nature biotechnology

Genome sequence of the model mushroom *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[3,18], 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]

Much remains to be learned about the biology of mushroom-forming fungi, which are an important source of food, secondary metabolites and industrial enzymes. The wood-degrading fungus *Schizophyllum commune* is both a genetically tractable model for studying mushroom development and a likely source of enzymes capable of efficient degradation of lignocellulosic biomass. Comparative analyses of its 38.5-megabase genome, which encodes 13,210 predicted genes, reveal the species's unique wood-degrading machinery. One-third of the 471 genes predicted to encode transcription factors are differentially expressed during sexual development of *S. commune*. Whereas inactivation of one of these, *fst4*, prevented mushroom formation, inactivation of another, *fst3*, resulted in more, albeit smaller, mushrooms than in the wild-type fungus. Antisense transcripts may also have a role in the formation of fruiting bodies. Better insight into the mechanisms underlying mushroom formation should affect commercial production of mushrooms and their industrial use for producing enzymes and pharmaceuticals.

© 2010 Nature America, Inc. All rights reserved.

The importance of mushroom-forming fungi in agriculture, human health and ecology underscores their biotechnological potential for a wide range of applications. The most conspicuous forms of these species, most of which are basidiomycetes, are their fleshy, spore-bearing fruiting bodies. Although these are primarily of economic value because of their use as food[1,2] (worldwide production of edible mushrooms amounts to ~2.5 million tons annually), mushrooms also produce anti-tumor and immunostimulatory molecules[1,2], as well as enzymes used for bioconversions[3]. Moreover, they have been identified as promising cell factories for the production of pharmaceutical proteins[4].

Despite their economic importance, relatively little is known about how mushroom-forming fungi obtain nutrients and how their fruiting bodies are formed. The vast majority of mushroom-forming fungi cannot be genetically modified, or even cultured under laboratory conditions. The basidiomycete *Schizophyllum commune*, which completes its life cycle in ~10 d, is a notable exception insofar as it can be cultured on defined media and there are a wealth of molecular tools to study its growth and development. It is the only mushroom-forming fungus for 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]. In contrast, constructs that have been developed for ascomycetes are often not functional in mushroom-forming basidiomycetes.

S. commune is one of the most commonly found fungi and can be isolated from all continents, except for Antarctica. *S. commune* has been reported to be a pathogen of humans and trees, but it mainly adopts a saprobic lifestyle 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[7]. The mushrooms of *S. commune* that form 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

[1]Department of Microbiology and Kluyver Centre for Genomics of Industrial Fermentation, Utrecht University, Utrecht, The Netherlands. [2]Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [3]Department of Microbiology, Friedrich Schiller University, Jena, Germany. [4]Department of Plant Pathology and Microbiology, University of California, Riverside, California, USA. [5]CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands. [6]INRA, Biotechnologie des Champignons Filamenteux, Marseille, France. [7]Universités Aix-Marseille I & II, Marseille, France. [8]Chemical and Biological Process Development Group, Pacific Northwest National Laboratory, Richland, Washington, USA. [9]Biology Department, Sacred Heart University, Fairfield, Connecticut, USA. [10]Architecture et Fonction des Macromolecules Biologiques, Université Aix-Marseille I & II, Marseille, France. [11]Department of Biological Sciences, Southern Illinois University, Edwardsville, Illinois, USA. [12]Department of Biology, Southeast Missouri State University, Cape Girardeau, Missouri, USA. [13]Division of Molecular Wood Biotechnology and Technical Mycology, Büsgen-Institute, University of Göttingen, Göttingen, Germany. [14]Department of Biochemistry and Food Chemistry, University of Turku, Biocity A, Turku, Finland. [15]Wood Protection & Biotechnology, Empa, Swiss Federal Laboratories for Materials Testing and Research, St. Gallen, Switzerland. [16]Molecular Microbiology, Institute of Biology, Leiden University, Leiden, The Netherlands. [17]Department of Biological Sciences, Union College, Schenectady, New York, USA. [18]Present address: Dartmouth Medical School, Hanover, New Hampshire, USA. Correspondence should be addressed to J.S.H. (hortons@union.edu), I.V.G. (IVGrigoriev@lbl.gov) or H.A.B.W. (h.a.b.wosten@uu.nl).

Received 14 April; accepted 12 May; published online 11 July 2010; doi:10.1038/nbt.1643

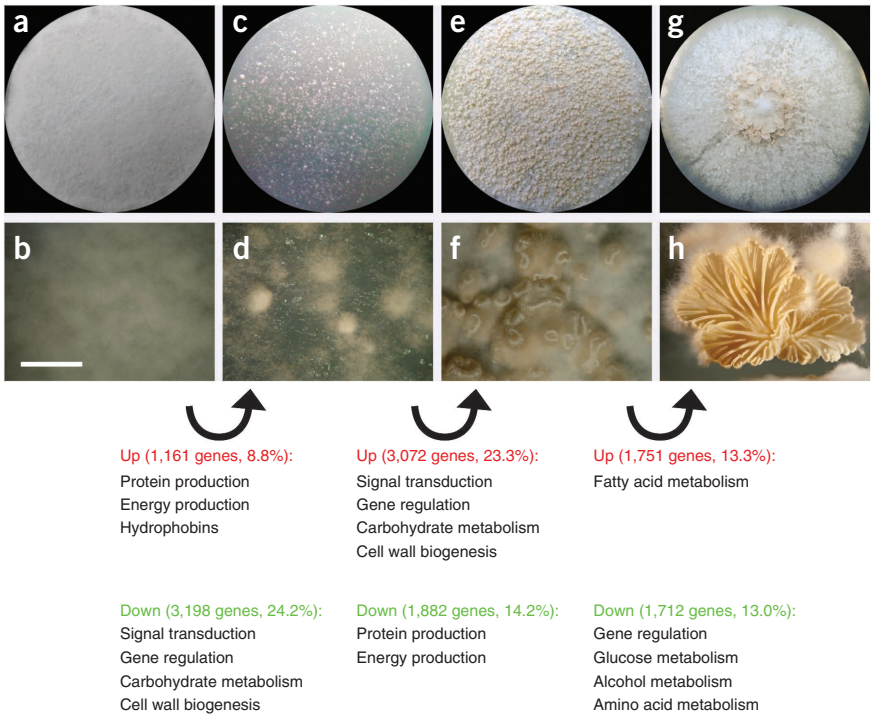

Figure 1 Development of *S. commune*. (**a**–**h**) Four-day-old (**a**–**f**) and 8-day-old (**g**,**h**) colonies grown from homogenates illustrate typical developmental stages in the life cycle of *S. commune*. A monokaryon generates sterile aerial hyphae that form a fluffy white layer on top of the vegetative mycelium (**a**,**b**). Aerial hyphae of a dikaryon interact with each other to form stage I aggregates (**c**,**d**), which, after a light stimulus, develop into stage II primordia (**e**,**f**). These primoridia further differentiate into sporulating mushrooms (**g**,**h**). Enrichment analysis shows that particular functional terms are over-represented in genes that are up- or downregulated during a developmental transition. These terms are indicated below the panels. **a**,**c**,**e**,**g** represent cultures grown in 9-cm Petri dishes, whereas **b**,**d**,**f**,**h** represent magnifications thereof. Scale bar, 1 cm (**h**), 2.5 mm (**b**,**d**) and 5 mm (**f**).

contains one nucleus. Initial growth of this mycelium occurs beneath the surface of the substrate, with formation of aerial hyphae a few days after germination (**Fig. 1a,b**). Monokaryons that encounter each other fuse, and a fertile dikaryon forms when the alleles of the mating-type loci *matA* and *matB* of the partners differ. A short exposure to light is essential for fruiting, whereas 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. 1c**,**d**) form fruiting-body primordia (**Fig. 1e**,**f**), which further develop into mature fruiting bodies (**Fig. 1g**,**h**). Karyogamy and meiosis occur in the basidia within the mature fruiting body, and 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 illustrate the potential of this basidiomycete as a model system to study mushroom formation. Besides the importance of understanding the sexual reproduction of *S. commune* for the commercial production of mushrooms, insight into the basis of this species' capacity to degrade lignocellulose may inspire more effective strategies to degrade lignocellulosic feedstocks for biofuel production.

RESULTS

The genome of *S. commune*

Sequencing of the genomic DNA of *S. commune* strain H4-8 with 8.29× coverage (**Supplementary Table 1**) revealed a 38.5-megabase genome assembly with 11.2% repeat content (**Supplementary Results 1**). The assembly is contained on 36 scaffolds (**Supplementary Table 2**), which represent 14 chromosomes[9]. We predict 13,210 gene models, with 42% supported by expressed sequenced tags (ESTs) and 69% similar to proteins from other organisms (**Supplementary Tables 3** and **4**). 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**) identifies 7,055 groups containing at least one *S. commune* protein (**Supplementary Table 5**). 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**). Notably, a similar percentage of proteins (36%) are unique to *S. commune*, as based on OrthoMCL analysis. Of these proteins, 46% have at least one inparalog (a gene resulting from a duplication within the genome) in *S. commune*. The uniqueness of the *S. commune* proteome is also illustrated by the over- and under-representation of protein family (PFAM) domains compared to other fungi (**Supplementary Results 2**) 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

We used massively parallel signature sequencing (MPSS) to compare whole-genome expression at the four developmental stages, defined by monokaryons, stage I aggregates, stage II primordia and mature fruiting bodies (**Fig. 1**). The majority of genes are either expressed in all four stages (4,859 genes) or not expressed in any of them (5,308 genes) (**Fig. 2** and **Supplementary Table 7**). Of the 13,210 predicted genes, 59.8% are expressed in at least one developmental stage (**Supplementary Table 7**). Fewer of the unique *S. commune* genes meet this criterion, whereas a higher percentage was observed for genes that share orthologs with Agaricomycetes or more distant fungi (**Supplementary Table 6**). This suggests that *S. commune* genes lacking homology to any reported sequences are more stringently regulated than orthologs of genes reported for other species. This is consistent with the observation that genes that are apparently unique to *S. commune* are over-represented in the pool of genes that are differentially expressed during the four developmental stages studied (**Supplementary Tables 8** and **9**).

Antisense transcription is a widespread phenomenon in *S. commune* (**Fig. 2b**,**c**). Of the tags that could be related to a gene model, 18.7% originate from an antisense transcript; and 42.3% of the predicted genes have antisense expression during one or more of the four developmental stages studied (**Supplementary Tables 7** and **10**). Northern hybridization with strand-specific probes confirmed the existence of antisense transcripts of *sc4* (DOE JGI Protein ID 73533; data not shown). Whereas a relatively large number of genes expressed in the antisense direction are uniquely expressed in stage II (2,888 genes), relatively few genes are expressed in the antisense direction in all stages (1,195 genes) (**Fig. 2b**). Our data suggest that 4,302 genes are expressed

© 2010 Nature America, Inc. All rights reserved.

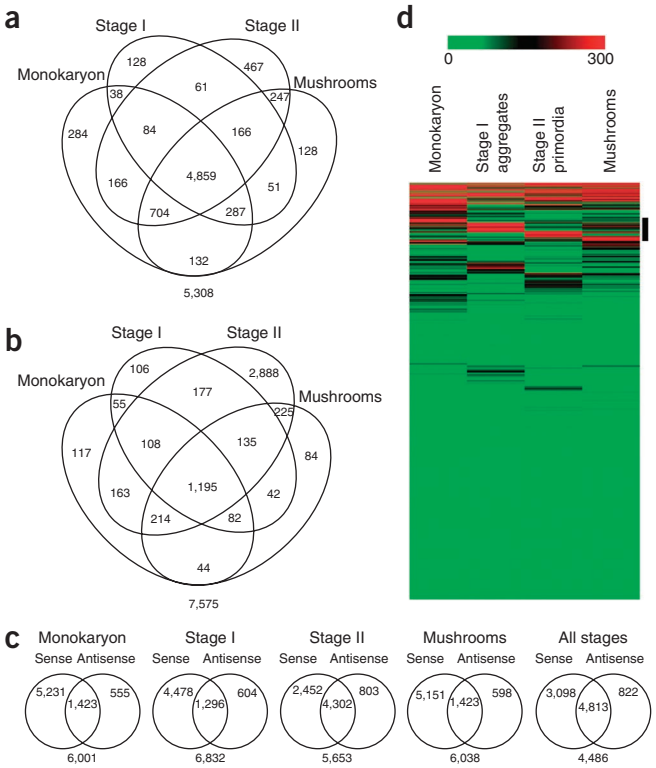

Figure 2 Gene expression in four developmental stages of *S. commune*. (**a**,**b**) The cutoff for expression is 4 tags per million (TPM). Venn diagrams show the overlap of genes expressed in the sense (**a**) and antisense (**b**) directions in the four developmental stages. For example, **a** shows that 61 genes are expressed in the sense direction in stage I and stage II, 4,859 genes are expressed in the sense direction in all stages, 132 genes are expressed in the sense direction in the monokaryon and mature fruiting bodies, and 5,308 genes are not expressed in the sense direction in any of the stages. (**c**) Venn diagrams of the overlap in genes that show sense and antisense expression in each developmental stage, and in all stages combined. (**d**) Heat map of expression of the *S. commune* genes in the four developmental stages. The bar at the top of the panel represents expression values between 0 and 300 TPM. Genes with expression values >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 for the genes in this cluster is given in **Supplementary Table 18**.

in both the sense and antisense directions during stage II (**Fig. 2c**). This overlap is larger for genes expressed during this phase of the life cycle than for the other developmental stages studied.

Fruiting-body development

We performed an enrichment analysis of functional annotation for the expression profiles of the developmental stages defined by monokaryons, stage I aggregates, stage II primordia and mature fruiting bodies. Functional terms involved in protein or energy production, or associated with hydrophobins, are over-represented in genes upregulated during formation of stage I aggregates (**Fig. 1** and **Supplementary Table 9**). Genes involved in signal transduction, regulation of gene expression, cell wall biogenesis and carbohydrate metabolism are enriched in the group of genes downregulated during the formation of stage I aggregates. These functional terms are enriched in the upregulated genes during formation of stage II primordia, whereas terms involved in protein and energy production are enriched in the downregulated genes (**Fig. 1** and **Supplementary Table 9**). Genes encoding transcription factors and genes involved in amino acid, glucose and alcohol metabolism are enriched in the group of genes downregulated during the formation of mature fruiting bodies.

As whole-genome expression was previously analyzed during mushroom formation in *Laccaria bicolor*[10], we next investigated whether the regulation of orthologous gene pairs of *L. bicolor* and *S. commune* might be correlated during fruiting. When we compared microarray expression profiles of free-living mycelium and mature fruiting bodies of *L. bicolor* to the MPSS expression profiles of monokaryotic mycelium and mature fruiting bodies of *S. commune*, we found that 6,751 expressed genes from *S. commune* had at least one expressed ortholog in *L. bicolor*. We determined the correlation of changes in expression of the functional annotation terms to which these orthologous pairs belong. There were 15 gene ontology terms, 2 KEGG terms, 4 KOG terms and 4 PFAM terms that showed a positive correlation in expression ($P < 0.01$; **Supplementary Table 11**). 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 α subunit). This indicates that regulation of these processes during mushroom formation is conserved in *S. commune* and *L. bicolor*.

Analysis of the matA and matB gene loci

Formation of a fertile dikaryon is regulated by the *matA* and *matB* mating-type loci. Proteins encoded in these loci activate signaling cascades (**Supplementary Results 3**) upstream of target genes. The target genes include those encoding enzymes and proteins that fulfill structural functions, such as hydrophobins (**Supplementary Results 4**), needed for the formation of fruiting bodies.

The *matA* locus of *S. commune* strain H4-8 appears to have more homeodomain genes than any fungal mating-type locus described thus far. This locus consists of two subloci, *Aα* and *Aβ*, which are separated by 550 kilobases (kb) on chromosome I of strain H4-8.

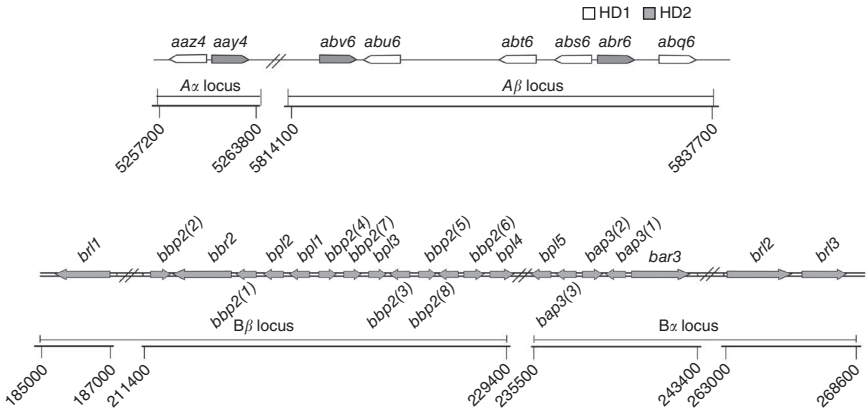

Figure 3 Distribution of genes encoding HD1 and HD2 homeodomain proteins in the *matA* locus and genes encoding pheromone receptors and pheromones in the *matB* locus of *S. commune* strain H4-8. The *matA* and *matB* loci are positioned on scaffolds 1 and 10, respectively. We identified an additional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

© 2010 Nature America, Inc. All rights reserved.

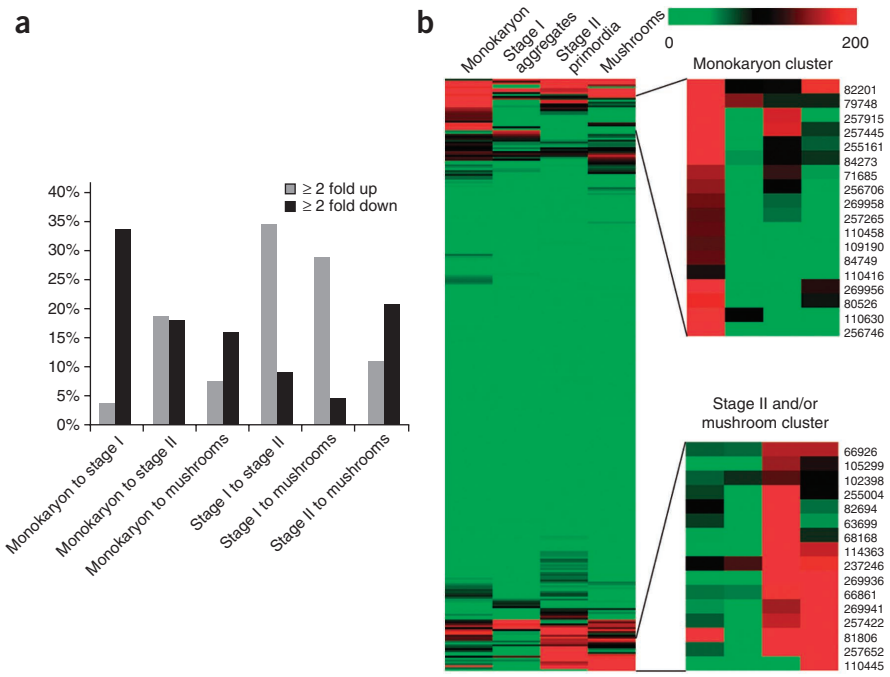

Figure 4 Expression of the 471 transcription factors in the genome of *S. commune*. (**a**) The histogram shows the percentage of transcription factor genes that are differentially expressed between stages of development. (**b**) The heat map shows a cluster containing predominantly monokaryon-specific transcription factors and a cluster containing predominantly stage II- and/or mushroom-specific transcription factors. These clusters are enlarged to the right of the heat map. The latter group contains two fungus-specific transcription factor genes, *fst3* and *fst4*.

Annotation revealed that the *Aα* locus of H4-8 contains two divergently transcribed genes, which encode the Y and Z homeodomain proteins of the HD2 and HD1 classes, respectively (**Fig. 3** and **Supplementary Table 12**). These two genes, *aay4* and *aaz4*, have been described previously[1]. A homeodomain gene has also been identified previously in the *Aβ* locus of H4-8 (ref. 11). Our genomic sequence revealed that this locus actually contains six predicted homeodomain genes: *abq6* (HD1), *abr6* (HD2), *abs6* (HD1), *abt6* (HD1, but lacking the nuclear localization signal), *abu6* (HD1) and *abv6* (HD2) (**Fig. 3** and **Supplementary Table 12**).

Annotation of the genomic sequence of *S. commune* reveals that the *matB* system contains more genes than previously envisioned. The *matB* locus comprises two linked loci, *Bα* and *Bβ*, which both encode pheromones and pheromone receptors[1] (**Fig. 3**). Previously, one pheromone receptor gene was identified in both *Bα3* and *Bβ2* of strain H4-8 (called *bar3* and *bbr2*, respectively)[12]. The genome sequence of *S. commune* reveals four additional genes with high sequence similarity to these pheromone receptor genes, which we call *B* receptor–like genes 1 to 4 (*brl1* to *brl4*; **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 shows that the *brl* genes are expressed (**Supplementary Table 13**). In fact, of all receptor and receptor-like genes, *brl3* shows the highest expression under the conditions tested.

Three and eight pheromone genes have previously been identified at the *Bα3* and *Bβ2* loci, respectively[13]. We identified one additional pheromone gene, named *B* pheromone–like-5 (*bpl5*), at the *Bα3* locus. Moreover, four additional pheromone-like genes were detected at the *Bβ2* locus, called *bpl1* to *bpl4* (**Fig. 3**). Of these, only *bpl2* showed no expression in MPSS analysis (**Supplementary Table 13**). The *Bα* gene *bpl5* and three of the new *Bβ* pheromone-like genes show deviations from the consensus farnesylation signal, CAAX (where C is cysteine, A is aliphatic and X is any residue), with the variant motifs CASR, CTIA, CRLT and CQLT for Bpl5, Bpl1, Bpl2 and Bpl3, respectively. Previously, one of the pheromone genes (*bbp2(6)*) was shown to function with the deviant 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

The genome of *S. commune* reveals genes encoding 471 putative transcription factors, of which 311 are expressed during at least one developmental stage (**Supplementary Table 14**). Of these genes, 56% are expressed in all developmental stages; 268 were expressed in the monokaryon, 200 during formation of stage I aggregates, 283 during formation of stage II aggregates and 253 during formation of mushrooms. We identified a cluster of monokaryon-specific transcription factors and a group of transcription factors upregulated in stage II primordia or in mature mushrooms, or both (**Fig. 4**). The latter group includes *fst3* (NCBI Protein ID: 257422) and *fst4* (NCBI Protein ID: 66861),

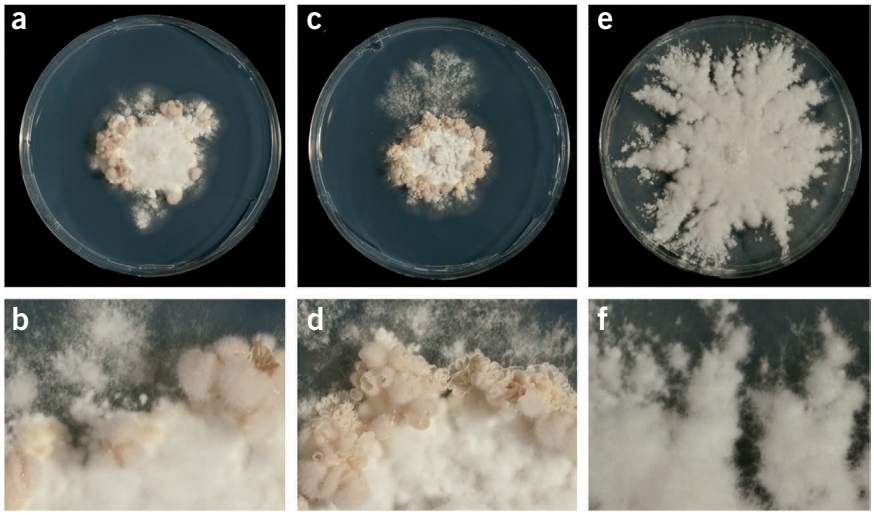

Figure 5 Transcription factors affecting fruiting body formation. (**a**,**b**) Wild-type dikaryon fruiting-body formation. (**c**–**f**) Fruiting-body formation in dikaryons in which *fst3* (**c**,**d**) or *fst4* (**e**,**f**) has been inactivated. Lower panels (**b**,**d**,**f**) show a magnification of part of the colonies shown in upper panels (**a**,**c**,**e**). Scale bar, 5 mm (**b**,**d**,**f**).

© 2010 Nature America, Inc. All rights reserved.

Table 1 Comparison of the number of FOLymes and CAZymes of *S. commune* with those of other fungi

	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

LO1, 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; LDA8, alcohol oxidases; GH, glycoside hydrolases; GT, glycosyl transferases; PL, polysaccharide lyases; CE, carbohydrate esterases.

which encode transcription factors that contain a fungus-specific Zn(II)2Cys6 zinc-finger DNA binding domain.

We inactivated the *fst3* and *fst4* genes via targeted gene deletions. The Δ*fst3* and Δ*fst4* monokaryons showed no phenotypic differences from the wild-type monokaryons. In contrast, the Δ*fst4* Δ*fst4* dikaryon did not fruit, but produced more aerial hyphae when compared to the wild type (**Fig. 5**). This suggests that Fst4 is crucial in the switch between the vegetative and reproductive phases of the *S. commune* life cycle. In contrast, the Δ*fst3* Δ*fst3* dikaryon formed more, albeit smaller, reproductive structures than those of the wild type (**Fig. 5**). As spatial and temporal regulation of fruiting-body formation and sporulation were not altered in the Δ*fst3* Δ*fst3* strain, we conclude that Fst3 inhibits the formation of clusters of mushrooms.

Wood degradation by *Schizophyllum commune*

As a white-rot fungus[6], *S. commune* degrades all woody cell wall components; in contrast, brown-rotters efficiently degrade cellulose but only modify lignin, leaving a polymeric residue. Lignin-degrading enzymes, which are commonly classified as FOLymes[14], comprise lignin oxidases (LO families) and lignin-degrading auxiliary enzymes that generate H_2O_2 for peroxidases (LDA families). The LO family consists of laccases (LO1), lignin peroxidases, manganese peroxidases, versatile peroxidases (LO2) and cellobiose dehydrogenases (CDHs; LO3). *S. commune* contains 16 FOLyme genes and 11 genes that encode enzymes distantly related to FOLyme enzymes (**Table 1** and **Supplementary Table 15**). The genome lacks genes encoding peroxidases of the LO2 family. However, it contains a CDH gene (LO3), two laccase genes (LO1) and 13 LDA genes, including four genes encoding glucose oxidases (LDA6) and benzoquinone reductases (LDA7) (**Table 1**).

S. commune appears to possess a more diverse assortment of FOLymes than the brown-rot fungus *Postia placenta* and the fungi that are known not to have ligninolytic activity (that is, *Ustilago maydis*, *Cryptococcus neoformans*, *Aspergillus nidulans*, *Neurospora crassa* and *Saccharomyces cerevisiae*; **Table 1**). In contrast, it has fewer FOLymes than either the coprophilic fungus *Coprinopsis cinerea* and the white-rot fungus *Phanerochaete chrysosporium*, which are predicted to possess 40 and 27 members, respectively[14].

Regarding polysaccharide degradation, *S. commune* has the most extensive machinery for degrading cellulose and hemicellulose of all of the basidiomycetes we examined. The Carbohydrate-Active Enzyme database (CAZy) identified 240 candidate glycoside hydrolases, 75 candidate glycosyl transferases, 16 candidate polysaccharide lyases and 30 candidate carbohydrate esterases encoded in the genome of *S. commune* (**Table 1** and **Supplementary Table 16**). Compared to the genomes of other basidiomycetes, *S. commune* has the highest number of glycoside hydrolases and polysaccharide lyases. *S. commune* is rich in genes encoding enzymes that degrade pectin, hemicellulose and cellulose (**Supplementary Table 17**). 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**). The pectinolytic capacity of *S. commune* is further complemented by the presence of pectin hydrolases from families 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 two class-level taxa (Wallemiomycetes and Entorrhizomycetes) and the subphyla Pucciniomycotina (rust), Ustilaginomycotina (smuts) and Agaricomycotina[16]. The Agaricomyotina include the mushroom- and puffball-forming fungi, crust fungi and jelly fungi. Genomic sequences are currently available for five members of the Agaricomycotina: *P. chrysosporium*[17], *L. bicolor*[10], *P. placenta*[18], *C. neoformans*[19] and *C. cinerea*[20]. Our 38.5-megabase assembly of the *S. commune* genome represents the first genomic sequence for a member of the family Schizophyllaceae. Thirty-six percent of the encoded proteins have no ortholog in other fungi. Only 43% of the predicted genes could be annotated with a gene ontology term, underscoring that much about the proteome of *S. commune* remains unknown. This percentage resembles that seen in other basidiomycetes: 30% in *L. bicolor*[10], 48% in *P. placenta*[18] and 49% in *P. chrysosporium*[17].

S. commune invades wood primarily by growing through the lumen of vessels, tracheids, fibers and xylem rays. Adjacent parenchymatic cells in the xylem tissue are invaded via simple and bordered pits. As a consequence of this approach to invasion, cellulose, hemicellulose or pectin can serve as the primary carbon source for *S. commune*. Indeed, the genome of *S. commune* probably encodes at least one gene in each family involved in the degradation of cellulose, hemicellulose and pectin. The large number of predicted pectinase genes is consistent with earlier studies describing *S. commune* as one of the best pectinase producers among the basidiomycetes[21]. *S. commune* also encodes carbohydrate-active enzymes that degrade other polymeric sugars, such as those acting on starch, mannans and inulins. Consistent with the wide variety of substrates that support its growth, *S. commune* has the most complete polysaccharide breakdown machinery of all basidiomycetes examined.

© 2010 Nature America, Inc. All rights reserved.

We know much less about how fungi degrade lignin than how they digest plant polysaccharides. Fungi are assumed to use FOLymes to degrade lignin[14]. Although members of the LO2 family of lignin oxidases are known to degrade lignin, it remains controversial whether laccases (LO1) and cellobiose dehydrogenases (CDHs; LO3) share this capacity. *S. commune* contains 16 genes encoding FOLymes. There are no members of the LO2 family, but the genome contains one CDH gene and two laccase genes. CDHs may participate in the degradation of cellulose, xylan and, possibly, lignin by generating hydroxyl radicals in a Fenton-type reaction. Laccases catalyze 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], but it is not clear whether the two *S. commune* gene products 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 four genes of the multicopper oxidase superfamily. In contrast, *P. placenta* contains two laccase-encoding genes (LO1) but lacks members of the LO2 and LO3 families. As *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 LDAs than *P. chrysosporium*. For instance, *S. commune* contains four glucose oxidase (LDA6) genes, whereas fungi seldom express more than one of these. 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 solubilizes inorganic phosphate and thus aids in the uptake of the nutrient[23].

The *matA* and *matB* mating-type loci of *S. commune* regulate the formation of a fertile dikaryon after the fusion of monokaryons that encounter one other. The genome sequence of this species now reveals that the mating type loci of *S. commune* contain the highest number of reported genes within such loci in the fungal kingdom. The *matB* locus comprises two linked loci, *Bα* and *Bβ*, which both 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 reported that the *Bα3* and *Bβ2* loci of H4-8 contain three and eight pheromone genes, respectively, and each contain one pheromone receptor gene[12,13]. We identified five additional pheromone genes and four additional pheromone receptor–like genes 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.J.F., unpublished data). This raises the question of whether the four receptor genes function in *matB*-regulated development. Expression of these genes, as discerned using MPSS, suggests that they do not represent pseudogenes.

The *matA* locus consists of two subloci, Aα and *Aβ*, of which 9 and 32 allelic specificities, respectively, are expected to occur in nature[1]. These loci are separated by 550 kb on chromosome I of strain H4-8. Such a large distance has not been found in other fungi that have a tetrapolar mating system. The functionally well-characterized *Aα* locus showed no substantial differences from the published descriptions[1]. It is composed of two genes encoding Y and Z homeodomain proteins of the HD2 and HD1 classes, 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, followed by the translocation of the active protein complex into the nucleus[1].

The *Aβ* locus of *S. commune* has been studied much less than the *Aα* locus. Notably, *Aβ* reflects the highest degree of homeodomain-gene complexity for any fungal mating-type locus described to date. It contains four homeodomain genes of the HD1 class and two of the HD2 class. The *Aβ* 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 *matAβ* would explain why recombination analyses predict as many as 32 mating specificities for this locus[26]. Overall, *S. commune* seems ideal for identifying the evolutionary pathways that have created high numbers of allelic specificities for enhancing outbreeding versus inbreeding rates.

As little is known about molecular processes that control formation of fruiting bodies in basidiomycetes, other than the role of the mating-type loci[8], we compared genome-wide expression profiles at four developmental stages. 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). Notably, 467 genes were specifically expressed in stage II primordia. This suggests that this stage represents a major developmental switch, an idea supported by the fact that genes involved in signal transduction and regulation of gene expression are enriched in the group of upregulated genes during formation of stage II primordia. A positive correlation of expression of these gene groups during mushroom formation in both *S. commune* and *L. bicolor* suggests that regulation of mushroom formation is a conserved process in the Agaricales.

Our analysis of gene expression in *S. commune* reveals a high frequency of antisense expression. About 20% of all sequenced mRNA tags originated from an antisense transcript, and >5,600 of the predicted genes showed antisense expression in one or more developmental stages. Antisense transcription was most pronounced in stage II primordia. At this stage, >4,300 genes were expressed in both the sense and antisense directions, and >800 genes were expressed in the antisense direction only. Previously, MPSS has revealed antisense 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 antisense transcript derived from a locus encoding a component of the circadian clock[28], possibly through RNA interference. It is tempting to speculate that antisense transcripts also regulate mRNA levels in *S. commune*. Natural antisense transcripts in eukaryotes have also been implicated in other processes, such as translational regulation, alternative splicing and RNA editing[29]. The antisense transcripts of *S. commune* may likewise have such functions. In all these cases, the antisense transcripts could function in a developmental switch that occurs when stage II primordia are formed.

The apparently high conservation of gene regulation in the Agaricales led us to study the 471 genes predicted to encode transcriptional regulators. Of these, 268 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 expressed during formation of stage II primordia again points to a major switch that probably occurs during this developmental stage.

We identified a group of monokaryon-specific transcription factors and a group of transcription factors that are upregulated in stage II primordia or mature mushrooms, or in both. The *fst3* and *fst4* genes encode transcriptional regulators belonging to the latter group. Growth and development were not affected in monokaryotic strains

© 2010 Nature America, Inc. All rights reserved.

in which *fst3* or *fst4* were inactivated. Phenotypic differences were, however, observed in the dikaryon. The Δ*fst4* Δ*fst4* dikaryon did not fruit but produced more aerial hyphae than the wild type. In contrast, the Δ*fst3* Δ*fst3* dikaryon formed more, albeit smaller, fruiting bodies than the wild type. This suggests 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 that sufficient energy is available for full development of fruiting bodies. As *fst3* and *fst4* have homologs in other mushroom-forming fungi, it is tempting to speculate that they have similar functions in these organisms. This is supported by the observation that the homologs of *fst3* and *fst4* are upregulated 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 expressed in the free-living mycelium.

In conclusion, 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 fruiting bodies. The large variety of genes that encode extracellular enzymes that act on polysaccharides probably explains why *S. commune* is so common in nature. Moreover, the genome sequence suggests that *S. commune* may have a unique mechanism to degrade lignin. Our MPSS data has provided leads on how mushroom formation is regulated, highlighting both the roles of certain transcription factors and the possible involvement of antisense transcription. Better understanding of the physiology and sexual reproduction of *S. commune* will probably have an impact on the commercial production of edible mushrooms and the use of mushrooms as cell factories.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

Data availability and accession codes. *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 number ADMJ00000000. MPSS data have been deposited in NCBI's Gene Expression Omnibus with accession number GSE21265.

Note: Supplementary information is available on the Nature Biotechnology website.

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, and Los Alamos National Laboratory under contract no. DE-AC02-06NA25396. The work was also supported by the Dutch Technology Foundation STW, the Applied Science division of the Netherlands Organization for Scientific Research and the Technology Program of the Dutch Ministry of Economic Affairs.

AUTHOR CONTRIBUTIONS
J.S.H., H.A.B.W., T.J.F., W.W.L., A.C.G., E.K. and S.E.B. wrote the proposal; L.G.L., J.F.d.J., E.L. and R.A.O. isolated RNA and DNA and made libraries; E.L. and S.L. coordinated sequencing of the genome; J.S. assembled the genome; I.V.G., R.A.O., K.A.B., J.E.S., S.E.B., E.K. and H.A.B.W. coordinated the annotation process; I.V.G., A.A., A.S., J.E.S., T.J.F., E.R., A.L., F.P., U.K., J.K.M., J.F.d.J., R.P.d.V., P.M.C., V.L., B.H., W.W.L., A.C.G., P.A.v.K., K.A.B., J.S.H., E.K., S.E., N.K., R.A.O. and M.R. annotated genes; R.A.O., J.F.d.J., F.W.M.R.S. and L.G.L. performed experiments; R.A.O., J.F.d.J., F.W.M.R.S., L.G.L. and H.A.B.W. interpreted and designed experiments; R.A.O., A.A., J.E.S., E.R., J.F.d.J., R.P.d.V., F.W.M.R.S., K.A.B., E.K., J.S.H., I.V.G. and H.A.B.W. wrote the paper; R.A.O. and H.A.B.W. coordinated writing of the paper.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.
This paper is distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike license, and is freely available to all readers at http://www.nature.com/naturebiotechnology/.

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, Univ. 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, 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 *Bβ* 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, UK, 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. USA* **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. Sci. USA* **107**, 11889–11894 (2010).
21. Xavier-Santos, S. *et al.* Screening for pectinolytic activity of wood-rotting basidiomycetes and characterization of the enzymes. *Folia Microbiol. (Praha)* **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**, 303–311 (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**, 6228–6233 (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* (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. *et al.* In search of antisense. *Trends Biochem. Sci.* **29**, 88–94 (2004).

© 2010 Nature America, Inc. All rights reserved.

ONLINE METHODS

Strains and culture conditions. *S. commune* was routinely grown at 25 °C on minimal medium (MM) with 1% (wt/vol) glucose and with or without 1.5% (wt/vol) agar[30]. Liquid cultures were shaken at 225 r.p.m. Glucose was replaced with 4% (wt/vol) glycerol for cultures used in the isolation of genomic DNA. All *S. commune* strains used were isogenic to strain 1-40 (ref. 31). Strain H4-8 (*matA43 matB41*; FGSC no. 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 (*matA41 matB41*; CBS 341.81) and 4-40 (*matA43 matB43*; CBS 340.81) were used for MPSS. 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 using a whole-genome shotgun strategy. All data were generated by paired-end sequencing of cloned inserts with six 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% (wt/vol) glucose using mycelial plugs as an inoculum. Strain H4-8 was grown for 4 d in the light, whereas the dikaryon H4-8 × H4-8.3 was grown for 4 d in the dark and 8 d in the light. Mycelia of the dikaryotic stages were combined and RNA was isolated as described[30]. The poly(A)$^+$ RNA fraction was obtained using the Absolutely mRNA Purification kit and manufacturer's instructions (Stratagene). cDNA synthesis and cloning followed the SuperScript plasmid system procedure with Gateway technology for cDNA synthesis and cloning (Invitrogen). For the monokaryon, two size ranges of cDNA were cut out of the gel to generate two cDNA libraries (JGI library codes CBXY for the range 0.6 kb–2 kb and CBXX for the range >2 kb). For the dikaryon, cDNA was used in the range >2 kb, 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). Subclone inserts were sequenced from both ends using Big Dye terminator chemistry and ABI 3730 instruments (Applied Biosystems).

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 genes, or used to extend predicted gene models into full-length genes by adding 5′ and/or 3′ UTRs. Because multiple gene models were generated for each locus, a single representative model at each locus was computationally selected on the basis of EST support and similarity to protein sequences in the NCBI nonredundant database. This resulted in a final set of 13,210 predicted genes, of which 1,314 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 the NCBI nonredundant set and classified according to Gene Ontology[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/).

Orthologs of *S. commune* proteins in the fungal kingdom. Proteins of *S. commune* were assigned to orthologous groups with OrthoMCL version 2.0 (ref. 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 version 2.2.20, with an *E* value of 10^{-5} as a cutoff. Custom scripts were used to further analyze 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**).

Representation analysis. FuncAssociate 2.0 (ref. 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 cutoff.

Protein families. The PFAM database version 24.0 (ref. 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 domains were determined using 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 five members in at least one of the fungi were analyzed. When *S. commune* was compared to the rest of the Dikarya, only groups with a minimum of five members in at least four of the fungi were analyzed. In all cases, a *P* value of 0.05 was used as a cutoff. Similar results were obtained using the nonparametric Mann-Whitney *U*-test.

CAZy annotation. Annotation of carbohydrate-related enzymes was performed using the CAZy annotation pipeline[46]. Ambiguous family attributions were processed manually along with all identified models that presented defects (such as deletions, insertions or splicing problems). 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 analyzed manually 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 characterized 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 °C in the dark was homogenized in 200 ml MM using a Waring blender for 1 min at low speed. Two milliliters of the homogenized mycelium was spread out over a polycarbonate membrane placed on top of solidified MM. Vegetative monokaryotic mycelium was grown for 4 d in the light. The dikaryon was grown for 2 and 4 d in the light to isolate mycelium with stage I aggregates and stage II primordia, respectively. Mature mushrooms 3 d old were picked from dikaryotic cultures that had grown for 8 d 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). Between 4.2 and 7.6 million tags of 20 bp were obtained for each of the stages. 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%

© 2010 Nature America, Inc. All rights reserved.

doi:10.1038/nbt.1643

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 untranslated region is incomplete or that the DpnII restriction site that defines the 5′ end of the tag is too close to the poly(A) 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 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[49]. For *L. bicolor*, the microarray values from replicates were averaged. Expression values of genes were increased by 1, and the ratio between monokaryon and mushrooms (for *S. commune*), and between free-living mycelium and mature fruiting bodies (for *L. bicolor*), was log-transformed. All expressed genes from *S. commune* that had at least one expressed ortholog in *L. bicolor* were taken into account, resulting in a total of 6,751 orthologous pairs. These pairs were classified on the basis of functional-annotation terms. Correlation of changes in expression of these gene classes was expressed as the Pearson correlation coefficient. Only gene ontology terms with 10–200 pairs were used in the analysis. In the case of PFAM domains, a minimum of ten ortholog pairs were used.

Deletion of transcription factors *fst3* and *fst4*. The transcription factor genes *fst3* (NCBI Protein ID: 257422) and *fst4* (NCBI Protein ID: 66861) were deleted using the 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.

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*. *Am. Nat.* **92**, 221–232 (1958).
32. Ohm, R.A. *et al.* An efficient gene deletion procedure for the mushroom-forming basidiomycete *Schizophyllum commune*. *World J. Microbiol. Biotechnol.* 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. *Nucleic Acids Res.* **32**, D277–D280 (2004).
39. Finn, R.D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222 (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**, 546–567 (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. *Nucleic Acids Res.* **37**, D233–D238 (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. McCarroll, S.A. *et al.* Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nat. Genet.* **36**, 197–204 (2004).

© 2010 Nature America, Inc. All rights reserved.