. Orton (hereafter referred to as *Laccaria*) is the largest sequenced fungal genome published so far[3,4,5,6,7] (Table 1). While no evidence for large scale duplications was observed within the *Laccaria* genome, tandem duplication occurred within multigene families (Supplementary Fig. 4). Transposable elements (TE) comprised a higher proportion (21%) than that identified in the other sequenced fungal genomes and may therefore account for the relatively large genome of *Laccaria* (Supplementary Table 3). Approximately 20,000 protein-coding genes were identified by combined gene predictions (Supplementary Information Section 2). Expression of nearly 80% (ca. 16,114) of the predicted genes was detected in either free-living mycelium, ectomycorrhizal root tips or fruiting bodies (Supplementary Table 4) using NimbleGen custom-oligoarrays (Supplementary Information Section 9). Most genes are activated in almost all tissues, whereas other more specialized genes were only activated in some specific developmental stages, such as free-living mycelium, ectomycorrhizae or fruiting body (Supplementary Table 5).

Only 14,464 of *Laccaria* proteins (70%) showed significant sequence similarity to documented proteins. Most homologs were found in the sequenced basidiomycetes *Phanerochaete chrysosporium*[4], *Cryptococcus neoformans*[5], *Ustilago maydis*[6], and *Coprinopsis cinerea*[7] (Supplementary Table 6). The percentage of proteins found in multigene families was related to genome size and was the largest in

Laccaria (Fig. 2). This was mainly due to the expansion of protein family size, but also due to the larger number of protein families in *Laccaria* when compared to the other basidiomycetes (Supplementary Table 7). Expansion of protein family sizes in *Laccaria* was prominent in the lineage-specific multigene families. Striking gene family expansions occurred in those genes predicted to have roles in protein-protein interactions (e.g. WD40) and signal transduction mechanisms (Supplementary Table 7). Two new classes of Gα genes were found and may be candidates for the complex communication that must occur between the mycobiont and its host-plant during mycorrhizae establishment (Supplementary Table 8). Several transcripts coding for expanded and lineage-specific gene families were upregulated in symbiotic and fruiting body tissues, suggesting a role in tissue differentiation (Supplementary Tables 5 & 9).

In our analysis of annotated genes, and in particular paralogous gene families, we highlighted processes which may be related to the biotrophic and saprotrophic lifestyles of *Laccaria*. Twelve predicted proteins showed a similarity to known haustoria-expressed secreted proteins (HESP) of the basidiomycetous rusts, *Uromyces fabea*[8] and *Melampsora lini*[9], which are involved in pathogenesis (Supplementary Table 10). Of the 2,931 proteins predicted to be secreted by *Laccaria*, most (67%) cannot be ascribed a function and 82% of these predicted proteins are specific to *Laccaria*. Within this set, we found a large number of genes that encode cysteine-rich products with a predicted size of <300 amino acids. Of these 278 small secreted proteins (SSP), 69% belong to multigene families, but only nine groups comprising a total of 33 SSP co-localized in the genome (Supplementary Fig. 5). The structure of two of these clusters is shown in Supplementary Fig. 6. Other SSP are scattered all over the genome and we found no correlation between SSP and TE genome localization (Supplementary Fig. 5). Transcript profiling revealed that the expression of several SSP genes is specifically induced upon in the symbiotic interaction (Table 1, Supplementary Fig. 10). Five of the 20 most highly upregulated fungal transcripts in ectomycorrhizal root tips code for SSP (Supplementary Table 5). These mycorrhiza-induced cysteine-rich SSP (MISSP) belong to *Laccaria*-specific orphan gene families. Within the MISSP, we found a family of secreted proteins with a CFEM domain (IPR014005) (Supplementary Fig. 7 & 8), as previously identified in the plant pathogenic fungi *M. lini*[9] and *M. grisea*[10] (Supplementary Table 10), and proteins with a gonadotropin- (IPR0001545) or snake toxin-like (SSF57302) domains related to the cysteine-knot domain. Expression of several SSP were downregulated in ectomycorrhizal root tips (cluster E in Supplementary Fig. 10) suggesting a complex interplay between these secreted proteins in symbiosis interaction.

The rich assortment of MISSP may therefore act as effector proteins to manipulate host cell signalling or suppress defence pathways during infection, as suggested for pathogenic rusts[8,9], smuts[6] (*U.*

maydis) and *Phytophthora*[11] species. To play a role in symbiosis development, MISSP should be expressed in *Laccaria* hyphae colonizing the root tips. To test this assertion, we determined the tissue distribution of the MISSP7 protein (ID 298595) showing the highest induction in ectomycorrhizal tips (Table 1, Supplementary Table 5). Two peptides located in the N-terminal and C-terminal parts of the mature protein were selected as antigens for the production of anti-MISSP7 antibodies. The selected peptides were not found in the deduced protein sequences of other *Laccaria* gene models nor in the *Populus trichocarpa* genome[12]. MISSP7 localization in *Laccaria/Populus* ectomycorrhizal root tips by indirect immunofluorescence is illustrated in Fig. 1 and Supplementary Fig. 11. Control images in which the ectomycorrhizae sections were obtained replacing primary anti-MISSP7 antibodies by pre-immune IgG are shown in Supplementary Fig. 12. Where ectomycorrhizae were treated with anti-MISSP7 antibody followed by fluorescent-labeled secondary antibody, fluorescence was localized in the hyphae colonizing short roots (Fig. 1, Supplementary Fig. 11) and not detected in the free-living mycelium (Supplementary Fig. 12). Although MISSP7 was detected in the hyphal mantle layers ensheating the root tips, the protein mainly accumulated in the finger-like, labyrinthine branch hyphal system (Hartig net) which provides a very large area of contact between cells of the two symbionts. It accumulated in the cytosol and cell wall of the fungal cells. The MISSP7 protein could thus interact with the plant components after secretion. MISSP7 shares no sequence similarity or protein motif with other SSP. Comparison of the MISSP sequences did not reveal a specific conserved motif, such as the RXLR motif[11] of phytopathogenic *Phytophthora* or the malaria parasite, that could potentially contribute to their function or to targeting to the host cell. Those SSP with an upregulated expression in fruiting body (Supplementary Table 5, Supplementary Fig. 10) may play a role in the differentiation of the sexual tissues and/or aggregation of sporophore tissues. Interestingly, they are a large set of SSP genes showing significant changes in gene expression in both ectomycorrhizal root tips and fruiting body (cluster A in Supplementary Fig. 10) suggesting that both developmental processes recruit similar gene networks (e.g., those involved in hyphal aggregation).

Host trees are able to harness the formidable web of mycorrhizal hyphae, that permeates the soil and leaf litter, for their nutritional benefit. A process that is pivotal to the success of ectomycorrhizal interactions is thus the equitable exchange of nutrients between the symbiont and its host-plant[1,2,13]. A comparison with other basidiomycetes (Supplementary Table 12) revealed that the total number of predicted transporters has been expanded in *Laccaria* compared to *C. cinerea* and *P. chrysosporium*. Interestingly, *Laccaria* has multiple ammonia transporters although it encodes a single nitrate permease. Ammonia is arguably the most important inorganic nitrogen source for ectomycorrhizal fungi[14]. One of the

ammonia transporters (*LbAMT2.2*), for instance, is greatly upregulated in ectomycorrhizae (Supplementary Table 5). *Laccaria,* thus, shows an increased genetic potential in terms of nitrogen uptake when compared to other basidiomycetes. These capabilities are consistent with *Laccaria* being exposed to a range of nitrogen sources from organic matter decay[15].

Although the *Laccaria* genome contains numerous genes coding for key hydrolytic enzymes, such as proteases and lipases, we observed an extreme reduction in the number of enzymes involved in the degradation of plant cell wall (PCW) oligo- and polysaccharides. Glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE) and their ancillary carbohydrate-binding modules (CBM) were identified using the carbohydrate-active enzyme (CAZyme) classification (http://www.cazy.org/). A comparison of the *Laccaria* candidate CAZymes with fungal phytopathogens confirms the adaptation of its enzyme repertoire to symbiosis and reveals the strategy used for the interaction with the host (Supplementary Tables 13 and 14). The reduction in PCW CAZymes affects almost all GH families culminating in the complete absence of several key families. For instance, there is only one candidate cellulase (GH5) appended to the sole fungal cellulose-binding module (CBM1) found in the genome and no cellulases from families GH6 and GH7 (Supplementary Table 14). Similar reductions or loss of hemicellulose and pectin degrading enzymes were also noted. These observations suggest that the inventory of *Laccaria* PCW degrading enzymes underwent massive gene loss as a result of its adaptation to a symbiotic lifestyle and that this species is now unable to use many PCW polysaccharides as a carbon source, including those found in soil and leaf litter. The remaining small set of secreted CAZymes with potential action on plant polysaccharides (e.g. GH28-polygalacturonases) is probably required for cell wall remodeling during fungal tissue differentiation as their expression was upregulated in both fruiting body and ectomycorrhizae (Supplementary Table 15, Supplementary Fig. 13). In contrast, transcripts coding for proteins with expansin domain were only induced in ectomycorrhizae suggesting they may be used by *Laccaria* for penetrating into the root apoplastic space. To survive before its mycorrhizal association with its host, *Laccaria* appears to have developed a capacity to degrade non-plant (e.g. animal, bacterial) oligo- and polysaccharides which is suggested by retention of CAZymes from families GH79, PL8, PL14 and GH88 (Supplementary Table 14). Interestingly, there is no invertase gene in the *Laccaria* genome, implying that this fungus is unable to directly use sucrose from the plant. This is consistent with earlier observations[16] that *Laccaria* depends on its host plant to provide glucose in exchange for nitrogen. We also noticed an expansion of CAZymes involved in the fungal cell wall biosynthesis and rearrangement, almost entirely due to an increased number of putative chitin synthases

and enzymes acting on β-glucans (Supplementary Table 14). Several of the corresponding genes are up- or downregulated upon developmental processes requiring cell wall alterations such as formation of fruiting bodies or mycorrhizae (Supplementary Table 15, Supplementary Fig. 13).

Ectomycorrhizal fungi play a significant role in mobilizing N from well-decomposed organic matter[2,15]. The hyphal network permeating the soil might therefore be expected to express a wide diversity of proteolytic enzymes. The total number of secreted proteases (116 members) identified (Supplementary Fig. S14) is relatively large compared with other sequenced saprotrophic basidiomycetes, such as *C. cinerea* and *P. chrysosporium*. Secreted aspartyl-, metallo- and serine-proteases may play a role in degradation of decomposing litter[15] confirming that *Laccaria* has also the ability to use nitrogen of animal-origin, as suggested previously[17]. They may also play a role in developmental processes as the expression of several secreted proteases is up- or downregulated in fruiting bodies and ectomycorrhizal root tips (Supplementary Table 16). Mycelial mats formed by *Laccaria* hyphae colonizing organic matter therefore possess the ability to degrade decomposing leaf litter.

Our analysis of the gene space reveals a multi-faceted mutualistic biotroph equipped to take advantage of transient occurrences of high-nutrient niches (living host roots and decaying soil organic matter) within a heterogeneous, low-nutrient environment. The availability of genomes from mutualistic, saprotrophic[4], and pathogenic[6] fungi, but also from the mycorrhizal tree *Populus trichocarpa*[12], now provides an unparalleled opportunity to develop a deeper understanding of the processes by which fungi colonize wood and soil litter, and also interact with living plants within their ecosystem in order to perform vital functions in the carbon and nitrogen cycles[2] that are fundamental to sustainable plant productivity.

METHODS SUMMARY

The Methods are described in Supplementary Information (www.nature.com/nature). The sections of the Supplementary Methods are arranged in the same order as the manuscript to facilitate cross-referencing. Here, we describe the datasets generated by this project and their availability.

Genomic sequence. The WGS project has been deposited at GenBank/EMBL/DDBJ under project accession ABFE00000000. The version described in this paper including assembly and annotation is the first version ABFE01000000. Scaffolds and assemblies for all genomic sequence generated by this project are also available from the JGI portal (http://genome.jgi-psf.org/Lacbi1/Lacbi1.download.ftp.html). A genome browser is available from JGI (www.jgi.doe.gov/laccaria). BLAST search of the genome is

available at JGI (www.jgi.doe.gov/laccaria) and INRA LaccariaDB (http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/blastlaccaria/blastlaccaria.php).

Predicted gene models. Consensus gene predictions, produced by combining several different gene predicters, are available from JGI (www.jgi.doe.gov/laccaria) as GFF files. These gene models can also be accessed from the Genome Browser in JGI Laccaria portal (http://genome.jgi-psf.org/cgi-bin/browserLoad/46b9a4360b37752a766008cb).

Gene annotations. Tables compiling KEGG, PFAM, KOG, and best BLAST hits for predicted gene models, transposable element and CAZyme data, and Tribe-MCL gene families are available from INRA LaccariaDB (http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/index.html).

Array data. The complete expression dataset is available as series (accession number # GSE9784) at the Gene Expression Omnibus at NCBI (http://www.ncbi.nlm.nih.gov/geo/).

References

1. Smith, S. E. & Read, D. J. Mycorrhizal Symbiosis (2nd edition, Academic Press, London) (1996).
2. Read, D. J. & Perez-Moreno, J. Mycorrhizas and nutrient cycling in ecosystems - a journey towards relevance? New Phytol. **157**, 475-492 (2003).
3. Galagan, J. E., Henn, M. R., Ma, L. J., Cuomo, C. A., Birren, B. Genomics of the fungal kingdom: insights into eukaryotic biology. Genome Res. **15**, 1620-1631 (2005).
4. Martinez, D. *et al.* Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. Nature Biotech. **22**, 695-700 (2004).
5. Loftus, B. J. *et al.* The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science **307**, 1321-1324 (2005).
6. Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature **444**, 97-101 (2006).
7. *Coprinus cinereus* Database: http://www.broad.mit.edu/annotation/genome/coprinus_cinereus/Home.html
8. Wirsel, S. G. R., Voegele, R. T., Mendgen, K. W. Differential regulation of gene expression in the obligate biotrophic interaction of *Uromyces fabae* with its host *Vicia faba*. Mol. Plant Microb. Int. 14, 1319-1326 (2001).
9. Catanzariti, A. M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A., Ellis, J. G. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. Plant Cell **18**, 243-256 (2006).
10. Kulkarni, R. D., Kelkar, H. S., Dean, R. A. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. Trends Bioch. Sci. **28**, 118-118 (2003).
11. Kamoun, S. A. Catalogue of the effector secretome of plant pathogenic oomycetes. Annu. Rev. Phytopathol. **44**, 41-60 (2006).

12. Tuskan, G. A. *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313**, 1596-1604 (2006).
13. Martin, F., Kohler, A., Duplessis, S. Living in harmony in the wood underground: ectomycorrhizal genomics. Curr. Opin. Plant Biol. **10**, 204-210 (2007).
14. Chalot, M., Blaudez, D., Brun, A. Ammonia: a candidate for nitrogen transfer at the mycorrhizal interface. Trends Plant Sci. **11**, 263-266 (2006).
15. Lindahl, B. D., Ihrmark, K., Boberg, J., Trumbore, S. E., Högberg, P., Stenlid, J., Finlay, R. D. Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. New Phytol. **173**, 611-620 (2007).
16. Nehls, U., Grunze, N., Willmann, M., Reich, M., Küster, H. Sugar for my honey: Carbohydrate partitioning in ectomycorrhizal symbiosis. Phytochem. **68**, 82-91 (2007).
17. Klironomos, J. N. & Hart, M. M. Animal nitrogen swap for plant carbon. Nature **410**, 651 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The genome sequencing of *Laccaria bicolor* was funded by the U.S. Department of Energy's Office of Science, Biological, and Environmental Research Program and by University of California, Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, and Los Alamos National Laboratory. Annotation and transcriptome analysis were supported by INRA, U.S. Department of Energy, U.S. National Science Foundation, European Commission, Région Lorraine and the Swedish Research Council. We would like to thank S Rombauts, L Sterck, K Vandepoele, G Werner and his colleagues, S Pitluck and K Zhou, B Hilselberger and J Gérard for their assistance. F.M. thanks Prof. N. Talbot for critical reading of an early draft of the manuscript.

Author Contributions is described in Supplementary data section 11.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.M. (fmartin@nancy.inra.fr).

Table 1. Genome characteristics of *Laccaria bicolor* and other basidiomycetes

Genome characteristics	***Laccaria***	***Coprinopsis***[7]	***Phanerochaete***[4]	***Cryptococcus***[5]	***Ustilago***[6]
Strain	S238N-H82	Okayama7#130	RP78	H99	521
Sequencing institution	JGI	Broad-MIT	JGI	Broad-MIT	Broad-MIT
Genome assembly (Mbp)	64.9	37.5	35.1	19.5	19.7
GC content (%)	46.6	51.6	53.2	48.2	54
Protein coding genes	20,614	13,544	10,048	7,302	6,522
CDS<300 bp	2,191	838	163	313	58
Average gene length (bp)	1,533	1,679	1,667	1,828	1,935
Average CDS length (bp)	1,134	1,352	1,366	1,502	1,840
Average exon length	210.1	251	232	253	1,051
Average intron length	92.7	75	117	66	127

Table 2. Changes in the expression of transcripts coding for mycorrhizae-induced cysteine-rich small secreted proteins

Protein ID	Family size	Length (AA)	Transcript Concentration (FLM)	*Pseudotsuga* ECM/FLM Ratio (fold)	*Populus* ECM/FLM Ratio (fold)	Features
298595	sc	68	nd	21877	12913	MISSP7
333839	5	129	nd	7844	1931	GPI-anchored
298667	2	70	nd	1906	1407	
332226	8	181	43	847	780	CFEM domain (IPR014005)
311468	2	59	nd	191	nd	
295737	8	288	131	171	252	
334759	sc	101	nd	109	18	
395403	4	121	24	103	93	
333423	9	120	6	102	72	Gonadotropin domain (IPR0001545)
312262	4	106	85	69	53	
295625	4	199	325	66	48	
325402	8	238	310	49	74	Snake toxin-like (SSF57302)
316998	sc	56	137	29	57	
333197	3	148	266	17	8	
327918	2	154	763	13	4	Homolog in *Coprinopsis cinerea*
307956	sc	74	336	13	90	Whey acidic domain (IPR008197)
327246	sc	194	1025	10	18	Homolog in *Coprinopsis cinerea*
303550	5	98	1365	10	14	
300377	2	291	5499	10	8	
293250	sc	224	127	9	10	Homolog in *Coprinopsis cinerea*
298648	sc	64	1108	8	12	
298646	2	73	1028	7	14	
293729	3	210	3000	7	7	

Transcript profiling was performed on free-living mycelium (FLM), and ectomycorrhizal root tips (ECM) of poplar (*Populus trichocarpa*) and Douglas fir (*Pseudotsuga menziesii*). See Supplementary Information section 9 for details. Abbreviations: AA, amino acids; nd, not detected; sc, single copy.

METHODS

Genome sequencing. The haploid genome of the strain S238N-H82 from *L. bicolor* (Maire) P.D. Orton was sequenced with the use of a whole-genome shotgun (WGS) strategy. All data were generated by paired-end sequencing of cloned inserts using Sanger technology on ABI3730xl sequencers. Supplementary Table 1 gives the number of reads obtained per library.

Genome assembly. The data was assembled using release 1.0.1b of JAZZ, a JGI WGS assembler. Based on the number of alignments per read, the main genome scaffolds were at a depth of 9.88. The amount of sequence in the unplaced reads was 6.5 Mbp, which is sufficient to cover the main-genome gaps to a mean depth of 9.9. A total of 64.9 Mbp are captured in the scaffold assembly (Supplementary Table 2).

Genome annotation. Gene models were predicted using FgenesH[17], homology-based FgenesH+[18], Genewise[19], as well as EuGène[20] and TwinScan[21], and alignments of several cDNA resources (Supplementary Information section 3). The JGI pipeline selected a best representative gene model for each locus based on EST support and similarity to known proteins from other organisms, and predicted 20,614 protein-coding gene models. All predicted genes were annotated using Gene Ontology[21], eukaryotic clusters of orthologous groups[22], and KEGG pathways[23]. Protein domains were predicted using InterProScan[24]. Signal peptides were predicted in 2,931 Laccaria proteins by both the hidden Markov and the neural network algorithms of SignalP[25]. After eliminating predicted transmembrane proteins and removal of transposable element fragments, we selected 278 cysteine-rich secreted proteins with a size <300 AA. Gene families were built from proteins in *Laccaria*, *C. cinerea, P. chrysosporium, C. neoformans* and *U. maydis* using Tribe-MCL tools[26] with default settings.

Indirect immunofluorescent localization of MISSP7. The peptides LRALGQASQGGDLHR and GPIPNAVFRRVPEPNF located in the N-terminal and C-terminal parts of the MISSP7 sequence (without the signal peptide) were synthesized and used as antigens for the generation of antibodies in rabbits according to the manufacturer's procedures (Eurogentec, Seraing, Belgium). The anti-MISSP7 IgG fraction was purified using MAbTrap kit (GE Healthcare) according to the manufacturer's recommendations. Subsequently, IgG-containing fraction was desalted using a HiTrap™ desalting column (GE Healthcare). The concentration of purified IgG from pre-immune serum was determined by Bradford assay using a Bio-

Rad protein assay. Final concentration of anti-MISSP7 IgG was 0.16 mg/ml. Immunolocalization was performed essentially as described by [27,28] with slight modifications (Supplementary section 10).

Gene expression. Average expression levels of genes in different tissues and conditions (SOM) were analyzed using CyberT statistical framework (http://www.igb.uci.edu/servers/cybert/) and hierarchical clustering with EPCLUST (http://ep.ebi.ac.uk/EP/EPCLUST/) (Supplementary section 8).

References

18. Salamov, A., Solovyev, V. *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Res **10**: 516-522 (2000).
19. Birney, E., Clamp, M., Durbin, R. GeneWise and genomewise. Genome Res **14**: 988-995 (2004).
20. Schiex, T., Moisan, A., Rouzé, P. EuGène: an eukaryotic gene finder that combines several sources of evidence. In: *Computational Biology,* Gascuel O, Sagot MF, eds, LNCS 2066, pp. 111-125 (2001).
21. Tenney, A.E. *et al.* Gene prediction and verification in a compact genome with numerous small introns. Genome Res **14:** 2330-2335 (2004).
22. Ashburner, M. et al. Gene ontology: tool for the unification of biology. Nature Genetics **25**: 25-29 (2000).
23. Koonin, E. V. et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biology **5**: R7 (2004).
24. Kanehisa, M, Goto, S, Kawashima, S, Okuno, Y, Hattori, M. The KEGG resource for deciphering the genome. Nucleic Acids Research **32**: D277-D280 (2004).
25. Zdobnov, EM, Apweiler, R. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17**: 847-848 (2001).
26. Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H. Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols **2**: 953-971 (2007).
27. Blancaflor, E.B., Zhao, L., Harrison, M.J. Microtubule organization in root cells of *Medicago truncatula* during development of an arbuscular mycorrhizal symbiosis with *Glomus versiforme*. Protoplasma **217**: 154–165 (2001).
28. Harrison, M.J., Dewbre, G.R., and Liu, J.Y. A phosphate transporter from *Medicago truncatula* involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. Plant Cell **14**: 2413–2429 (2002).

Legends of Figures

Figure 1 The *Laccaria bicolor* ectomycorrhizal symbiosis and the immunofluorescent localization of the small secreted protein MISSP7 in ectomycorrhizae.

a, Fruiting bodies of *L. bicolor* colonizing seedlings of Douglas fir (*Pseudotsuga menziesii*). The subterranean mycelial web has developed symbiotic ectomycorrhizal tissues on host root tips and has produced fruiting bodies above ground (Photograph courtesy of D. Vairelles, INRA-Nancy). **b**, Laser scanning confocal microscopy image of a transverse section of *P. menziesii–L. bicolor* ectomycorrhizal root tips showing extramatrical mycelium (em), aggregated hyphae of the mantle sheath (m), hyphae proliferating between the epidermal (ec), tannin (tc) and cortical (cc) of the host root to form the symbiotic Hartig net (hn). Bar = 10 µm. **c-f**, Indirect immunofluorescent localization of MISSP7. Transverse (**c, e**) and longitudinal (**d, f**) sections of *Populus trichocarpa-L. bicolor* ectomycorrhizal tips. MISSP7 was detected with anti-MISSP7 IgG and secondary antibody conjugated with AlexaFluor 488 in the hyphae of the mantle (m) and the uniseriate Hartig net (hn) ensheathing the epidermal cells (ec) of the colonized roots. Rectangle in panels (d) and (f) show the finger-like, labyrinthine hyphal system accumulating large amount of MISSP7. (e) and (f), phase contrast images. Bar = 10 µm.

Figure 2 Expansion of protein families in *Laccaria bicolor.*

a, Relationship between genome size and number of protein families. **b**, Relationship between genome size and protein family sizes in five sequenced basidiomycetes. Protein sequences predicted from the genome sequences of *Laccaria bicolor*, *Coprinopsis cinerea*, *Phanerochaete chrysosporium*, *Cryptococcus neoformans* and *Ustilago maydis* were clustered into families using the TRIBE-MCL algorithm (see Supplementary Information section 5 for details).

Figure 1

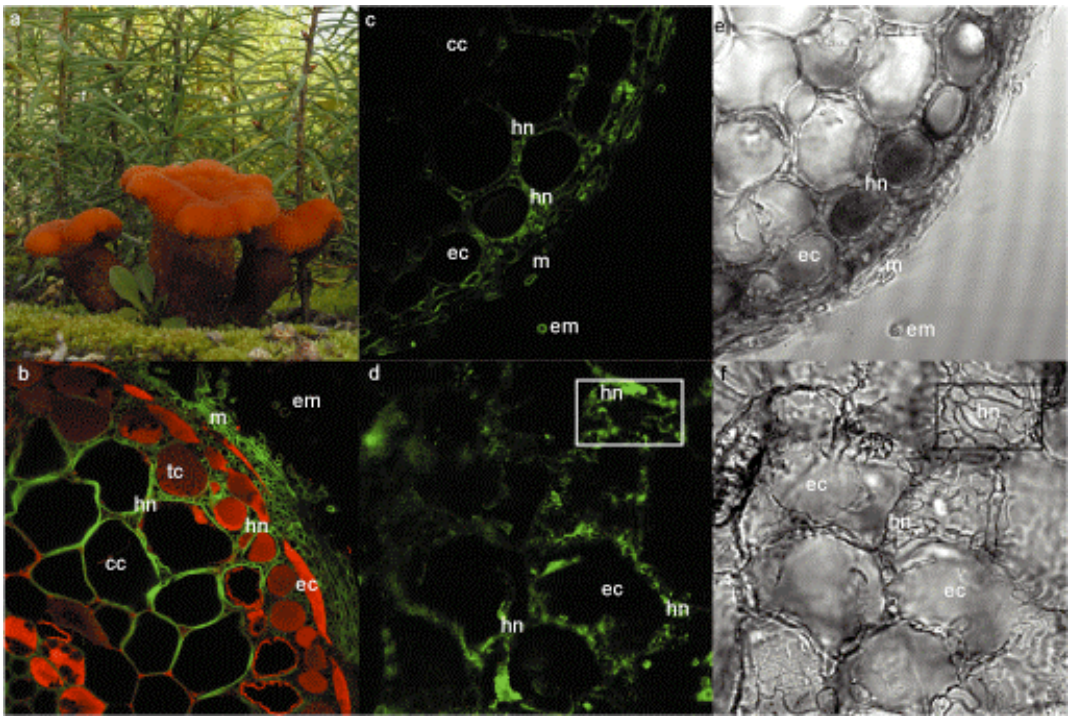
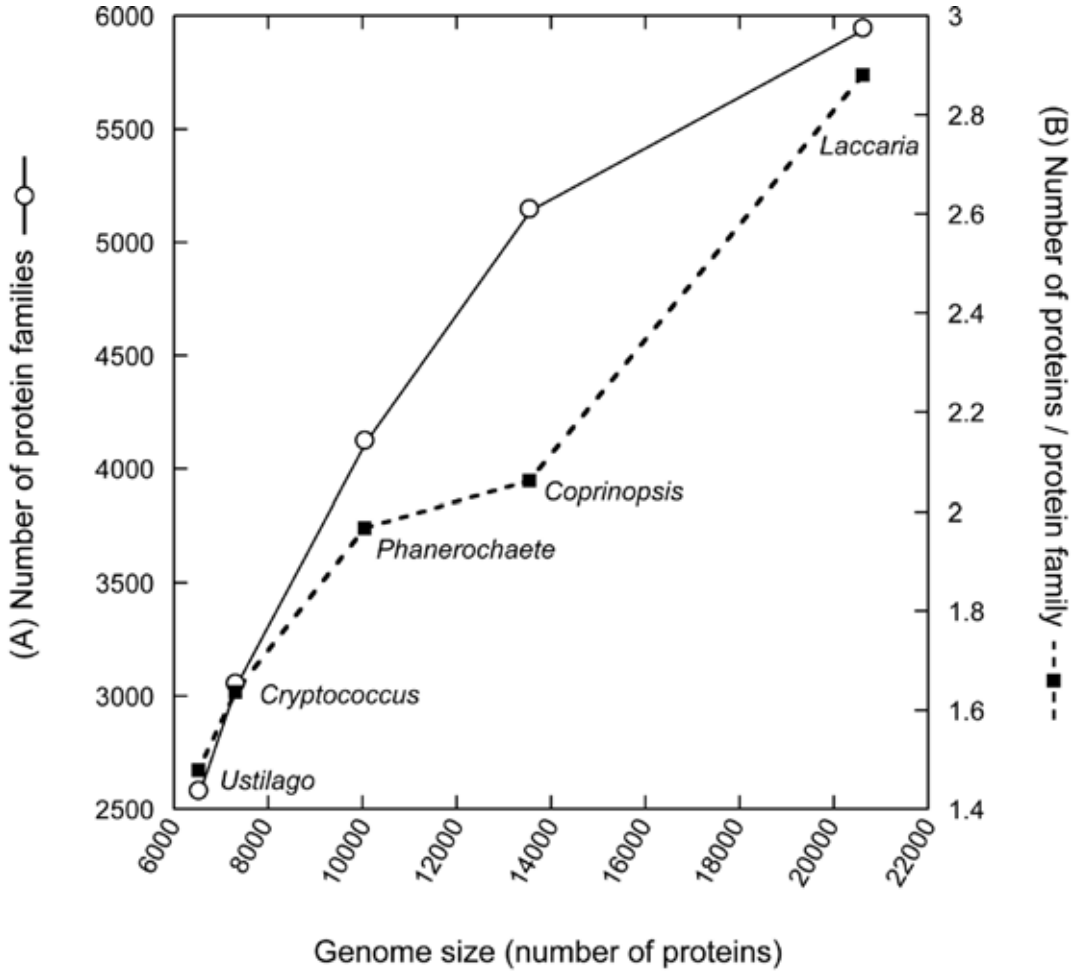

a
c
cc
hn
hn
ec
m
em
e
hn
ec
m
em
b
em
m
tc
hn
hn
cc
ec
d
hn
ec
hn
hn
f
hn
ec
hn
ec

Figure 2.


(A) Number of protein families
(B) Number of proteins / protein family
Genome size (number of proteins)
Laccaria
Coprinopsis
Phanerochaete
Cryptococcus
Ustilago
6000
5500
5000
4500
4000
3500
3000
2500
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
6000
8000
10000
12000
14000
16000
18000
20000
22000