LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-485265

Formation of mushrooms and lignocellulose degradation encoded in the genomes sequence of Schizophyllum commune

S. Lucas, I. Grigoriev, R. Ohm, H. Wosten, J. deJong, L. Lugones, A. Aerts, E. Kothe, J. Stajich, R. deVries, E. Record, A. Levasseur, S. Baker, K. Bartholomew, P. Coutinho, S. Erdmann, T. Fowler, A. Gathman, V. Lombard, B. Henrissat, N. Knabe, U. Kues, W. Lilly, E. Lindquist, J. Magnuson, F. Schwarze, M. Raudaskoski, F. Piumi, A. Salamov, J. Schmutz, P. vanKuyk, S. Horton

May 31, 2011

Nature BioTechnology

Disclaimer

LLNL-JRNL-485265

Formation of mushrooms and lignocellulose degradation encoded in the genome sequence of *Schizophyllum commune*

Robin A. Ohm[1], Jan F. de Jong[1], Luis G. Lugones[1], Andrea Aerts[2], Erika Kothe[3], Jason E. Stajich[4], Ronald P. de Vries[1,5], Eric Record[6,7], Anthony Levasseur[6,7], Scott E. Baker[2,8], Kirk A. Bartholomew[9], Pedro M. Coutinho[10], Susann Erdmann[3], Thomas J. Fowler[11], Allen C. Gathman[12], Vincent Lombard[10], Bernard Henrissat[10], Nicole Knabe[3]*, Ursula Kües[13], Walt W. Lilly[12], Erika Lindquist[2], Susan Lucas[2], Jon K. Magnuson[8], François Piumi[6,7], Marjatta Raudaskoski[14], Asaf Salamov[2], Jeremy Schmutz[2], Francis W.M.R. Schwarze[15], Patricia A. vanKuyk[16], J. Stephen Horton[17**], Igor V. Grigoriev[2**], Han A.B. Wösten[1**]

[1]Department of Microbiology and Kluyver Centre for Genomics of Industrial Fermentations, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

[2]DOE Joint Genome Institute, Walnut Creek, CA 94598, USA.

[3]Department of Microbiology, Friedrich Schiller University, Neugasse 25, 07743 Jena, Germany.

[4]Department of Plant Pathology and Microbiology, 900 University Ave, University of California, Riverside, California 92521, USA.

[5]CBS-KNAW Fungal Biodiversity Centre, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

[6]INRA, UMR1163, Biotechnologie des Champignons Filamenteux, Case 932, 163 Avenue de Luminy, F-13288 Marseille cedex 09, France.

[7]Universités Aix-Marseille I & II, UMR1163, BCF, Case 932, 163 Avenue de Luminy, F-13288 Marseille cedex 09, France.

[8]Chemical and Biological Process Development Group, Pacific Northwest National Laboratory, 902 Battelle Boulevard, P.O. Box 999, MSIN P8-60 Richland, WA 99352 USA.

[9]Biology Department, Sacred Heart University, 5151 Park Ave, Fairfield, CT 06457 USA.

[10]Architecture et Fonction des Macromolecules Biologiques, UMR6098, CNRS, Univ. Aix-Marseille I & II, 163 Avenue de Luminy, 13288 Marseille, France.

[11]Department of Biological Sciences, Southern Illinois University Edwardsville, 8 Circle Drive, Edwardsville IL 62026 USA.

[12]Department of Biology, Southeast Missouri State University, Cape Girardeau, MO 63701, USA.

[13]Division of Molecular Wood Biotechnology and Technical Mycology, Büsgen-Institute, University of Göttingen, Büsgenweg 2, 37077 Göttingen, Germany.

[14]Department of Biochemistry and Food Chemistry, University of Turku, Biocity A, Tykistönkatu 6A, FI-20520 Turku, Finland.

[15]Wood Protection & Biotechnology, Empa, Swiss Federal Laboratories for Materials Testing and Research, Lerchenfeldstrasse 5, CH-9014 St. Gallen, Switzerland.

[16]Molecular Microbiology, Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands.

[17]Department of Biological Sciences, Union College, Schenectady, NY 12308, USA.

*Current address : Dartmouth Medical School, 1 Rope Ferry Road, Hanover, NH 03755-1404, USA

** Corresponding authors

[1]*To whom correspondence may be addressed. E-mail*: hortons@union.edu; 'H.A.B.Wosten@uu.nl; and ivgrigoriev@lbl.gov

May 19, 2011

ACKNOWLEDGMENTS:

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract No. DE-AC52-07NA27344, Los Alamos National Laboratory under contract No. DE-AC02-06NA25396. The work was also supported by the Dutch Technology

LLNL-JRNL-485265

Foundation STW, Applied Science division of NWO and the Technology Program of the Ministry of Economic Affairs.

DISCLAIMER:

Formation of mushrooms and lignocellulose degradation encoded in the genome sequence of *Schizophyllum commune*

Robin A. Ohm[1], Jan F. de Jong[1], Luis G. Lugones[1], Andrea Aerts[2], Erika Kothe[3], Jason E. Stajich[4], Ronald P. de Vries[1,5], Eric Record[6,7], Anthony Levasseur[6,7], Scott E. Baker[2,8], Kirk A. Bartholomew[9], Pedro M. Coutinho[10], Susann Erdmann[3], Thomas J. Fowler[11], Allen C. Gathman[12], Vincent Lombard[10], Bernard Henrissat[10], Nicole Knabe[3]*, Ursula Kües[13], Walt W. Lilly[12], Erika Lindquist[2], Susan Lucas[2], Jon K. Magnuson[8], François Piumi[6,7], Marjatta Raudaskoski[14], Asaf Salamov[2], Jeremy Schmutz[2], Francis W.M.R. Schwarze[15], Patricia A. vanKuyk[16], J. Stephen Horton[17**], Igor V. Grigoriev[2**], Han A.B. Wösten[1**]

[1]Department of Microbiology and Kluyver Centre for Genomics of Industrial Fermentations, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

[2]DOE Joint Genome Institute, Walnut Creek, CA 94598, USA.

[3]Department of Microbiology, Friedrich Schiller University, Neugasse 25, 07743 Jena, Germany.

[4]Department of Plant Pathology and Microbiology, 900 University Ave, University of California, Riverside, California 92521, USA.

[5]CBS-KNAW Fungal Biodiversity Centre, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

[6]INRA, UMR1163, Biotechnologie des Champignons Filamenteux, Case 932, 163 Avenue de Luminy, F-13288 Marseille cedex 09, France.

[7]Universités Aix-Marseille I & II, UMR1163, BCF, Case 932, 163 Avenue de Luminy, F-13288 Marseille cedex 09, France.

[8]Chemical and Biological Process Development Group, Pacific Northwest National Laboratory, 902 Battelle Boulevard, P.O. Box 999, MSIN P8-60 Richland, WA 99352 USA.

[9]Biology Department, Sacred Heart University, 5151 Park Ave, Fairfield, CT 06457 USA.

[10]Architecture et Fonction des Macromolecules Biologiques, UMR6098, CNRS, Univ. Aix-Marseille I & II, 163 Avenue de Luminy, 13288 Marseille, France.

[11]Department of Biological Sciences, Southern Illinois University Edwardsville, 8 Circle Drive, Edwardsville IL 62026 USA.

[12]Department of Biology, Southeast Missouri State University, Cape Giraradeau, MO 63701, USA.

[13]Division of Molecular Wood Biotechnology and Technical Mycology, Büsgen-Institute, University of Göttingen, Büsgenweg 2, 37077 Göttingen, Germany.

[14]Department of Biochemistry and Food Chemistry, University of Turku, Biocity A, Tykistönkatu 6A, FI-20520 Turku, Finland.

[15]Wood Protection & Biotechnology, Empa, Swiss Federal Laboratories for Materials Testing and Research, Lerchenfeldstrasse 5, CH-9014 St. Gallen, Switzerland.

[16]Molecular Microbiology, Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands.

[17]Department of Biological Sciences, Union College, Schenectady, NY 12308, USA.

*Current address : Dartmouth Medical School, 1 Rope Ferry Road, Hanover, NH 03755-1404, USA

** Corresponding authors

The authors have no competing interests as defined by Nature Publishing Group, or other interests that might be perceived to influence the results and discussion reported in this paper.

The wood degrading fungus *Schizophyllum commune* is a model system for mushroom development. Here, we describe the 38.5 Mb assembled genome of this basidiomycete and application of whole genome expression analysis to study the 13,210 predicted genes. Comparative analyses of the *S. commune* genome revealed unique wood degrading machinery and mating type loci with the highest number of reported genes. Gene expression analyses revealed that one third of the 471 identified transcription factor genes were differentially expressed during sexual development. Two of these transcription factor genes were deleted. Inactivation of *fst4* resulted in the inability to form mushrooms, whereas inactivation of *fst3* resulted in more but smaller mushrooms than wild-type. These data illustrate that mechanisms underlying mushroom formation can be dissected using *S. commune* as a model. This will impact commercial production of mushrooms and the industrial use of these fruiting bodies to produce enzymes and pharmaceuticals.

The fungal kingdom comprises diverse and important organisms that impact agriculture, human health, carbon cycling, and biotechnology. The mushroom fruiting body is the most conspicuous form of the fungi and is found primarily in the basidiomycete group. Mushrooms produce anti-tumor and immuno-stimulatory molecules[1,2] and enzymes that can be used for bioconversions[3]. Moreover, they have been identified as promising cell factories for the production of pharmaceutical proteins[4]. The main economic value of mushrooms, however, is their use as food[1,2]. The world-wide production of edible mushrooms amounts approximately 2.5 million tons annually. Despite their economical interest, relatively little is known about how mushroom-forming fungi obtain nutrients and how fruiting bodies are formed. Many mushroom-forming fungi cannot be cultured in the lab nor genetically modified. The basidiomycete *Schizophyllum commune* is one of the notable exceptions. It can be cultured on defined media and it completes its life cycle in approximately 10 days. Moreover, molecular tools to study growth and development of *S. commune* have been developed. In fact, it is the only mushroom-forming fungus in which genes

have been inactivated by homologous recombination. The importance of *S. commune* as a model system is also exemplified by the fact that its recombinant DNA constructs will express in other mushroom-forming fungi[5].

S. commune is one of the most commonly found fungi and can be isolated from all continents except for the arctic regions. *S. commune* has been reported to be a pathogen of humans and trees but it mainly adopts a saprobic life style by causing white rot[6]. It is predominantly found on fallen branches and timber of deciduous trees. At least 150 genera of woody plants are substrates for *S. commune*, but it also colonizes softwood and grass silage (see[7]). The mushrooms of *S. commune* that are formed on these substrates are used as a food source in Africa and Asia.

In the life cycle of *S. commune*[8] meiospores germinate to form a sterile monokaryotic mycelium, in which each hyphal compartment contains one nucleus. This mycelium grows initially submerged but after a few days aerial hyphae are formed (**Fig. 1A, E**). Monokaryons that are confronted with each other will fuse. A fertile dikaryon is formed when the alleles of the mating-type loci *matA* and *matB* of the partners differ. A short exposure to light is essential for fruiting, while a high concentration of carbon dioxide and high temperatures (30-37°C) are inhibitory. Mushroom formation is initiated with the aggregation of aerial dikaryotic hyphae. These aggregates (**Fig. 1B, F**) form fruiting body primordia (**Fig. 1C, G**), which further develop into mature fruiting bodies (**Fig. 1D, H**). Karyogamy and meiosis take place in the basidia within the mature fruiting body. The resulting basidiospores can give rise to new monokaryotic mycelia.

Here, we report the genomic sequence of the monokaryotic *S. commune* strain H4-8 and demonstrate the role of this basidiomycete as a model system to study mushroom formation.

RESULTS

Genome of *S. commune*

Sequencing genomic DNA of *S. commune* strain H4-8 with 8.29x coverage (**Supplementary Table 1** online) revealed 38.5 Megabase genome assembly with 11.2% repeat content (**Supplementary Text1** on line). The assembly is contained on 36 scaffolds (**Supplementary Table 2** online), which represent 14 chromosomes[9]. A total of 13,210 gene models are predicted, with 42% supported by expressed sequenced tags (ESTs) and 69% being similar to proteins from other organisms (**Supplementary Tables 3** and **4** online). Clustering of the proteins of *S. commune* with those of other sequenced fungi (a phylogenetic tree of the organisms used in the analysis is shown in **Supplementary Fig. 1** online) resulted in 7055 groups containing at least one *S. commune* protein (**Supplementary Table 5** online). Analysis of these clusters suggested that 39% of the *S. commune* proteins have orthologs in the Dikarya and are thus conserved in the Basidiomycota and Ascomycota (**Supplementary Table 6** online). Intriguingly, a similar percentage of proteins (36%) are unique to *S. commune.* Of these proteins, 46% have at least one inparalog (*i.e.* a gene resulting from a duplication within the genome) in *S. commune.* The uniqueness of the *S. commune* proteome is also illustrated by the presence of protein family (PFAM) domains (**Supplementary Text2** online) and the fact that only 43% of the predicted genes (5,703 out of the 13,210) could be annotated with a gene ontology (GO) term.

Global gene expression analysis

Whole genome expression was analysed in four developmental stages (monokaryon, stage I aggregates, stage II primordia, mature fruiting bodies; see **Fig. 1**) using Massively Parallel Signature Sequencing (MPSS). The majority of genes are either expressed in all stages (4859 genes) or not expressed (5308 genes) (**Fig. 2A,D, Supplementary Table 7** online). 59.8% of the 13,210 predicted genes are expressed in at least 1 developmental stage (**Supplementary Table 7** online). Fewer of the unique *S. commune* genes meet this threshold, whereas a higher percentage is observed for genes that share orthologs with Agaricomycetes or more distant fungi (**Supplementary Table 6** online). This suggests that genes specific to *S. commune* are subject to a

more stringent regulation. This is consistent with the observation that *S. commune* specific genes are over-represented in the pool of genes that are differentially expressed during the four developmental stages (**Supplementary Tables 8** and **9** online).

Anti-sense transcription is a widespread phenomenon in *S. commune*. 18.7% of the tags that could be related to a gene model originate from an anti-sense transcript. 42.3% of the predicted genes have anti-sense expression in one or more developmental stages (**Supplementary Tables 7** and **10** online). Northern hybridization with strand-specific probes confirmed the existence of anti-sense transcripts of *sc4* (Protein ID 73533) (data not shown). In the anti-sense direction, a relatively large number of genes are uniquely expressed in stage II (2888 genes) and relatively few genes are expressed in all stages (1195 genes) (**Fig. 2B**). In stage II, 4302 genes are expressed in both sense and anti-sense direction (**Fig. 2C**). This overlap is large compared to the other developmental stages.

Fruiting body development

An enrichment analysis of functional annotation was performed on the expression profiles of the four developmental stages (monokaryon, stage I aggregates, stage II primordia, mature fruiting bodies). Functional terms involved in protein production, energy production and hydrophobins are over-represented in genes that were up-regulated during formation of stage I aggregates (**Fig. 1** and **Supplementary Table 9** online). Genes involved in signal transduction, regulation of gene expression, cell wall biogenesis and carbohydrate metabolism are enriched in the group of down-regulated genes during formation of stage I aggregates. These functional terms are enriched in the up-regulated genes during formation of stage II primordia, whereas terms involved in protein and energy production are enriched in the down-regulated genes (**Fig. 1** and **Supplementary Table 9** online). During formation of mature fruiting bodies, genes encoding transcription factors, and genes involved in amino acid, glucose and alcohol metabolism are enriched in the group of down-regulated genes.

Whole genome expression analysis during mushroom formation has also been performed in *Laccaria bicolor*[10]. Regulation of orthologous gene pairs of *L. bicolor* and *S. commune* could therefore be correlated during fruiting. To this end, microarray expression profiles of free-living mycelium and mature fruiting bodies of *L. bicolor* were compared to the MPSS expression profiles of monokaryotic mycelium and mature fruiting bodies of *S. commune.* 6751 expressed genes from *S. commune* had at least 1 expressed ortholog in *L. bicolor*. The correlation of changes in expression of the functional annotation terms to which these orthologous pairs belong was determined. There were 15 GO terms, 2 KEGG terms, 4 KOG terms and 4 PFAM terms that showed a positive correlation in expression ($p < 0.01$) (**Supplementary Table 11** online). These terms include metabolic pathways (such as valine, leucine and isoleucine biosynthesis) and regulatory mechanisms (such as transcriptional regulation by transcription factors and signal transduction by G-protein alpha subunit). This indicates that regulation of these processes during mushroom formation is conserved in *S. commune* and *L. bicolor*.

Analysis of specific gene groups

Formation of a fertile dikaryon is regulated by the *matA* and *matB* mating type loci. Proteins encoded in these loci activate signalling cascades (see **Supplementary Text3** online) thus regulating target genes. These target genes include proteins that fulfil structural functions such as hydrophobins (see **Supplementary Text4** online) and enzymes. As a result, fruiting bodies are formed.

matA

The *matA* locus of strain H4-8 appears to have the highest homeodomain gene number in a fungal mating type locus described so far. This locus consists of two subloci, *Aα* and *Aß*, which are separated by 550 kb on chromosome I of strain H4-8. Annotation revealed that the *Aα* locus of H4-8 contains two divergently transcribed genes encoding Y and Z homeodomain proteins of the HD2

and HD1 class, respectively (**Fig. 3; Supplementary Table 12** online). These two genes, *aay4* and *aaz4*, have been previously described[1]. One homeodomain gene had also been identified in the *Aß* locus of H4-8[11]. The genomic sequence revealed that this locus actually contains six predicted homeodomain genes *abq6* (HD1), *abr6* (HD2), *abs6* (HD1), *abt6* (HD1, but lacking the NLS), *abu6* (HD1) and *abv6* (HD2) (**Fig. 3; Supplementary Table 12** online).

matB

Annotation of the genomic sequence of *S. commune* has revealed that the *matB* system contains more genes than previously envisioned. The *matB* locus comprises two linked loci *Bα* and *Bß*, which encode pheromones and pheromone receptors (**Fig. 3**; see[1]). Previously, one pheromone receptor gene was identified both in *Bα3* and *Bß2* of strain H4-8 (called *bar3* and *bbr2,* respectively)[12]. The genome sequence revealed four additional genes with high sequence similarity to these pheromone receptor genes, which we call *B* receptor-like genes 1-4 (*brl1-4*) (**Fig. 3**). Three of these genes are located near *bar3* and *bbr2* on scaffold 10, whereas one (*brl4*) is located on scaffold 8. MPSS analysis showed that the *brl* genes are expressed (**Supplementary Table 13** online). In fact, of all receptor and receptor-like genes, *brl3* shows the highest expression under the conditions tested.

Three and eight pheromone genes have been previously identified in the *Bα3* and *Bß2* loci, respectively[13]. One additional pheromone gene, *bpl5* (*B* pheromone-like), has been identified in the *Bα3* locus. Moreover, four additional pheromone-like genes were detected in the *Bß2* locus, called *bpl1-4* (**Fig. 3**). Based on the MPSS analysis, only *bpl2* failed to show expression (**Supplementary Table 13** online). The *Bα* gene *bpl5* and three of the new *Bß* pheromone-like genes show deviations from the consensus farnesylation signal with CASR for Bpl5, CTIA for Bpl1, CRLT for Bpl2 and CQLT for Bpl3. Previously, one of the pheromone genes (*bbp2(6)*) was shown to function with the deviating farnesylation signal CEVM[12]. This suggests that in *S. commune* only one amino acid residue in the consensus sequence of the farnesylation signal needs to be aliphatic.

Transcription factors

Genes for 471 putative transcription factors have been identified in the genome of *S. commune*, of which 311 are expressed in at least one developmental stage (**Supplementary Table 14** online). Of these genes, 56% are expressed in all developmental stages. 268, 200, 283 and 253 of these transcription factor genes were expressed, respectively, in the monokaryon, and during formation of stage I aggregates, stage II primordia and mushrooms. Interestingly, a cluster of monokaryotic specific transcription factors and a cluster of transcription factors that are upregulated in stage II primordia and/or in mature mushrooms were identified (**Fig. 4**). The latter cluster includes *fst3* (Protein ID: 257422) and *fst4* (Protein ID: 66861). These genes encode transcription factors that contain a fungal specific Zn(II)2Cys6 zinc finger DNA binding domain.

Genes *fst3* and *fst4* were inactivated by targeted gene deletions. The Δ*fst3* and Δ*fst4* monokaryons showed no phenotypic differences when compared to the wild-type. In contrast, the Δ*fst4*Δ*fst4* dikaryon did not fruit but produced more aerial hyphae when compared to the wild-type (**Fig. 5**). Apparently, Fst4 is involved in the switch between the vegetative phase and the reproductive phase. The Δ*fst3*Δ*fst3* dikaryon did form fruiting bodies. In fact, the mutant formed more, but smaller reproductive structures than those of the wild type (**Fig. 5**). Spatial and temporal regulation of fruiting body formation and sporulation was not altered in the Δ*fst3*Δ*fst3* strain. From these data we conclude that Fst3 inhibits formation of clusters of mushrooms.

Wood degradation by *Schizophyllum commune*

Degradation of lignin

S. commune has evolved its own set of enzymes that may degrade lignin. Such enzymes are classified as FOLymes[14]. These oxidative enzymes consist of lignin oxidases (LO families) and lignin-degrading auxiliary enzymes that generate H_2O_2 for peroxidises (LDA families). The LO

family consists of laccases (LO1), lignin peroxidases, manganese peroxidases, and versatile peroxidases (LO2) and cellobiose dehydrogenases (CDHs; LO3). *S. commune* contains 16 FOLyme genes and 11 genes that encode enzymes that are distantly related to these enzymes (**Table 1** and **Supplementary Table 15** online). The genome lacks genes encoding peroxidases of the LO2 family. However, it contains a CDH gene (LO3), 2 laccase genes (LO1), and a total of 13 LDA genes including 4 genes encoding glucose oxidases (LDA6), and benzoquinone reductases (LDA7) (**Table 1**).

Ustilago maydis, Cryptococcus neoformans, Aspergillus nidulans, Neurospora crassa and *Saccharomyces cerevisiae***,** which have not been reported to have ligninolytic activity, and the brown rot fungus *Postia placenta* contain a lower diversity of FOLymes compared to *S. commune* (**Table 1**). In contrast, the coprophillic fungus *Coprinopsis cinerea* and the white rot fungus *P. chrysosporium* contain many more FOLymes than *S. commune* with 40 and 27 members, respectively[14].

Polysaccharide degradation

S. commune has the most complete polysaccharide breakdown machinery of all basidiomycetes examined. The Carbohydrate-Active Enzyme database (CAZy) identified 240 candidate glycoside hydrolases (GH), 75 candidate glycosyl transferases (GT), 16 candidate polysaccharide lyases (PL), and 30 candidate carbohydrate esterases (CE) that are encoded in the genome of *S. commune* (**Table 1**, **Supplementary Table 16** online). Compared to the genomes of other basidiomycetes, *S. commune* has the highest number of GHs, and PLs. *S. commune* is rich in genes encoding pectin, hemicellulose and cellulose degrading enzymes (**Supplementary Table 17** online). In fact, *S. commune* has genes in each family involved in the degradation of these plant cell wall polysaccharides. The *S. commune* genome is particularly rich in members of the glycosyl hydrolase families GH93 (hemicellulose degradation), and GH43 (hemicellulose and pectin degradation), and the lyase families PL1, PL3, and PL4 (pectin degradation) (**Supplementary Table 17** online). The

pectinolytic capacity of *S. commune* is complemented with pectin hydrolases from family GH28, GH88 and GH105.

DISCUSSION

The phylum Basidiomycota contains roughly 30,000 described species, accounting for 37% of the true fungi[15]. The Basidiomycota comprises of two class level taxa (Wallemiomycetes, Entorrhizomycetes) and the subphyla Pucciniomycotina (rust), Ustilaginomycotina (smuts), and Agaricomycotina[16]. The Agaricomyoctina include the mushroom and puffball forming fungi, crust fungi, and jelly fungi. Currently five genomic sequences of Agaricomycotina are available: *P. chrysosporium*[17], *L. bicolor*[10], *P. placenta*[18], *C. neoformans*[19] and *C. cinerea*[20]. We here report the 38.5 Megabase genome assembly of *S. commune,* which represents the first genomic sequence of the family of the Schizophyllaceae. The genome of *S. commune* is predicted to have 13,210 genes. 36% of the encoded proteins have no ortholog in other fungi. The fact that much about the proteome of *S. commune* is still unknown is also illustrated by the fact that only 43% of the predicted genes could be annotated with a gene ontology (GO) term. This is a similar percentage as seen in other basidiomycetes: 30% in *L. bicolor*[10], 48% for *P. placenta*[18] and 49% for *P. chrysosporium*[17].

Substrate utilization

S. commune is reported to be a white rot fungus[6]. White-rot fungi degrade all woody cell-wall components, including the recalcitrant lignin. In contrast, brown-rotters efficiently degrade cellulose but only modify lignin, leaving a polymeric residue. *S. commune* primarily invades wood by growing through the lumen of vessels, tracheids, fibres and xylem rays. Adjacent parenchymatic cells in the xylem tissue are invaded via simple and bordered pits. As a consequence of this way of invasion, cellulose, hemicellulose, and pectin can serve as the primary carbon source for *S. commune*. Indeed, the genome of *S. commune* is rich in genes that encode enzymes that are

involved in the degradation of these polysaccharides. *S. commune* has genes in each family involved in the degradation of cellulose, hemicellulose, and pectin. The high number of pectinase genes correlates with earlier studies describing *S. commune* as one of the best basidiomycete pectinase producers[21]. *S. commune* also encodes carbohydrate active enzymes that degrade other polymeric sugars (*e.g.* those that act on starch, mannan and inulin). In fact, *S. commune* has the most complete polysaccharide breakdown machinery of all basidiomycetes examined. This complete machinery is consistent with the wide variety of substrates that support growth of *S. commune*.

Compared to plant polysaccharides, relatively little is known about how fungi degrade lignin. Fungi are assumed to degrade lignin with FOLymes[14]. These oxidative enzymes consist of lignin oxidases (LO families) and lignin-degrading auxiliary enzymes (LDA families). Evidence has shown that at least the LO2 family is involved in lignin degradation. This family consists of the lignin peroxidases, manganese peroxidases and versatile peroxidases. Involvement of laccases (LO1) and cellobiose dehydrogenases (CDHs; LO3) is still controversial. *S. commune* contains 16 genes encoding FOLymes. Genes of the LO2 family are lacking but the genome contains one CDH gene and 2 laccase genes. CDHs may participate in the degradation of cellulose, xylan and, possibly, lignin by generating hydroxyl radicals in a Fenton-type reaction. Laccases catalyse the one-electron oxidation of phenolic, aromatic amines and other electron-rich substrates with the concomitant reduction of O_2 to H_2O. They are classified as having either low or high redox potential[22]. It is not clear whether they belong to the high or low redox potential enzyme categories.

When the genomes of the white rot fungi *S. commune* and *P. chrysosporium*[17] and the brown rot fungus *P. placenta*[18] are compared, it is clear that *S. commune* has evolved its own set of FOLymes. *P. chrysosporium* lacks genes encoding laccases (LO1). It is thought to degrade lignin with the enzymes encoded by 16 isogenes of peroxidases (LO2), one CDH gene (LO3) and 4 genes of the multicopper oxidase (MCO) super family. In contrast, *P. placenta* contains 2 laccase-encoding genes (LO1) but lacks members of the LO2 and LO3 families. Since *S. commune* and *P.*

placenta lack true LO2 FOLymes, one would expect a low number of LDAs that are responsible for H_2O_2 production for the peroxidases. This is not the case. *S. commune* contains more and a higher diversity of LDAs when compared to *P. chrysosporium*. For instance, *S. commune* contains four glucose oxidase (LDA6) genes, whereas one or none are normally present in fungi. In the absence of peroxidases of the LO2 family, it is expected that the glucose oxidases of *S. commune* serve another function. Glucose oxidases convert glucose into gluconic acid. This acid solubilises inorganic phosphate and thus aids in the uptake of this nutrient[23].

Mushroom formation

The mating type loci

Monokaryons of *S. commune* will fuse when they are confronted with each other. Formation of a fertile dikaryon is regulated by the *matA* and *matB* mating type loci. The genome sequence has revealed that *matA* and *matB* of *S. commune* represent the fungal mating type loci that contain the highest number of genes. The *matB* locus comprises two linked loci *Bα* and *Bβ*, which encode pheromones and pheromone receptors[1]. Nine allelic specificities have been identified for both loci, resulting in 81 different mating types for *matB*. It was previously described that the *Bα3* and *Bβ2* loci of H4-8 each contain one pheromone receptor gene and three and eight pheromone genes, respectively[12,13]. 5 additional pheromone genes and 4 pheromone receptor-like genes were identified in the genome of H4-8. These newly identified receptor-like genes are present in a *matB* deletion strain, which has no pheromone response with any mate (T. Fowler, unpublished results). This raises the question whether the four receptor genes are functional in *matB*-regulated development. MPSS analysis showed that they are all expressed, which suggests that they do not represent pseudogenes.

The *matA* locus consists of two subloci, Aα and *Aβ*, of which 9 and 32 alellic specificities are expected to occur in nature[1]. These loci are separated by 550 kb on chromosome I of strain H4-

8. This large distance has not been found in other fungi that have a tetrapolar mating system. The functionally well-characterized *Aα* locus showed no significant differences from the published descriptions[1]. It is composed of two genes encoding Y and Z homeodomain proteins of the HD2 and HD1 class, respectively. The Y and Z proteins, as in other basidiomycetes, interact in non-self combinations to activate the A-pathway of sexual development[1,24]. Notably, a nuclear localization signal is present in Y but not in Z. This is consistent with non-self interaction of the two proteins taking place in the cytosol, which is followed by the translocation of the active protein complex into the nucleus[1].

Aß of *S. commune* has been studied much less compared to the *Aα* locus. Interestingly, *Aß* reflects the highest homeodomain gene complexity of a fungal mating type locus described to date. It contains 4 homeodomain genes of the HD1 class and two of the HD2 class. The *Aß* locus of *S. commune* thus resembles that of *C. cinerea*, which consists of two pairs of functional HD1 and HD2 homeodomain genes (b and d)[25]. The large number of genes in *matAß* would explain why recombination analyses predict as many as 32 mating specificities for this locus[26]. Taken together, *S. commune* seems ideal to identify the evolutionary pathways for creating high numbers of allelic specificities for enhancing outbreeding versus inbreeding rates.

Whole genome expression

Little is known about molecular processes that control formation of fruiting bodies in basidiomycetes other than the role of the mating type loci[8]. Therefore, we performed a whole genome expression analysis in four developmental stages (*i.e.* in the sterile monokaryon, and in stage I aggregates, stage II primordia, and mature fruiting bodies of the dikaryon; see **Fig. 1**). MPSS showed that relatively few genes were specifically expressed in the monokaryon (284 genes) and in stage I aggregates and the mature mushrooms (128 genes in both cases). Interestingly, 467 genes were specifically expressed in stage II primordia. This suggests that this stage represents a major developmental switch. This is supported by the fact that genes involved in signal

transduction and regulation of gene expression are enriched in the group of up-regulated genes during formation of stage II primordia. Of interest, expression of these gene groups, among others, show a positive correlation during mushroom formation of *S. commune* and *L. bicolor*. This suggests that regulation of mushroom formation is a conserved process in the Agaricales.

Anti-sense expression was found to occur widely in *S. commune*. About 20% of all sequenced mRNA tags originated from an anti-sense transcript and more than 5600 of the predicted genes showed anti-sense expression in one or more developmental stages. Anti-sense transcription was most pronounced in stage II primordia. In this stage, more than 4300 genes were expressed in both sense and anti-sense direction and more than 800 genes were expressed in the anti-sense direction only. Previously, MPSS showed anti-sense transcripts in *Magnaporthe grisea*[27]. Little is known about the function of these transcripts in fungi. The circadian clock of *N. crassa* is entrained in part by the action of an anti-sense transcript of a clock component locus[28], possible by RNA interference. It is tempting to speculate that anti-sense transcripts also regulate mRNA levels in *S. commune*. Natural anti-sense transcripts in eukaryotes have also been implicated in other processes such as translational regulation, alternative splicing and RNA editing[29]. The anti-sense transcripts of *S. commune* may also have such functions. In all these cases, the anti-sense transcripts could be functional in the developmental switch that would occur when stage II primordia are formed.

Transcriptional regulators

The fact that gene regulation seems to be conserved in the Agaricales made us decide to study the transcriptional regulators in more detail. 471 genes were identified that are predicted to encode a transcription factor. 268 of these transcription factors were expressed in the monokaryon, whereas 200, 283 and 253 were expressed during formation of stage I aggregates, stage II primordia and mushrooms, respectively. The relatively high number of transcription factors that are expressed during formation of stage II primordia again point to a major switch that takes place during this developmental stage.

A cluster of monokaryotic specific transcription factors and a cluster of transcription factors that are up-regulated in stage II primordia and/or in mature mushrooms were identified. Genes *fst3* and *fst4* that encode transcriptional regulators with a fungal specific Zn(II)2Cys6 zinc finger DNA binding domain belong to the latter cluster and were inactivated by targeted gene deletion. Growth and development of monokaryotic strains in which *fst3* or *fst4* were inactivated were not affected. Phenotypic differences, however, were observed in the dikaryon. The Δ*fst4*Δ*fst4* dikaryon did not fruit but produced more aerial hyphae when compared to the wild-type. On the other hand, the Δ*fst3*Δ*fst3* dikaryon formed more, but smaller, fruiting bodies than the wild-type. Taken together, we conclude that Fst4 is involved in the switch between the vegetative and the reproductive phase and that Fst3 inhibits formation of clusters of mushrooms. Inhibition of such clusters could be important in a natural environment to ensure sufficient energy is available for full fruiting body development. Interestingly, *fst3* and *fst4* have homologs in other mushroom forming fungi and it is tempting to speculate that they have a similar function in these organisms. This is supported by the observation that the homologs of *fst3* and *fst4* are up-regulated in young fruiting bodies of *L. bicolor* compared to free-living mycelium[10]. In mature fruiting bodies of *L. bicolor*, the expression level of the homolog of *fst3* remains constant compared to young fruiting bodies, whereas the *fst4* homolog returns to the level of expression in the free-living mycelium.

Summary

The genomic sequence of *S. commune* will be an essential tool to unravel mechanisms by which mushroom forming fungi degrade their natural substrates and form their reproductive structures. The large variety of genes that encode extracellular enzymes that act on polysaccharides likely explains why *S. commune* is so common in nature. Moreover, the genome sequence suggests that *S. commune* has a unique mechanism for lignin degradation. The deep capture of gene expression via MPSS has provided us with leads on how mushroom formation is regulated. In addition to certain transcription factors, anti-sense transcription may play an important role in this process. Without

doubt, the understanding of physiology and sexual reproduction of *S. commune* will have an impact on the commercial production of edible mushrooms and the use of mushrooms as a cell factory.

ACKNOWLEDGEMENTS

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract No. DE-AC52-07NA27344, Los Alamos National Laboratory under contract No. DE-AC02-06NA25396. The work was also supported by the Dutch Technology Foundation STW, Applied Science division of NWO and the Technology Program of the Ministry of Economic Affairs.

CONTRIBUTION TO THE MANUSCRIPT

JSH, HW, TF, WL, AG, EK, and SB wrote the proposal; LL, JdJ, EL and RO isolated RNA and DNA and made libraries; EL and SL coordinated sequencing of the genome; JSc assembled the genome; IG, RO, KB, JSt, SB, EK and HW coordinated the annotation process; IG, AA, AS, JSt, TF, ER, AL, FP, UK, JM, JdJ, RdV, PC, VL, BH, WL, AG, PvK, KB, JSH, EK, SE, NK, RO, MR annotated genes; RO, JdJ, FS, and LL performed experiments; RO, JdJ, FS, LL and HW interpreted and designed experiments; RO, AA, JSt, ER, JdJ, RdV, FS, KB, EK, JSH, IG, SH, and HW wrote the paper; RO and HW coordinated writing of the paper.

LEGENDS

Figure 1 Development of *S. commune*. Four-day-old (A-C; E-G) and eight-day-old (D, H) colonies grown from homogenates showing typical developmental stages in the life-cycle of *S. commune*. A monokaryon forms sterile aerial hyphae that form a fluffy white layer on top of the vegetative mycelium (A, E). Aerial hyphae of a dikaryon interact with each other to form stage I aggregates

(B, F) which, after a light stimulus, develop into stage II primordia (C, G). These primoridia further differentiate into sporulating mushrooms (D, H). An enrichment analysis shows that particular functional terms are over-represented in genes that are up- or down-regulated during a developmental transition. These terms are indicted below panels A-H. A-D represent cultures grown in 9 cm Petri-dishes, whereas E-H represent magnifications thereof. Bar represents 1 mm (H), 2.5 mm (E, F) and 5 mm (G).

Figure 2 Gene expression in four developmental stages of *S. commune* illustrated by VENN diagrams (A-C) and a heat map (D). The cut-off for expression is 4 TPM. VENN diagrams in (A) and (B) show the overlap of genes that are expressed in sense and anti-sense direction in the four developmental stages, respectively. As an example, in (A) 61 genes are expressed in sense direction in stage I and stage II, 4859 genes are expressed in all stages, 132 genes are expressed in the monokaryon and mature fruiting bodies and 5308 genes are not expressed in any of the stages. (C) VENN diagram of the overlap in genes that show sense and anti-sense expression in a particular developmental stage and with all stages combined. (D) Heat map of expression of the *S. commune* genes in the four developmental stages. Bar on top of the panel represents expression values between 0 and 300 TPM. Genes with expression values higher than 300 TPM are also indicated in red. The bar on the right indicates a cluster of 366 highly expressed and differentially regulated genes. Annotation information of the genes in this cluster is given in **Supplementary Table 18** online.

Figure 3 Distribution of genes encoding HD1 and HD2 homeodomain proteins in the *matA* locus and genes encoding pheromone receptors and pheromones in the *matB* locus of *S. commune* strain H4-8. The *matA* and *matB* loci are positioned on scaffold 1 and 10, respectively. One additional pheromone receptor gene, *brl4*, has been identified on scaffold 8.

Figure 4 Histogram (A) and heat map (B) of expression of the 471 transcription factors in the genome of *S. commune.* The histogram (A) shows the percentage of transcription factor genes that are differentially expressed between stages of development. The heat map (B) shows a cluster containing predominantly monokaryon specific transcription factors and a cluster containing predominantly stage II and/or mushroom specific transcription factors. These clusters are enlarged on the right part of the heat map. The latter group contains two fungal specific transcription factor genes, called *fst3* and *fst4* (protein IDs 257422 and 66861, respectively).

Figure 5 Fruiting body formation in a wild-type dikaryon (A, D) and in dikaryons in which *fst3* (B, E) and *fst4* (C, F) have been inactivated. D-F show a magnification of part of the colonies shown in A-C. Bar represents 5 mm (D-F).

Table 1. Comparison of the number of FOLymes and CAZymes of *S. commune* with those of other fungi. LO1 represents the laccases, LO2 peroxidases, LO3 cellobiose dehydrogenases, LDA1 aryl alcohol oxidases, LDA2 vanillyl-alcohol oxidases, LDA3 glyoxal oxidases, LDA4 pyranose oxidases, LDA5 galactose oxidases, LDA6 glucose oxidases, LDA7 benzoquinone reductases, and LDA8 alcohol oxidases. GH represent the glycoside hydrolases, GT glycosyl transferases, PL polysaccharide lyases, and CE carbohydrate esterases.

	FOLymes											CAZymes			
Species	LO1	LO2	LO3	LDA1	LDA2	LDA3	LDA4	LDA5	LDA6	LDA7	LDA8	GH	GT	PL	CE
S. commune	2	0	1	1	0	2	1	0	4	4	1	240	75	16	30
C. cinerea	17	1	1	18	0	0	0	0	1	2	0	211	71	13	54
L. bicolor	9	1	0	4	0	0	0	0	3	2	0	163	88	7	20
P. placenta	2	0	0	3	0	0	0	0	0	1	2	124	51	4	13
P. chrysosporium	0	16	1	3	0	1	1	0	1	4	0	181	66	4	20
C. neoformans	0	0	0	0	0	0	0	0	0	0	0	75	64	3	8
U. maydis	0	0	0	0	0	1	0	1	1	1	0	101	64	1	19
S. cerevisiae	0	0	0	0	0	0	0	0	0	3	0	46	68	0	3
A. nidulans	1	0	1	0	1	0	0	0	1	0	0	250	91	21	32
N. crassa	5	0	2	1	0	0	0	1	1	1	1	173	76	4	23

METHODS

Strains and Culture conditions

S. commune was routinely grown at 25 °C on minimal medium (MM) with 1% glucose and with or without 1.5% agar[30]. Liquid cultures were shaken at 225 rpm. Glucose was replaced with 4% glycerol for cultures used in the isolation of genomic DNA. All *S. commune* strains used were isogenic to strain 1-40[31]. Strain H4-8 (*matA43matB41*; FGSC #9210) was used for sequencing. EST libraries were generated from H4-8 and from a dikaryon that resulted from a cross between H4-8 and strain H4-8b (*matA4 matB43*)[32]. Strains 4-39 (*matA41matB41;* CBS 341.81) and 4-40 (*matA43matB43;* CBS 340.81) were used for MPSS analysis. These strains show a more synchronized fruiting compared to a cross between H4-8 and H4-8b. Partial sequencing of the haploid genome revealed that strains 4-40 and 4-39 have minor sequence differences (< 0.2%) with strain H4-8 (data not shown).

Isolation of genomic DNA, genome sequencing and assembly

Genomic DNA of *S. commune* was isolated as described[30] and sequenced with the use of a whole-genome shotgun strategy. All data were generated by paired-end sequencing of cloned inserts with 6 different insert sizes using Sanger technology on ABI3730xl sequencers. The data were assembled using the whole-genome shotgun assembler Arachne (http://www.broad.mit.edu/wga/).

EST library construction and sequencing

Cultures were inoculated on MM plates with 1% glucose using mycelial plugs as an inoculum. Strain H4-8 was grown for 4 days in the light, whereas the dikaryon H4-8 x H4-8.3 was grown for 4 days in the dark and 8 days in the light. Mycelium of the dikaryotic stages was combined and RNA was isolated as described[30]. The PolyA+ RNA fraction was obtained using the Absolutely mRNA Purification kit and manufacturer's instructions (Stratagene, La Jolla, CA). cDNA synthesis and cloning followed the SuperScript plasmid system procedure with Gateway technology for

cDNA synthesis and cloning (Invitrogen, Carsbad, CA). For the monokaryon, two size ranges of cDNA were cut out of the gel to generate two cDNA libraries (JGI library codes CBXY for range 0.6k-2kb and CBXX for the range >2kb). For the dikaryon, cDNA was used in the range >2kb, resulting in library CBXZ. The cDNA inserts were directionally ligated into vector pCMVsport6 (Invitrogen) and introduced into ElectroMAX T1 DH10B cells (Invitrogen). Plasmid DNA for sequencing was produced by rolling circle amplification (Templiphi, GE Healthcare, Piscataway, NJ). Subclone inserts were sequenced from both ends using Big Dye terminator chemistry and ABI 3730 instruments (Applied Biosystems, Foster City, CA).

Annotation methods

Gene models in the genome of *S. commune* were predicted using Fgenesh[33], Fgenesh+[33], Genewise[34] and Augustus[35]. Fgenesh was trained for *S. commune* with a sensitivity of 72% and a specificity of 74%. Augustus *ab initio* gene predictions were generated with parameters based on *C. cinerea* gene models[20]. In addition, about 31,000 *S. commune* ESTs were clustered into nearly 9,000 groups. These groups were either directly mapped to the genomic sequence with a threshold of 80% coverage and 95% identity, included as putative full-length (FL) genes, or used to extend predicted gene models into FL genes by adding 5' and/or 3' UTRs. Since multiple gene models were generated for each locus, a single representative model at each locus was computationally selected based on EST support and similarity to protein sequences in the NCBI non-redundant database. This resulted in a final set of 13,210 predicted genes, of which 1314 genes have been manually curated. In 66 cases, models were created or coordinates were changed.

All predicted gene models were functionally annotated by homology to annotated genes from NCBI non-redundant set and classified according to Gene Ontology (GO)[36], eukaryotic orthologous groups (KOGs)[37], KEGG metabolic pathways[38], and Protein Family (PFAM) domains[39].

Repeat content

RepeatModeler 1.0.3 (http://www.repeatmasker.org/RepeatModeler.html) was used to generate *de novo* repeat sequence predictions for *S. commune*. Repeats were classified by comparison to the RepBase database (http://www.girinst.org/repbase/index.html). RepeatModeler produced 76 families of repeats used as a search library in RepeatMasker (http://www.repeatmasker.org).

Data availability

S. commune assemblies, annotations, and analyses are available through the interactive JGI Genome Portal at http://jgi.doe.gov/Scommune. Genome assemblies together with predicted gene models and annotations were also deposited at DDBJ/EMBL/GenBank under the project accession ADMJ00000000.

Orthologs of *S. commune* proteins in the fungal kingdom

Proteins of *S. commune* were assigned to orthologous groups with OrthoMCL V2.0[40] with an inflation value of 1.5. Members of such groups were assigned as orthologs (in the case of proteins from another species) or inparalogs (in the case of proteins from *S. commune*). Orthologs were determined in *C. cinerea*[20], *L. bicolor*[10], *P. placenta*[18], *P. chrysosporium*[17], *C. neoformans*[19], *U. maydis*[41], *S. cerevisiae*[42], *A. nidulans*[43] and *N. crassa*[44]. All versus all BLASTp analysis was performed using NCBI standalone BLAST v2.2.20 with an E-value of 1e-5 as a cut-off. Custom scripts were used to further analyse the orthologous groups resulting from the OrthoMCL analysis. The evolutionary conservation for each orthologous group was expressed as the taxon this orthologous group was most specifically confined to (see **Supplementary Fig. 1** online).

Representation analysis

FuncAssociate 2.0[45] was used to study over- and under-representation of taxon-specific genes and of functional annotation terms in sets of differentially regulated genes. Default settings were used with a p-value of 0.05 or 0.01 as the cut off.

Protein families

The PFAM database version 24.0[39] was used to identify PFAM protein families. Custom scripts in Python were written to group genes on basis of their PFAM domains. Differences in the number of predicted proteins belonging to a PFAM family across the fungal domain was determined using the Student's t-test. When Agaricales were compared to the rest of the Dikarya or when *S. commune* was compared to the Agaricales, only groups with a minimum of 5 members in at least one of the fungi were analysed. In the case *S. commune* was compared to the rest of the Dikarya, only groups with a minimum of 5 members in at least four of the fungi were analysed. In all cases, a p value of 0.05 was used as a cut off. Similar results were obtained using the non-parametric Mann-Whitney-U test.

CAZy annotation

Annotation of carbohydrate-related enzymes was performed using the Carbohydrate-Active Enzyme database (CAZy) annotation pipeline[46]. Ambiguous family attributions were processed manually along with all identified models that presented defects (deletions, insertions, splicing problems, etc.). Each protein was also compared to a library of experimentally characterized proteins found in CAZy to provide a functional description.

FOLy annotation

Lignin oxidative enzymes (FOLymes)[14] were identified by BLASTP analysis of the *S. commune* gene models against a library of FOLy modules using an e-value <0.1. The resulting 68 protein models were manually analysed using the BLASTP results as well as multiple sequence alignments and functional inference based on phylogeny[47]. Basically, a protein was identified as a FOLyme when it showed a similarity score above 50% with sequences of biochemically characterised

enzymes. When the similarity score was <50% the proteins were scored as a FOLyme related protein.

MPSS expression analysis

Total RNA was isolated from the monokaryotic strain 4-40 and from the dikaryon resulting from a cross between 4-40 and 4-39. A 7-day-old colony grown on solid MM at 30 ^{o}C in the dark was homogenized in 200 ml MM using a Waring blender for 1 min at low speed. 2 ml of the homogenized mycelium was spread out over a polycarbonate membrane that was placed on top of solidified MM. Vegetative monokaryotic mycelium was grown for 4 days in the light. The dikaryon was grown for 2 and 4 days in the light to isolate mycelium with stage I aggregates and stage II primordia, respectively. Mature mushrooms of 3 days old were picked from dikaryotic cultures that had grown for 8 days in the light. RNA was isolated as described[30]. MPSS was performed essentially as described[48] except that after DpnII digestion MmeI was used to generate 20 bp tags. Tags were sequenced using the Clonal Single Molecule Array technique (Illumina, Hayward, CA, US). Between 4.2 and 7.6 million tags of 20 bp were obtained for each of the stages. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus[49] and are accessible through GEO Series accession number GSE21265 (http://www.ncbi.nlm.nih.gov/geo/). Programs were developed in the programming language Python to analyze the data. Tag counts were normalized to tags per million (TPM). Those with a maximum of < 4 TPM in all developmental stages were removed from the data set. This data set consisted of a total of 40,791 unique tags. Of these tags, 61.7% and 58.6% could be mapped to the genome sequence and the predicted transcripts, respectively, using a perfect match as the criterion. The mapped tags accounted for 71.4% and 70.8% of the total number of tags, respectively. For comparison, 97.4% of the ESTs from *S. commune* strain H4-8 could be mapped to the assembly. Unmapped tags can be explained by sequencing errors in either tag or genomic DNA. Moreover, RNA editing may have altered the transcript sequencing to produce tags that do not match the genome perfectly. It may

also be that the assigned UTR is incomplete or that the DpnII restriction site that defines the 5' end of the tag is too close to the polyA-tail of the mRNA. TPM values of tags originating from the same transcript were summed to assess their expression levels. A transcript is defined as the predicted coding sequence (CDS) extended with 400 bp flanking regions at both sides.

Comparison of gene expression in *L. bicolor* and *S. commune*

Whole genome expression analysis of *L. bicolor*[10] and *S. commune* was done essentially as described[50]. For *L. bicolor*, the microarray values from replicates were averaged. Expression values of genes were increased by one and the ratio between monokaryon and mushrooms (for *S. commune*) and free-living mycelium and mature fruiting bodies (for *L. bicolor*) was log-transformed. All expressed genes from *S. commune* that had at least 1 expressed ortholog in *L. bicolor* were taken into account, resulting in a total of 6751 orthologous pairs. These pairs were classified on basis of functional annotation terms. Correlation of changes in expression of these gene classes was expressed as the Pearson correlation coefficient. Only GO-terms with 10-200 pairs were used in the analysis. In the case of PFAM domains, a minimum number of 10 ortholog pairs were used.

Deletion of transcription factors *fst3* and *fst4*

The transcription factor genes *fst3* (proteinID: 257422) and *fst4* (proteinID: 66861) were deleted using vector pDelcas[32]. Transformation of *S. commune* strain H4-8 was done as described[30]. Regeneration medium contained no antibiotic, whereas selection plates contained 20 μg ml^{-1} nourseothricin. Deletion of the target gene was confirmed by PCR. Compatible monokaryons with a gene deletion were selected from spores originating from a cross of the mutant strains with wild-type strain H4-8.3.

REFERENCES

1) Kothe, E. Mating-type genes for basidiomycete strain improvement in mushroom farming. *Appl. Microbiol. Biotechnol.* **56**, 602-612 (2001).

2) Kües, U. & Liu, Y. Fruiting body production in basidiomycetes. *Appl. Microbiol. Biotechnol.* **54**, 141-152 (2000).

3) Lomascolo, A., Stentelaire, C., Asther, M. & Lesage-Meessen, L. Basidiomycetes as new biotechnological tools to generate natural aromatic flavours for the food industry. *Trends Biotechnol.* **17**, 282-289 (1999).

4) Berends, E., Scholtmeijer, K., Wösten, H.A.B., Bosch, D. & Lugones, L.G. The use of mushroom-forming fungi for the production of N-glycosylated therapeutic proteins. *Trends Microbiol.* **17**, 439-443 (2009).

5) Alves, A.M. *et al.* Highly efficient production of laccase by the basidiomycete *Pycnoporus cinnabarinus*. *Appl. Environ. Microbiol.* **70**, 6379-6384 (2004).

6) Schmidt, O. & Liese, W. Variability of wood degrading enzymes of *Schizophyllum commune*. *Holzforschung* **34**, 67-72 (1980).

7) de Jong, J.F. Aerial hyphae of *Schizophyllum commune*: their function and formation. PhD Thesis University of Utrecht (2006).

8) Wösten, H.A.B. & Wessels, J.G.H. The emergence of fruiting bodies in basidiomycetes. In *The Mycota. Part I: Growth, Differentiation and Sexuality* (eds. Kües, U. & Fisher, R.) 393-414, (Springer Verlag, Berlin, 2006).

9) Asgeirsdottir, S.A., Schuren, F.H.J. & Wessels, J.G.H. Assignment of genes to pulse-field separated chromosomes of *Schizophyllum commune*. *Mycol. Res*. **98**, 689-693 (1994).

10) Martin, F. *et al*. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**, 88-92 (2008).

11) Shen, G.P. *et al.* The Aalpha6 locus: its relation to mating-type regulation of sexual development in *Schizophyllum commune*. *Curr. Genet.* **39**, 340-345 (2001).

12) Fowler, T.J., Mitton, M.F., Vaillancourt, L.J. & Raper, C.A. Changes in mate recognition through alterations of pheromones and receptors in the multisexual mushroom fungus *Schizophyllum commune*. *Genetics* **158**, 1491–1503 (2001).

13) Fowler, T.J., Mitton, M.F., Rees, E.I. & Raper, C.A. Crossing the boundary between the *Bα* and *Bß* mating-type loci in *Schizophyllum commune*. *Fungal Genet. Biol.* **41**, 89–101 (2004).

14) Levasseur, A. *et al.* FOLy: an integrated database for the classification and functional annotation of fungal oxidoreductases potentially involved in the degradation of lignin and related aromatic compounds. *Fungal Genet. Biol*. **45**, 638-645 (2008).

15) Kirk, P.M., Cannon, P.F., David, J.C. & Stalpers, J.A. *Ainsworth and Bisby's Dictionary of the Fungi.* (CAB International, Wallingford, 2001).

16) Hibbett, D.S. *et al.* A higher-level phylogenetic classification of the Fungi. *Mycol. Res*. **111**, 509-547 (2007).

17) Martinez, D. *et al.* Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat. Biotechnol.* **22**, 695-700 (2004).

18) Martinez, D. *et al*. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc. Natl. Acad. Sci. U S A* **106**, 1954-1959 (2009).

19) Loftus, B.J. *et al.* The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**, 1321-1324 (2005).

20) Stajich, J.S. *et al.* Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc. Natl. Acad. S.c U S A* (in press).

21) Xavier-Santos, S. *et al.* Screening for pectinolytic activity of wood-rotting basidiomycetes and characterization of the enzymes. *Folia Microbiol.* **49**, 46-52 (2004).

22) Xu, F. *et al.* A study of a series of recombinant fungal laccases and bilirubin oxidase that exhibit significant differences in redox potential, substrate specificity, and stability. *Biochim. Biophys. Acta.* **1292**, 303-11 (1996).

23) Xiao, C. *et al.* Isolation of phosphate-solubilizing fungi from phosphate mines and their effect on wheat seedling growth. *Appl. Biochem. Biotechnol*. **159**, 330-342 (2009).

24) Spit, A., Hyland, R.H., Mellor, E.J. & Casselton, L.A. A role for heterodimerization in nuclear localization of a homeodomain protein. *Proc. Natl. Acad. Sci. USA* 95:6228-6633 (1998).

25) Casselton, L.A. & Olesnicky, N.S. Molecular genetics of mating recognition in basidiomycete fungi. *Microbiol. Mol. Biol. Rev.* **62**, 55-70 (1998).

26) Raper, J. *Genetics of Sexuality of Higher Fungi* 1-283, (The Roland Press, New York. 1966).

27) Gowda, M. *et al.* Deep and comparative analysis of the mycelium and appressorium transcriptomes of *Magnaporthe grisea* using MPSS, RL-SAGE, and oligoarray methods. *BMC Genomics* **7**, 310 (2006).

28) Kramer, C., Loros, J.J., Dunlap, J.C. & Crosthwaite, S.K. Role for antisense RNA in regulating circadian clock function in *Neurospora crassa. Nature* **421**, 948-952 (2003).

29) Lavorgna, G., Dahary, D., Lehner, B., Sorek, R., Sanderson, C.M. & Casari, G. In search of antisense. *Trends Biochem. Sc.* **29**, 88-94 (2004).

30) van Peer, A.F., de Bekker, C., Vinck, A., Wösten, H.A.B. & Lugones, L.G. Phleomycin increases transformation efficiency and promotes single integrations in *Schizophyllum commune*. *Appl. Environ. Microbiol.* **75**, 1243-1247 (2009).

31) Raper, J.R., Krongelb, G.S. & Baxter, M.G. The number and distribution of incompatibility factors in *Schizophyllum*. *Amer. Nat.* **92**, 221-232 (1958).

32) Ohm, R.A., de Jong, J.F., Berends, E., Wang, F., Wösten, H.A.B. & Lugones, L.G. An efficient gene deletion procedure for the mushroom-forming basidiomycete *Schizophyllum commune*. *World J. Microbiol. Biotechnol.* doi:10.1007/s11274-010-0356-0

33) Salamov, A.A. & Solovyev, V.V. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516-522 (2000).

34) Birney, E. & Durbin, R. Using GeneWise in the *Drosophila* annotation experiment. *Genome Res.* **10**, 547-548 (2000).

35) Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, 215-225 (2003).

36) Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25-29 (2000).

37) Koonin, E.V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7 (2004).

38) Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori M. The KEGG resource for deciphering the genome. *Nucl. Acids Res*. **32**, D277-80 (2004).

39) Finn, R.D. *et al.* The Pfam protein families database. *Nucl. Acids Res.* **38**, D211-222 (2010).

40) Li, L., Stoeckert, C.J. & Roos, D.S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178-2189 (2003).

41) Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97-101 (2006).

42) Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 563-567 (1996).

43) Galagan, J.E. *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**, 1105-1115 (2005).

44) Galagan, J.E. *et al.* The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**, 859-868 (2003).

45) Berriz, G. F., Beaver, J.E., Cenik, C., Tasan, M. & Roth, F.P. Next generation software for functional trend analysis. *Bioinformatics* **25**, 3043-3044 (2009).

46) Cantarel, B.L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucl. Acids Res.* **37**, D233-238 (2009).

47) Gouret, P. *et al.* FIGENIX: Intelligent automation of genomic annotation: expertise integration in a new software platform. *BMC Bioinformatics* **6**, 198 (2005).

48) Brenner, S. *et al.* Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**, 630-634 (2000).

49) Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. **30**, 207-210 (2002).

50) McCarroll S.A. *et al.* Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nat. Genet.* **36**, 197-204 (2004).

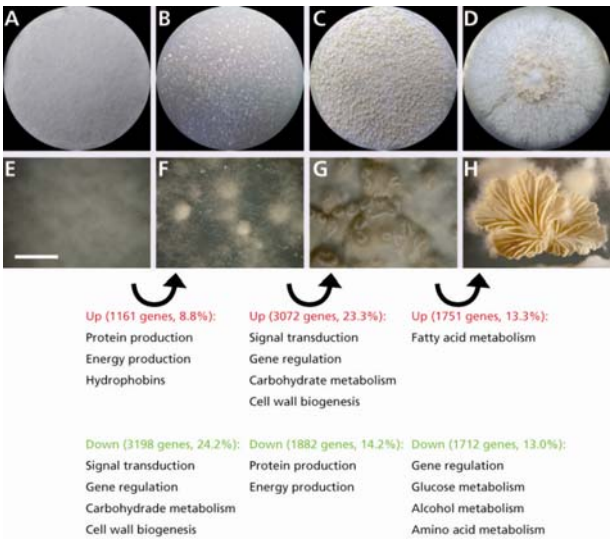
A
B
C
D
E
F
G
H
Up (1161 genes, 8.8%):
Protein production
Energy production
Hydrophobins
Up (3072 genes, 23.3%):
Signal transduction
Gene regulation
Carbohydrate metabolism
Cell wall biogenesis
Up (1751 genes, 13.3%):
Fatty acid metabolism
Down (3198 genes, 24.2%):
Signal transduction
Gene regulation
Carbohydrade metabolism
Cell wall biogenesis
Down (1882 genes, 14.2%):
Protein production
Energy production
Down (1712 genes, 13.0%):
Gene regulation
Glucose metabolism
Alcohol metabolism
Amino acid metabolism

Figure 1.

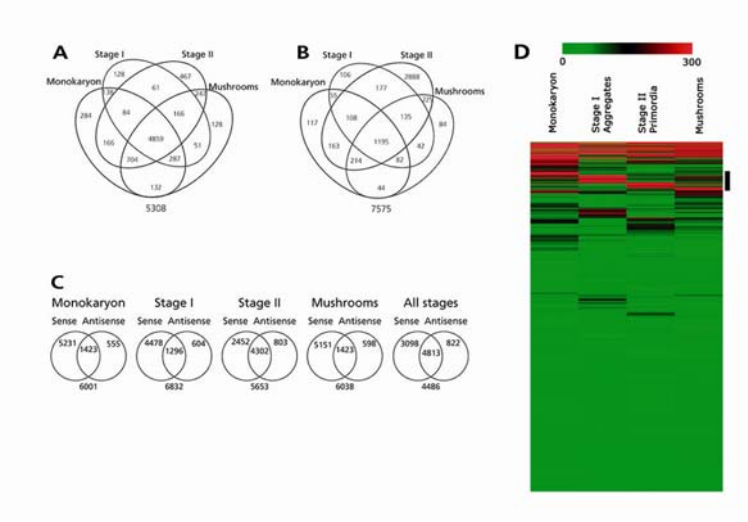
A
Stage I
Stage II
Monokaryon
Mushrooms
5308
B
Stage I
Stage II
Monokaryon
Mushrooms
7575
C
Monokaryon
Stage I
Stage II
Mushrooms
All stages
Sense
Antisense
D
0
300
Monokaryon
Stage I Aggregates
Stage II Primordia
Mushrooms

Figure 2.

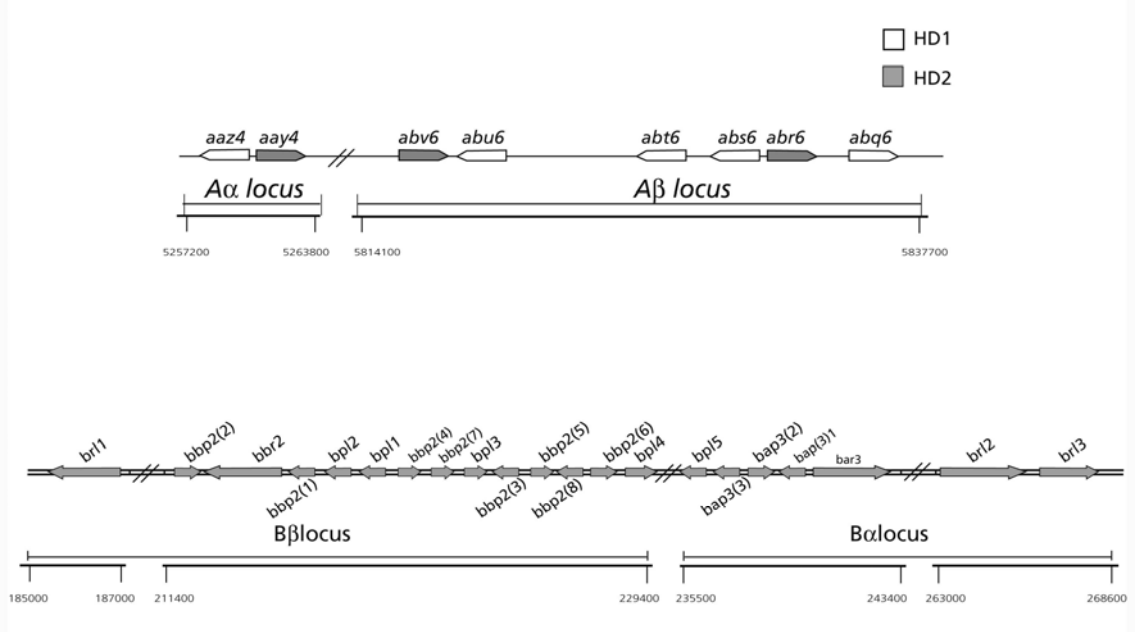
HD1
HD2
aaz4
aay4
abv6
abu6
abt6
abs6
abr6
abq6
Aα locus
Aβ locus
5257200
5263800
5814100
5837700
brl1
bbp2(2)
bbr2
bbp2(1)
bpl2
bpl1
bbp2(4)
bbp2(7)
bpl3
bbp2(3)
bbp2(5)
bbp2(6)
bbp2(8)
bpl4
bpl5
bap3(3)
bap3(2)
bapl(3)1
bar3
brl2
brl3
Bβlocus
Bαlocus
185000
187000
211400
229400
235500
243400
263000
268600

Figure 3.

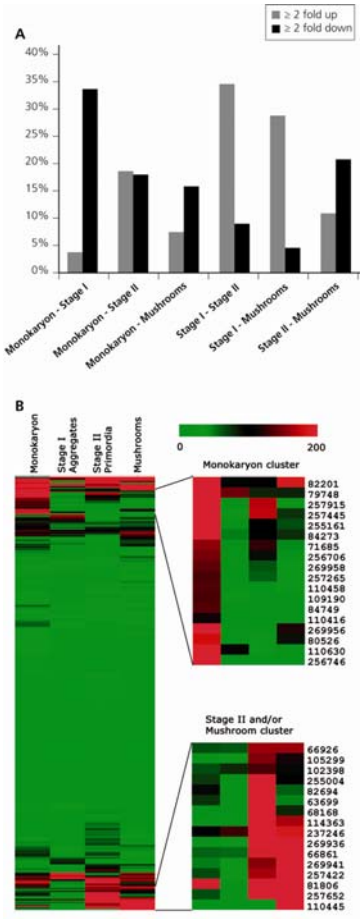
A
≥ 2 fold up
≥ 2 fold down
40%
35%
30%
25%
20%
15%
10%
5%
0%
Monokaryon - Stage I
Monokaryon - Stage II
Monokaryon - Mushrooms
Stage I - Stage II
Stage I - Mushrooms
Stage II - Mushrooms
B
Monokaryon
Stage I Aggregates
Stage II Primordia
Mushrooms
0
200
Monokaryon cluster
82201
79748
257915
257445
255161
84273
71685
256706
269958
257265
110458
109190
84749
110416
269956
80526
110630
256746
Stage II and/or Mushroom cluster
66926
105299
102398
255004
82694
63699
68168
114363
237246
269936
66861
269941
257422
81806
257652
110445

Figure 4.

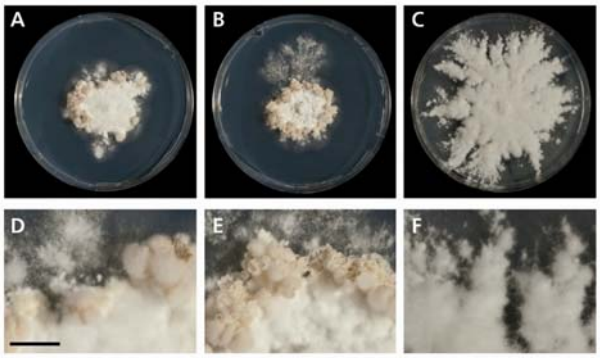

Figure 5.