Supporting Information
Appendix

I. Genome assembly and annotation

IA. *C. subvermispora* strain

A pure whole genome shotgun approach was used to sequence *C. subvermispora* 105752 monokaryotic strain B (1) (USDA, Forest Mycology Center, Madison, WI). Aspen (*Populus*) wood block decay studies were conducted using a 1:1:1 mixture of top soil, vermiculite and peat was placed in glass Petri plates, hydrated and autoclaved at 120 °C for 60 minutes. After the soil mixture had cooled, aspen (*Populus*) wood wafers that had been hydrated and autoclave sterilized were placed on the soil and inoculated with *C. subvermispora* parental strain 105752 or the monokaryon strain. Additional aspen wood wafers were hydrated to 80 - 100%, autoclaved for 60 minutes at 120 C and placed on feeder strips 7 days after initial inoculation. Some wafers were removed 30, 60 or 90 days later, weighed and percent weight loss was determined:

	% weight loss of colonized wood	
	105752 (parental)	105752 (RP-B)
30 days	10.24	6.12
60 days	41.73	40.7
90 days	56.25	66.63

Additional wafers were removed at the same time period, immediately frozen to -20 °C and prepared for scanning electron microscopy as previously described (2). These results (Fig. S1), together with previous enzyme activity measurments (1), showed the monokaryon comparable to the parent dikaryon strain 105752.

IB. Genome assembly

The *C. subvermispora* B genome was sequenced using a combination of (Titanium unpaired, 3 kb, 6 kb, 9 kb, 23 kb Titanium paired end), and Sanger (39kb Fosmids) sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website (http://www.jgi.doe.gov/). A list of data to be excluded from the draft assembly was created by identifying possible contaminant data in preliminary Newbler assemblies of 454 and Sanger data. The resulting screened data were assembled with the Newbler assembler, 2.4pre (091202_1007) with the following options; -fe removed_dir/reads_removed.FQC -consed -finish -nrm -info -rip -e 45 -a 50 -l 350 -g -ml 20 -mi 98, to a final estimated assembled coverage of 57x with 740 scaffolds with an N/L50 of 8/1.8 Mb, and 1644 contigs with an L/N50 of 77/200 Kb. One round of automated gap closure using the JGI in house gapResolution tool resulted in a final assembly with 1161 contigs with an N/L50 of 14/239 Kb. Newbler-assembled consensus EST sequence data was used to quickly assess the completeness of the final assembly FASTA sequence with alignment using 90% identity and 85% coverage thresholds. This resulted in 99.88% placement. Statistics on the genome assembly are shown in below.

Genome assembly coverage statistics

Library	Library Type	Raw Reads	Raw Bases	Trimmed Bases	Assem Reads	Assem Bases	Coverage	Insert	Std Dev
GCXP	454	3,489,121	1,351,201,484	1,348,788,042	3,428,012	1,323,486,621	34.96x		
GOOX	454PE	1,452,560	324,050,098	291,500,305	1,366,912	284,076,891	7.50x	2953	738
GOOY	454PE	1,121,105	248,556,948	224,830,585	1,042,448	217,684,418	5.75x	2954	738
GHOB	454PE	863,104	184,016,503	166,773,536	825,717	161,271,611	4.26x	6348	1586
GCNW	454PE	278,137	107,264,405	91,945,328	266,977	89,751,833	2.37x	22827	5706
GGOH	454PE	107,818	32,759,134	26,631,934	102,887	25,779,722	0.68x	8659	2165
GBSF	SANG	53,408	46,054,288	40,282,491	53,031	39,949,246	1.06x	38627	9656
Total		7,365,253	2,293,902,860	2,190,752,221	7,085,984	2,142,000,342	56.57x		

Genome assembly statistics

# scaffolds	740
# contigs	1154
Scaffold sequence length	39.0 Mb
Contig sequence length	37.9 Mb (-> 2.8% gap)
Scaffold N/L50	8/1.7 Mb
Contig N/L50	45/238.9 kb
# scaffolds > 50 kb	42
% main genome in scaffolds > 50 kb	96.9%
Repetitive sequences	0.48 Mb (1.2%)

IC. Genome Annotation

The *C. subvermispora* genome was annotated using the JGI annotation pipeline, which takes multiple inputs (scaffolds, ESTs, and known genes) and runs several analytical tools for gene prediction and annotation, and deposits the results in the JGI Genome Portal (http://www.jgi.doe.gov/Ceriporiopsis) for further analysis and manual curation.

Genomic assembly scaffolds were masked using RepeatMasker and the RepBase library of 234 fungal repeats (3). tRNAs were predicted using tRNAscan-SE (4). Using the repeat-masked assembly, several gene prediction programs falling into three general categories were used: 1) *ab initio* - FGENESH (5); GeneMark (6), 2) *homology-based* - FGENESH+; Genewise (7) seeded by BLASTx alignments against GenBank's database of non-redundant proteins (NR: http://www.ncbi.nlm.nih.gov/BLAST/), and 3) *EST-based* - EST_map (http://www.softberry.com/), seeded by EST contigs. Genewise models were extended where possible using scaffold data to find start and stop codons. EST BLAT alignments (8) of two million ESTs sequenced with Roche 454 technology and 15,000 EST clusters assembled with Newbler were used to extend, verify, and complete the predicted gene models. The resulting set of gene models was then filtered for the best models, based on EST and homology support, to produce a non-redundant representative set of 12125 gene models. This representative set was subject to further analysis and manual curation. Measures of model quality include proportions of the models complete with start and stop codons (91% of models), consistent with ESTs (78% of models covered over ≥75% of exon length), supported by similarity with proteins from the NCBI NR database (76% of models). General statistics on the filtered set of gene models follow.

Gene model statistics

	Average
Gene length (bp)	2006
Transcript length (bp)	1605
Protein length (aa)	426
Exons per gene	6.67
Exon length (bp)	241
Intron length (bp)	72.8

All predicted gene models were functionally annotated using SignalP (9), TMHMM (10) InterProScan (11), BLASTp (12) against NR, and hardware-accelerated double-affine Smith-Waterman alignments (deCypherSW; http://www.timelogic.com/decypher_sw.html) against SwissProt (http://www.expasy.org/sprot/), KEGG (13) and KOG (14). KEGG hits were used to assign EC numbers (http://www.expasy.org/enzyme/), and Interpro and SwissProt hits were used to map GO terms (http://www.geneontology.org/). Multigene families were predicted with the Markov clustering algorithm (MCL (15)) to cluster the proteins using BLASTp alignment scores between proteins as a similarity metric. Functional annotations are summarized below. Manual curation of the automated annotations was performed by using the web-based interactive editing tools of the JGI Genome Portal to assess predicted gene structures, assign gene functions, and report supporting evidence.

Functional annotation of proteins

	Ceriporiopsis subvermispora **annotation**
Protein models	12125
Proteins assigned to a KOG	6353
KOG categories genome-wide	2970
Proteins assigned a GO term	5659
GO terms genome-wide	2170
Proteins assigned an EC number	2372
EC numbers genome-wide	646
Proteins assigned a Pfam domain	5965
Pfam domains genome wide	2196

Summary of of Pfam domains in *C. subvermispora*

Protein families with copy-number expansions relative to other published Agaricomycetes included membrane proteins, oxidoreductases, secondary metabolism, membrane transporters, and hydrolases. The expansion of the *peroxidase* Pfam, which contains the ligninolytic peroxidases, is particularly notable with some 17 copies in *C. subvermispora* and *Phanerochaete chyrsosporium* compared to one or two copies in *Postia placenta*, *Serpula lacrymans, Schizophyllum commune,* and *Laccaria bicolor*.

Twenty-five most abundant Pfam domains in *C.subvermispora*. *Cs = C. subvermispora*; *Lb = L. bicolor*; *Pc = P. chrysosporium*; *Pp = P. placenta*; *Sc = S. commune*; *Sl = S. lacrymans*.

Pfam domain	Description	*Cs*	*Lb*	*Pc*	*Pp*	*Sc*	*Sl*
WD40	WD domain, G-beta repeat	863	565	287	148	342	335
p450	Cytochrome P450	192	69	113	65	107	137
MFS_1	Major Facilitator Superfamily	149	100	141	30	181	124
Pkinase	Protein kinase domain	145	120	106	56	116	95
Mito_carr	Mitochondrial carrier protein	94	95	90	18	90	90
ABC_tran	ABC transporter	81	65	78	21	74	60
Helicase_C	Helicase conserved C-terminal domain	79	92	57	28	81	70
adh_short	short chain dehydrogenase	78	39	70	50	105	74
RRM_1	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	77	87	70	17	92	64
Aldo_ket_red	Aldo/keto reductase family	57	20	52	32	55	43
DEAD	DEAD/DEAH box helicase	54	57	34	14	65	43
Pkinase_Tyr	Protein tyrosine kinase	53	35	19	9	22	22
ADH_zinc_N	Zinc-binding dehydrogenase	52	21	52	33	44	37
ABC_membrane	ABC transporter transmembrane region	52	29	30	5	39	29
ADH_N	Alcohol dehydrogenase GroES-like domain	49	23	45	22	35	34
AAA	ATPase family associated with various cellular activities (AAA)	45	39	46	17	47	44
Zn_clus	Fungal Zn(2)-Cys(6) binuclear cluster domain	43	41	24	10	56	15
Asp	Eukaryotic aspartyl protease	41	23	52	18	27	22
Sugar_tr	Sugar (and other) transporter	37	37	39	9	53	39
zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	36	28	13	7	37	28
Methyltransf_11	Methyltransferase domain	33	27	28	12	39	34
COesterase	Carboxylesterase	33	14	9	18	23	25
zf-C2H2	Zinc finger, C2H2 type	32	43	36	42	60	25
Methyltransf_12	Methyltransferase domain	31	27	28	13	34	26
SH3_1	SH3 domain	30	41	26	12	28	15

The inferred protein family expansion in *C. subvermispora* is tabulated below. Counts of proteins with a given Pfam domain were found to be significantly expanded in *C. subvermispora* if a *z*-test against the counts in the other organisms gave a *p*-value <0.5. Expanded domains include GO terms for membrane (Bac_rhodopsin, Cation_ATPase_C, Cation_ATPase_N, DUF6,E1-E2_ATPase, EHN, Ion_trans, MIP, MS_channel, Syntaxin, Voltage_CLC), oxidoreductase activity (2OG-FeII_Oxy, ADH_N, ADH_zinc_N, Aldedh,

Aldo_ket_red, FAD_binding_1, NAD_binding_1, NIR_SIR, NIR_SIR_ferr, Tyrosinase), metabolic process (Abhydrolase_3, Acetyltransf_1, Aldedh, CoA_trans, CoA_transf_3, DHDPS, PLDc, Tyrosinase, UDPGP), transmembrane transport (ABC_membrane, C4dic_mal_tran, Ion_trans, MS_channel, Mem_trans, Sulfate_transp, Voltage_CLC), hydrolase activity (Abhydrolase_2, Abhydrolase_3, Amidohydro_1, DLH, Metallophos, SurE), and heme binding (Cytochrom_C, NIR_SIR, p450, peroxidase). Notice the particularly dramatic expansion in the peroxidase family, which has a likely role in lignin degradation.

Pfam domain	***Cs***	***Pc***	***Pp***	***Sl***	***Lb***	***Sc***
WD40	863	287	148	335	565	342
p450	192	113	65	137	69	107
Pkinase	145	106	56	95	120	116
ABC_tran	81	78	21	60	65	74
Aldo_ket_red	57	52	32	43	20	55
Pkinase_Tyr	53	19	9	22	35	22
Asp	41	52	18	22	23	27
ADH_zinc_N	52	52	33	37	21	44
ABC_membrane	52	30	5	29	29	39
ADH_N	49	45	22	34	23	35
zf-C3HC4	36	13	7	28	28	37
COesterase	33	9	18	25	14	23
Metallophos	25	19	14	21	29	20
TPR_1	28	24	11	16	17	21
SNF2_N	24	12	9	22	20	21
Hydrophobin	24	16	2	18	16	12
Acetyltransf_1	23	19	11	21	21	20
Aldedh	23	19	4	15	16	16
Abhydrolase_3	23	8	23	12	4	12
TPR_2	18	21	7	8	11	15
2OG-FeII_Oxy	21	14	7	14	10	12
ICMT	21	8	10	8	4	4
Pro-kuma_activ	19	9	10	9	6	5
Cupin_1	18	16	4	6	2	10
Aminotran_1_2	16	14	5	13	12	17
Cupin_2	17	16	4	7	5	11
peroxidase	17	17	1	1	2	1
PHD	16	11	8	15	15	16
E1-E2_ATPase	16	11	7	11	16	9
HEAT	16	10	4	11	12	12
GST_N	16	12	11	12	13	11
Peptidase_C14	13	3	6	3	15	9
Pro_isomerase	15	14	7	11	11	14
DSPc	13	7	8	8	14	10
NAD_binding_1	11	13	2	7	7	9
PX	13	11	2	9	13	10
Methyltransf_2	13	9	7	6	5	8
Amidase	11	12	1	11	5	7
RhoGEF	12	8	6	9	11	12

RhoGAP	12	10	8	9	11	11
hATC	12	4	1	1	9	3
3Beta_HSD	10	9	3	11	8	7
Amidohydro_1	11	8	5	11	7	11
BRCT	11	7	10	6	10	11
Cation_ATPase_N	9	4	3	4	11	5
GST_C	11	10	8	8	11	5
FHA	11	10	4	5	8	10
LIM	9	6	7	4	10	7
Ribonuclease_3	10	3	4	7	7	8
Exonuc_X-T	10	4	0	6	6	6
DUF1793	10	1	0	1	0	5
PGAM	8	7	2	6	6	9
Mov34	9	7	1	7	8	8
DLH	8	5	4	5	9	6
ZZ	9	9	0	5	4	4
Cation_ATPase_C	9	4	1	5	5	4
NIR_SIR_ferr	9	3	2	3	3	2
Kelch_1	6	2	3	8	4	4
Peptidase_M16_C	8	8	0	8	5	6
PQ-loop	8	5	3	8	7	8
MIP	8	8	1	3	7	2
CMAS	8	5	2	7	7	6
FAD_binding_1	7	8	2	3	3	3
SNARE_assoc	8	5	0	4	5	7
MCM	8	6	3	6	5	6
HET	8	0	0	4	0	0
bZIP_1	8	6	5	6	5	6
Glyco_hydro_28	6	3	6	7	3	3
LysM	6	4	0	7	3	4
APH	7	7	4	3	5	7
RasGEF	7	6	1	6	6	5
IBN_N	7	5	0	5	6	6
CBFD_NFYB_HMF	7	5	1	6	6	5
JmjC	7	6	1	4	5	5
EHN	7	4	1	3	0	6
HMA	7	3	6	3	3	4
C1_1	7	6	4	4	4	4
Ubie_methyltran	5	6	2	4	1	3
Aa_trans	6	6	1	5	3	5
Glyco_hydro_10	6	6	4	1	0	5
Peptidase_S8	6	4	0	1	5	2
Amino_oxidase	6	4	2	3	5	4
Subtilisin_N	6	1	0	2	2	2
SurE	6	3	1	0	1	1
Bac_rhodopsin	3	5	0	1	0	0
Sec7	5	2	3	4	5	5
NIF	5	4	1	4	4	5
PBD	4	5	1	1	3	2

Ion_trans	5	4	3	5	5	3
Chitin_synth_1	5	4	0	4	4	4
Chitin_synth_1N	5	4	0	4	4	4
Band_7	5	5	2	3	5	3
FYVE	5	4	1	3	5	4
Sulfate_transp	5	4	1	3	5	4
PLDc	5	3	1	4	2	4
DUF6	5	2	0	3	4	3
RGS	5	3	0	2	2	3
Kelch_2	5	4	2	3	2	2
CRAL_TRIO_N	5	4	3	3	3	4
Glyco_hydro_53	5	1	0	1	0	1
DHDPS	3	4	0	2	1	2
CUE	3	2	2	1	4	1
ENTH	4	4	0	3	2	4
Pro_CA	4	4	0	4	3	2
CoA_transf_3	4	3	0	3	4	3
Oxidored_molyb	4	4	1	4	3	1
TB2_DP1_HVA22	4	3	1	1	4	4
UDPGP	4	4	0	3	3	3
Complex1_LYR	4	2	3	4	4	2
FF	4	4	2	2	3	4
Syntaxin	4	3	2	4	4	2
Arginase	4	3	0	3	2	4
DUF803	4	4	1	3	3	3
BAR	4	3	0	2	3	3
CoA_trans	4	2	0	4	2	2
Mem_trans	4	2	2	1	2	4
SGL	4	1	1	1	2	4
Tyrosinase	4	0	0	3	3	1
CPSF_A	4	3	1	3	3	2
MutS_I	4	4	3	3	3	3
Sod_Fe_N	4	3	1	2	3	2
Ribonuclease	4	1	1	0	3	2
Melibiase	4	2	2	3	1	1
Abhydrolase_2	4	2	0	2	2	2
Citrate_synt	4	3	2	3	3	3
STAS	4	2	0	2	1	2
CBM_20	4	2	2	2	1	1
NIR_SIR	4	2	1	2	2	1
TENA_THI-4	4	1	0	1	1	1
zf-PARP	4	2	1	1	2	1
BCS1_N	3	2	0	2	3	3
FATC	3	2	0	2	3	3
Glyco_transf_22	3	2	0	3	2	3
Pex2_Pex12	3	2	0	2	3	3
TFIIS_C	3	3	0	2	3	2
Voltage_CLC	3	2	0	2	3	3
tRNA-synt_1c	3	2	0	3	2	3

DUF1680	2	0	1	1	0	3
Gly_transf_sug	2	1	0	0	1	3
DUF1000	3	1	0	2	3	3
DUF155	3	1	0	3	2	3
PWWP	2	0	0	0	3	1
L51_S25_CI-B8	3	3	1	2	3	2
GAF	3	1	2	3	1	3
HORMA	3	1	1	2	3	3
MA3	3	3	1	1	2	3
PLA2_B	3	3	3	1	0	0
GDPD	3	2	0	2	2	3
Glyco_hydro_1	3	2	2	2	0	3
HD	3	3	2	2	0	2
SRPRB	3	2	2	2	3	3
HSP90	3	3	0	1	2	2
MS_channel	3	2	0	1	2	3
AIG2	3	1	0	1	3	2
IPK	3	2	1	2	3	2
C4dic_mal_tran	3	2	1	2	1	3
RA	3	2	1	1	3	2
CPSase_L_D3	3	2	0	2	2	2
Trp_halogenase	3	1	1	2	1	3
FAT	3	1	0	2	2	2
Glyco_hydro_15	3	2	2	2	2	3
PhoD	3	2	0	0	1	2
MutS_IV	3	2	1	2	2	1
FolB	3	0	2	1	1	0
TIP49	3	2	1	2	2	2
HAT	3	1	0	1	2	1
PROCN	3	2	0	1	1	1
YCII	3	2	0	0	0	0
Cytochrom_C	3	1	2	2	1	1
PROCT	3	1	0	1	1	1
RNA_pol_Rpb1_6	3	1	0	0	1	1
RNA_pol_Rpb1_7	3	1	1	0	1	1

II. Telomeric regions and Helitrons

i. Telomeric sequences in *C. subvermispora*

Un-assembled reads and assembled scaffolds (unmasked) were used to perform the screening of telomeric sequences in *C. subvermispora* using the Tandem Repeat Finder software (http://tandem.bu.edu/trf/trf.html)

All the unassembled scaffolds were screened for telomeric sequences and 207 out of them harboured five to 23 tandemly repeated copies of the telomeric sequence TTAGGG with a modal repeat number of 17-19 copies (82 % of the scaffolds) and 18.7 repeats per scaffold on average. The analysis also revealed that the telomeric sequence CCCTAA was present in 187 scaffolds with a copy number varying from seven to 22 and a modal repeat number of 17 to 19 (84%) and a repeat mean of 19.2 copies. In total, the average number of telomeric repeats in *C. subvermispora* is 19.0.

The genome of *C. subvermispora* was assembled in 740 scaffolds. Forty-two of them contained telomeric sequences: 22 harbouring TTAGGG sequence (Dataset S1 column 1, scaffolds: 3, 4, 7, 8, 25, 35, 41, 43, 138, 145, 310, 360, 363, 293, 444, 450, 474, 476, 571, 610, 624, and 663). The telomeric region TTAGGG was placed at the bottom (3') end of the chromosome in 21 out of the 22 scaffolds: The only exception was scaffold 3 in which this sequence was interstitial.

Twenty out of 42 scaffolds harboured the sequence CCCTAA (Dataset S1 column 1, scaffolds: 1, 2, 3 ,5 ,7 ,13 ,16, 22, 27, 28, 40, 368, 439, 440, 564, 609, 635 ,656, 671, and 673). The telomeric region CCCTAA was placed at the upper (5') end of the chromosome in 19 out of 20 scaffolds while it was found to be interstitial in scaffold 3.

These results suggest an incorrect assembling of scaffold 3 because both the direct $TTAGGG_{21}$ (TTAGGG, scaffold 3: 1205371-1205497) and the reverse $CCCTAA_{21}$ (CCCTAA, scaffold 3: 1206763-1205497) telomeric sequences are separated by a 1265 bp gap.

The telomeres found in scaffolds 5 and 7 appear to have a complex structure. The telomere of scaffold 5 contained 22 and 20 copies of the CCCTAA sequence separated by a gap of 193 bp. In the case of scaffold 7, 20 and 18 copies of this sequence appeared separated by a gap of 1680 bp. We presume that these complex structures reflect mis-assemblages of the telomeric sequences in these two scaffolds.

Summarizing, we identified 42 regions containing telomeric repeats. Twenty-two contained direct telomeric sequences TTAGGG present in 22 scaffolds, and 22 reverse telomeric regions CCCTAA present in 20 scaffolds. Scaffolds 5 and 7 harbored two telomere reverse sequences at the upper (5') end. Thus, the *Ceriporiopsis* genome is arranged in 11 linkage groups similar to that of the agaric *Pleurotus ostreatus,* a distantly related white-rot fungus. Scaffold 7 may represent a fully assembled *C. subvermispora* chromosome as it features telomeric repeats at both ends.

ii. Analysis of subtelomeric sequences in *C. subvermispora*

Twenty-two scaffolds harboring subtelomeric regions adjacent to the telomeric repeat TTAGGG and 20 adjacent to the telomeric repeat CCCTAA were studied.

In the case of the subtelomeric regions adjacent to the TTAGGG repeat, a 100 kb sequence region was studied in the long scaffolds: 3, 4, 7, 8, 25 and 35. For the scaffolds shorter than 100 kb (scaffolds: 41, 43, 138, 145, 310, 360, 363, 393, 444, 450, 474, 476, 571, 610, 624 and 663) the whole scaffold was studied (Dataset S1 columns 1, 5 and 6). In the case of the subtelomeric region adjacent to the CCCTAA repeat, a segment of 100 kb was studied in the long scaffolds 1, 2, 3, 7, 13, 16, 22, 27 and 28; while the whole scaffold was analysed in those shorter than 100 kb (scaffolds 40, 368, 439, 440, 564, 609, 635, 656, 671 and 673). (Dataset S1 columns1, 5 and 6).

Five hundred and fourteen filtered model genes were found in the subtelomeric regions studied. For a further comparative analysis, only those *C. subvermispora* genes found on scaffolds larger than 100 kb (466 gene models) were considered. It was observed that 21 to 32 gene models were found in these 100 kb sequences (Dataset S1, column 7). Filtered model genes were used to search for homologies against the NCBI database (Table S3, columns 8, 9, 10, 11), and the *Pleurotus ostreatus* annotated genome (Dataset S1, columns 12, 13, 14 15). In both cases, a threshold E-value $<10^{-20}$ was set with the purpose of identifying predicted proteins, hypothetical proteins and genes present.

The comparison against the NCBI data base revealed the following figures per 100 kb long sequence: four to 22 gene models with no homology to a database entry (column 8) and five to 27 gene models with homology (column 9). Among the gene models with homology to database entries, two to 23 gene models showed homology to predicted and hypothetical proteins (column 10), and two to nine gene models were homologous to identified genes (column 11).

The comparisons between the *Ceriporiopsis* subtelomeric regions and the *P. ostreatus* annotated genome yielded the following results: 5 – 23 filtered gene models with no hits (column 12), and seven to 24 models with homology to *P. ostreatus* proteins described in NCBI data base (column 13). Among the gene models with homology to *P. ostreatus* models three to 21 were homologous to predicted and hypothetical proteins (column 14) and 1 to8 were homologous to *P. ostreatus* gene models with assigned function in the NCBI database (column 15). A significant percentage (63%) of the predicted and hypothetical proteins present in the sub-telomeric regions of *Ceriporiopsis* showed conserved domains characteristic of secondary metabolites.

iii. Synteny of the subtelomeric regions in *C. subvermispora*

Most of the *C. subvermispora* gene-models present in the sub-telomeric regions were found to be present in interstitial locations of four *P. ostreatus* linkage groups. The sub-telomeric gene-models found in the *C. subvermispora* scaffold 2, however, were found to have counterparts in sub-telomeric regions of the *P. ostreatus* linkage group 4 (Dataset S1, columns, 16, 17).

iv. Helitrons in *C. subvermispora*

In order to look for the *C. subvermispora* helitrons, we used the Helsearch program (16) to analyse the assembled scaffolds. This program was developed using the tiny structural features of *Helitrons*, and a requirement for at least two identical 3′ ends between separate elements.

The program "HelSearch" was designed to search for the CTRRt sequence in the genome sequence. To narrow the results, an insertion site T was included in the search. The proposed *Helitron* end (helend) structure is composed of a minimum of six hairpin pairs (2 mismatches allowed) upstream of the CTRR, a 2- to 4-bp loop, and 5–8 bp between the hairpin and CTRR. Identified candidate helends were grouped together by their hairpin structure. Flanking sequences were obtained for each helend, and multiple alignments by CLUSTALW were performed for those sequences within each group (16).

Eleven different helitrons were found by the Helsearch program in the *C. subvermispora* genome, the same number as the putative helitrons found in the *P. ostreatus* genome (11 in the PC15 monokaryon)(Dataset S1). In total, 49 helitron insertions were identified. The number of insertions ranged from two (five helitrons) to 10 (two helitrons). Nine insertions were found in scaffold 1 and seven of them correspond to insertions of the same helitrons into Lys-tRNA genes. The second helitrons-richest scaffold is number 15, containing five insertions, three of which correspond to insertions in protein-kinase genes. Four insertions of two different helitrons occur in scaffold 9.

Neither of the helitrons found in *Ceriporiopsis* had perfectly conserved hairpins with the *Pleurotus* helitrons. However, many of the positions where the helitrons were found in *C. subvermispora* contained genes similar (or with similar recurrent domains) to those found linked to helitrons in *P. ostreatus* or in *Tremella*. For instance, the following domains are common to *Ceriporiopsis, Pleurotus* and *Tremella*: Cytochrome P_{450}, PFAM DUF 889, Helicase domains, Proteophosphoglycan and Protein kinase.

The helitron with hairpin TCCCGT_CAAC_ACGGGA was found in two adjacent positions in scaffold_5 suggesting the occurrence of a duplication rolling circle event. The helitron with hairpin **CCCGTGG_TGT_CCACGGG** was not found in the *Pleurotus* genome. However, a helitron with a very similar sequence (**CCCGTGC_GAA_GCACGGG)** was found in *Pleurotus*. Both contained genes with the same protein domains: the PFAM DUFF889 helicase domain and proteophosphoglycan. It appears that these genes were contained in a helitron that could be transferred horizontally between species. A similar study done in *Tremella mesenterica* yielded similar results for these domains.

III. Oxidative systems potentially involved in lignin degradation

IIIA. Heme peroxidases in the *C. subvermispora* genome

i. Manual annotation of the different peroxidase gene models. A preliminary screening of the automatically-annotated genome of *C. subvermispora* was performed using the Advanced Search option ("peroxidase" as search term) at the JGI web-site, and forty-four gene models were initially identified. Then, a sequence-by-sequence exhaustive analysis revealed that only twenty-six of the above gene models encode heme peroxidases.

These could be classified into four different groups (after amino acid sequence alignment and detection of characteristic amino acid residues specific for each group) corresponding to classical families and recently described superfamilies, as follows: **i)** Cytochrome *c* peroxidase (CCP) (1 model, Cesubv83438); **ii)** Ligninolytic peroxidases (15 models), including twelve typical "long" manganese peroxidases (MnP) (Cesubv49863, Cesubv50297, Cesubv50686, Cesubv94398, Cesubv105539, Cesubv114036, Cesubv114076, Cesubv116608, Cesubv117436, Cesubv139965, Cesubv143390, Cesubv157986) specific for Mn^{2+} and one "short" MnP (Cesubv124076) also able to oxidize phenols and ABTS in the absence of Mn^{2+} and evolutionarily more related to lignin peroxidases (LiP) and versatile peroxidases (VP) than to the classical long MnPs (3), one putative VP (Cesubv99382) and one putative LiP (Cesubv118677); **iii)** "Low redox-potential" peroxidase (one model, Cesubv112162); and **iv)** Heme-thiolate peroxidase/ peroxygenases (chloroperoxidase/aromatic peroxygenase, CPO/APO, type enzymes) (nine models, Cesubv80799, Cesubv81391, Cesubv114787, Cesubv114799, Cesubv115079, Cesubv115379, Cesubv118102, Cesubv121474, Cesubv122198).

Among the nine putative heme-thiolate peroxidases (HTPs) some may be able to hydroxylate both aromatic and aliphatic compounds, among other reactions (17, 18). In contrast with other white-rot genomes, those of *P. chrysosporium* and *C. subvermispora* completely lack genes encoding the so-called dye-decolorizing peroxidases (DyP) that are able to degrade lignin model dimers (19).

Classification of the 15 ligninolytic peroxidases as LiP, MnP and VP was made on the basis of amino acid residues at their respective active center and substrate oxidation sites after homology modeling, using related crystal structures as templates and programs implemented by the automated protein homology modeling server "SWISS-MODEL" (20).

ii. Comparison with other basidiomycete peroxidases. Heme peroxidase gene models in the *C. subvermispora* genome were aligned with 107 basidiomycete and 2 ascomycete heme peroxidase protein sequences, and theoretical molecular structures for selected peroxidases were examined (Figure S2**)**. This was followed by an analysis of the evolutionary relationships among all the basidiomycete heme peroxidases described to date, where the positions of the 26 gene models from the *C. subvermispora* genome were compared.

The dendrogram (Figure S3) shows: **i)** A large cluster containing all the ligninolytic peroxidases (a total of 90 including LiP, MnP, VP - most of them included in compressed sub-trees - and a few related sequences described with different names) together with five generic peroxidases (putative low redox-potential enzymes) from *Coprinopsis cinerea*, *Coprinellus*

disseminatus, *P. placenta,* the so-called NOPA peroxidase identified in the genome of *P. chrysosporium* and *C. subvermispora* Cesubv112162 found in the *C. subvermispora* genome; **ii)** The dye-decolorizing peroxidases (DyP) sub-tree (formed by 14 peroxidases from *Pleurotus ostreatus*, *Marasmius scorodonius*, *Bjerkandera adusta*, *Termitomyces albuminosus* and *Polyporus* sp.; **iii)** The group of APO/CPO-type heme-thiolate peroxidases including the *Agrocybe aegerita* APO and related enzymes from *P. chrysosporium* and *Agaricus bisporus*, as well as peroxidases recently reported for the first time in the genome of *Pleurotus ostreatus* and now in the genome of *C. subvermispora* where nine gene models of this peroxidase superfamily were localized (unexpectedly, the related *Leptoxyphium fumago* CPO did not fit in this cluster); and **iv)** a group of evolutionarily distant peroxidases formed by CCPs, including one model from *C. subvermispora* not related with the previous ones.

A closer view of the ligninolytic peroxidase cluster is also shown in Figure S3**.** The structural-functional classification based on the presence of a Mn^{2+}-oxidation site and/or a catalytic tryptophan, previously used to classify the *C. subvermispora* ligninolytic peroxidase genes, is extended here to all the 90 basidiomycete peroxidases described to date (the presence of a catalytic tyrosine in a *Trametes cervina* LiP is also indicated). Several of these enzymes were not highly expressed and/or regulated under the narrow set of culture conditions employed here (main text Fig. 1; GSE34636_si_Table1) and additional investigation is needed to more fully describe the expression of all the *C. subvermispora* peroxidases. On the other hand, we also observed that a MnP-encoding gene was among the most strongly upregulated in aspen medium (Cesubv117436, >6.5-fold). Interestingly, the three peroxidase proteins detected in wood medium (Cesubv157986, Cesubv116608 and Cesubv50297) correspond to the abovementioned "extra long" MnPs, whose high stability has been reported in *D. squalens* (21).

In this expanded dendrogram, typical “long” MnPs (from *Lentinus edodes*, *Ceriporiopsis rivulosa, C. subvermispora, Dichomitus squalens, P. chrysosporium, Phanerochaete sordida, Phlebia radiata* and an unidentified basidiomycete) form a large and homogeneous cluster clearly separated from the rest of ligninolytic peroxidases, all of them exhibiting a shorter C-terminal tail. By contrast, VP, LiP and “short” MnPs seem to have divergent and recurrent evolution pathways. Peroxidases Cesubv118677 and Cesubv99382 seem to represent transition states of these evolutionary processes, not only by their position in the dendrogram but also by their catalytic properties, as discussed below.

iii. Heterologous expression and purification of *C. subvermispora* peroxidases Following careful inspection and revision of automated annotations, three *C. subvermispora* genes were selected for *Escherichia coli* expression; putative VP (Cesubv99382), LiP (Cesubv118677) and MnP (Cesubv117436). The latter exhibited the highest induction levels in medium containing ball-milled aspen versus glucose as sole carbon source (main text Figure 1). The precise positions of introns were determined by comparison with other known ligninolytic peroxidase genes, and the corresponding cDNA sequences were synthesized (ATG-Biosynthetics). The synthesized cDNAs coding for the mature proteins were cloned into the pFLAG1 expression vector under the control of the IPTG-inducible *tac* promoter. The enzymes were overexpressed in *E. coli* W3110, and SDS-PAGE analysis of the cell extracts showed accumulation of the recombinant peroxidases in the insoluble fraction containing the

inclusion bodies. The proteins were subsequently activated *in vitro* under conditions previously optimized for *P. eryngii* VP (22). Direct veratryl alcohol (VA) oxidation was measured to check the activity of peroxidases in the refolding mixtures confirming enzyme activation. Then, the enzymes were purified to homogeneity using an anion exchange column (Resource Q from GE Healthcare). Finally, their molar extinction coefficients at the absorption maximum of the Soret band (407 nm) were determined.

LiP from *P. chrysosporium* (isoenzyme LiPH8) and VP from *P. eryngii* (isoenzyme VPL) were also expressed in *E. coli* and purified for comparison with the *C. subvermispora* peroxidases, after *E. coli* expression and *in vitro* activation following the protocols of Doyle and Smith (23) and Pérez-Boada *et al*. (22) respectively.

iv. Enzyme characterization. Oxidation of three representative substrates, namely Mn^{2+}, and high redox-potential veratryl alcohol (VA) and Reactive Black 5 (RB5), by both recombinant peroxidases were investigated under steady-state conditions. The corresponding steady-state kinetic constants were compared with those of *P. eryngii* VP and *P. chrysosporium* LiP produced by this method (main text Table 1).

As described in the main text, Cesubv118677 could not oxidize Mn^{2+} as expected according to the absence of a typical manganese oxidation site in its theoretical molecular structure (Figure S2A)**.** Surprisingly, Cesubv99382, tentatively identified as a VP, was also not able to oxidize Mn^{2+}, even though a typical manganese oxidation site is predicted in its structural model (Figure S2C).

Typical VPs and MnPs oxidizing Mn^{2+} do not include an acidic residue contiguous to the aspartate forming the manganese binding site, whereas Cesubv99382 presents a glutamic acid at the corresponding position. Typical LiPs, unable to oxidize Mn^{2+}, exhibit an acidic residue at an equivalent position. Effective transformation of LiP into MnP have been obtained by removing this acidic residue, in addition to introducing the corresponding aspartic acid at the manganese binding site (24).

IIIB. Multicopper oxidases and iron homeostasis

i. Multicopper oxidases (MCOs) and related proteins involved in iron metabolism Using the *P. chrysosporium* MCO1 and Fet3 protein sequences, as well as the *C. subvermispora* Lcs-1 protein (25), BLASTP searches of the *Ceriporiopsis* genome revealed various multicopper oxidase-encoding genes, at least seven of which were classified as laccases *sensu stricto* (26) (Figure 2 main text; GSE34636_si_Table1.xls).

The *C. subvermispora* laccases are related (25) (main text, Fig. 2) to those previously identified in *Postia placenta*, a polypore fungus producing brown rot of wood (27), however the number and phylogenetic diversity of laccases in *C. subvermispora* is higher. Bulk lignin is not removed during *P. placenta* wood decay, but substantial modification occurs, including some scission (28), and laccase has been implicated in the production of possibly responsible reactive oxygen species (29). Consistent with a role in lignocellulose modification, transcript levels corresponding to *C. subvermispora* laccase Cesubv118801 and *P. placenta*

Pospl111314 (30) were significantly upregulated (>3-fold, $P<0.01$) in media containing ball milled *Populus grandidentata* wood (aspen) relative to glucose medium (main text, Fig. 2).

ClustalW analysis of all *C. subvermispora* MCOs, plus all five MCOs from *P. chrysosporium* (MCO1 to MCO4, plus Fet3) showed the presence of the glutamic acid involved in iron oxidation in the Fet3 protein from *S. cerevisiae* (analogous to the E214 in *P. chrysosporium* MCO1) in the *Ceriporiopsis* Cesubv51376 protein. Interestingly, in the surroundings of the mentioned amino acid, there is an insertion not present in laccases.

The latter observation suggests that the *C. subvermispora* genome encodes MCOs which are different from laccases, and that these are more similar to those observed and described originally in *P. chrysosporium*. Thus, we analyzed the four laccase signatures previously defined by Kumar et al., 2003 (31). These signatures are shared among laccase proteins, but they are absent in *P. chrysosporium* MCO proteins, as well as in Fet3 proteins. These signatures are shown in four different alignments, including the multicopper oxidases identified in the *C. subvermispora* genome.

L1 signature

```
L1         HWHG xxxxxxxxx DG xxxxx QCPI
Cs 84170   HWHG IFQQGTAWA DG PAFVT QCPI  laccase
Cs 88089   HWHG IFQHGTTWA DG PAFVS QCPI  laccase
Cs 108852  HWHG IFQQHTNWA DG AAMVS QCPI  laccase
Cs 115063  HWHG IFQHTTAWA DG PAFVT QCPI  laccase
Cs 115068  HWHG ILQHTTAWA DG PAFVT QCPI  laccase
Cs 118801  HWHG LFQHGTTWA DG PAFVS QCPI  laccase
Cs 137686  HWHG LFQEGTTWA DG AAFVS QCPI  laccase
Pc mco1    HWHG IPQNGTAYY DG TAGIT ECGI
Cs 51376   HWHG LYQRGTNYY DG TAAIT QCGI  PcMCO-like
Pc mco2    HWHG LFQNQTNYY DG TAGIT ECGI
Pc mco3    HWHG LFHNGTNYY DG TAAIT ECGI
Pc mco4    HWHG LYQNSTNYY DG TAGVT ECGI
Pc Fet3    HHHG MFFNSTSWM DG ALAIS QCGV
Cs 67172   HHHG MFFNSTSWM DG ALGVS QCGI  Fet3
```

The Cesubv51376 L1 signature is more closely related to *P. chrysosporium* MCOs, and distinct from laccases and Fet3 ferroxidases. The Cesubv67172 L1 signature is similar to PcFet3. Letters in orange denote differences from the consensus sequence (top).

L2 signature

```
L2         GTxWYHSHxxxQYCDGLxGx(FLIM)
Cs 84170   GTYWYHSHLATQYCDGLRGPL    laccase
Cs 88089   GTFWYHSHLAAQYCDGLRGPL    laccase
Cs 108852  GTFWYHSHLKQQYCDGLRGPL    laccase
Cs 115063  GTFWYHSHLATQYCDGLRGPL    laccase
Cs 115068  GTFWYHSHLVDQYCDGLRGPL    laccase
Cs 118801  GTFWYHSHLATQYCDGLRGPL    laccase
Cs 137686  GTFWYHSHLATQYCDGLRGPL    laccase
Pc mco1    GTTWWHSHYDTQYTDGVTGAL
Cs 51376   GSTWWHAHYSTQYTDGITGAL    PcMCO-like
Pc mco2    GTTWWHAHYSTQYTDGITGAL
Pc mco3    GTTWWHAHYSTQYTDGITGAL
Pc mco4    GTTWWHAHYDTQYTDGVTGAL
Pc Fet3    GTYWVHSHASGQYVDGLRAPV
Cs 67172   GTYWWHAHAKGQYVNGLRAPL    Fet3
```

The Cesubv51376 L2 signature is similar to PcMCO2 and MCO3, but different from MCO1 and MCO4 (Fig. 2, main text). It also possesses a different amino acid in the second position within the signature, not observed among all proteins analyzed. Cesubv67172 (orthologue to Fet3) shares similarities with PcFet3 and PcMCOs. Letters in orange denote differences from the consensus sequence (top).

L3 signature

```
L3         HPxHLHGH
Cs 84170   HPFHLHGH
Cs 88089   HPFHLHGH
Cs 108852  HPLHLHGH
Cs 115063  HPFHLHGH
Cs 115068  HPFHLHGH
Cs 118801  HPFHLHGH
Cs 137686  HPFHLHGH
Pc mco1    HPFHLHGY
Cs 51376   HPFHLHGH
Pc mco2    HPFHLHGH
Pc mco3    HPFHLHGH
Pc mco4    HPFHLHGH
Pc Fet3    HPFHLHGH
Cs 67172   HPFHLHGH
```

Only PcMCO1 has a difference in the L3 signature. The letter in orange denote a difference from the consensus sequence (top).

L4 signature

```
L4         G(PA)W x(LFV)HCHI(DAE)x H xxx G(LMF)xxx(LFM)
Cs 84170   G P  W F L   HCHI D   W H LAA G L   AIV L
Cs 88089   G P  W F L   HCHI D   W H LQA G   F AIV  F
Cs 108852  G P  W F L   HCHI D   S H LNA G   F AIV  F
Cs 115063  G P  W F L   HCHI D   Y H LNA G   F AIV  F
Cs 115068  G P  W F L   HCHI D   Y H LTT G   F AVV  F
Cs 118801  G P  W F  F  HCHI D   W H LQA G   F AIV  F
Cs 137686  G P  W F L   HCHI D   W H LNA G   F AIV  F
Pc mco1    G  A W T L   HCHI S   W H MSA G L   LMQ  F
Cs 51376   G Y  W A  F  HCHI Q   W H MAA G L   LFQ  F
Pc mco2    G L  W A  F  HCHL  A  W H MAA G L   LMQ I
Pc mco3    G L  W A  F  HCHL  A  W H MAA G  M  LMQ V
Pc mco4    G I  W T L   HCHI  A  W H MAA G L   MMQ I
Pc Fet3    G  A W I  F  HCHI   E W H LQA G L   AVT  F
Cs 67172   G V  W F  F  HCHI   E W H LEV G L   AVQ  F
```

The L4 signature was found in all laccase-encoding genes, but not in PcMCOs, Cesubv67172 (*C. subvermispora* Fet3) and Cesubv51376, the *Ceriporiopsis* PcMCO-like gene. Phylogenetic analysis showed that Cesubv51376 clusters together with the MCO2 and MCO3 from *P. chrysosporium*, in a separate branch from laccases (main text Figure 2).

In aggregate, these results indicate that the *C. subvermispora* genome encodes for at least seven laccases, one Fet3 and a protein similar to *P. chrysosporium* MCOs. This latter protein is most similar to PcMCO3, and based on the glutamic acid conservation, it should have a ferroxidase activity similar to that experimentally determined for PcMCO1.

Interestingly, five MCOs encoding genes are clustered together on scaffold_7 (four laccases and the MCO3-like encoding gene). The only previously described laccase gene from *Ceriporiopsis* (*lcs-1*, (25)) is also located within a cluster on scaffold_21, composed of three laccase encoding genes. Relative to glucose-containing medium, transcripts corresponding to *lcs-1* accumulate at significant levels in medium containing ball-milled aspen as sole carbon source (>6-fold $P<0.01$) (Figure 2 main text; GSE34636_si_Table1.xls).

ii. Non-reductive pathways potentially involved in iron metabolism

i) Sid1-like encoding genes Siderophore-iron transporters compose the non-reductive iron uptake system pathway. Sid1 in *Ustilago maydis* encodes the first committed step (first enzyme) in hydroxamate siderophore synthesis. There is one gene model for *sid1* in *Ceriporiopsis* (Cesubv113443).

ii) Non-ribosomal peptide synthetase. As expected, several NRPS are present in the *Ceriporiopsis* genome. These gene models include Cesubv58842, Cesubv74328, Cesubv94730, Cesubv95639, Cesubv106744, Cesubv106870, Cesubv107385, Cesubv112105, Cesubv115161, Cesubv150972, Cesubv153005, Cesubv157295 and Cesubv162022. Additional information regarding NRPSs is presented in section IX (see below).

iii) NRAMP family iron transporters. Known as Smf 1 to 3 in *S. cerevisiae*, the NRAMP proteins play an important role in Mn^{2+} and iron transport. Two genes have been identified (Cesubv128550 and Cesubv102620).

IIIC. Peroxide generation and accessory enzymes

Numerous proteins have been implicated in lignin degradation via generation of reduced iron and peroxide. Among the latter were four *C. subvermispora* copper radical oxidase genes (*cros*), one of which (Cesubv155904) was significantly ($P<0.05$) upregulated (>2-fold) in in aspen wood relative to a glucose medium (GSE34636_si_Table 1 and SI appendix). In contrast, the *P. chrysosporium* genome contains seven CROs and none are significantly upregulated under identical conditions. The classification of upregulated *C. subvermispora* gene Cesubv155904 was established by comparisons with glyoxal oxidase of *P. chrysosporium* (AAA87594.1). The experimentally identified catalytic residues (32) C70, Y135, YH377-378 and H471 were easily identified. Cesubv115979 contains four WSC domains preceding the CRO alignment. The function of this highly conserved domain is unknown.

Other upregulated *C. subvermispora* genes likely involved in redox cycling include aldehyde reductases (Cesubv113709, Cesubv118401), aldehyde dehydrogenase (Cesubv87228), arabinitol 2-dehydrogenase (Cesubv152859) and quinone oxidoreductase (Cesubv111965). Of these, the *P. chrysosporium* aldehyde reductase homolog Phchr10221 was also upregulated (>2-fold, $P<0.05$).

A single copy of cellobiose dehydrogenase (CDH) with substantial transcript accumulation (>3-fold, $P<0.01$) was observed for both fungi grown in media with aspen as compared to glucose medium. The CDH gene of *C. subvermispora* (Cersub84792) corresponds to those previously cloned and sequenced (aminoacid and nucleotide GenBank accession numbers ACF60617 and EU660051, respectively). Sequence alignments and modeling revealed the conserved substrate binding pocket (33) and secretion signal. The precise role of CDH is unclear, but the observed iron reduction and peroxide generation by this bifunctional enzyme suggest an important role in Fenton chemistry (34). More recently, CDH has been shown to enhance cellulose depolymerization in concert with members of the 'glycoside hydrolase' family 61 (35, 36) and, consistent with this, multiple GH61 genes are upregulated in *P. chrysosporium* and in *C. subvermispora* aspen cultures (GSE34636_si_Table 1).

Other H_2O_2-producing GMC oxidoreductases include five and three AAO-encoding genes in *C. subvermispora*, and in *P. chrysosporium*, respectively (Cesubv117387, Cesubv137959, Cesubv118493, Cesubv84544 and Cesubv107983; Phchr135972, Phchr6199, and Phchr37188). Extracellular hydrogen peroxide by redox-cycling of aromatic aldehydes is often associated wirh aryl-alcohol dehydrogenases (AAD) (37). Both *C. subvermispora* (Cesubv114803, Cesubv153097) and *P. chrysosporium* (Phchr128103, Phchr11055 (38)) feature two putative aryl-alchol dehydrogenases genes, all of which are expressed at modest levels. Four methanol oxidase genes in *C. subvermispora,* including Cesubv80773, were strongly down regulated, whereas the *P. chrysosporium* methanol oxidase gene encoding Phchr126879 was highly expressed in aspen-containing medium. A second *P. chrysosporium* methanol oxidase, Phchr5574, is down regulated under these conditions (GSE34636_si_Table 1).

A previously unknown GMC oxidase, Cesubv47053, was aligned (39) to glucose oxidase

from *Botryotinia fuckeliana* (GOX_Botry), aryl-alcohol oxidase from *Pleurotus eryngii* (AAO_pe), cellobiose dehydrogenase from *Pycnoporus cinnabarinus* (Celldhy_AAC32197) and pyranose oxidase from *Trametes versicolor* (P79076). Highly conserved histidine/asparagine active site residues (40) were observed as conserved motifs in the GMC family (41). The most closely related *P. chrysosporium* protein, Phchr127396, was not significantly upregulated in aspen medium (GSE34636_si_Table1.xls).

Protein ID	**Putative function**	**Glu log_2**	**BMA log_2**	**BMA/Glu Fold change**
84792	Cellobiose dehydrogenase	10.34	14.45	**17.32**
117387	Aryl-alcohol oxidase	10.81	10.89	1.05
137959	Aryl-alcohol oxidase	11.55	11.39	0.90
118493	Aryl-alcohol oxidase	10.35	9.69	0.63
84544	Aryl-alcohol oxidase	12.79	12.74	0.97
107938	Aryl-alcohol oxidase	11.21	10.89	0.80
80773	Methanol oxidase	12.76	10.36	0.19
139982	Methanol oxidase	11.46	11.28	0.88
86996	Methanol oxidase	11.49	10.99	0.71
117749	Methanol oxidase	11.81	11.18	0.64
114803	Aryl-alcohol dehydrogenase	11.20	11.17	0.98
153097	Aryl-alcohol dehydrogenase	12.02	11.80	0.86
	Automatic annotation			
47053	GMC	10.74	12.95	**4.63**
86745	Alcohol dehydrogenase	13.09	13.68	1.50
116159	GMC like Glucose oxidase	11.72	11.91	1.14
84557	GMC like glucose oxidase	13.55	13.59	1.03
91114	GMC	11.30	11.26	0.97
73736	GMC	11.45	11.40	0.97
52107	GMC	11.53	11.41	0.92
87836	GMC	12.19	11.97	0.85
91113	GMC	11.47	11.18	0.82
79287	GMC	11.10	10.75	0.79
105302	GMC	11.67	11.28	0.77
117749	GMC	11.81	11.18	0.65
115252	GMC	12.16	11.52	0.64
84404	GMC	10.87	10.23	0.64
	Copper radical oxidases			
48364	PcCRO6	10.44	10.41	0.980
115979	PcCRO5 with 4 WSC domains	14.51	12.42	**0.224**
155904	Two PcCRO2-like genes: 74273 and 74274. Microarray signals confined to 74273.	10.32	12.52	**4.604**

IV. Carbohydrate active enzymes (CAZys).

As described in the main text, the number and expression patterns of potential cellulase genes differ beween *C. subvermispora* and *P. chrysosporium*. For example, *C. subvermispora* contained three predicted proteins belonging to glycoside hydrolase family 7 (GH7) whereas six GH7 protein models were predicted for *P. chrysosporium.* Generally considered an important group featuring exocellobiohydrolases, four of the *P. chrysosporium* GH7 genes were significantly upregulated (>2-fold; $P<0.01$) in medium containing ball-milled aspen compared to glucose medium. The same *P. chrysosporium* cultures also featured significant accumulations of transcripts corresponding to two GH5 β-1-4-endoglucanases, and two GH12 endoglucanases (main text Table 3). Under identical culture conditions, upregulated *C. subvermispora* cellulases genes included a single exocellobiohydrolase (GH7), a β-1-4 endoglucanase (GH5), and a GH12 endoglucanase. Also possibly relevant, the *P. chrysosporium* genome encoded 11 GH3 family members and 31 proteins with predicted cellulose binding domains (carbohydrate binding module family 1) compared to six GH3s and 16 CBM1s in *C. subvermispora*. Mass spectrometry identified 18 and three glycoside hydrolase proteins in *P. chrysosporium* and *C. subvermispora* filtrates from aspen cultures, respectively.

In addition to the canonical cellulases, several, but not all, *C. subvermispora* hemicellulases also showed relatively low expression on aspen medium (Table S1). Both white rot fungi exhibited regulated expression of three GH10-encoding genes. *P.chrysosporium* GH11 xylanase model Phchr133788, GH43 endo-1-5-α-arabinosidase model Phchr4822, GH53 Endo-1-4-β-galactanase model Phchr138710 and GH95 α-fucosidase model Phchr6997 all showed significant transcript accumulation, whereas the *C. subvermispora* orthologs showed no accumulation in aspen. More striking, all four *P. chrysosporium* GH74-encoding genes were significantly upregulated in aspen, but transcript levels of the single *C. subvermispora* ortholog were steady in both media. With regard to hemicellulose degradation, we observed that putative glucuronyl esterases Cesubv118322 and Cesubv21396, both members of Carbohydrate Esterase family 15 (CE15), showed only modest transcript accumulation in aspen medium (GSE34636_si_Table1.xls). The *P. chrysosporiuum* homologs Phchr130517 and Phchr6482 have been implicated in the cleavage of lignin-hemicellulose esters (42) and, in distinct contrast to *C. subvermispora*, both are significantly upregulated in aspen medium relative to glucose (GSE34636_si_Table1.xls). The increased expression of CE15 genes in *P. chrysosporium* may indicate their ability to facilitate simultaneous utilization of both the lignin and hemicellulose components. Both fungi have several Carbohydrate Esterase Family 1 (CE1), four (CE4) and 16 (CE16) esterases. Transcription of the *P. chrysosporium* CE1 model Phchr126075, annotated as a feruloyl esterase, is increased 17-fold on aspen medium relative to glucose and a similar, but less pronounced increase occurs for the *C. subvermispora* ortholog. Similar to the presumed role of the glucuronoyl esterase (CE15), increased expression of this putative feruloyl esterase may benefit a lignocellulose utilization strategy which efficiently co-utilizes both the lignin and carbohydrate components by helping to disengage the hemicellulose from the lignin. *P. chrysosporium* and *C. subvermispora* have four and two more genes annotated as CE1 enzymes, respectively, but transcript levels of these are not increased on aspen medium. The much smaller CE16 esterase family may be more specific for acetyl group hydrolysis (43). *C. subvermispora* has several genes annotated in this newly defined esterase family. Transcripton of one is upregulated 6.5-fold during

growth on aspen medium while the others are not regulated. This gene was not annotated in *P. chrysosporium*, but a BLASTp search reveals only two homologs which show no significant regulation. Both *C. subvermispora* and *P. chrysosporium* have four CE4 genes that are likely involved in deacetlylation of complex carbohydrates. However, none show any transcript accumulation growing on aspen medium relative to glucose.

Beyond well established CAZy families, *P. chrysosporium* and *C. subvermispora* genomes encode multiple GH61 family members. Recently characterized as metalloenzymes (44), these enzymes may work together with cellulase (35, 36). Consistent with such a role, significant upregulation was observed for *P. chrysosporium* and *C. subvermispora* representatives (GSE34636_si_Table1.xls). Sequence comparisons show greater diversity among the 13 *P. chrysosporium* GH61 genes, but no clear relationship between primary amino acid sequence and regulation was observed.

Analysis of growth profiles in correlation with CAZy annotation

In agar medium, *C. subvermispora* grew as thin colonies on most carbon sources tested (Fig. S1). Growth on cellulose and beechwood xylan was poor compared to glucose. In contrast, growth on pectin, galactomannan (Guar gum), starch and inulin was comparable to or slightly better than glucose.

All fungi are able to use starch as a carbon source, which is reflected by the presence of GH13, GH15 and GH31 candidates in all genomes. The same was observed for galactomannan, which can be linked to the presence of GH5, GH26, GH27 and GH36 candidates in the genomes. Interestingly, the basidiomycetes do not contain GH26 endomannanases (with the exception of *S. commune*) or GH36 α-galactosidases, which suggest that their strategy for degrading galactomannan is different from the ascomycetes, but apparently equally successful based on their growth profiles.

Growth on pectin (relative to glucose) is better for *C. subvermispora* than for *C. cinerea*, *S. commune* and *P. anserina*. The main difference with respect to the CAZy content appears to be in the number of pectin hydrolases of family GH28, which is higher in *C. subvermispora* than in the other species. No significant differences in other pectin-related families were detected, suggesting that the additional GH28 enzymes in *C. subvermispora* have a significant influence on pectin hydrolysis.

Poor growth on xylan cannot be easily explained by the CAZy annotation because the number of xylan-related genes in *C. subvermispora* is similar to *P. chrysosporium* and other fungi that grow well on this substrate. In addtion, reduced growth on xylose, and therefore the quality of the main monosaccharide component of xylan as a carbon source, does not suggest an explanation as this was also observed for *P. chrysosporium* and some of the other species. Without doubt, culture conditions, especially substrate composition, will significantly alter expression patterns. For example, GH10 endo xylanase-encoding genes are among those most highly upregulated in both *P. chrysosporium* and *C. subvermispora* grown on aspen (Table S1). Further complicating interpretation, we observed that the multiple GH74 xyloglucanases were highly upregulated in *P. chrysosporium* growing on ball-milled aspen relative to glucose (>2-fold transcript accumulation, $P<0.01$). The single *C. subvermispora* ortholog, Cesubv79435, showed minimal transcript levels and no significant transcriptional regulation under identical conditions (GSE34636_si_Table1.xls).

Growth on cellulose is poor for all fungi, most likely due to the fact that during isolation of cellulose the natural structure is affected making it more difficult to degrade enzymatically. Nevertheless, cellulose specialists are able to grow on this substrate (e.g. *P. anserina*, and other species at www.fung-groth.org). *C. subvermispora* in that sense is not a cellulose specialist, an observation consistent with CAZy annotations, expression patterns, and numerous studies of decay processes (45-49).

V. Monosaccharide catabolism and gene regulation regulation

VA. Monosaccharide and CAZy regulation

i. Methods. Several strategies were used to refine annotations. First, all predicted CDSs were submitted to the KEGG Automatic Annotation Server which applies a best-best algorithm to associate the submitted sequence to a known KO number, COG, EC number, GO number and biochemical reactions. Second, CDSs were analyzed using HT-GO-FAT (High Throughput Gene Ontology Functional Annotation Toolkit), another useful software toolkit that utilizes a custom-curated BLAST database to annotate sequences by GO, EC number, KEGG pathways and so on. EC numbers can be deduced from the associated GO numbers by this program. Third, text mining was used to assign an EC number when an obvious enzyme could not be associated with an EC number or a complete EC number through the above-mentioned methods. For this purpose, the name of the enzyme was queried in the KEGG Ligand database for synonyms. Cis-acting binding sites were identified using RSAT-tools (50).

DNA and protein sequences were visually aligned using Genedoc 2.6 (http://www.psc.edu/biomed/genedoc). Phylogenetic trees were constructed by the neighbor-joining (NJ) method (51) using the computer program MEGA, Version 4.0 (52). Unalignable N- and C-terminal regions in the amino acid sequences were omitted from the analyses, and gaps and missing data were pairwise deleted. In apparent pseudogenes containing a premature translational stop, a full-length deduced amino acid sequence was re-created by replacing it with an unidentified residue (X). In genes where single-base frameshifts were evident in the genome sequence, a continuous reading frame was created either by deleting the superfluous base or by adding one base of unknown identity (N).

ii. Results Seventy-nine gene models that were verified are listed below. The genes encoding all enzymes of glycolysis were identified. *C. subvermispora* contains a single hexo- and a single glucokinase, which were identified by phylogenetic analysis and similarity to the corresponding proteins in other fungi. As in *P. ostreatus*, multiple phosphoglycerate mutases were found, and some genes were reminiscent of ancient duplications and horizontal gene transfer events or deletion in other fungi.

The genes encoding enzymes of the pentose phosphate pathway were also partially amplified as two 6-phosphogluconate dehydrogenases, transketolases and transaldolases were found. In addition, we found a putative gluconate kinase gene, and a gene encoding a putative glucose oxidase of the glucose-methanol-choline oxidoreductase family, implying that *C. subvermispora* has the capacity for oxidative glucose breakdown. The glucose oxidase gene was identified by BLAST search with the *Schizophyllum commune* glucose oxidase. This led to several hits with expectancy scores of $< 10^{-35}$. The best hit (Cesubv116159; 2E-47) had 28 % amino acid sequence identity to the glucose oxidase from *Coccidioides immitis*, with an

FAD-binding region and a signal peptide. The other hits could not unequivocally be attributed to a glucose oxidase.

The *C. subvermispora* genes required for catabolism of other monosaccharides derived from hemicellulose were also found: the Leloir pathway for D-galactose catabolism is present. Since this pathway exclusively requires the α-anomer, genes encoding two intracellular mutarotases were also identified, indicating that D-galactose likely arises intracellularily (e.g., by uptake of oligosaccharides and intracellular hydrolysis). It is of further interest that *C. subvermispora,* like *P. ostreatus*, has two galactokinases. While one appears to be an orthologue of the other basidiomycete galactokinases, the second has orthologues only in ascomycetes.

We also detected the genes encoding the enzymes necessary for degradation of D-xylose and L-arabinose. Interestingly, *C. subvermispora* lacks the multiplicity in these genes that is seen in *P. ostreatus* and ascomycetes. Thus, the arsenal of genes for pentose catabolism is narrower in comparison to other fungi.

Eleven of the 22 glycolytic genes acted at the C3 level of carbohydrate breakdown. Together with the finding of eight aquaporin genes this may indicate the rapid formation and turn-over of glycerol as an osmotic stabilizer has an important function in *C. subvermispora*.

The relative expression of these genes was examined during growth in media containing ball-milled aspen or glucose as sole carbon sources. Only 17 genes were shown to be upregulated on aspen by a factor of >two-fold. This suggests that most of the genes of carbohydrate metabolism (e.g., those of glycolysis or the pentose phosphate pathway) are already transcribed at a sufficient level on glucose, or are absent under these conditions (e.g., genes involved in chitin monomer metabolism). However, all genes involved in pentose catabolism (xylose reductase, L-arabinitol dehydrogenase, xylitol dehydrogenase) were found to be upregulated.

iii.Regulation of CAZy gene expression For identification of possible transcription factor binding sites regulating cellulase and hemicellulase expression in *Ceriporiopsis*, all CAZys and putative monosaccharide transporters (27 genes) with transcripts accumulating >two-fold in aspen relative to glucose were selected (Figure S4). The RSAT tool was used to search for motifs that were abundant in 5'-untranslated sequences (1000bp preceding the start codon ATG). Searches were conducted for direct motifs with a maximum of 8 nt and dyads of 3-4 nt.

No binding sites for the ascomycete cellulase and xylanase regulator XYR1/XlnR (5'-GGSGAAT/AA-3') were found. Consistent with this observation, a BLAST search, using only the XYR1 Zn-finger as a query identified Zn(2)Cys(6) proteins from *Ceriporiopsis* with very low probability ($E>10^{-5}$), which, when BLASTed back, identified non-XYR1 proteins with higher similarity. Regulation by XYR1/XlnR seems to be absent from *Ceriporiopsis*.

The search for an 8 nt direct motif yielded the consensus sequence 5'GCGGGGAA3' that was present in eight of the 27 genes. This sequence is similar to the binding motif for the catabolite repressor protein MIG1/CRE1/CreA. In order to test whether *Ceriporiopsis* indeed

contains an orthologue of MIG1/CreA/Cre1, we used 145 amino acids of the C2H2 Zn finger in a BLAST search. Cesubv122216 was obtained as 'best hit' (119 amino acids; $E=10^{-27}$). Its identity with MIG1/CreA/Cre1 was further supported by a comparison of the aa-sequence of the C2H2 zinc finger of various CRE1/CreA/MIG1 proteins (Figure S4). Both spacing and occurrence of the amino acids critical for binding (RxExxxR in the first, and RxxExxR in the second) are highly conserved. It is therefore concluded that Cesubv122216 is an orthologue of the ascomycete *MIG1/CRE1/creA* gene.

A search for motifs with dyad symmetry that are enriched in the 27 upregulated genes identified the sequence CGR-(N7)-YCG in 22 from 27 most upregulated genes ($P < 0.05$), with a mean occurrence of 3.15 copies per promoter (Figure S4). A search through YEASTRACT did not identify a protein known to bind this motif.

.

Monosaccharide metabolism		
Protein ID[1]		**Pentose phosphate pathway**
111274	KOG0563	Glucose-6-phosphate 1-dehydrogenase
79320	KOG2653	6-phosphogluconate dehydrogenase
106675	KOG2653	6-phosphogluconate dehydrogenase
110270	KOG3147	6-phosphogluconolactonase
62010	KOG2772	Transaldolase
84369	KOG2772	Transaldolase
56207	KOG0523	Transketolase
113807	KOG0523	Transketolase
83888	KOG2855	Ribokinase
115134	KOG3075	Ribose 5-phosphate isomerase
81796	KOG2517	Ribulose kinase ???
109209	KOG2517	Ribulose kinase
112824	KOG2531	Sugar (pentulose and hexulose) kinases
120051	KOG2531	Sugar (pentulose and hexulose) kinases
61952	KOG3111	D-ribulose-5-phosphate 3-epimerase
114314	KOG3354	Gluconate kinase
116159	KOG1238	Glucose oxidase
		Pentose catabolism
113811	KOG0725	polyol dehydrogenase?
152859	KOG0725	D-arabinitol-2-dehydrogenase
90509	KOG1577	xylose reductase
111362	KOG0024	L-arabinitol-4-dehydrogenase ("sorbitol dehydrogenase")
106920	KOG0024	xylitol dehydrogenase ("sorbitol dehydrogenase")
		Glycolysis
80619	KOG1369	Glucokinase

114817	KOG1369	Hexokinase
120206	KOG2446	Glucose-6-phosphate isomerase
85419	KOG0234	Fructose-6-phosphate 2-kinase/fructose-2,6-biphosphatase
107622	KOG0234	Fructose-6-phosphate 2-kinase/fructose-2,6-biphosphatase
126750	KOG2854	Possible pfkB family carbohydrate kinase
117565	KOG2440	Pyrophosphate-dependent phosphofructo-1-kinase
85265	KOG4153	Fructose 1,6-bisphosphate aldolase
115642	KOG1643	Triosephosphate isomerase
81702	KOG2426	Dihydroxyacetone kinase/glycerone kinase
60983	KOG0657	Glyceraldehyde 3-phosphate dehydrogenase
139989	KOG0657	Glyceraldehyde 3-phosphate dehydrogenase
79772	KOG0235	Phosphoglycerate mutase
112131	KOG0235	Phosphoglycerate mutase
118541	KOG4513	Phosphoglycerate mutase
117324	KOG1367	3-phosphoglycerate kinase
102220	KOG3734	Predicted phosphoglycerate mutase
144593	KOG4754	Predicted phosphoglycerate mutase
151079	KOG4754	Predicted phosphoglycerate mutase
87979	KOG2670	Enolase
159707	KOG2670	Enolase
110535	KOG2323	Pyruvate kinase
82587	KOG1458	Fructose-1,6-bisphosphatase
		N-acetylglucosamine catabolism
118153	KOG1794	N-Acetylglucosamine kinase
118257	KOG3892	N-acetyl-glucosamine-6-phosphate deacetylase
89886	KOG4135	Predicted phosphoglucosamine acetyltransferase
117862	KOG3148	Glucosamine-6-phosphate isomerase
		galactose catabolism
97848	KOG0631	Galactokinase
119060	KOG0631	Galactokinase
58921	KOG2638	UDP-glucose pyrophosphorylase
88357	KOG2638	UDP-glucose pyrophosphorylase
88416	KOG2638	UDP-glucose pyrophosphorylase
99718	KOG2638	UDP-glucose pyrophosphorylase
99732	KOG2638	UDP-glucose pyrophosphorylase
162082	KOG2638	UDP-glucose pyrophosphorylase
114627	KOG0625	Phosphoglucomutase
62068	KOG2537	Phosphoglucomutase/phosphomannomutase
113859	KOG1220	Phosphoglucomutase/phosphomannomutase
110596	KOG1594	Uncharacterized enzymes related to aldose 1-epimerase
127368	KOG1604	Predicted mutarotase

		mannose catabolism
115637	KOG2757	Mannose-6-phosphate isomerase
126570	KOG2757	Mannose-6-phosphate isomerase
		sugar acid metabolism
111564	KOG1429	dTDP-glucose 4-6-dehydratase/UDP-glucuronic acid decarboxylase
141261	KOG1429	dTDP-glucose 4-6-dehydratase/UDP-glucuronic acid decarboxylase
114974	KOG1372	GDP-mannose 4,6 dehydratase
115055	KOG2666	UDP-glucose/GDP-mannose dehydrogenase
		storage polysaccharides
116670	KOG2099	Glycogen phosphorylase
66259	KOG3742	Glycogen synthase
104682	KOG0658	Glycogen synthase kinase-3
89195	KOG1050	Trehalose-6-phosphate synthase component TPS1 and related subunits
107076	KOG1050	Trehalose-6-phosphate synthase component TPS1 and related subunits
116417	KOG1050	Trehalose-6-phosphate synthase component TPS1 and related subunits
		osmotic balance
56474	KOG0223	Aquaporin (major intrinsic protein family)
83345	KOG0224	Aquaporin (major intrinsic protein family)
92283	KOG0224	Aquaporin (major intrinsic protein family)
95824	KOG0223	Aquaporin (major intrinsic protein family)
112306	KOG0224	Aquaporin (major intrinsic protein family)
119849	KOG0223	Aquaporin (major intrinsic protein family)
124356	KOG0224	Aquaporin (major intrinsic protein family)
128807	KOG0224	Aquaporin (major intrinsic protein family)

[1]Protein ID numbers in bold font correspond to genes with transcripts accumulating >two-fold in ball-milled aspen versus glucose (GSE34636_si_Table1.xls).

VI. Environmental sensing

Regulation of gene expression under different environmental conditions represents a specific reaction of the fungus to the environment which has evolved to optimize fitness and competitiveness in nature. Also expression of plant cell wall degrading enzymes reflects a reaction to extracellular signals. The receptors of the heterotrimeric G-protein pathway as well as membrane-bound members of the two component phosphorelay pathway are among the best candidates for reception of signals triggering expression of degradative enzymes in *C. subvermispora.*

VIa. Heterotrimeric G-protein signalling in *C. subvermispora*

Heterotrimeric G-protein signalling represents one of the major pathways for reception and transmission of extracellular signals. For the prolific cellulase producer *Trichoderma reesei*, an involvement of two G-protein alpha subunits in the regulation of cellulase gene expression has been reported (53, 54). Moreover, considerable light dependent regulation of numerous glycoside hydrolase genes by the G-protein beta and gamma subunit, as well as their regulator PhLP1 (encoding a class I phosducin-like protein) has been observed (D. Tisch and M. Schmoll, manuscript in preparation). We therefore investigated whether the capability or preference for degradation of cellulose would be reflected in the genomes of *C. subvermispora*, *P. chrysosporium* and *P. placenta*.

i. G-protein coupled receptors As previously detected in *P. placenta* (27), *C. subvermispora* contains multiple *STE3*-type pheromone receptor genes clustered in one locus, albeit with five genes (Cesubv54303, Cesubv116643, Cesubv116644, Cesubv54036 and Cesubv116650) there are fewer than in *P.placenta*. However, only the latter two (Cesubv54036 and Cesubv116650) have the required seven transmembrane domains, while the others only have six transmembrane domains.

Search for additional G-protein coupled receptors revealed only few genes to be related to those described for ascomycetes (55). We found that *C. subvermispora* contains two microbial opsins (Cesubv113708 and Cesubv112229), both related to *N. crassa* ORP-1 (NCU01735) and two mPR-type GPCRs, one related to *M. grisea* MG04679 (Cesubv113945) and one related to MG05072 (Cesubv66653).

Besides these clearly defined homologues (as confirmed by bidirectional blast searches), we found additional seven transmembrane domain proteins, for which a function as GPCR cannot be excluded. Their assignment to this group should be treated with caution. Here two genes (Cesubv110652 and Cesubv87441) with distant similarity to rhodopsin-like GPCRs (low similarity to InterPro domain IPR000276) were detected. Additionally, protein Cesubv158481 shows distant similarity to pth11-like GPCRs.

ii.G-protein alpha subunits The genome of *C. subvermispora* comprises nine genes encoding putative G-protein alpha subunits. Phylogenetic analysis with G-protein alpha subunits from several fungi, including the brown rot fungus *P. placenta* and *P. chrysosporium* revealed that these fungi possess orthologues of the three G-alpha subunits well characterized in

filamentous ascomycetes (55-57). The *C. subvermispora* proteins are Cesubv87056, Cesubv79105 and Cesubv118319.

Moreover, as already observed in *P. placenta* (27), *C. subvermispora* and *P. chrysosporium* have several more putative G-alpha subunits related to *Ustilago maydis* GBA4 (58). If the respective characteristic in terms of cellulose or lignin degradation would be reflected in the number and/or stucture of the major components of the heterotrimeric G-protein pathway, i.e. the G-protein alpha subunits, a characteristic phylogenetic clustering or corresponding numbers of G-alpha proteins encoded in the genome would be expected. Therefore this analysis should reflect more similar characteristics of *C. subvermispora* with *P. chrysosporium* than with *P. placenta*. However, the phylogenetic tree revealed no indications that this would be the case. The preference for degradation of cellulose or lignin is not reflected in the characteristics of the G-alpha proteins (Figure S6A).

iii.G-protein beta and gamma subunits Despite the presence of numerous proteins comprising G-protein beta-like WD40 repeats with relatively low similarity, only one gene model could be unequivocally identified to represent a G-protein beta subunit (encoding Cesubv87872).

Surprisingly, we found two putative G-protein gamma subunits to be encoded next to each other in the *C. subvermispora* genome (Cesubv82306 and Cesubv64347). The sequence similarity of these proteins is unexpectedly high considering the normally low sequence conservation of G-gamma subunits (55). Nevertheless, a duplication in this area seems unlikely since sequence identity at the nucleotide level only covers part of the protein and is only up to 68 %.

Interestingly, we detected a similar genomic locus for G-protein gamma subunit encoding genes in *P. chrysosporium* (Phchr82306 and protein ID 72084 in v1.0). In *P. placenta* the same situation was found, although in the first genome annotation (27) only one G-protein gamma subunit was reported. However, this may be due to the fact that the second model was removed from the final gene catalogue and despite high similarity with the respective proteins of *C. subvermispora* and *P. chrysosporium* (tblastn search), no model is present in the homologous genomic area. Unfortunately the gene model shown from version 1 at this locus is no longer accessible. We conclude that all three fungi possess two G-protein gamma subunits.

iv. Regulator of G-protein signaling (RGS) Comparison of genes encoding RGS (regulator of G-protein) domains of *C. subvermispora*, *P. placenta* and *P. chrysosporium* revealed that in contrast to *T. reesei*, neither of them has a homologue of *A. nidulans* RgsA (59). However, in *C. subvermispora* two homologues of FlbA, which is involved in regulation of hyphal proliferation, development and biosynthesis of secondary metabolites in *A. nidulans* (60) were found. Interestingly, neither *C. subvermispora*, nor *P. placenta* and *P. chrysosporium* have a gene encoding a GprK-type G-protein coupled receptor (which comprises an RGS-domain). Therefore the respective homologues of *T. reesei* and *A. nidulans* were omitted in the phylogenetic analysis (Figure S6B).

v. Phosducin like proteins

In contrast to *T. reesei*, no class I phosducin-like proteins were found in *C. subvermispora, P. chrysosporium* or *P. placenta*. On the other hand, class III phosducin-like proteins were detected in these three basidiomycetes :

	class I	class II	class III
T. reesei	TR_58856	TR_80048	
C. subvermispora		Cesubv117248	Cesubv11373
P. placenta		Ppl_126398	Ppl_119112
P. chrysosporium		Pchr_891	Pchr_1497

While class I phosducin-like proteins have been found to act as cochaperones for folding of the G-protein beta and gamma subunits (61, 62), members of class II are reported to have no role in G-protein signaling, but assist in folding of proteins essential for regulation of cell cycle progression as well as actin and tubulin (62). Therefore it can be concluded that in these three basidiomycetes regulation of the efficiency of heterotrimeric G-protein signaling does not involve the function of phosducin-like proteins.

VIb. Two component phosphorelay systems (histidine kinases)

The genomic inventory of *C. subvermispora* with respect to two component phosphorelay systems largely resembles that of *P. placenta* (27). While all histidine kinases found in *P. placenta* also have a homologue in *C. subvermispora*, the response regulator receiver proteins (RR) comprise an additional member related to *S. cerevisiae* Ssk1p, which targets the osmoregulatory Hog1 MAPkinase pathway. The respective *N. crassa* homologue, RRG-1 is involved in control of vegetative cell integrity, hyperosmotic sensitivity, fungicide resistance and protoperithecial development through control of the osmosensitivity MAPkinase pathway (63). No evidence for duplication of the histidine phosphotranspherase (HPT), as hypothesized for *P. placenta,* was found.

class	**protein ID**	**transmembrane domains**	**putative function**
I	Cesubv52469	5	
III	Cesubv15951	none	putative osmosensor
VI	Cesubv17595	none	
VI	Cesubv18540	4	
VIII	Cesubv11992	none	putative phytochrome
X	Cesubv10006	none	
RR	Cesubv11391	related to Rim15p	
	Cesubv52649	related to Skn7p	
	Cesubv11187	related to Ssk1p	
	Cesubv94747	related to Ssk1p	
HPT	Cesubv81109		

VII. Oxidative phosphorylation (OXPHOS)

The identification of *C. subvermispora* nuclear genes coding for mitochondrial proteins that participate in the OXPHOS pathway (Complexes I-V, alternative oxidases and alternative NAD(P)H dehydrogenases) was done using a similar approach to that described previously [27, 64]. The *C. subvermispora* nuclear genome encodes at least 60 OXPHOS subunits: the seven central subunits of the eukaryotic Complex I core (NDUFS1, NDUFS2, NDUFS3, NDUFS7, NDUFS8, NDUFV1 and NDUFV2), 20 Complex I accessory subunits (NDUFS4, NDUFS6, NDUFA1/MWFE, NUDFA2, NDUFA4, NDUFA5, NDUFA6, NDUFA8, NDUFA9, NDUFA11, NDUFA12/DAP13, NDUFA13/GRIM19, NDUFAB1/ACP, NDUFB7, NDUFB8, NDUFB9, NDUFB11, NUXM, NUZM and NI9M), an alternative oxidase, three alternative NAD(P)H dehydrogenases, the four Complex II subunits (SDH1-SDH4), Complex III core proteins QCR1 and QCR2, the cytochrome c_1 (CYT1), the Ryeske iron-sulfur protein (RIP1), fourComplex III additional subunits (QCR6-QCR9), a cytochrome *c* (CYTC), six Complex IV subunits (COX4, COX5A, COX5B, COX6A, COX6B and COX9), eight Complex V essential subunits (ATP1, ATP2, ATP3, ATP4, ATP5, ATP7, ATP16 and ATP17), and two Complex V additional subunits (ATP18 and ATP20).

Comparison of the repertoires of OXPHOS proteins encoded in nuclear genes of basidiomycetes shows that they are highly similar. However, there are some distinctive OXPHOS subunits in the basidiomycetes: NDUFA4 appears to be a Complex I accessory subunit specific to basidiomycetes and zygomycetes [64, 65]; and the accessory subunits NURM, NUVM and 10.4 are fungus-specific subunits of Complex I [66] but they are absent from all the basidiomycetes. In addition, all the nuclear genomes of basidiomycetes lack several other OXPHOS subunits: the Complex I accessory subunit NUWM (excepting *P. placenta*) (listed below), and the Complex V essential subunit ATP14 and additional subunits ATP19, STF1 and STF2 [27]. No significant up- or down-regulation of the OXPHOS genes was observed in the conditions analyzed except the genes encoding the alternative oxidase (Cesubv92418), one alternative NAD(P)H dehydrogenases (Cesubv118779) and ATP1 (Cesubv112572) that are significantly up-regulated in media containing glucose (exhibiting >two-fold transcript accumulation on glucose versus ball-milled aspen) (GSE34636_si_Table1.xls1).

Regulated expression of OXPHOS proteins in *C. subvermispora*

Subunit	Protein ID	Fold change Glu/BMA	*P* value Glu/BMA
Complex I			
NDUFS1	–	–	–
NDUFS2	87840	0.93	2.86E-01
NDUFS3	–	–	–
NDUFS4	81217	0.97	6.17E-01
NDUFS6	120929	1.01	8.72E-01
NDUFS7	114465	1.16	1.05E-01
NDUFS8	62472	1.27	3.31E-02
NDUFV1	114392	0.88	1.05E-01
NDUFV2	55069	1.14	2.85E-01
NDUFA1/MWFE	–	–	–
NDUFA2	161771	1.00	9.67E-01
NDUFA4	111569	1.39	1.33E-01
NDUFA5	110397	1.04	4.87E-01
NDUFA6	139986	0.97	6.17E-01
NDUFA8	111175	1.09	2.52E-01
NDUFA9	112285	1.57	4.83E-02
NDUFA11	–	–	–
NDUFA12/DAP13	–	–	–
NDUFA13/GRIM19	111080	1.16	2.22E-01
NDUFAB1/ACP	110769	1.44	3.85E-02
NDUFB7	–	–	–
NDUFB8	–	–	–
NDUFB9	44083	1.23	1.57E-01
NDUFB11	–	–	–
NUXM	–	–	–
NUZM	154260	0.98	8.70E-01
NI9M	84814	1.00	9.70E-01
Alternative oxidases	92418	3.62	2.82E-04
Alternative NADP(H) dehydrogenases[a]	113234	1.14	1.40E-01
	114957	1.01	8.66E-01
	118779	2.37	6.78E-04
Complex II			
SDH1	117293	1.75	4.16E-03
SDH2	84567	1.27	1.03E-01
SDH3	117293	1.75	4.16E-03
SDH4	160305	1.17	2.47E-01
Complex III			
CYT1	81697	1.17	9.67E-02
RIP1	110112	1.74	1.84E-02
QCR1	111982	1.62	1.61E-02
QCR2	85735	1.31	1.46E-01
QCR6	125105	1.37	9.62E-02
QCR7	–	–	–
QCR8	51699	1.05	4.45E-01
QCR9	–	–	–
Cytochrome *c*			
CYTC	88982	1.52	1.56E-02

Complex IV			
COX4	82331	1.64	5.44E-02
COX5A	143623	0.98	7.83E-01
COX5B	26055	1.34	2.18E-01
COX6A	110532	1.10	6.00E-01
COX6B	117921	1.22	1.61E-01
COX9	81752	1.10	3.55E-01
Complex V			
ATP4	113003	1.04	4.80E-01
ATP5	114086	1.43	1.87E-01
ATP7	80618	1.22	7.79E-02
ATP17	80845	1.20	1.91E-01
ATP1	112572	2.15	5.38E-03
ATP2	113308	1.88	1.23E-02
ATP3	83893	1.26	1.67E-01
ATP16	113978	1.33	1.32E-01
ATP18	48950	1.21	3.23E-01
ATP20	–	–	–

VIII. P450 genes of *C. subvermispora*

Summary

In Basidiomycota, cytochrome P450 monooxygenases are known to play important role in lignin and xenobiotic degradation. The genome of the white rot basidiomycete *C. subvermispora* features 222 P450 genes, a number that is closer to that reported for the brown rot basidiomycete *P. placenta* (250 P450s) and greater than that for the model white rot basidiomycete *P. chrysosporium* (149 P450s). Among the 222 P450s, 205 were considered as authentic P450s whereas the remaining 17 P450s were grouped as tentative P450s as the latter showed only one of the two conserved P450 signature domains in the available sequence length. The authentic P450 group could be classified into 11 known fungal clans (CYP7, CYP51, CYP52, CYP53, CYP54, CYP61, CYP64, CYP67, CYP505, CYP534, and CYP547) and one un-assigned clan (UA). Among the clans, CYP64 showed the highest number P450s (106 P450s) whereas clans CYP7 and CYP61 showed the lowest (1 P450 each). The authentic P450s were classified into 32 families and 47 subfamilies. In comparison to the *P. chrysosporium* P450ome*, C. subvermispora* showed two new clans (CYP7 and the un-assigned clan) and expansion of member P450s for clans CYP51, CYP53, CYP64, CYP534, and CYP547 and reduction of member P450s for clans CYP52, CYP67, and CYP505. CYP61 clan remained the same with a single P450 between the two basidiomycete species. Family-level comparison showed eight new families **(**CYP66, CYP537, CYP638, CYP5027, CYP5046, CYP5065, CYP5222, and CYP6001**)** and expansion of several families viz. CYP63, CYP5137, CYP53, CYP5143, CYP512, CYP502, CYP5037, CYP5144, CYP5152, CYP5158, CYP5139, and CYP5136. Interestingly, the *C. subvermispora* genome showed extensive clustering of the P450 genes, with a total of nine genomic clusters spread across eightdifferent genome scaffolds. Three P450 genes were differentially up-regulated, whereas five genes were down-regulated when grown on ball-milled aspen (BMA) versus glucose as the sole carbon sources (GSE34636X_si_Table1). Overall, it appeared that the entire P450ome in this fungus is expressible but is not tightly regulated by the nature of the carbon source.

Introduction

Cytochrome P450 monooxygenases, also known as CYP enzymes, are heme-thiolate proteins that are spread across all biological kingdoms. Until now, more than 18,000 P450 genes were known, of which around 12,000 have been classified based on the existing P450 nomenclature rules (67). Cytochrome P450 enzymes perform a wide range of biological reactions including both biosynthesis and biodegradation reactions. While P450s have several known functions in fungi, in general, the P450 proteins/enzymes in the Basidiomycota are known to be involved in lignin and xenobiotic degradation. Basidiomycetes are characterized by an extraordinarily large P450 contingent (designated P450ome) ranging from ~150 P450s in the white rot fungus *P.chrysosporium* to ~ 250 P450s in the brown rot fungus *P. placenta* (27, 68). Here we report the identification, annotation, and phylogenetic classification of the P450ome in the white rot basidiomycete *C. subvermispora*.

Methodology

Initial determination of the putative cytochrome P450 gene models in C. *subvermispora* genome was made by searching the JGI whole genome database for the 'P450' hits. The resulting sequences were subjected to BLAST analysis and searched for the presence of the conserved P450 signature domains. P450s that showed both the oxygen- binding and the heme-binding domains were considered authentic P450s. Truncated P450 sequences were manually annotated for authenticity. Genescan was used to extend the questionable gene models in terms of the upstream and/or down stream regions to detect the presence of the P450 domains. The mRNA-to-genomic DNA alignment program Spidey was used to predict the intron-exon junctions. The P450s that showed only one of the P450 signature domains were considered as tentative P450s. Authentic P450s were grouped into families and subfamilies based on the existing International Nomenclature Committee criteria of > 40% homology for assigning a family and > 55% for a subfamily. P450 Superfamily Nomenclature rules were followed for assigning the family, sub-family and clan classification as earlier applied for P450ome classification in the model white-rot fungus, *P. chrysosporium*. P450s that did not have *P. chrysosporium* P450 homologues were annotated based on the phylogenetic alignment with other P450s on the phylogenetic tree (Fig. S5); the tree was constructed using Mega 4 software. P450s showing both the conserved domains and a reasonable deduced protein length (≥ 330 aa) were used for the tree construction.

Results and Discussion

P450 gene identification, phylogenetic analysis and genome organization:

The *C. subvermispora* genome yielded 230 P450 hits. Among these hits, eight sequences showed homology to non-P450s and were dropped from further analysis. Of the remaining 222 putative sequences, 183 sequences that contained both the P450 signature domains were considered as authentic P450s. From the remaining 39 putative P450 sequences, we successfully deduced the P450 signature domains for 22 sequences by manual annotation and grouped these sequences under authentic P450s. Hence the total count of authentic P450s increased to 205. Seventeen sequences showed one of the P450 signature domains, and hence, these sequences were classified as tentative P450s. Taken together, our analysis showed that the *C. subvermispora* P450ome consists of 222 P450s.

The authentic P450s could be classified into 11 known fungal clans (CYP7, CYP51, CYP52, CYP53, CYP54, CYP61, CYP64, CYP67, CYP505, CYP534, and CYP547) and one un-assigned clan (UA). Among the clans, CYP64 showed the highest number (106 P450s), whereas, clans CYP7 and CYP61 showed the lowest (1 P450 each). The authentic P450s (205) were subjected to family and sub-family classification. Using the existing P450 nomenclature criteria, these P450s could be grouped into 32 families and 47 sub-families (Fig. S5). Among the 32 families, the CYP5144 family contained the

highest number of member P450s (56 P450s) followed by CYP5158 (17 P450s), CYP5037 (15 P450s) and CYP5139 (14 P450s).

The *C. subvermispora* genome showed extensive clustering of P450 genes. Specifically, there are nine genomic clusters spread across eight different genome scaffolds namely 1, 6, 7, 9, 16, 19, 25 and 31 (listed below). The genomic clusters contained four to eight P450 members per cluster and spanned a region of 13 Kb (scaffold 1) to 3.2 Mb (scaffold 7). Genome-wide analysis revealed that the P450 genes contained as high as 27 introns, with the majority of gene models containing nine to 13 introns. The P450 GC content varied from 47% to 64%.

Differential expression of P450 genes

Microarray analysis of the *C. subvermispora* mycelia grown using glucose or ball-milled aspen (BMA) wood as the sole carbon source revealed expression of all P450 genes albeit with no dramatic differential fold changes (GSE34636_si_Table1). The majority of the P450 genes remained within the cut-off limit of ±2.0 fold change except eight P450 genes that showed differential regulation to an extent. Of the differentially expressed genes, three genes showed statistically significant up-regulation (2.0 to 2.4 fold change), whereas five genes showed down-regulation under the BMA growth conditions (-0.31 to -0.46 fold change). Overall, it appeared that the P450ome in this fungus is actively expressed but is not tightly regulated by the nature of the carbon source (wood polysaccharides versus glucose).

Comparative P450omics

In comparison to the model white rot (*P. chrysosporium*) genome that contains 149 P450s, *C. subvermispora* showed a higher number of P450s (222 P450s). *C. subvermispora* showed two new clans (CYP7 and un-assigned clan) and expansion of member P450s for the clans CYP51, CYP53, CYP64, CYP534, and CYP547 and reduction of member P450s for the clans CYP52, CYP67, and CYP505. CYP61 clan remained the same with a single member P450 between the two basidiomycete species. Although the P450ome is groupable into 32 families (listed below) in both white-rot fungi, differences were observed in the families. Both the genomes showed 24 common families. Therefore, eight new families **(**CYP66, CYP537, CYP638, CYP5027, CYP5046, CYP5065, CYP5222, and CYP6001**)** were observed in the *C. subvermispora* P450ome. In addition, *C. subvermispora* genome showed an expansion of member P450s in certain families, viz, CYP63, CYP5137, CYP53, CYP5143, CYP512, CYP502, CYP5037, CYP5144, CYP5152, CYP5158, CYP5139, CYP5136.

P450ome classification in *C. subvermispora*

Clan	Family	Sub-family	No. of P450s sub-family	family	clan
CYP7	CYP638	NS	1	1	**1**
CYP51	CYP51	F	1	1	**14**
	CYP5156	A	1	1	
	CYP63	A	8	12	
		B	1		
		C	3		
CYP52	CYP5137	A	4	4	**8**
	CYP5141	A	2	2	
	CYP5151	A	1	2	
		B	1		
CYP53	CYP53	C	4	4	**12**
	CYP537	B	2	2	
	CYP5140	A	1	1	
	CYP5143	A	4	5	
		NS	1		
CYP54	CYP512	A	9	11	**11**
		B	1		
		G	1		
CYP61	CYP61	A	1	1	**1**
CYP64	CYP5065	A	2	3	**106**
		NS	1		
	CYP66	A	1	1	
	CYP502	A	4	5	
		B	1		
	CYP5037	A	1	15	
		B	12		
		NS	2		
	CYP5046	A	1	1	
	CYP5144	A	2	56	
		C	44		
		D	1		
		F	6		
		J	1		
		NS	2		
	CYP5027	A	2	2	
	CYP5152	A	6	6	
	CYP5158	A	17	17	
CYP67	CYP5035	A	6	9	**9**
		B	3		

CYP505	CYP505	D	4	4	**4**
CYP534	CYP5138	NS	1	1	**25**
	CYP5139	A	8	14	
		NS	6		
	CYP5150	A	8	8	
	CYP5154	A	1	1	
	CYP5155	A	1	1	
CYP547	CYP5136	A	12	12	**12**
UA	CYP5222	NS	1	1	**2**
	CYP6001	NS	1	1	
		Authentic P450s			**205**
		Tentative P450s			**17**
		Total No. of P450s			**222**

Abbreviations: NS, New-subfamily; UA, unassigned

Genomic cluster analysis of cytochrome P450 genes in *C. subvermispora*

Cluster No.	Scaffold	No. of P450s	No. of families in each cluster	Clan(s)
1	1	4	1	CYP54
2	6	5	3	CYP64 and CYP547
3	6	6	1	CYP64
4	7	5	3	CYP64 and CYP547
5	9	8	2	CYP53
6	16	6	2	CYP547
7	19	7	6	CYP52 and CYP64
8	25	5	3	CYP547 and CYP64
9	31	4	3	CYP67 and CYP64

IX. Natural Products and Secondary metabolism

The overwhelming majority of natural product metabolites are created by either terpene cyclases or multidomain enzymes, i.e., polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs), or hybrids thereof. Various classes of enzymes, such as monooxygenases, halogenases, and methyltransferases, summarized as tailoring enzymes, diversify these primary natural product compounds.

Few studies have investigated *C. subvermispora* as a producer of secondary metabolites. Except ceriporic acid and its derivatives (69, 70) which represent fatty acid-like compounds and apparently interfere with the Fenton chemistry, no reports pertaining to secondary metabolism exist. The minimal set of *Ceriporiopsis* metabolites, identified by chemical means, is contrasted by multiple putative secondary metabolite pathways, encoded in the *C. subvermispora* genome. Based on the analysis of the genomic data, we have annotated 42 genes putatively related to secondary metabolism. In particular, we expect a rich terpene metabolism, along with polyketide- and peptide-derived products. We identified genes for two PKSs (one non-reducing), three NRPS-like enzymes, and one PKS/NRPS-hybrid (listed below; domain organization shown in Figure S6C). We expect two products of NRPS genes participate in ´ primary metabolism, as they resemble a Lys2-like α-aminoadipate semialdehyde reductase (*nps3*) typically involved in L-lysine biosynthesis, and a siderophore-synthesizing enzyme (*nps2*). An impressive total of 14 genes for terpene cyclases were found, some or all of which may catalyze secondary product formation.

Putative siderophore synthesis locus - The product of the putative siderophore gene *nps2* deviates in its domain setup from the canonical SidC/ferrichrome synthetase architecture in that the first two modules (A-PCP-C) are absent. Typically, ferrichrome synthetases include at least two functional A-domains (71) which select, activate and load the monomeric building blocks onto the enzymatic assembly line. Addtional enzymatic activites needed to complete ferrichrome synthesis are monooxygenase and acetyltransferase. The genetic basis has been shown with *Omphalotus olearius* (72). In *C. subvermispora*, a monooxygenase gene was found next to *nps2*. Unlike in *Omphalotus*, the acetyltransferase gene is absent. A situation similar to *C. subvermispora* is found in the *Coprinopsis* genome. Plausible explanations include that the single A-domain in Nps2 is very unspecific and can activate all monomers necessary for ferrichrome synthesis. Alternatively, *Ceriporiopsis* siderophores may deviate structurally from the ferrichromes and can be assembled with just one A-domain.

Tailoring enzymes - Numerous reading frames were identified that may encode tailoring enzymes. As some of them are located in the vicinity of NRPS and PKS genes it appears *Ceriporiopsis* follows the biosynthesis gene cluster paradigm, as is the case for numerous other filamentous fungi [(73) and references therein]. We also found transporter genes clustered with natural product genes. These transporters may function as molecular pumps to secrete the metabolites into the extracellular space and thus, confer resistance to the producer. Halogenated natural products have not yet been described from

Ceriporiopsis, however, two genes (*hal1* and *hal2*), which very likely code for flavin-dependent halogenases, indicate the capacity to synthesize as yet unknown halogenated products.

Natural product biosynthesis is subject to complex regulatory cascades, so the realm of *C. subvermispora* metabolic abilities may not be fully explorable under laboratory conditions. Still, based on the genomic evidence we expect a diverse secondary metabolome which will help understand the fungus´ chemical ecology, and may represent a new source of biologically active compounds.

Putative natural product genes in the *C. subvermispora* genome based on sequence homology.

Predicted protein function	**Name**	**Protein ID**	**Scaffold**	**Nucleotide range**	**Protein length (aa)**	**Domain architecture**	**Notes**	**Clusters with**
PKS-NRPS-Hybrid	*pks1*	77565	22	325094-336915	3587	KS-AT-DH-ER-KR-ACP-C-A-PCP-TE		*mat3*
Nonribosomal synthetase	*nps1*	71694	3	1323048-1327101	1049	A-PCP-Red		
	nps2	172109	5	1057757-1065923	2434	A-PCP-C-PCP-C-PCP-C	*sidC* homolog siderophore biosynthesis	*smo1*
	nps3	107385	1	846668-851180	1417	(C)-A-PCP-Red	L-aminoadipiate semialdehyde reductase, *lys2* homolog	*deh1, acl5*
Polyketide synthase	*pks2*	38187	1	1385018-1392048	2031	SAT-KS-AT-PT-ACP-TE	Noncanonical (GC-AG) intron expected	*mat1, mat2*
	pks3	96782	10	1380477-1389343	2519	KS-AT-DH-KR-ACP-Red		*smo2*
Halogenase	*hal1*	75071	1	773806-775951	524		Flavin dependent halogenase	*omt7, pal1, acl6*
	hal2	68862	2	433357-435539	529		Flavin dependent halogenase 63% identical aa with hal1	
Transporter	*mat1*	110134	1	1368770-1375971	1615			*mat2*

	mat2	110141	1	1376809-1383959	1596			
	mat3	118967	22	340622-348053	1983			
CoA-Ligases	*acl1*	64238	4	2292299-2295626	588			
	acl2	81293	3	1743589-1745969	570			
	acl3	29600	11	291100-293344	680			
	acl4	139051	10	357717-364243	563			
	acl5	116973	12	841184-844032	429			
	acl6	116230	10	767328-769837	569			
Monooxygenase	*smo1*	113443	5	1066864-1068914	542		flavin dependent enzyme	
	smo2	85632	10	1376693-1379133	520		cytochrom P450 enzyme	
Dehydrogenase	*deh1*	116975	12	844667-846349				
O-Methyl-transferase	*omt1*	151553	4	429422-431966	464			
	omt2	152302	4	1640057-1641967	463			
	omt3	92888	4	434836-436308	342			
	omt4	47409	4	1060570-1062419	465			
	omt5	114284	6	1352972-1355357	466			
	omt6	151192	3	2377538-2379552	470			
	omt7	116232	10	776925-778999	465			
Phenylalanine-amonia lyase	*pal1*	116229	10	761014-766319	733			*omt7, acl6*
Terpene synthase								
	ter1	63236	3	628413-629525	310			
	ter2	113927	6	111363-114094	323			*ter7*
	ter3	83368	6	1290967-1293995	322			*ter5*
	ter4	78286	30	249385-250557	330			
	ter5	83362	6	1281068-1282200	323			
	ter6	162851	28	22342-23768	334			*ter10*

	ter7	105307	6	103652-105138	293			
	ter8	74510	8	1561738-1564339	444			*ter9*
	ter9	66140	8	1576387-1577580	326			
	ter10	162846	28	17520-18711	316			
	ter11	69301	28	220342-221448	309			
	ter12	78106	28	143447-144576	332			*ter13*
	ter13	78107	28	147556-148639	293			
	ter14	116142	10	353532-356289	287			*acl4*

Abbreviations for polyketide synthase or nonribosomal peptide synthetase domains are:
KS: keto synthase
AT: acyl transferase
DH: dehydratase
ER: enoyl reductase
KR: keto reductase
ACP: acyl carrier protein
TE: thioesterase
SAT: starter unit: ACP-transacylase
PT: product template domain
C: condensation domain
A: adenylation domain
PCP: peptidyl carrier protein
Red: reductase domain

X. RNA Silencing Proteins in *Ceriporiopsis subvermispora*

RNA interference (RNAi) is a naturally occurring post-transcriptional gene-silencing (PTGS) phenomenon, in which sequence-specific double-stranded RNA (dsRNA) triggers the degradation of homologous mRNA, thereby reducing gene expression. Within the fungal kingdom, the mechanistic aspects of RNAi were first thoroughly analyzed in *N. crassa* (74-77). Argonaute, Dicer and RNA-dependent RNA polymerase (RdRP) have been identified as the three fundamental components of the RNA-silencing machinery. Since then, RNAi machinery has been identified in a wide range of fungi including ascomycetes, basidiomycetes, and zygomycetes, many of which harbor multiple RNA-silencing components in the genome, whereas a portion of ascomycete and basidiomycete fungi apparently lack the entire array of components or some of them (78-80).

The automatically-annotated genome of *C. subvermispora* (http://genome.jgi-psf.org/Cersu1/Cersu1.home.html) was screened for candidate genes encoding orthologues of the three fundamental components of the RNA-mediated gene silencing machinery (Argonaute, Dicer and RNA-dependent RNA polymerase (RdRP)), using similar parameters previously defined by Nakayashiki *et al.* (78) and Salame *et al.* (79). BLAST probing of the *C. subvermispora* genome databases, was performed with the tblastn program using conserved protein sequences for the Piwi domain in Argonaute – *N. crassa* Qde-2 (NCU04730) and Sms-2 (NCU09434), for the RNase III domain in Dicer – *N. crassa* Dcl-1/Sms-3 (NCU08270) and Dcl-2 (NCU06766), and for the RdRP domain in RdRP – *N. crassa* Qde1 (NCU07534) and RRP3 (NCU08435). The results, listed below, demonstrate the presence of the silencing machinery in the *C. subvermispora* genome.

RNA silencing proteins in *C. subvermispora* genome

Predicted protein	Model name*	Protein ID	Genome coordinates (Location on scaffold)	Best Hit (NCBI blastp)	E-Value
Argonaute	estExt_fgenesh1_kg.C_70432	114999	scaffold_7:1596003-1599809	gi|170095037|ref|XP_001878739.1| argonaute-like protein [Laccaria bicolor S238N-H82] >gi|164646043|gb|EDR10289.1| argonaute-like protein [Laccaria bicolor S238N-H82]	0.0
	estExt_fgenesh1_kg.C_90032	115695	scaffold_9:136607-139353	gi|170116966|ref|XP_001889672.1| argonaute-like protein [Laccaria bicolor S238N-H82] >gi|164635387|gb|EDQ99695.1| argonaute-like protein [Laccaria bicolor S238N-H82]	0.0
Dicer	estExt_fgenesh1_kg.C_30635	112373	scaffold_3:2307787-2313888	gi|170099920|ref|XP_001881178.1| predicted protein [Laccaria bicolor S238N-H82] >gi|164643857|gb|EDR08108.1| predicted protein [Laccaria bicolor S238N-H82]	0.0
RNA-dependent RNA polymerase (RdRP)	estExt_fgenesh1_pg.C_200084	126709	scaffold_20:339143-344506	gi|169845505|ref|XP_001829472.1| predicted protein [Coprinopsis cinerea okayama7#130] >gi|116509537|gb|EAU92432.1| predicted protein [Coprinopsis cinerea okayama7#130]	0.0
	estExt_fgenesh1_pm.C_180016	108431	scaffold_18:81222-87375	gi|170117206|ref|XP_001889791.1| RNA-directed RNA polymerase [Laccaria bicolor S238N-H82] >gi|164635257|gb|EDQ99567.1| RNA-directed RNA polymerase [Laccaria bicolor S238N-H82]	0.0

* Complete transcript evidence was confirmed for all gene models indicated.

XI. *C. subvermispora* autophagy-related and nitrogen sensing genes

Forest-inhabiting wood decay saprophytes have evolved to exploit habitats where nutrition is unevenly distributed. On the discovery of a new food source resources are remobilised from under-supplied parts of the colony to exploit this new resource. Nutrient sensing and autophagy are important processes in the scavenging morphology of wood decay basidiomycetes.

Autophagy is a process conserved in eukaryotes for the degradation and recycling of cellular components and has been implicated in responses to starvation and other stresses, pathogenesis and development of multicellular structures (81, 82). Autophagy occurs in double-membraned vesicles called autophagosomes. Studies in *S. cerevisiae* have identified approximately 30 genes necessary for autophagy (AuTophaGy-response genes ATG) and regulatory kinase induction pathways, e.g. TOR (83, 84). Autophagy may also be selective, such as in the degradation of peroxisomes, endoplasmic reticulum or mitochondria. In yeast a selective autophagy known as biosynthetic cytoplasm-to-vacuole targeting (Cvt) pathway is described, but has not been identified in filamentous fungi.

Nitrogen availability is limited in forest soils and amino acids are actively translocated through the fungal colony to ensure active regions are fully resourced (85). Signalling kinases associated with the TOR pathway and small GTPases related to the yeast gene Rhb1 are important elements in this process.

Analysis of the *C. subvermispora* genome and comparison with other sequenced basidiomycetes indicates the presence of a conserved non-selective autophagic pathway similar to that described in *S. cerevisiae* (83). Overall, the gene composition was very similar for all the basidiomycete genomes analysed and those ATG genes absent from *C. subvermispora* were also absent from other basidiomycetes (ATG10, 14, 19, 21, 23, 28, 29, 31). A previous study on autophagy genes during the infection of rice by *Magnaporthe grisea* showed that genes ATG14 and 31 were not necessary for maintaining autophagy in the blast fungus, while the absence of genes involved in selective autophagy (Cvt pathway), i.e. ATG 19, 20, 21 and 23, reinforced the view that such pathways were specific to yeasts and absent in filamentous fungi (82). The data from the basidiomycete genomes agree with these conclusions and presumably this is why ATG28 and 29 are also absent. Intriguingly, an ortholog of ATG10 thought to be necessary for autophagy was not identified in any basidiomycete genome and further functional studies are required to determine whether it is necessary for autophagy in basidiomycetes or has been replaced.

Orthologs of autophagy initiation and nutrient sensing genes (TOR, Vps34 and Rhb1) were identified in the genome of each basidiomycete fungus, with *C. subvermispora* containing two TOR genes in close proximity to one another on scaffold 4.

***S. cerevisiae* autophagy-related and nitrogen sensing genes present in the *C. subvermispora* genome**

Autophagy-related gene	Gene function*	*C. subvermispora* protein ID and scaffold location
ATG1	Protein serine/threonine kinase, required for autophagy and for the cytoplasm-to-vacuole targeting (Cvt) pathway	51841 7:882799-884297
ATG2	Peripheral membrane protein required for the formation of cytosolic sequestering vesicles involved in vacuolar import through both the Cvt pathway and autophagy; interacts with Atg9p and is necessary for its trafficking	113111 5:187298-195831
ATG3	Protein involved in autophagy; E2-like enzyme that plays a role in formation of Atg8p-phosphatidylethanolamine conjugates, which are involved in membrane dynamics during autophagy	114329 6:1546872-1548613
ATG4	Cysteine protease required for autophagy; cleaves Atg8p to a form required for autophagosome and Cvt vesicle generation; mediates attachment of autophagosomes to microtubules through interactions with Tub1p and Tub2p	172141 1:494112-497365
ATG5	Conserved autophagy-related protein that undergoes conjugation with Atg12p and then associates with Atg16p to form a cytosolic complex essential for autophagosome formation	113081 5:6865-11052
ATG6	Protein that forms a membrane-associated complex with Apg14p that is essential for autophagy; involved in a retrieval step of the carboxypeptidase Y receptor, Vps10p, to the late Golgi from the endosome; involved in vacuolar protein sorting	113949 6:189205-191519
ATG7	Autophagy-related protein and dual specificity member of the E1 family of ubiquitin-activating enzymes; mediates the conjugation of Atg12p with Atg5p and Atg8p with phosphatidylethanolamine, required steps in autophagosome formation	114575 7:24618-27508
ATG8	Protein required for autophagy; modified by the serial action of Atg4p, Atg7p, and Atg3p, and conjugated at the C terminus with phosphatidylethanolamine, to become the form essential for generation of autophagosomes	113797 5:2048298-2049551
ATG9	Transmembrane protein involved in formation of Cvt and autophagic vesicles; cycles between the pre-autophagosomal structure and other	82398 5:692725-696147

	cytosolic punctate structures, not found in autophagosomes	
ATG10	E2-like conjugating enzyme that mediates formation of the Atg12p-Atg5p conjugate, which is a critical step in autophagy	No hit
ATG11	Peripheral membrane protein required for delivery of aminopeptidase I (Lap4p) to the vacuole in the cytoplasm-to-vacuole targeting pathway; also required for peroxisomal degradation (pexophagy)	116294 10:1002446-1008040
ATG12	Ubiquitin-like modifier, conjugated via an isopeptide bond to a lysine residue of Atg5p by the E1 enzyme, Atg7p, and the E2 enzyme, Atg10p, a step that is essential for autophagy	52358 8:1461657-1462007
ATG13	Phosphorylated protein that interacts with Vac8p, required for the cytoplasm-to-vacuole targeting (Cvt) pathway and autophagy	115375 8:802384-805499 115163 8:252062-255456
ATG14	Subunit of an autophagy-specific phosphatidylinositol 3-kinase complex (with Vps34p,Vps15p, and Vps30p) required for organization of a pre-autophagosomal structure; ATG14 transcription is activated by Gln3p during nitrogen starvation	No hit
ATG15	Lipase, required for intravacuolar lysis of autophagic bodies; located in the endoplasmic reticulum membrane and targeted to intravacuolar vesicles during autophagy via the multivesicular body (MVB) pathway	117171 13:330576-332311 117477 14:321603-323269
ATG16	Protein that interacts with the Atg12p-Atg5p conjugate during formation of the pre-autophagosomal structure; essential for autophagy	110249 1:1707810-1709254
ATG17	Scaffold protein responsible for pre-autophagosomal structure organization; interacts with and is required for activation of Apg1p protein kinase; involved in autophagy but not in the Cvt (cytoplasm to vacuole targeting) pathway	107843 14:768067-770041
ATG18	Phosphatidylinositol 3,5-bisphosphate-binding protein of the vacuolar membrane, predicted to fold as a seven-bladed beta-propeller; required for recycling of Atg9p through the pre-autophagosomal structure	140175 12:301182-303228
ATG19	Protein involved in the cytoplasm-to-vacuole targeting pathway and in autophagy, recognizes cargo proteins and delivers them to the	No hit

	preautophagosomal structure for eventual engulfment by the autophagosome and degradation	
ATG20	Protein required for transport of aminopeptidase I (Lap4p) through the cytoplasm-to vacuole targeting pathway; binds phosphatidylinositol-3-phosphate, involved in localization of membranes to the preautophagosome, potential Cdc28p substrate	112741 4:1050167-1052610
ATG21	Similar to ATG18. Phosphatidylinositol 3,5-bisphosphate-binding protein required for maturation of pro-aminopeptidase I, predicted to fold as a seven-bladed beta-propeller; displays punctate cytoplasmic localization	More similar to ATG18
ATG22	Protein required for the breakdown of autophagic vesicles in the vacuole during autophagy, putative integral membrane protein that localizes to vacuolar membranes and punctuate structures attached to the vacuole	119991 32:78055-80610
ATG23	Peripheral membrane protein, required for autophagy and for the cytoplasm-to-vacuole targeting (Cvt) pathway	No hit
ATG24	Sorting nexin, involved in the retrieval of late-Golgi SNAREs from the post-Golgi endosome to the trans-Golgi network and in cytoplasm to vacuole transport; contains a PX domain; forms complex with Snx41p and Atg20p	110448 1:2338768-2341106
ATG26	UDP-glucose:sterol glucosyltransferase, conserved enzyme involved in synthesis of sterol glucoside membrane lipids, involved in autophagy	90212 1:1574429-1579696
ATG27	Type II membrane protein involved in autophagy; binds phosphatidylinositol 3-phosphate, required for the cytoplasm-to-vacuole targeting (Cvt) pathway	114489 6:2035545-2036861
ATG28	Involved in degradation of peroxisomes	No hit
ATG29	Protein specifically required for autophagy; may function in autophagosome formation at the pre-autophagosomal structure in collaboration with other autophagy proteins	No hit
ATG31	May form a complex with Atg17p and Atg29p that localizes other proteins to the pre-autophagosomal structure	No hit
COG2	Subunit of the conserved oligomeric Golgi complex that is required for maintaining normal structure and activity of the Golgi complex, interacts with the USO1 vesicle docking protein and may be necessary for normal Golgi ribbon formation and trafficking of Golgi enzymes.	50066 6:1423008-1426462
TOR	Protein kinase that integrates signals from several pathways to	93164

	regulate transcription/growth, implicated in the regulation of autophagy	4:1303637-1305908 81897 4:1325360-1334958
Vsp34	Class 3 phosphatylinositol 3-kinase which may act as a transducer of nutrient availability signals to TOR which in turn can activate several transcriptional changes	56755 16:675665-678833
Rhb1	Small GTPase involved in the amino acid sensing response	87377 16:916107-917784

*ATG Gene function description taken from http://yeastgenome.org/.

Comparison of autophagy-related gene compliment from sequenced basidiomycetes and *M. grisea*

Autophagy-related gene	CS	AB	CC	CN	HA	LB	PC	PP	SC	SL	UM	MG
ATG1	1	1	1	1	1	1	1	1	1	1	1	1
ATG2	1	1	1	1	1	1	1	1	1	1	1	1
ATG3	1	1	1	1	1	1	1	2	1	1	1	1
ATG4	1	1	1	1	1	1	1	1	2	1	1	1
ATG5	1	1	1	1	1	1	1	1	1	1	1	1
ATG6	1	1	1	1	1	1	1	1	1	1	1	1
ATG7	1	1	1	1	1	1	1	1	1	1	1	1
ATG8	1	1	1	1	1	1	1	1	1	1	1	1
ATG9	1	1	1	1	1	1	1	1	1	1	1	1
ATG10	0	0	0	0	0	0	0	0	0	0	0	1
ATG11	1	1	1	1	1	1	1	1	1	1	1	1
ATG12	1	1	1	1	1	1	1	1	1	1	1	1
ATG13	1	2	1	0	1	1	1	1	0	1	0	1
ATG14	0	0	0	0	0	0	0	0	0	0	0	0
ATG15	2	4	3	1	2	1	3	2	3	3	1	1
ATG16	1	1	1	1	1	1	1	1	1	1	1	1
ATG17	1	1	1	1	1	1	1	1	1	1	0	1
ATG18	1	1	1	1	1	1	1	1	1	1	1	1
ATG19	0	0	0	0	0	0	0	0	0	0	0	0
ATG20	1	1	1	1	1	1	1	1	1	1	1	0
ATG21	0	0	0	0	0	0	0	0	0	0	0	0
ATG22	1	1	1	1	1	1	1	1	1	1	1	1
ATG23	0	0	0	0	0	0	0	0	0	0	0	0
ATG24	1	1	1	1	1	1	1	1	1	1	1	1
ATG26	1	2	1	1	1	1	1	1	3	1	1	1
ATG27	1	1	1	1	1	1	1	1	1	1	1	1
ATG28	0	0	0	0	0	0	0	0	0	0	0	1
ATG29	0	0	0	0	0	0	0	0	0	0	0	1
ATG31	0	0	0	0	0	0	0	0	0	0	0	0
COG2	1	1	1	1	1	1	1	1	1	1	1	1
TOR	2	2	1	2	1	1	2	1	1	1	2	1

Vps34	1	1	1	1	1	1	1	1	1	1	1	1
Rhb1	1	1	2	1	1	2	1	1	2	2	1	2

CS = *Ceriporiopsis subvermispora* B; AB = *Agaricus bisporus*; CC = *Coprinopsis cinerea*; CN = *Cryptococcus neoformans*; HA = *Heterobasidion annosum*; LB = *Laccaria bicolor*; PC = *Phanerochaete chrysosporium*; PP = *Postia placenta*; SC = *Schizophyllum commune*; SL = *Serpula lacrymans*; UM = *Ustilago maydis*; MG = *Magnaporthe grisea.*

XII. Mating type

C. subvermispora has a bipolar mating system (86, 87). Like other basidiomycetes, *C. subvermispora* has two loci with putative mating type genes. These loci are present on two scaffolds (scaffold 1 and scaffold 11). Since they are apparently unlinked, one of the two loci is expected to have lost mating type function. It remains to be shown whether this is the putative *B* locus as reported in a few other Agaricomycetes (88, 89). Both loci structurally resemble mating type loci of other species, and they contain expressed genes for proteins whose sequences do not indicate any easy recognisable mechanism of self-compatibility.

A typical *A*-mating-type locus for two intact genes for HD1 and HD2 homeodomain transcription factors and one putative pseudogene is found on scaffold 1. The locus is flanked by the expressed gene *mip* (2381465-2378962) for a mitochondrial intermediate peptidase and an expressed gene *β-fg* (beta-flanking gene; 2368171-2369230) for a conserved fungal protein of unknown function. The putative *A* mating type locus contains one divergently transcribed gene pair for an HD1 and an HD2 homeodomain transcription factor (*a1-1*, position 2372117-2369766; *a2-1*, 2372612-2374566). The *HD1* gene *a1-1* has six introns, but EST sequences suggest alternative splicing of transcripts. Only when all six introns are spliced, a full length protein of 674 aa (ID 102296) will be generated. Proteins from not fully spliced transcripts would have a length of 344 aa and 381 aa, respectively. All have a complete TALE-type homeodomain with three extra aa in between helix 1 and helix 2 and the atypical WFKTT motif in the third helix of the homeodomain. Moreover, they all have a potential bipartite nuclear localization signal (KRKRGICDDSRRVAFRRR). The *HD2* gene *a2-1* has three introns and its product (ID 110462) is 602 aa long with the classical DNA-binding motif WFNQR in the helix 3 of the homeodomain. The position of the *a1-1*/*a2-1* gene pair relative to the *mip* and the *β-fg* genes is inversely oriented compared to most other Agaricomycetes species, with the exceptions of *P. chrysosporium*, *P. placenta* and *Flammulina velutipes* (88, 90). Although being transcribed as suggested from EST sequences, the third gene within the *A* mating type locus (2375939-2375233) is possibly a non-functional relic of an *HD1* gene. The first 137 aa of its 210 aa short protein product (ID 110463) has 43-58 % similarity (28-34% identity) to the N-terminal recognition domains of HD1 proteins of *Serpula lacrymans*, *P. chrysosporium* and *Laccaria bicolor*. A homeodomain is not present in this protein. It is not possible to create from the DNA-sequence a model for a protein being longer than the N-terminal recognition motif. This probable pseudogene is transcribed in the same direction as its neighboring gene *mip* and is thus also inversely orientated relative to *HD1* genes next to *mip* genes in most other Agaricomycetes.

A larger region for potential *B* mating type genes is located on scaffold 11 (1014004-1085135). Pfam HMM alignments recognized seven regions for fungal mating pheromones (PF08015). For all, genes had to be manually defined even though for four of them, there is evidence for expression by EST sequences. Furthermore, five other pheromone genes were newly defined under support of EST sequences. The predicted pheromone precursors of five genes at the one border of the region (CsPh1 to CsPh5) and the predicted pheromone

precursor of a gene located close to the other end of the region (CsPh12) can be grouped by higher similarity (Fig. S6D). Some of these are depicted in BlastX searches at lowest stringency by members of the large family of non-mating-type pheromone-precursors found in *Coprinopsis cinerea* (89). Evidence for genes encoding pheromone-precursors outside the *B* mating type-like locus was not found in this analysis. The predicted pheromone precursors of the remaining six genes (CsPh6 to CsPh11) are more divergent in their sequences as is typical for pheromone precursors encoded in *B* mating type loci (89). Still, CsPh7 and CsPh9 are very similar to each other as well as CsPh8 and CsPh10 (Fig. S6D). In the *B* mating-type-like locus of *C. subvermispora* there are also six genes for G-protein-coupled seven-transmembrane pheromone receptors of the *STE3*-type; outside of this region there are none. The genes for CsPh6 to CsPh11 intermingle with the genes of pheromone receptors. Gene *CsSTE3.1* has one downstream pheromone gene (*CsPh6*), *CsSTE3.2* has one upstream (*CsPh7*) and one downstream pheromone gene (*CsPh8*), *CsSTE3.3*, *CcSTE3.4, CcSTE3.5*, and *CcSTE3.6* have each one upstream pheromone gene (*CsPh9*, *CsPh10*, *CsPh11*, and *CsPh12*, respectively). The order of pheromone receptor genes partially resembles the order of the gene in the *B* mating type locus of *C. cinerea* (89, 91). CsSTE3.1 is most similar to *C. cinerea* STE3.1, CsSTE3.2 and CsSTE3.4 to *C. cinerea* STE3.2a, CsSTE3.3 to *C. cinerea* STE3.3, CsSTE3.5 to *C. cinerea* STE3.2b and CsSTE3.5 to *C. cinerea* STE3.2151. The gene of the later has in *C. cinerea* no linked pheromone precursor gene, is located about 30 kb downstream of the *B* mating type locus and has possibly no mating type function (89, 91). In between in the *C. cinerea B* locus, among other genes are two genes for multidrug transporters of the multi-facilitator superfamily (91). These two genes are also found in *C. subvermispora* but here at the outer border of the *B* mating type-like region beyond gene *CsSTE3.5*. In conclusion, if all pheromone receptor genes with linked pheromone precursor genes would be active, duplications of modules must have happened as compared to *C. cinerea*. In the normal *B* mating type situation, products from the same locus will not interact with each other but products from genes from allelic loci (89). The similarity of certain pheromone precursors and of certain pheromone receptors in *C. subvermispora* calls for the question whether by multiplication of related genes a situation has possibly been generated in which interactive products are encoded by the chromosomal locus.

As a final interesting observation regarding the *B*-mating type-like locus on *C. subvermispora*, within the locus there is a *cla4* gene for a PAK family kinase. In other basidiomycete species, a *cla4* gene was found either linked to a mating type locus for pheromones or pheromone receptors or linked to non-mating-type pheromone receptor genes residing outside of a mating type locus (89).

B mating type locus listing:

CsPh1
Protein ID:172056 Location:scaffold_11:1013885-1014033

CsPh2
Protein ID:172072 Location:scaffold_11:1015013-1015512

CsPh3
Protein ID: 172065 Location: scaffold_11:1023188-1023482

CsPh4
Protein ID:172052 Location:scaffold_11:1025672-1025794

Protein ID:97099 Location:scaffold_11:1026854-1030622
Unknown protein

Protein ID:116639 Location:scaffold_11:1031436-1032306
Unknown protein

Protein ID:97101 Location:scaffold_11:1035292-1036322
Unknown protein

CsPh5
Protein ID:172070 Location:scaffold_11:1037397-1037630

Protein ID:116640 Location:scaffold_11:1038206-1039817
Unknown protein

Receptor CsSTE3.1 most similar to CC1G 02129 = STE3.1
Protein ID:107158 Location:scaffold_11:1039281-1041259

CsPh6
Protein ID:172166 Location:scaffold_11:1041449-1042014

CsPh7
Protein ID:172169 Location:scaffold_11:1044175-1044712

Receptor CsSTE3.2 most similar to CC1G 15338= STE3.2a
Protein ID:116643 Location:scaffold_11:1045305-1047056

CsPh8
Protein ID:172055 Location:scaffold_11:1047894-1048007

CsPh9
Protein ID:172168 Location:scaffold_11:1048981-1050093

Receptor CsSTE3.3 most similar to CC1G 02137 = STE3.3
Protein ID:116644 Location:scaffold_11:1050370-1051890

CsPh10

Protein ID:172049 Location:scaffold_11:1052166-1052871

Receptor CsSTE3.4 most similar to pheromone receptor Rcb2 B43 CC1G 15338=STE3.2a
Protein ID:172172 Location:scaffold_11:1053637-1055163

Protein ID:116646 Location:scaffold_11:1056223-1057511
Unknown protein

CsPh11
Protein ID:172053 Location:scaffold_11:1058177-1058426

Protein ID:116647 Location:scaffold_11:1058591-1059085
Unknown protein

Receptor CsSTE3.5 most similar to pheromone receptor CC1G 15339=STE3.2b
Protein ID:172080 Location:scaffold_11:1059962-1061923

Protein ID:172061 Location:scaffold_11:1062269-1065783
Chitin synthase

Protein ID:172068 Location:scaffold_11:1067977-1069978
Unknown protein

CsPh12
Protein ID:172067 Location:scaffold_11:1070109-1070625

Receptor CsSTE3.6 most similar to pheromone receptor CC1G 02151 =STE3.2151
Protein ID:116650 Location:scaffold_11:1072152-1074219

Protein ID:116651 Location:scaffold_11:1074661-1079830
Glycoside hydrolase family 1 protein

Protein ID:116654 Location:scaffold_11:1080028-1082688
Msf1

Protein ID:116655 Location:scaffold_11:1083279-1085249
Msf2

Protein ID:116656 Location:scaffold_11:1087963-1093727
Unknown protein

Protein ID:116657 Location:scaffold_11:1094916-1096212
Glucose/ribitol dehydrogenase

A mating type locus listing:

a2-1, ID 110462
MHTSLSCETKVLHRVFTLAQHIEDRITASKRFIPPKSSQHLESPCSLYSHYILPEPAPILSELMSLDLDR
EAAIIFSDAYMRAAERIKATCEIQYGRVHRATIQRCKTSSPDIQKTAHTLQLAYVTNYMRMLKTWFNIIV

KCYVPRALHHKGQSPGQPETSGSRRPFTSAVVAVLEAFFVENAFPTRDEKHELAAETHMDYRQIHVWFQN
RRNRSREKGKAVKNGVQAQLPSDLEEAMTTILEKQWKDENTDLGLEDAFFDTSCIPRTLACDILDRDAPP
HAYPTTYPPSCDYQPFPILEGHHAFSTPWARKPVTTSGCMTSSIDVSSLTEMFANLSLEAHSSSGHKMRR
IASRETSTRGSSADWRVAVAPVAPLAALQVSSYPRRNAGNPSSPINCALPKTAHSLPTRSAICSSTRTRT
SRTRVRKPMALPRRLPKSQSRSHAKNQLSGPPTLDSRPDAVDPQSSSHSDTSSSGPHRYIKPYSRPISPP
SYPASCGRKAPKLYRQQGLNNEHGASPPGEASSFGCSASRSASLTSSSSVSLSSCSETESPLATPPSESI
PLFMPKLQLSQDWLNALDVTLANLDIPSNLLGTTFEESVLST*

>jgi|Cersu1|102296|estExt_fgenesh1_pm.C_10546
a1-1
MPLSMISSIQQRLLRSYDDFLAAVISGSTTLETFCGDWLRLHSDIVVMFTSGRLDVATMSLAQTTYSRLA
IYASTFHQLSADREQLTEKLMDEAQSMLSRMSISDDPPAVCFNIEDSRHLRPHMPDICTSSFEEVALSWL
MVNLNNPYPSDSVRLHLAQCYGIPLSAVSQWFKTTRRKIGWTSLCRGHFRGSRRRIIDDATIAFTGKHDQ
PLAAQFESIQDNVYRMFCERDQSELNDSAEVIAQALVKPIGHNERSSQKSPCRDYSHHGVNGHDQGSGAR
NPECDRVWSPSNCKSDEPGKLKRKRGICDDSRRVAFRRRLSEPQVGSSVLESHQNSSPYMTQDLELHSYY
GYLTGTPDERGSYEDSGSPESMAYFRVGEKRSHHSAALTDASVSVGLTDLDDDVHGSLRANSKQIPLSSV
HASRPTGLSTGLERRLDDAHDDGLIVEAETHQVVGKQQDTNSTSSWALPMLSMTQAHVDEDLGPTSLLDV
ENFYDWCSLGIYSQNCSSQPTDSSQLFDFKNDWEATHPPSADVTTLPSMTPTTGLQDFLQTILGSDCDSI
LSTVSASSNFYSELGALAYTWSSSCRSPSPILHDTPVTATEQAMLNPASRLGELDVSPACPPAPDAEIRT
SDHDQSLVIDKFLPSQCAAHPPSLYAPIINAHDHHHEGSLWQVS*

alternative a1-1 forms:

MPLSMISSIQQRLLRSYDDFLAAVISGSTTLETFCGDWLRLHSDIVVMFTSGRLDVATMSLAQTTYSRLA
IYASTFHQLSADREQLTEKLMDEAQSMLSRMSISDDPPAVCFNIEDSRHLRPHMPDICTSSFEEVALSWL
MVNLNNPYPSDSVRLHLAQCYGIPLSAVSQWFKTTRRKIGWTSLCRGHFRGSRRRIIDDATIAFTGKHDQ
PLAAQFESIQDNVYRMFCERDQSELNDSAEVIAQALVKPIGHNERSSQKSPCRDYSHHGVNGHDQGSGAR
NPECDRVWSPSNCKSDEPGKLKRKRGICDDSRRVAFRRRLSEPQVGSSVLESHQNSSPYMTQDL

MPLSMISSIQQRLLRSYDDFLAAVISGSTTLETFCGDWLRLHSDIVVMFTSGRLDVATMSLAQTTYSRLA
IYASTFHQLSADREQLTEKLMDEAQSMLSRMSISDDPPAVCFNIEDSRHLRPHMPDICTSSFEEVALSWL
MVNLNNPYPSDSVRLHLAQCYGIPLSAVSQWFKTTRRKIGWTSLCRGHFRGSRRRIIDDATIAFTGKHDQ
PLAAQFESIQDNVYRMFCERDQSELNDSAEVIAQALVKPIGHNERSSQKSPCRDYSHHGVNGHDQGSGAR
NPECDRVWSPSNCKSDEPGKLKRKRGICDDSRRVAFRRRLSEPQVGSSVLESHQNSSPYMTQDLELHSYY
GYLTGTPDERGSYEDSGSPESMAYFRVGEKR

Relic
>jgi|Cersu1|110463|estExt_fgenesh1_kg.C_10660
MSTIRDRLQAAEEELLLAVNAGNNSGALQRFDERWSALQAEVLAASQAGTIDADTIQLAQAVSGRVAIIA
QAFIDLDETAERLGQQLARKVDRIVNQATQGPPGLPSRLRGAAPGMNLGPQAGSSQAPAQPPYPAPYHIP
PTLQSSIQATYGTSGMSTAPGSLAQSSSTAGGTAATSASGSANAARQDEGASNATEGSRKSKSKKGKGNA
*

XIII. LC-MS/MS and microarray methods

Mass spectrometric analysis. "In Liquid" digestion and mass spectrometric analysis was done at the Mass Spectrometry Facility (Biotechnology Center, University of Wisconsin-Madison). In short, extracellular proteins from 250-ml culture filtrates were precipitated by direct addition of solid trichloroacetic acid (TCA) to 10% (wt/vol). Following overnight storage at -20°C, the precipitate was centrifuged and the pellet washed several times with cold acetone. Pelleted proteins (estimated at ~300 μg) were re-solubilized and denatured in 40µl of 8M Urea / 100 mM NH_4HCO_3 for 10 minutes then diluted to 200µl for tryptic digestion with: 10µl of 25 mM DTT, 10µl ACN, 110µl 25 mM NH_4HCO_3 pH8.2 and 30µl trypsin solution (100 ng/µl *Trypsin Gold* from PROMEGA Corp. in 25 mM NH_4HCO_3). Digestion was conducted overnight at 37°C then terminated by acidification with 2.5% TFA (Trifluoroacetic Acid) to 0.3% final and 8 µl loaded for nanoLC-MS/MS analysis.

NanoLC-MS/MS

Peptides were analyzed by nanoLC-MS/MS using the Agilent 1100 nanoflow system (Agilent Technologies, Palo Alto, CA) connected to a hybrid linear ion trap-orbitrap mass spectrometer (LTQ-Orbitrap XL, Thermo Fisher Scientific, San Jose, CA) equipped with a nanoelectrospray ion source. Capillary high-performance liquid chromatography (HPLC) was performed using an in-house fabricated column with integrated electrospray emitter essentially as described (92) but using 360μm x 75μm fused silica tubing. The column was packed with Jupiter 4 μm C_{12} particles (Phenomenex Inc., Torrance, CA) to approximately 12 cm. Sample loading (8 μl) and desalting were achieved using a trapping column in line with the autosampler (Zorbax 300SB-C18, 5 μM, 5 x 0.3 mm; Agilent). HPLC solvents were as follows: Isocratic loading: 1% (v/v) ACN, 0.1% formic acid; Gradient elution: Buffer A: 0.1% formic acid in water, and Buffer B: 95% (v/v) acetonitrile, 0.1% formic acid in water. Sample loading and desalting were done at 10 µL/min, whereas, gradient elution was performed at 200nL/min and increasing %B from A of 0 to 40% in 75 minutes, 40 to 60% in 20 minutes, and 60 to 100% in 5 min. The LTQ-Orbitrap was set to acquire MS/MS spectra in data-dependent mode as follows: MS survey scans from m/z 300 to 2000 were collected in centroid mode at a resolving power of 100,000. MS/MS spectra were collected on the five most-abundant signals in each survey scan. Dynamic exclusion was employed to increase dynamic range and maximize peptide identifications. This feature excluded precursors up to 0.55 m/z below and 1.05 m/z above previously selected precursors. Precursors remained on the exclusion list for 15 sec. Singly-charged ions and ions for which the charge state could not be assigned were rejected from consideration for MS/MS. Raw MS/MS data was searched against a total of 12,125 predicted and annotated gene models of *C. subvermispora* using in-house *Mascot* search engine with Methionine oxidation and Glutamine, Asparagine deamidation as variable modifications, peptide mass tolerance was set at 20 ppm and fragment mass at 0.6 Da. Nearly identical methodology was previously used to characterize the *P. chrysosporium* and *P. placenta* secretomes (93).

Protein annotation and significance of identification was done with help of Scaffold software (version 3.00.07, Proteome Software Inc., Portland, OR). Peptide identifications

were accepted within 0.9% False Discovery Rate as specified by the Peptide Prophet algorithm. Protein identifications were accepted if they could be established at greater than 95.0% probability, as assigned by the Protein Prophet algorithm (94), and contained at least two identified peptides.

Expression microarrays. From the data set of 12,125 *C. subvermispora* genes, each Roche NimbleGen (Madison, WI) array targeted 12083 genes, and all but 15 genes were represented by eight unique 60mer probes per gene. All probes were replicated four times per chip. No probes could be designed for 14 sequences and 15 sequences were represented by one to seven probes. Twenty-five sequences groups of sequences shared all of their probes. In these rare cases, only one set of probes (in four copies) were placed on the array to represent all of the sequences in a group, named with the seq ID of the exemplar sequence for the group. Complete design details are available as platform GPL15051 within NCBI GEO series GSE34636. The design platform and data for *P. chrysosporium* are deposited under GPL8022 and series GSE14736 (30).

Total RNA was purified from 5-day old cultures containing ball-milled *Populus* as sole carbon source. In short, cultures were harvested by filtering through Miracloth (Calbiochem, EMD Biosciences, Gibbstown, NJ), squeeze dried and snap frozen in liquid nitrogen. Pellets were stored at -80 C until use. Extraction buffer was prepared by combining 10 ml 690 mM para-aminosalicylic acid (sodium salt) (Sigma-Aldrich, St. Louis, MO) with 10 ml 56 mM triisopropylnapthalene sulfonic acid (sodium salt) (Sigma-Aldrich), and placed on ice. To this was added 5 ml 5X RNB (1.0 M Tris, 1.25 M NaCl, 0.25 M EGTA). The pH of the 5X RNB was adjusted to 8.5 with NaOH. The mixture was kept on ice and shaken just before use.

Frozen fungal pellets were ground to a fine powder with liquid nitrogen in an acid washed, pre-chilled mortar and pestle. The ground mycelia were transferred to Falcon 2059 tubes (VWR International, West Chester, PA), and extraction buffer was added to make a thick slurry. The samples were vortexed vigorously and placed on ice until all samples were processed. One half volume TE-saturated phenol (Sigma-Aldrich) and ¼ volume chloroform (Sigma-Aldrich) were added to each sample and vortexed vigorously. Samples were spun at 2940 x *g* in a fixed-angle rotor for five minutes. The aqueous layer was removed to a new tube, and phenol:chloroform extractions were repeated until the interface between the aqueous and organic layers was clear. The final aqueous extractions were placed in clean 2059 tubes, to which was added 0.1 volume 3M sodium acetate, pH 5.2, (DEPC-treated) and two volumes absolute ethanol. The tubes were shaken vigorously and stored overnight at -20 ^{0}C.

The tubes were spun 1 hour at 2940 x *g*, the supernatants were decanted, and the pellets were resuspended in 4 ml RNase-free H_2O. Total RNA was purified using the RNeasy Maxi kit (Qiagen, Valencia, CA) according to the manufacturer's protocol for RNA cleanup. RNAs were eluted from the RNeasy spin columns using two spins, for a final volume of 2 ml. The eluted RNAs were ethanol precipitated and stored overnight at -20C. The RNAs were spun one hour at 2940 x *g*, washed once with 70% ethanol, and

resuspended in 50-100 µl RNase-free H_2O. Three biological replicates per medium were used (six separate arrays).

RNA was converted to double-strand cDNA using Invitrogen's ds cDNA Synthesis kit (Carlsbad, CA) at the Univeristy of Wisconsin-Madison Biotechnology Center and labeled with the Cy3 fluorophore sample for hybridization to the array by Roche NimbleGen (Iceland). In brief, 10 µg of total RNA was incubated with 1X first strand buffer, 10 mM DTT, 0.5 mM dNTPs, 100pM oligo T7 $d(T)_{24}$ primer, and 400U of SuperScript II (Invitrogen) for 60 min at 42°C. Second strand cDNA was synthesized by incubation with 1X second strand buffer, 0.2 mM dNTPs, 0.07 U/ul DNA ligase (Invitrogen, Carlsbad, CA), 0.27 U/ul DNA polymerase I (Invitrogen, Carlsbad, CA), 0.013 U/ul RNase H (Invitrogen), at 16°C for two hours. Immediately following, 10U T4 DNA polymerase (Invitrogen) was added for an additional five minute incubation at 16°C. Double-stranded cDNA was treated with 27 ng/ul of RNase A (EpiCentre Technologies, Madison, WI) for 10 min at 37°C. Treated cDNA was purified using an equal volume of phenol:chloroform:isoamyl alcohol (Ambion/Life Technologies, Grand Island, NY), ethanol precipitated, washed with 80% ethanol, and resuspended in 20 µl water. One µg of each cDNA sample was amplified and labeled with one unit per µl of Klenow Fragment (New England BioLabs, UK) and one OD unit of Cy3 fluorophore (TriLink Biotechnologies, Inc., San Diego, CA) for two hours at 37°C. Array hybridization was carried out with 6 µg of labeled cDNA suspended in NimbleGen hybridization solution for 17 hours at 42°C. Arrays were scanned on the Axon 4000B Scanner (Molecular Devices, Sunnyvale, CA) and data was extracted from the scanned image using NimbleScan v2.4. DNASTAR ArrayStar v4 (Madison, WI) software was used to quantify and visualize data. All MIAME compliant (95) microarray expression data has been deposited in NCBI's Gene Expression Omnibus (96) and accessible through GEO Series accession number GSE34636.

IX. References

1. Tello M*, et al.* (2001) Isolation and characterization of homokaryotic strains from the ligninolytic basidiomycete *Ceriporiopsis subvermispora*. *FEMS Microbiol Lett* 199(1):91-96.
2. Blanchette RA*, et al.* (2010) An Antarctic hot spot for fungi at Shackleton's historic hut on Cape Royds. *Microb Ecol* 60(1):29-38 .
3. Jurka J*, et al.* (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1-4):462-467 .
4. Lowe TM & Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25(5):955-964 .
5. Salamov AA & Solovyev VV (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res* 10(4):516-522.
6. Isono K, McIninch JD, & Borodovsky M (1994) Characteristic features of the nucleotide sequences of yeast mitochondrial ribosomal protein genes as analyzed by computer program GeneMark. *DNA Res* 1(6):263-269 .
7. Birney E & Durbin R (2000) Using GeneWise in the *Drosophila* annotation experiment. *Genome Res* 10(4):547-548.
8. Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12(4):656-664 .
9. Nielsen H, Engelbrecht J, Brunak S, & von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* 10:1-6.
10. Melen K, Krogh A, & von Heijne G (2003) Reliability measures for membrane protein topology prediction algorithms. *J Mol Biol* 327(3):735-744 .
11. Zdobnov EM & Apweiler R (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17(9):847-848 .
12. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
13. Kanehisa M, Goto S, Kawashima S, Okuno Y, & Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32(Database issue):D277-280.
14. Koonin EV*, et al.* (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol* 5(2):R7.
15. Enright AJ, Van Dongen S, & Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30(7):1575-1584.
16. Yang L & Bennetzen JL (2009) Structure-based discovery and description of plant and animal Helitrons. *Proc Natl Acad Sci U S A* 106(31):12832-12837 .
17. Gutierrez A*, et al.* (2011) Regioselective oxygenation of fatty acids, fatty alcohols and other aliphatic compounds by a basidiomycete heme-thiolate peroxidase. *Arch Biochem Biophys* 514(1-2):33-43 .
18. Hofrichter M, Ullrich R, Pecyna MJ, Liers C, & Lundell T (2010) New and classic families of secreted fungal heme peroxidases. *Appl Microbiol Biotechnol* 87(3):871-897 .

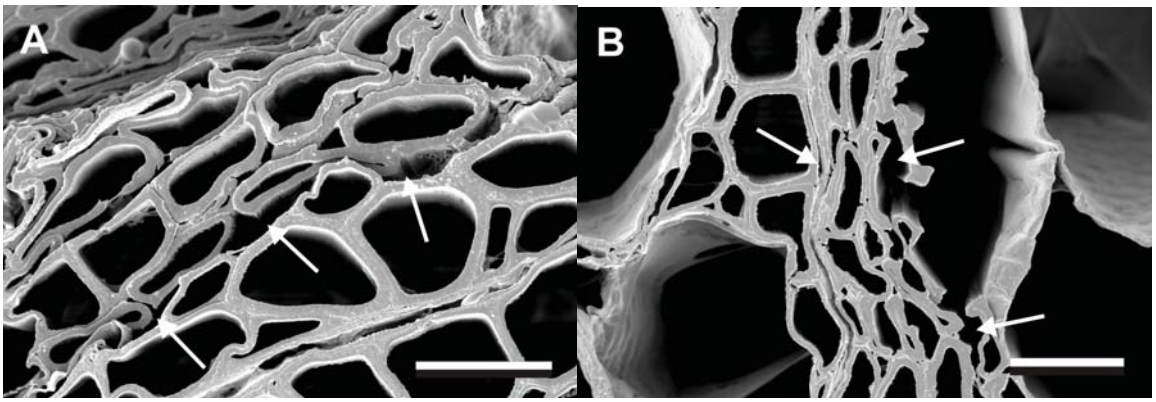
19. Liers C, Bobeth C, Pecyna M, Ullrich R, & Hofrichter M (2010) DyP-like peroxidases of the jelly fungus *Auricularia auricula-judae* oxidize nonphenolic lignin model compounds and high-redox potential dyes. *Appl Microbiol Biotechnol* 85(6):1869-1879 .
20. Bordoli L*, et al.* (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 4(1):1-13 .
21. Li D, Youngs HL, & Gold MH (2001) Heterologous expression of a thermostable manganese peroxidase from *Dichomitus squalens* in *Phanerochaete chrysosporium*. *Arch Biochem Biophys* 385(2):348-356.
22. Perez-Boada*, et al.* (2002) Expression of *Pleurotus eryngii* versatile peroxidase in *Escherichia coli* and optimization of *in vitro* folding. *Enzyme Microb Technol* 30:518-524.
23. Doyle WA & Smith AT (1996) Expression of lignin peroxidase H8 in *Escherichia coli*: folding and activation of the recombinant enzyme with Ca^{2+} and haem. *Biochem J* 315:15-19.
24. Mester T & Tien M (2001) Engineering of a manganese-binding site in lignin peroxidase isozyme H8 from *Phanerochaete chrysosporium*. *Biochem. Biophys. Res. Commun.* 284(3):723-728.
25. Karahanian E, Corsini G, Lobos S, & Vicuna R (1998) Structure and expression of a laccase gene from the ligninolytic basidiomycete *Ceriporiopsis subvermispora*. *Biochim Biophys Acta* 1443(1-2):65-74.
26. Hoegger PJ, Kilaru S, James TY, Thacker JR, & Kües U (2006) Phylogenetic comparison and classification of laccase and related multicopper oxidase protein sequences. *FEBS J* 273(10):2308-2326.
27. Martinez D*, et al.* (2009) 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(6):1954-1959.
28. Yelle DJ, Wei D, Ralph J, & Hammel KE (2011) Multidimensional NMR analysis reveals truncated lignin structures in wood decayed by the brown rot basidiomycete *Postia placenta*. *Environ Microbiol* 13(4):1091-1100 .
29. Wei D*, et al.* (2009) Laccase and its role in production of extracellular reactive oxygen species during wood decay by the brown rot basidiomycete *Postia placenta*. *Appl Environ Microbiol* 76(7):2091-2097.
30. Vanden Wymelenberg A*, et al.* (2010) Comparative transcriptome and secretome analysis of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium*. *Appl Environ Microbiol* 76:3599-3610.
31. Kumar SV, Phale PS, Durani S, & Wangikar PP (2003) Combined sequence and structure analysis of the fungal laccase family. *Biotechnol Bioeng* 83(4):386-394 .
32. Whittaker MM, Kersten PJ, Cullen D, & Whittaker JW (1999) Identification of catalytic residues in glyoxal oxidase by targeted mutagenesis. *J Biol Chem* 274(51):36226-36232.
33. Hallberg BM, Henriksson G, Pettersson G, Vasella A, & Divne C (2003) Mechanism of the reductive half-reaction in cellobiose dehydrogenase. *J Biol Chem* 278(9):7160-7166.

34. Baldrian P & Valaskova V (2008) Degradation of cellulose by basidiomycetous fungi. *FEMS Microbiol. Rev.* 32(3):501-521.
35. Harris PV*, et al.* (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* 49(15):3305-3316.
36. Langston JA*, et al.* (2011) Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl Environ Microbiol* 77:7007-7015 .
37. Guillen F & Evans CS (1994) Anisaldehyde and veratraldehyde acting as redox cycling agents for H_2O_2 production by *Pleurotus eryngii*. *Appl Environ Microbiol* 60(8):2811-2817.
38. Reiser J, Muheim A, Hardegger M, Frank G, & Feichter A (1994) Aryl-alcohol dehydrogenase from the white-rot fungus *Phanerochaete chrysosporium* : gene cloning, sequence analysis, expression and purification of recombinant protein. *J. Biol. Chem* 269:28152-28159.
39. Thompson JD, Higgins DG, & Gibson TJ (1994) ClustalW improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:2552-2556.
40. Ferreira P, Ruiz-Duenas FJ, Martinez MJ, van Berkel WJ, & Martinez AT (2006) Site-directed mutagenesis of selected residues at the active site of aryl-alcohol oxidase, an H_2O_2-producing ligninolytic enzyme. *FEBS J* 273(21):4878-4888.
41. Cavener DR (1992) GMC oxidoreductases. A newly defined family of homologous proteins with diverse catalytic activities. *J Mol Biol* 223(3):811-814 .
42. Duranova M, Spanikova S, Wösten HA, Biely P, & de Vries RP (2009) Two glucuronoyl esterases of *Phanerochaete chrysosporium*. *Arch. Microbiol.* 191(2):133-140.
43. Li XL, Skory CD, Cotta MA, Puchart V, & Biely P (2008) Novel family of carbohydrate esterases, based on identification of the *Hypocrea jecorina* acetyl esterase gene. *Appl Environ Microbiol* 74(24):7482-7489 .
44. Quinlan RJ*, et al.* (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci U S A* 108(37):15079-15084 .
45. Akhtar M, Attridge MC, Myers GC, Kirk TK, & Blanchette RA (1992) Biomechanical pulping of loblolly pine with different strains of the white rot fungus *Ceriporiopsis subvermispora*. *TAPPI* 75:105-109.
46. Blanchette R (1991) Delignification by wood-decay fungi. *Ann Rev Phytopath* 29:381-398.
47. Blanchette RA, Burnes TA, Eerdmans MM, & Akhtar M (1992) Evaluating isolates of *Phanerochaete chrysosporium* and *Ceriporiopsis subvermispora* for use in biological pulping processes. *Holzforschung* 46:109-115.
48. Eriksson K-EL, Blanchette RA, & Ander P (1990) *Microbial and enzymatic degradation of wood and wood components* (Springer-Verlag, Berlin).
49. Otjen L, Blanchette R, Effland M, & Leatham G (1987) Assessment of 30 white rot basidiomycetes for selective lignin degradation. *Holzforschung* 41:343-349.

50. Thomas-Chollier M, *et al.* (2008) RSAT: regulatory sequence analysis tools. *Nucleic Acids Res* 36(Web Server issue):W119-127 .
51. Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4):406-425.
52. Kumar S, Tamura K, & Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5(2):150-163 .
53. Seibel C, *et al.* (2009) Light-dependent roles of the G-protein alpha subunit GNA1 of *Hypocrea jecorina* (anamorph *Trichoderma reesei*). *BMC Biol* 7:58 .
54. Schmoll M, Schuster A, Silva Rdo N, & Kubicek CP (2009) The G-alpha protein GNA3 of *Hypocrea jecorina (Anamorph Trichoderma reesei*) regulates cellulase gene expression in the presence of light. *Eukaryot Cell* 8(3):410-420 .
55. Li L, Wright SJ, Krystofova S, Park G, & Borkovich KA (2007) Heterotrimeric G protein signaling in filamentous fungi. *Annu Rev Microbiol* 61:423-452 .
56. Lafon A, Han KH, Seo JA, Yu JH, & d'Enfert C (2006) G-protein and cAMP-mediated signaling in aspergilli: a genomic perspective. *Fungal Genet Biol* 43(7):490-502 .
57. Schmoll M (2008) The information highways of a biotechnological workhorse--signal transduction in *Hypocrea jecorina*. *BMC Genomics* 9:430 .
58. Regenfelder E, *et al.* (1997) G proteins in *Ustilago maydis*: transmission of multiple signals? *EMBO J* 16(8):1934-1942.
59. Han KH, Seo JA, & Yu JH (2004) Regulators of G-protein signalling in *Aspergillus nidulans*: RgsA downregulates stress response and stimulates asexual sporulation through attenuation of GanB (Galpha) signalling. *Mol Microbiol* 53(2):529-540 .
60. Yu JH, Mah JH, & Seo JA (2006) Growth and developmental control in the model and pathogenic aspergilli. *Eukaryot Cell* 5(10):1577-1584 .
61. Seo JA & Yu JH (2006) The phosducin-like protein PhnA is required for Gbetagamma-mediated signaling for vegetative growth, developmental control, and toxin biosynthesis in *Aspergillus nidulans*. *Eukaryot Cell* 5(2):400-410 .
62. Willardson BM & Howlett AC (2007) Function of phosducin-like proteins in G protein signaling and chaperone-assisted protein folding. *Cell Signal* 19(12):2417-2427 .
63. Jones CA, Greer-Phillips SE, & Borkovich KA (2007) The response regulator RRG-1 functions upstream of a mitogen-activated protein kinase pathway impacting asexual development, female fertility, osmotic stress, and fungicide resistance in *Neurospora crassa*. *Mol Biol Cell* 18(6):2123-2136 .
64. Lavin JL, Oguiza JA, Ramirez L, & Pisabarro A (2008) Comparative genomics of oxidative phosphorylation systems in fungi. *Fungal Genet Biol* 45:1248-1256.
65. Huynen MA, de Hollander M, & Szklarczyk R (2009) Mitochondrial proteome evolution and genetic disease. *Biochim Biophys Acta* 1792(12):1122-1129 .
66. Abdrakhmanova A (2005) Accessory subunits of complex I from *Yarrowia lipolytica*. PhD (Johann Wolfgang Goethe-Universität Frankfurt).
67. Nelson DR (2011) Progress in tracing the evolutionary paths of cytochrome P450. *Biochim Biophys Acta* In Press.

68. Martinez D*, et al.* (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol* 22:695-700.
69. Nishimura H*, et al.* (2008) De novo synthesis of (Z)- and (E)-7-hexadecenylitaconic acids by a selective lignin-degrading fungus, *Ceriporiopsis subvermispora*. *Phytochemistry* 69(14):2593-2602 .
70. Rahmawati N, Ohashi Y, Watanabe T, Honda Y, & Watanabe T (2005) Ceriporic acid B, an extracellular metabolite of *Ceriporiopsis subvermispora*, suppresses the depolymerization of cellulose by the Fenton reaction. *Biomacromolecules* 6(5):2851-2856.
71. Schwecke T*, et al.* (2006) Nonribosomal peptide synthesis in *Schizosaccharomyces pombe* and the architectures of ferrichrome-type siderophore synthetases in fungi. *Chembiochem* 7(4):612-622 .
72. Welzel K, Eisfeld K, Antelo L, Anke T, & Anke H (2005) Characterization of the ferrichrome A biosynthetic gene cluster in the homobasidiomycete *Omphalotus olearius*. *FEMS Microbiol Lett* 249(1):157-163 .
73. Hoffmeister D & Keller NP (2007) Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat Prod Rep* 24(2):393-416 .
74. Catalanotto C, Azzalin G, Macino G, & Cogoni C (2002) Involvement of small RNAs and role of the qde genes in the gene silencing pathway in *Neurospora*. *Genes Dev* 16(7):790-795 .
75. Cogoni C & Macino G (1999) Posttranscriptional gene silencing in *Neurospora* by a RecQ DNA helicase. *Science* 286(5448):2342-2344 .
76. Cogoni C & Macino G (1999) Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* 399(6732):166-169 .
77. Romano N & Macino G (1992) Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol Microbiol* 6(22):3343-3353 .
78. Nakayashiki H, Kadotani N, & Mayama S (2006) Evolution and diversification of RNA silencing proteins in fungi. *J Mol Evol* 63(1):127-135 .
79. Salame TM, Ziv C, Hadar Y, & Yarden O (2011) RNAi as a potential tool for biotechnological applications in fungi. *Appl Microbiol Biotechnol* 89(3):501-512 .
80. Li L, Chang SS, & Liu Y (2010) RNA interference pathways in filamentous fungi. *Cell Mol Life Sci* 67(22):3849-3863 .
81. Cao Y, Cheong H, Song H, & Klionsky DJ (2008) In vivo reconstitution of autophagy in *Saccharomyces cerevisiae*. *J Cell Biol* 182(4):703-713 .
82. Kershaw MJ & Talbot NJ (2009) Genome-wide functional analysis reveals that infection-associated fungal autophagy is necessary for rice blast disease. *Proc Natl Acad Sci U S A* 106(37):15967-15972 .
83. Klionsky DJ*, et al.* (2003) A unified nomenclature for yeast autophagy-related genes. *Dev Cell* 5(4):539-545 .
84. Noda T & Ohsumi Y (1998) Tor, a phosphatidylinositol kinase homologue, controls autophagy in yeast. *J Biol Chem* 273(7):3963-3966 .

85. Tlalka M, Fricker M, & Watkinson S (2008) Imaging of long-distance alpha-aminoisobutyric acid translocation dynamics during resource capture by *Serpula lacrymans*. *Appl Environ Microbiol* 74(9):2700-2708 .
86. Domanski S (1969) Grzyby zasiedlajace drewno w *Puszczy Bialieskiej* X. *Fibuloporia subvermispora* (Pilát) Doman., comb. nov. i jej rozpoznanie. . *Acta Societas Botanicorum Poloniae* 38:453-464.
87. Nobles MK, Macrae R, & Tomlin BP (1957) Results of infertility test on some species of hymenomycetes. *Can J Bot* 35:377-387.
88. James TY, Lee M, & van Diepen LT (2011) A single mating-type locus composed of homeodomain genes promotes nuclear migration and heterokaryosis in the white-rot fungus *Phanerochaete chrysosporium*. *Eukaryot Cell* 10(2):249-261 .
89. Kües U, James T, & Heitman J (2011) Mating type in Basidiomycetes: Unipolar, bipolar and tetrapolar patterns of sexuality. *The Mycota*, eds Pöggeler S & Wöstemeyer J (Springer, Heidelberg), Vol 14, pp 97-160.
90. van Peer AF*, et al.* (2011) Comparative genomics of the mating-type loci of the mushroom *Flammulina velutipe*s reveals widespread synteny and recent inversions. *PLoS One* 6(7):e22249 .
91. Stajich JE*, et al.* (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea (Coprinus cinereus)*. *Proc Natl Acad Sci U S A* 107(26):11889-11894 .
92. Martin SE, Shabanowitz J, Hunt DF, & Marto JA (2000) Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 72:4266-4274.
93. Vanden Wymelenberg A*, et al.* (2011) Significant alteration of gene expression in wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* by plant species. *Appl Environ Microbiol* 77(13):4499-4507 .
94. Nesvizhskii AI, Keller A, Kolker E, & Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 75(17):4646-4658.
95. Brazma A*, et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 29(4):365-371.
96. Edgar R, Domrachev M, & Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30(1):207-210.


A
B

C

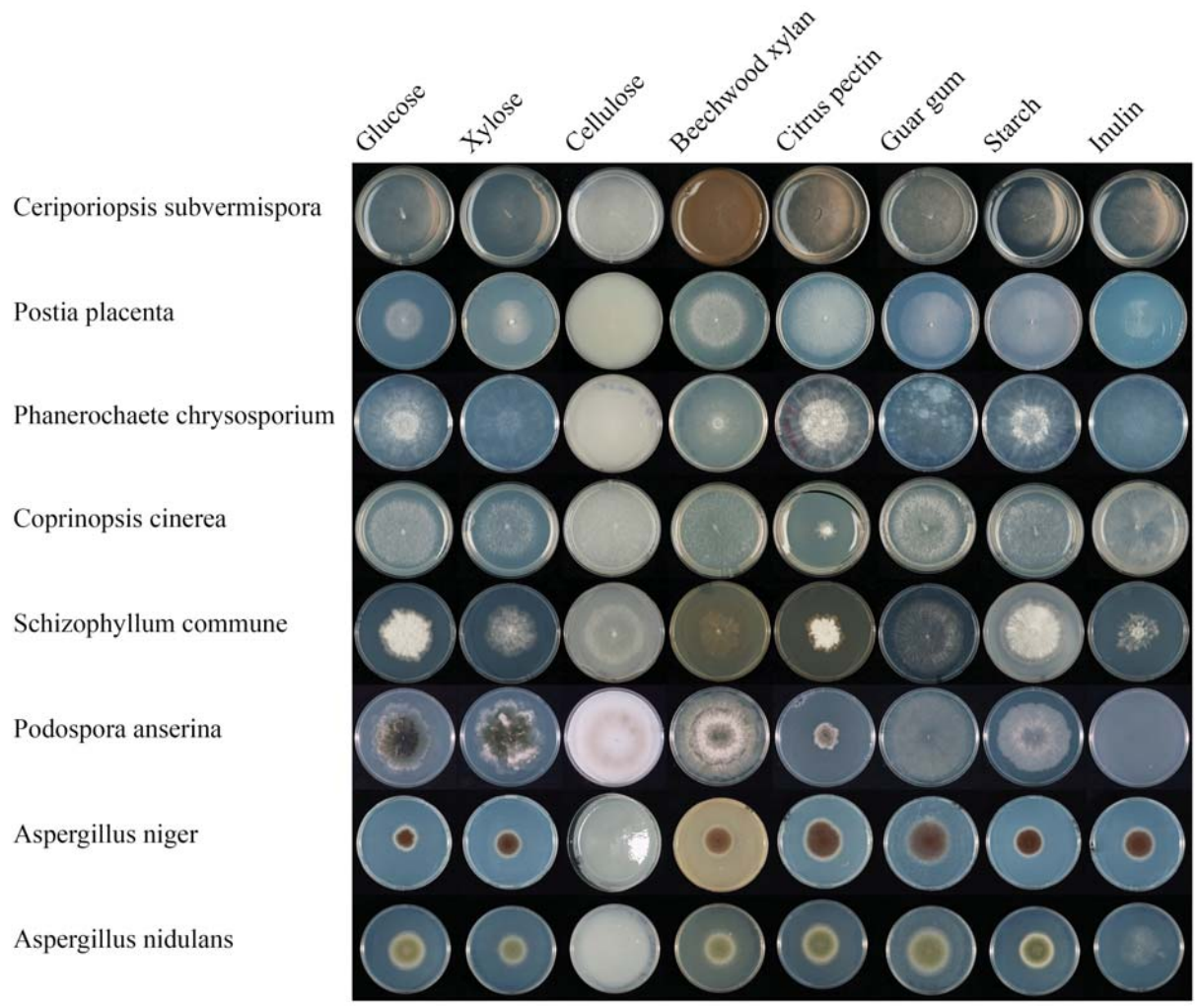

Glucose
Xylose
Cellulose
Beechwood xylan
Citrus pectin
Guar gum
Starch
Inulin
Ceriporiopsis subvermispora
Postia placenta
Phanerochaete chrysosporium
Coprinopsis cinerea
Schizophyllum commune
Podospora anserina
Aspergillus niger
Aspergillus nidulans

Figure S1. Growth characteristics of *C. subvermispora*. Weight loss and scanning electron micrographs of transverse sections from aspen wood after 60 days of decay by the parent dikayotic strain (A) and the monokaryon (B). Wood fiber cells show areas where lignin has been extensively removed (arrows) leaving only secondary wall layers. The removal of the lignin-rich middle lamella between cells caused cells to detach from one another and resulted in the collapse of some secondary walls. (C) Growth profiles on monosaccharides (25 mM final concentration) and polysaccharides (1% final concentration). Strains were grown until the largest colony was just reaching the edge of the plate. For growth comparison, glucose was used as a standard for all fungi because it is the best monosaccharide for growth. Complete growth profiles and media composition can be found at www.fung-growth.org.

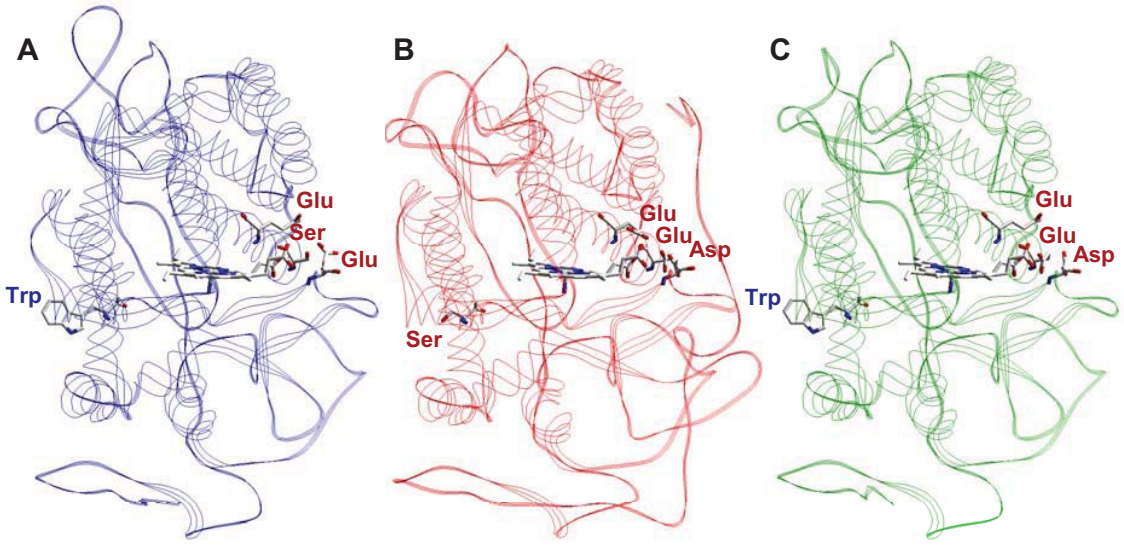

A
Glu
Ser
Glu
Trp
B
Glu
Glu
Asp
Ser
C
Glu
Glu
Asp
Trp

Figure S2. Homology-based molecular models of three *C. subvermispora* peroxidases were obtained automatically using *Pleurotus ostreatus* VP and *P. chrysosporium* MnP crystal structures (PDB entries 2BOQ and 3M5Q, respectively) as templates. **A)** LiP-type model **(**Cesubv118677) including an exposed tryptophan potentially involved in oxidation of high redox-potential substrates; **B)** MnP-type model (Cesubv 117436) including a putative Mn^{2+} oxidation site (formed by two glutamates and one aspartate); and **C)** VP-type model (Cesubv99382) including both putative catalytic sites described in **A** and **B**. Note that a serine residue in the LiP-type model occupies the position of one of the two catalytic glutamates involved in Mn^{2+} oxidation in the MnP and VP-type models. Another serine residue occupies the position of the putative catalytic tryptophan present in the LiP and VP-type models.

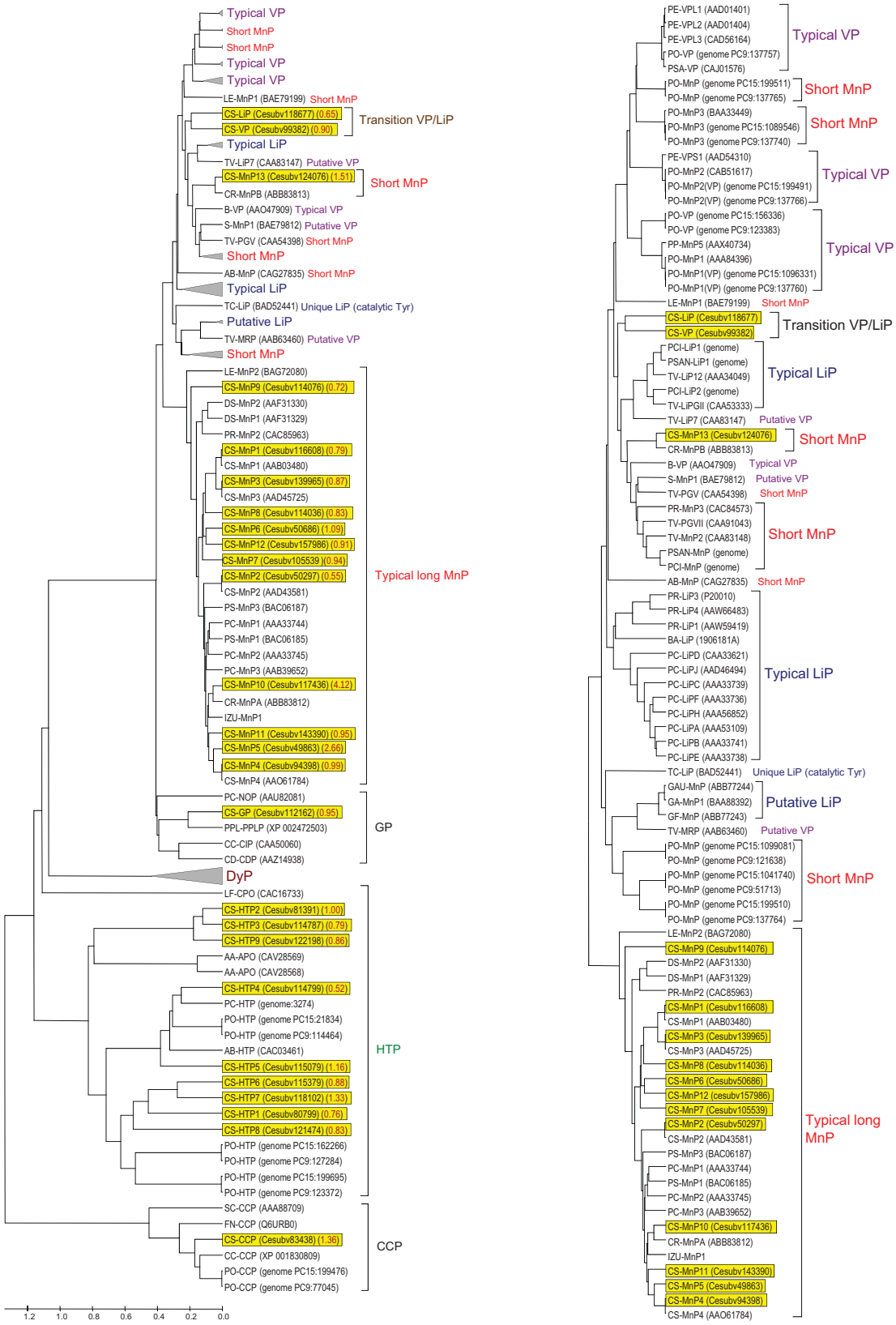
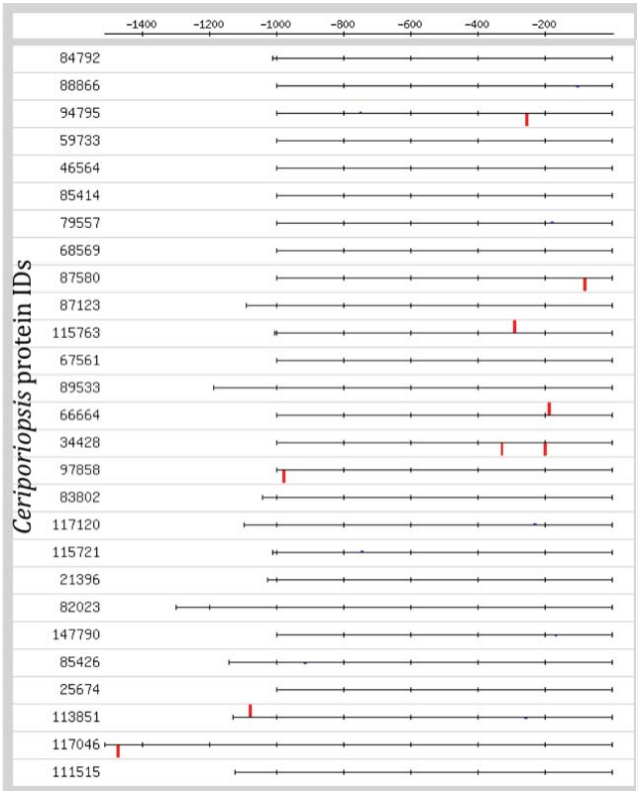

Typical VP
Short MnP
Short MnP
Typical VP
Typical VP
LE-MnP1 (BAE79199) Short MnP
CS-LiP (Cesubv118677) (0.65)
CS-VP (Cesubv99382) (0.90)
Transition VP/LiP
Typical LiP
TV-LiP7 (CAA83147) Putative VP
CS-MnP13 (Cesubv124076) (1.51)
CR-MnPB (ABB83813)
Short MnP
B-VP (AAO47909) Typical VP
S-MnP1 (BAE79812) Putative VP
TV-PGV (CAA54398) Short MnP
Short MnP
AB-MnP (CAG27835) Short MnP
Typical LiP
TC-LiP (BAD52441) Unique LiP (catalytic Tyr)
Putative LiP
TV-MRP (AAB63460) Putative VP
Short MnP
LE-MnP2 (BAG72080)
CS-MnP9 (Cesubv114076) (0.72)
DS-MnP2 (AAF31330)
DS-MnP1 (AAF31329)
PR-MnP2 (CAC85963)
CS-MnP1 (Cesubv116608) (0.79)
CS-MnP1 (AAB03480)
CS-MnP3 (Cesubv139965) (0.87)
CS-MnP3 (AAD45725)
CS-MnP8 (Cesubv114036) (0.83)
CS-MnP6 (Cesubv50686) (1.09)
CS-MnP12 (Cesubv157986) (0.91)
CS-MnP7 (Cesubv105539) (0.94)
CS-MnP2 (Cesubv50297) (0.55)
CS-MnP2 (AAD43581)
PS-MnP3 (BAC06187)
PC-MnP1 (AAA33744)
PS-MnP1 (BAC06185)
PC-MnP2 (AAA33745)
PC-MnP3 (AAB39652)
CS-MnP10 (Cesubv117436) (4.12)
CR-MnPA (ABB83812)
IZU-MnP1
CS-MnP11 (Cesubv143390) (0.95)
CS-MnP5 (Cesubv49863) (2.66)
CS-MnP4 (Cesubv94398) (0.99)
CS-MnP4 (AAO61784)
Typical long MnP
PC-NOP (AAU82081)
CS-GP (Cesubv112162) (0.95)
PPL-PPLP (XP 002472503)
CC-CIP (CAA50060)
CD-CDP (AAZ14938)
GP
DyP
LF-CPO (CAC16733)
CS-HTP2 (Cesubv81391) (1.00)
CS-HTP3 (Cesubv114787) (0.79)
CS-HTP9 (Cesubv122198) (0.86)
AA-APO (CAV28569)
AA-APO (CAV28568)
CS-HTP4 (Cesubv114799) (0.52)
PC-HTP (genome:3274)
PO-HTP (genome PC15:21834)
PO-HTP (genome PC9:114464)
AB-HTP (CAC03461)
CS-HTP5 (Cesubv115079) (1.16)
CS-HTP6 (Cesubv115379) (0.88)
CS-HTP7 (Cesubv118102) (1.33)
CS-HTP1 (Cesubv80799) (0.76)
CS-HTP8 (Cesubv121474) (0.83)
PO-HTP (genome PC15:162266)
PO-HTP (genome PC9:127284)
PO-HTP (genome PC15:199695)
PO-HTP (genome PC9:123372)
HTP
SC-CCP (AAA88709)
FN-CCP (Q6URB0)
CS-CCP (Cesubv83438) (1.36)
CC-CCP (XP 001830809)
PO-CCP (genome PC15:199476)
PO-CCP (genome PC9:77045)
CCP
1.2
1.0
0.8
0.6
0.4
0.2
0.0
PE-VPL1 (AAD01401)
PE-VPL2 (AAD01404)
PE-VPL3 (CAD56164)
PO-VP (genome PC9:137757)
PSA-VP (CAJ01576)
Typical VP
PO-MnP (genome PC15:199511)
PO-MnP (genome PC9:137765)
Short MnP
PO-MnP3 (BAA33449)
PO-MnP3 (genome PC15:1089546)
PO-MnP3 (genome PC9:137740)
Short MnP
PE-VPS1 (AAD54310)
PO-MnP2 (CAB51617)
PO-MnP2(VP) (genome PC15:199491)
PO-MnP2(VP) (genome PC9:137766)
Typical VP
PO-VP (genome PC15:156336)
PO-VP (genome PC9:123383)
PP-MnP5 (AAX40734)
PO-MnP1 (AAA84396)
PO-MnP1(VP) (genome PC15:1096331)
PO-MnP1(VP) (genome PC9:137760)
Typical VP
LE-MnP1 (BAE79199) Short MnP
CS-LiP (Cesubv118677)
CS-VP (Cesubv99382)
Transition VP/LiP
PCI-LiP1 (genome)
PSAN-LiP1 (genome)
TV-LiP12 (AAA34049)
PCI-LiP2 (genome)
TV-LiPGII (CAA53333)
Typical LiP
TV-LiP7 (CAA83147) Putative VP
CS-MnP13 (Cesubv124076)
CR-MnPB (ABB83813)
Short MnP
B-VP (AAO47909) Typical VP
S-MnP1 (BAE79812) Putative VP
TV-PGV (CAA54398) Short MnP
PR-MnP3 (CAC84573)
TV-PGVII (CAA91043)
TV-MnP2 (CAA83148)
PSAN-MnP (genome)
PCI-MnP (genome)
Short MnP
AB-MnP (CAG27835) Short MnP
PR-LiP3 (P20010)
PR-LiP4 (AAW66483)
PR-LiP1 (AAW59419)
BA-LiP (1906181A)
PC-LiPD (CAA33621)
PC-LiPJ (AAD46494)
PC-LiPC (AAA33739)
PC-LiPF (AAA33736)
PC-LiPH (AAA56852)
PC-LiPA (AAA53109)
PC-LiPB (AAA33741)
PC-LiPE (AAA33738)
Typical LiP
TC-LiP (BAD52441) Unique LiP (catalytic Tyr)
GAU-MnP (ABB77244)
GA-MnP1 (BAA88392)
GF-MnP (ABB77243)
Putative LiP
TV-MRP (AAB63460) Putative VP
PO-MnP (genome PC15:1099081)
PO-MnP (genome PC9:121638)
PO-MnP (genome PC15:1041740)
PO-MnP (genome PC9:51713)
PO-MnP (genome PC15:199510)
PO-MnP (genome PC9:137764)
Short MnP
LE-MnP2 (BAG72080)
CS-MnP9 (Cesubv114076)
DS-MnP2 (AAF31330)
DS-MnP1 (AAF31329)
PR-MnP2 (CAC85963)
CS-MnP1 (Cesubv116608)
CS-MnP1 (AAB03480)
CS-MnP3 (Cesubv139965)
CS-MnP3 (AAD45725)
CS-MnP8 (Cesubv114036)
CS-MnP6 (Cesubv50686)
CS-MnP12 (cesubv157986)
CS-MnP7 (Cesubv105539)
CS-MnP2 (Cesubv50297)
CS-MnP2 (AAD43581)
PS-MnP3 (BAC06187)
PC-MnP1 (AAA33744)
PS-MnP1 (BAC06185)
PC-MnP2 (AAA33745)
PC-MnP3 (AAB39652)
CS-MnP10 (Cesubv117436)
CR-MnPA (ABB83812)
IZU-MnP1
CS-MnP11 (Cesubv143390)
CS-MnP5 (Cesubv49863)
CS-MnP4 (Cesubv94398)
CS-MnP4 (AAO61784)
Typical long MnP

Figure S3. Comparative analysis of basidiomycete peroxidases. Left, relationships among 133 basidiomycete (and two reference ascomycetes, *Leptoxyphium fumago* and *S. cerevisiae*) peroxidases (GeneBank references in parentheses, and *C. subvermispora* protein models in yellow). Amino-acid sequence comparisons as Poisson distances and clustering based on UPGMA and "pair-wise deletion" option of MEGA4. Compressed sub-trees are shown to facilitate the *C. subvermispora* peroxidases analysis. Induction levels in ball milled aspen (BMA) *vs* glucose as sole carbon source for each peroxidase are shown in red in brackets. Abbrev: AA, *Agrocybe aegerita*; AB, *Agaricus bisporus*; B, *Bjerkandera* sp; BA, *Bjerkandera adusta*; CC, *Coprinopsis cinerea;* CD, *Coprinellus disseminatus*; CR, *Ceriporiopsis rivulosa*; CS, *C. subvermispora*; DS, *Dichomitus squalens*; FN, *Filobasidiella neoformans*; GA, *Ganoderma applanatum*; GAU, *Ganoderma australe*; GF, *Ganoderma formosanum*; IZU, basidiomycete IZU-154; LE, *Lentinula edodes*; LF, *Leptoxyphium fumago*; MS, *Marasmius scorodonius*; P, *Polyporus* sp; PC, *P. chrysosporium*; PCI, *Pycnoporus cinnabarinus*; PE, *Pleurotus eryngii*; PO, *Pleurotus ostreatus*; PP, *Pleurotus pulmonarius*; PPL, *Postia placenta*; PR, *Phlebia radiata*; PS, *Phanerochaete sordida*; PSA, *Pleurotus sapidus*; S, *Spongipellis* sp; SC, *Saccharomyces cerevisiae*; TA, *Termitomyces albuminosus*; TC, *Trametes cervina*; and TV, *Trametes versicolor*. CCP, cytochrome *c* peroxidase; CDP, *Coprinellus disseminatus* peroxidase; CIP, *Coprinopsis cinerea* peroxidase; DyP, dye-decolorizing peroxidase; HTP, heme-thiolate peroxidase; LiP, lignin peroxidase; MnP, manganese peroxidase; MRP, manganese-repressed peroxidase from *T. versicolor*; NOP, *P. chrysosporium* hypothetical (generic) peroxidase; PGV, *T. versicolor* hypothetical peroxidase; PGVII, *T. versicolor* hypothetical peroxidase; PPLP, *P. placenta* peroxidase; VP, versatile peroxidase. Right side dendrogram focuses on 90 ligninolytic peroxidases showing putative evolutionary relationships and structural-functional classification. “Short” and “long” MnPs have an experimentally confirmed or a putative Mn^{2+}-oxidation site formed by two glutamic and one aspartic residues, and differ in the length of the C-terminal tail. Typical LiPs contain an experimentally confirmed or a putative catalytic tryptophan whereas TC-LiP is a “unique” ligninolytic peroxidase with a catalytic tyrosine. Typical VPs harbor catalytic sites described above for both MnPs and Typical LiPs. Two *C. subvermispora* peroxidases initially annotated as LiP (Cesubv118677) and VP (Cesubv99382) are in fact a new type of “hybrid” peroxidase sharing catalytic properties with VP and LiP being also unable to oxidize Mn^{2+} (Main text Table 1), in spite of exhibiting (Cesubv99382) a typical Mn^{2+} oxidation site.

A


-1400
-1200
-1000
-800
-600
-400
-200
Ceriporiopsis protein IDs
84792
88866
94795
59733
46564
85414
79557
68569
87580
87123
115763
67561
89533
66664
34428
97858
83802
117120
115721
21396
82023
147790
85426
25674
113851
117046
111515

B

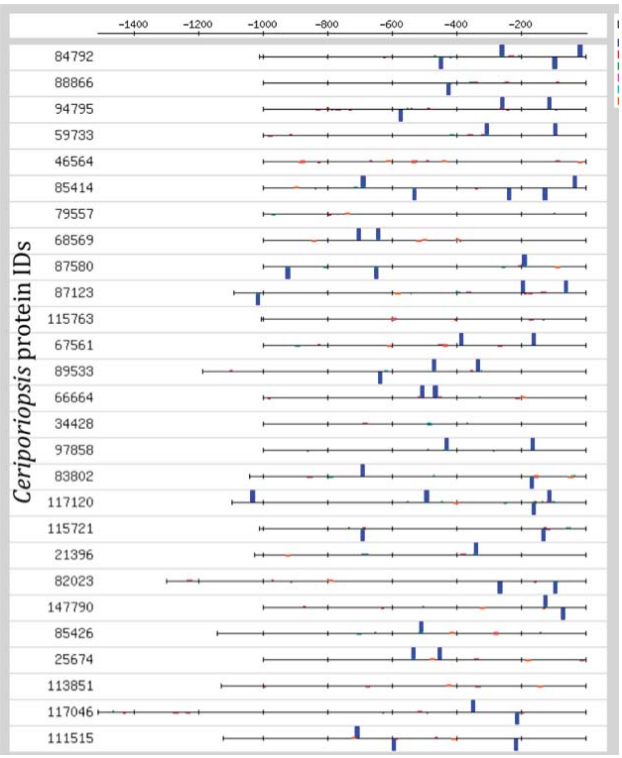

-1400
-1200
-1000
-800
-600
-400
-200
Ceriporiopsis protein IDs
84792
88866
94795
59733
46564
85414
79557
68569
87580
87123
115763
67561
89533
66664
34428
97858
83802
117120
115721
21396
82023
147790
85426
25674
113851
117046
111515

Figure S4. (A) Identification of a putative CRE1-binding motif in promoters regions of CAZy-encoding genes. Read marks denote the cis-regulatory motif, while horizontal axes represent the upstream region of each gene indicated at the left (starting at -1 relative to the translation start site). (B), Enrichment of UAS motifs in 27 CAZy-encoding genes that are most strongly upregulated. The blue bars specify the position of the CGA-N7-CCG motif.

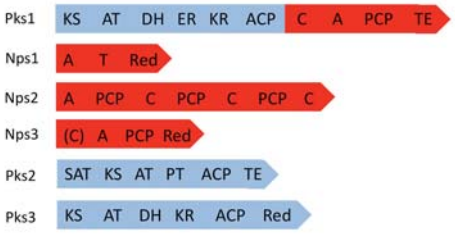

96245(CYP5158A)
52757(CYP5158A)
52911(CYP5158A)
52252(CYP5158A)
65977(CYP5158A)
85632(CYP5037B)
110763(CYP5144C)
115906(CYP5158A)
115908(CYP5158A)
80072(CYP5158A)
100017(CYP5158A)
116918(CYP5065A)
45215(CYP5037A)
84543(CYP5037B)
96232(CYP5158A)
138759(CYP5158A)
96265(CYP5158A)
53073(CYP5158A)
116679(CYP5037B)
107188(CYP5046A)
120119(CYP5027A)
60535(CYP5027A)
138257(CYP5144C)
118484(CYP5037B)
84764(CYP5037B)
157706(CYP5037NS)
73707(CYP5065A)
114955(CYP5037B)
159814(CYP5037B)
80124(CYP5065NS)
118063(CYP5144C)
96445(CYP5144A)
95844(CYP5144C)
95847(CYP5144C3)
115592(CYP5144C)
107728(CYP5144C)
157065(CYP5158A)
155335(CYP502A)
58021(CYP502A)
161182(CYP502B)
115561(CYP502A)
118480(CYP502A)
143292(CYP5144C)
118062(CYP5037B)
153900(CYP5144C)
115591(CYP5144A)
157058(CYP5144C)
53277(CYP5144C)
52754(CYP5158A)
115946(CYP5158A)
96243(CYP5158A)
141279(CYP5144F)
94793(CYP5144C)
154684(CYP5144C)
113891(CYP5144F)
138007(CYP5144C)
105266(CYP5144F)
110546(CYP5144F)
153819(CYP5144F)
83037(CYP5144C)
94290(CYP5144C)
153831(CYP5144F)
50620(CYP5144C)
53222(CYP5144C)
50154(CYP5144C)
114152(CYP5144C)
65155(CYP5144C)
78948(CYP5037NS)
100592(CYP5144C)
142250(CYP5037B)
66107(CYP5144C)
67997(CYP5144C)
118759(CYP5144C)
81801(CYP5144J)
141153(CYP5144C)
112052(CYP5144C)
109980(CYP5144C)
151116(CYP5144C)
63625(CYP5144C)
117766(CYP5144C)
93086(CYP5144C)
151145(CYP5144C)
141146(CYP5144C)
87014(CYP5144C)
141968(CYP5144C)
135729(CYP5144C)
132072(CYP5144C)
73399(CYP5144C)
133760(CYP5144C)
93317(CYP5144NS)
104492(CYP5144C)
81800(CYP5144C)
63783(CYP5144C)
97741(CYP5136A)
72211(CYP5136A)
126579(CYP5136A)
77670(CYP5136A)
126990(CYP5136A)
159220(CYP5136A)
54984(CYP5136A)
113940(CYP5136A)
116112(CYP5136A)
115023(CYP5136A)
115012(CYP5137A)
152151(CYP5136A)
116271(CYP5141A)
147048(CYP5141A)
97529(CYP5035B)
107483(CYP5035A)
111756(CYP5035A)
60322(CYP5035A)
127530(CYP5035A)
119940(CYP5035A)
119768(CYP5035B)
131750(CYP5035B)
77592(CYP5035A)
95659(CYP63A)
115922(CYP63A)
115537(CYP63B)
111472(CYP63A)
153962(CYP63A)
118474(CYP63A)
111473(CYP63A)
120204(CYP63A)
112411(CYP63A)
87770(CYP63A)
118491(CYP63C)
109895(CYP63C)
89930(CYP63C)
91160(CYP5143A)
113894(CYP5143A)
67761(CYP5150A)
82603(CYP5150A)
88470(CYP5150A)
154296(CYP5150A)
94672(CYP5150A)
103315(CYP5150A)
115731(CYP5150A)
80974(CYP537B)
150666(CYP537B)
100013(CYP5139NS)
119356(CYP5139NS)
118598(CYP53C2)
112429(CYP53C)
151209(CYP53C)
116910(CYP53C)
118078(CYP5139A)
116306(CYP5139NS)
112008(CYP5138NS)
115038(CYP5151B)
118080(CYP5139A)
115058(CYP5139A)
118076(CYP5139A)
117989(CYP5139A)
160359(CYP5139A)
160445(CYP5139A)
87345(CYP5154A)
116316(CYP5152A)
158908(CYP5152A)
121955(CYP5152A)
97680(CYP5152A)
140923(CYP5152A)
120902(CYP5140A)
86534(CYP61A)
148226(CYP51F)
110011(CYP5155A)
119908(CYP6001NS)
111948(CYP5122C)
112882(CYP5122A)
115604(CYP5122A)
58933(CYP5122A)
79328(CYP5122A)
79865(CYP5122A)
79864(CYP5122A)
41330(CYP5122A)
110800(CYP5122B)
116788(CYP5122A)
86021(CYP5122A)
45258(CYP5122A)
158140(CYP5222NS)
119728(CYP5156A)
109437(CYP638NS)
116287(CYP505D)
158100(CYP505D)
75518(CYP505D)
139249(CYP505D)
116692(CYP505D)

Figure S5. Phylogenetic tree of the *C. subvermispora* P450ome. The tree was constructed using the bootstrap Neighbor-Joining method. A total of 190 P450s grouped under 32 families and 47 sub-families were included in the analysis. Each P450 on the tree is shown with a tentative CYP name (along with the family and sub-family names) suffixed to the protein ID. CYP names were assigned based on homology with the annotated P450s from other fungal organisms.

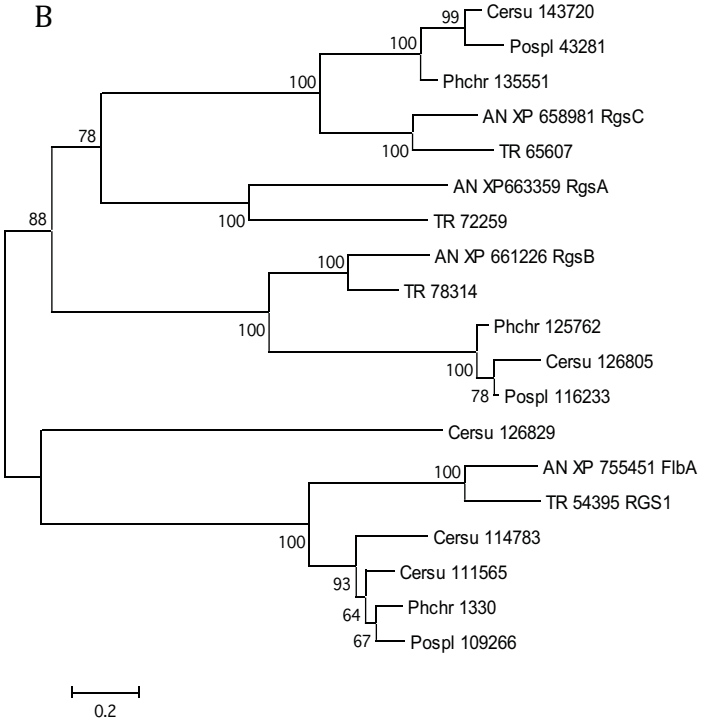
D

```
CsPh1   MDAF.ILA.........SPVPAE.EIPTTQVPVDSEQTG.GSN.TSQHCIIA*
CsPh2   MDAFFTLA.........APVPSADET.EIQVPVDEEQF..GSSTTSQHCIIA*
CsPh3   MDAFFTIA.........SPVPSE.E..PSQIPMDYEDTGSGSS.TSQHCIIA*
CsPh4   MDTFFALA.........APVPNE.ETPNPEVPADSEQF..GSNTTSQHCIIA*
CsPh5   MDAFLTVA.........EPVPAEHEIVEIEVPTDEEHSNIQS.FVSQHCIIA*
CsPh6   MDLFADLESFL....DVPDSQASTDITSSSPSTPQELERMDGNTSFSWCSIA*
CsPh7   MDEFLDIEHLL..........PPSPSSSSDATALPMDEDSAWYRPGGFCTIT*
CsPh8   MDAFEPLSM..............SIPEDMPDVPRDAEAPGTGGMSMFCTVT*
CsPh9   MYEFATPDF....PDYLMLAHRDEPSSPDEAEVTYADLDSGTGY.GGVCIIS*
Csph10  MDFFDSFEPLPVSEPTELSPCEHSTSSATDELSIPADFEHDTSSGHFFCVIA*
CsPh11  MDRFEILSLL..........ASLDDSLQPTDDVPLDAEAPSTGGSSTFCVVA*
CsPh12  MDAFFTIA..........DVEPTTTGADSPLVNEE.TQTNS.FVSQHCIIA*
```

C



B



A

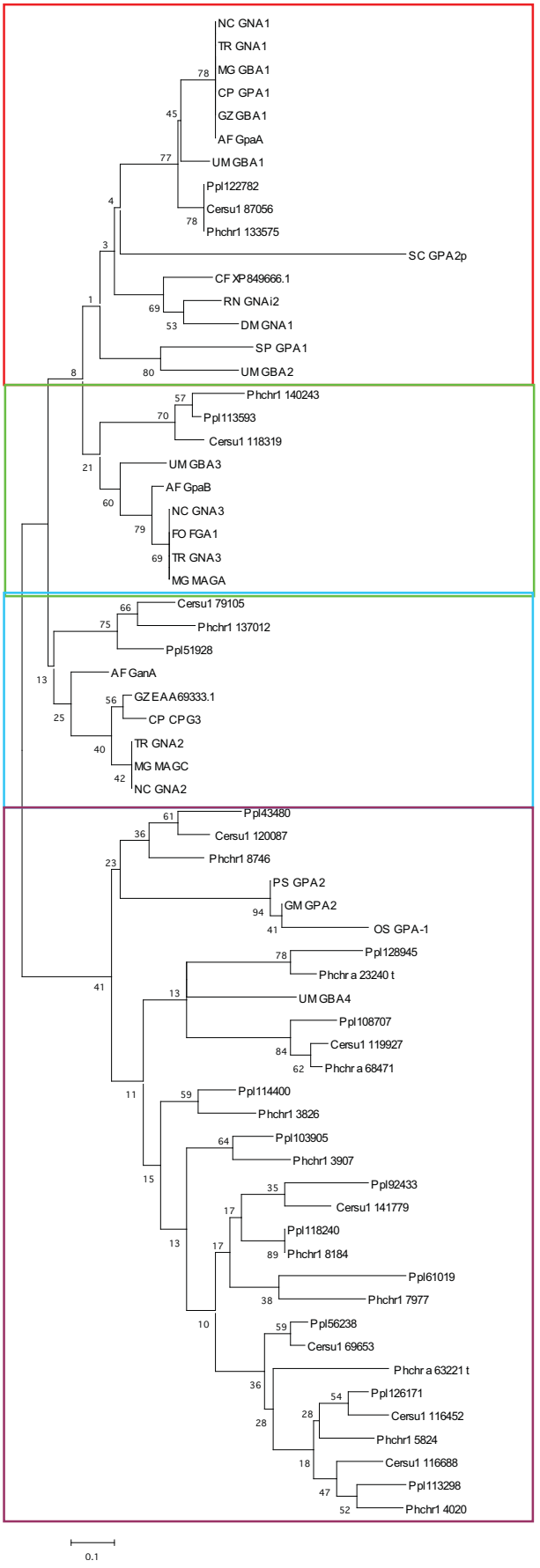


Figure S6. **Panel A**, G-alpha protein evolutionary history was inferred using the Minimum Evolution method. The optimal tree with the sum of branch length = 9.27509441 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighbor-joining algorithm was used to generate the initial tree. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 16 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4. **Panel B**, Evolutionary relationships of 19 RGS-domain proteins. Evolutionary history was inferred using the Minimum Evolution method. The optimal tree with the sum of branch length = 9.61251549 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighbor-joining algorithm was used to generate the initial tree. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 199 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.**Panel C**, Domain organization of predicted polyketide synthase (blue), nonribosomal peptide synthetase (red), and hybrid gene products (blue/red) in the *C. subvermispora* genome. Domain abbreviations: KS, keto synthase; AT, acyl transferase; DH, dehydratase; ER, enoyl reductase; KR, keto reductase; ACP, acyl carrier protein; TE, thioesterase; SAT, starter unit: ACP-transacylase;PT, product template domain; C, condensation domain; A, adenylation domain; PCP, peptidyl carrier protein; Red, reductase domain. **Panel D**, Comparison of pheromone precursor sequences. Grey, yellow and green shading mark groupings by similarity.

Table S1. Expression of putative hemicellulases of *C. subvermispora* and *P. chrysosporium*

	C. subvermispora							*P. chrysosporium*						
		LC-MS/MS[b]		Microarrays[a]					LC-MS/MS		Microarrays			
Putative activity/family		Unique Peptides		Signal (log_2)		Ratio	Prob.		Unique Peptides		Signal (log_2)		Ratio	Prob
	ID No.	Glc	BMA	Glc	BMA	BMA/Glc		ID No.	Glc	BMA	Glc	BMA	BMA/Glc	
Endo-1,4-β-mannanase/GH5	94795	-	-	9.54	14.4	**28.1**	<0.01	140501	-	**3**	9.04	12.9	**14.4**	<0.01
Endo-1,4-β-mannanase/GH5	147790	-	-	10.04	11.69	**3.15**	<0.01	5115	-	-	11.4	12.4	**2.07**	<0.01
Endo-1,4-β-xylanase/GH10	59733	-	-	9.24	13.69	**21.9**	<0.01	7045	-		9.24	9.67	1.35	0.16
Endo-1,4-β-xylanase/GH10	67561			10.30	13.71	**10.7**	<0.01	139732	-		11.21	11.40	1.14	0.13
Endo-1,4-β-xylanase/GH10	97858			9.52	11.85	**5.00**	<0.01	138345	-	-	9.23	10.43	**2.29**	0.01
Endo-1,4-β-xylanase/GH10	116326			10.69	11.20	1.43	<0.01	138715	-	-	11.16	13.54	**5.20**	<0.01
Endo-1,4-β-xylanase/GH10	85545			10.57	10.96	1.31	0.08	125669			13.38	13.84	1.37	0.02
Endo-1,4-β-xylanase/GH10	53412	-	-	10.03	9.91	0.92	0.49	7852	-	-	10.23	12.01	**3.44**	<0.01
Endo-1,4-β-xylanase/GH11	121827	-	-	10.9	11.4	1.47	0.02	133788	-		8.62	13.23	**24.6**	<0.01
Endo-1,5- α-arabinosidase/GH43	81178			10.62	10.49	0.91	0.59	4822			11.77	13.74	**3.91**	<0.01
Endo-1,5- α-arabinosidase/GH43	111990			9.51	9.40	0.93	<0.01	133070			12.05	12.72	1.59	<0.01
1,4-β-D-xylan xylohydrolase /GH43								333			11.83	12.67	1.80	<0.01
Galactan 1,3-β-galactosidase/GH43								297			10.64	11.10	1.38	0.06
α-arabinofuranosidase/GH51	162360	**6**	**2**	10.06	10.32	1.20	0.14	38548			12.38	13.06	1.61	<0.01
α-arabinofuranosidase/GH51	139314			11.52	11.59	1.05	0.62	3651		**2**	12.20	12.96	1.70	<0.01
Endo-1,4-β-	91809			8.85	8.91	1.05	0.31				12.30	14.05	**3.37**	<0.01

galactanase/GH53								138710						
Endo-1,4-β-galactanase/GH53	27988			9.88	9.85	0.98	0.80							
Endo-1,4-β-galactanase/GH53	101748			9.52	9.48	0.97	0.79							
Endo-1,4-β-galactanase/GH53	62991			11.10	11.05	0.97	0.51							
Endo-1,4-β-galactanase/GH53	61193			10.48	10.37	0.93	0.30							
Endo-1,4-β-galactanase/GH53	61145			11.69	11.56	0.92	0.25							
Xyloglucanase/GH74	79435			9.34	9.21	0.92	0.074	138266			11.24	13.55	**4.97**	<0.01
Xyloglucanase/GH74								134556			9.22	11.97	**6.73**	<0.01
Xyloglucanase/GH74								28013			10.43	13.50	**8.43**	<0.01
Xyloglucanase/GH74								9770			11.43	13.22	**3.47**	<0.01
α-fucosidase/GH95	115143			11.41	11.53	1.09	0.41	6997			12.47	13.77	**2.47**	<0.01

[a]Normalized microarray data is presented as $\log_2$ signal strength average of three fully replicated experiments. Significant accumulation (B/G ratio) of transcripts in ball milled aspen (BMA) relative to glucose grown (Glc) cultures was determined using the Moderated t-Test and associated false detection rates (Prob).

[b]Number of unique peptides detected by LC-MS/MS after 5 days growth on BMA, or Glucose medium. Complete microarray and MS/MS results are listed in GSE34636_si_Table1. Complete *P. chrysosporium* microarray and LC-MS/MS data available under GSE14736 and GSE29659, respectively. Significant ratios and/or peptide scores are highlighted in bold type.