Supplemental Information

Draft genome of *Omphalotus olearius* provides a predictive framework for sesquiterpenoid natural product biosynthesis in Basidiomycota

Grayson T. Wawrzyn, Maureen B. Quin, Swati Choudhary, Fernando López-Gallego and Claudia Schmidt-Dannert

Table of Contents

SUPPLEMENTAL DATA 39
- Supplemental Tables 39
  - Table S1 39
  - Table S2 40
  - Table S3 41
  - Table S4 42
  - Table S5 44
  - Table S6 47
- Supplemental Figures 48
  - Figure S1 48
  - Figure S2 49
  - Figure S3 50
  - Figure S4 51
  - Figure S5 52

KEY FOR SUPPLEMENTAL DATA SETS 1-3 53

SUPPLEMENTAL EXPERIMENTAL PROCEDURES 54
- *Cloning of sesquiterpene synthases* 54
- *Overexpression and purification of Omp6 and Omp7* 54
- *Homology searches, phylogenetic tree construction and biosynthetic cluster prediction* 55
- *Quantitative RT-PCR* 57

SUPPLEMENTAL REFERENCES 58

SUPPLEMENTAL DATA

Supplemental Tables

Table S1 Assembly summary for the genome of *O. olearius* (DC.) Singer strain VT-653.13. See **Figure 2** for *in vivo* production of sesquiterpenes by *O. olearius*.

Parameter	
Genome size	28.15 Mbp
Scaffold number	868
Scaffold N50	199,357 bp
Scaffold N90	22,541 bp
Estimated genome coverage	94.16 %
GC% transcript	50.08 %
GC% intergenic	45.23 %
Predicted protein coding genes	8,172
Percent of total genome	44.24 %
Mean gene length	1,523 bp

Table S2 Sequence details for cloned *O. olearius* sesquiterpene synthases. Peptide reference numbers refer to predicted sequences in **Data S1**. Gene and transcript length listed are measured in base pairs and protein length is measured in amino acids. Graphical representations of intron/exon organization were created using GSDS (Guo et al., 2007). See **Figure 3** for sesquiterpenes produced by each gene.

Gene	Peptide reference	Scaffold number	Gene start	Gene stop	Gene length	# of introns	Transcript length	Protein length
Omp1	MUStwsD_GLEAN_10001317	131	23448	22256	1192	3	999	332
Omp2	MUStwsD_GLEAN_10002575	63	27510	26270	1240	4	1002	333
Omp3	MUStwsD_GLEAN_10003938	33	137438	135914	1524	6	1104	367
Omp4	MUStwsD_GLEAN_10005581	14	149118	150371	1253	4	1035	344
Omp5a	MUStwsD_GLEAN_10000810	182	25829	24552	1277	4	1038	345
Omp5b	MUStwsD_GLEAN_10000811	182	28337	27066	1271	4	1038	345
Omp6	MUStwsD_GLEAN_10003820	35	114288	113036	1252	3	1074	357
Omp7	MUStwsD_GLEAN_10000831	179	28809	30073	1264	4	1038	345
Omp8	MUStwsD_GLEAN_10000534	223	21771	20858	913	2	1044	347
Omp9	MUStwsD_GLEAN_10000543	223	11690	10443	1247	3	1053	350
Omp10	MUStwsD_GLEAN_10000292	285	10962	9810	1152	3	987	328

Confirmed intron/exon organization

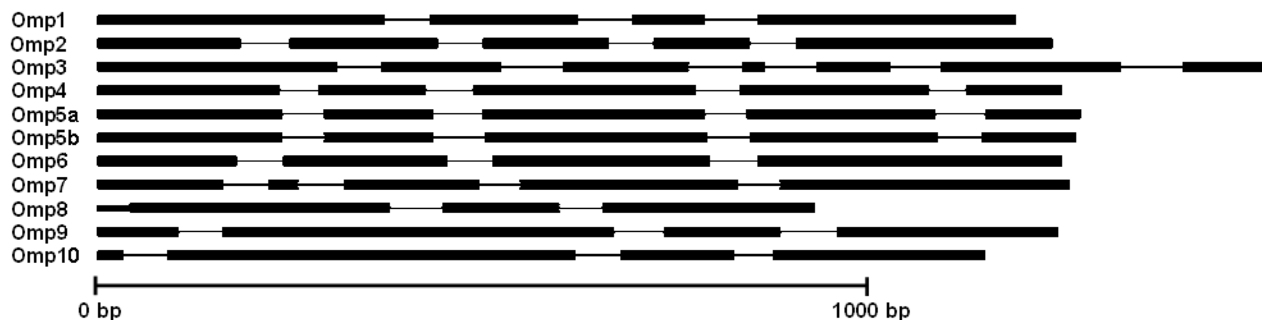


Table S3 Sequence details for cloned putative biosynthetic genes co-located in clusters with Omp6 or Omp7. Gene numbers refer to putative protein predictions in **Figure 5**. Peptide reference numbers refer to predicted sequences in **Data S1**. Gene and transcript length listed are measured in base pairs and protein length is measured in amino acids. Graphical representations of intron/exon organization were created using GSDS (Guo et al., 2007).

Gene	Peptide reference	Scaffold number	Gene start	Gene stop	Gene length	# of introns	Transcript length	Protein length
Omp6								
6-h	MUStwsD_GLEAN_10003819	35	107497	109609	2112	6	1737	579
6-i	MUStwsD_GLEAN_10003819	35	110221	112383	2162	9	1602	534
6-j	MUStwsD_GLEAN_10003821	35	115022	116998	1976	7	1599	533
6-k	MUStwsD_GLEAN_10003822	35	117853	120355	2502	11	1908	635
6-l	MUStwsD_GLEAN_10003823	35	121648	120433	1215	3	1050	349
Omp7								
7-b	MUStwsD_GLEAN_10000830	179	28175	26111	2064	8	1620	539

Confirmed intron/exon organization

Table S4 Sesquiterpene synthases (STS) identified in analyzed Basidiomycota genomes. Initial BLAST results were aligned and erroneous/incomplete sequences removed. Shown are total number of identified, putative STS (initial) and number of eliminated sequences lacking conserved active site motifs (NSE/DTE or DDxxDD) or failing to fall in the 250-425 amino acid length range expected for STS. The sequences removed accounted for ~28 % of the initial BLAST results. The final set of sequences was used for phylogenetic analysis shown in **Figure 7**. The full phylogram, including STS sequence identifiers, is provided in **Data S2**. Genome abbreviations are those used in the fungal genome database at the Joint Genome Institute (http://genome.jgi-psf.org/programs/fungi/index.jsf).

Abbreviation	Basidiomycota Reference Genome	Initial	NSE/DTE	DDxxD	Length	Final
Agabi_varbisH97_2	*Agaricus bisporus var bisporus (H97) v2.0*	9	1		3	5
Agabi_varbur_1	*Agaricus bisporus var. burnettii JB137-S8*	9			3	6
Aurde1	*Auricularia delicata SS-5 v1.0*	19	3		1	15
Bjead1_1	*Bjerkandera adusta v1.0*	12		1		11
Botbo1	*Botryobasidium botryosum v1.0*	6				6
Cersu1	*Ceriporiopsis subvermispora B*	27	3	3	3	18
Conpu1	*Coniophora puteana v1.0*	25		2	1	22
Cop	*Coprinopsis cinerea*	6				6
Cryne_H99_1	*Cryptococcus neoformans var. grubii H99*	0				0
Dacsp1	*Dacryopinax sp. DJM 731 SSP-1 v1.0*	5	2			3
Dicsq1	*Dichomitus squalens v1.0*	20	2	1	3	14
Fomme1	*Fomitiporia mediterranea v1.0*	18	2			16
Fompi1	*Fomitopsis pinicola SS1 v1.0*	22	1	2	2	17
Gansp1	*Ganoderma sp. 10597 SS1 v1.0*	27	2	1	3	21
Glotr1_1	*Gloeophyllum trabeum v1.0*	14	2	2		10
Gymlu1	*Gymnopus luxurians v1.0*	24	3	4		17
Hetan2	*Heterobasidion annosum v2.0*	16	1			15
Hypsu1	*Hypholoma sublateritium v1.0*	14	1	1	1	11
Jaaar1	*Jaapia argillacea v1.0*	12		1		11
Lacbi1	*Laccaria bicolor v2.0*	27	6	7	5	9
Malgl1	*Malassezia globosa*	0				0
Mellp1	*Melampsora laricis-populina v1.0*	0				0
Omp	***Omphalotus olearius VT-653.13***	**11**				**11**
Paxin1	*Paxillus involutus ATCC 200175 v1.0*	16	1	4		11
Phaca1	*Phanerochaete carnosa HHB-10118-Sp v1.0*	20	1	2	7	10
Phchr1	*Phanerochaete chrysosporium v2.0*	10	3		2	5
Phlbr1	*Phlebia brevispora HHB-7030 SS6 v1.0*	18	1	3	3	11
Phlgi1	*Phlebiopsis gigantea v1.0*	9		1		8
PleosPC15_2	*Pleurotus ostreatus PC15 v2.0*	16		1		15
Pospl1	*Postia placenta MAD-698*	41	11	3	5	22
Pucgr1	*Puccinia graminis*	0				0

Abbreviation	Basidiomycota Reference Genome	Initial	NSE/DTE	DDxxD	Length	Final
Punst1	*Punctularia strigosozonata v1.0*	12	1			11
Rhoba1_1	*Rhodotorula graminis strain WP1 v1.1*	0				0
Schco2	*Schizophyllum commune v2.0*	4		1		3
SerlaS7_3_2	*Serpula lacrymans S7.3 v2.0*	13	1	3	1	8
Sporo1	*Sporobolomyces roseus v1.0*	0				0
Stehi1	*Stereum hirsutum FP-91666 SS1 v1.0*	18	3	2	1	12
Trave1	*Trametes versicolor v1.0*	20	4	1		15
Treme1	*Tremella mesenterica Fries v1.0*	0				0
Ustma1	*Ustilago maydis*	0				0
Walse1	*Wallemia sebi v1.0*	0				0
Wolco1	*Wolfiporia cocos MD-104 SS10 v1.0*	22	2	2	1	17
		542	**55**	**46**	**44**	**392**

Table S5 Predicted biosynthetic gene clusters in *O. olearius* using SMURF (http://jcvi.org/smurf/index.php). Corresponding peptide reference numbers (**Data S1**) are included for each putative gene. Cluster position is indicated with the putative PKS/NRPS with a position of zero. Genes in the window are tagged as domain positive (1) if they contain at least one of the secondary metabolite domains recognized by SMURF, or domain negative (0) if they do not. The top conserved domain hit (CDD) at NCBI is listed in parentheses next to each ORFs putative function. The previously cloned *O. olearius fso1* (Welzel et al., 2005) and the fatty acid synthase (Antelo et al., 2009) are indicated. See **Figure 6** for the illudin biosynthetic cluster.

Cluster - Scaffold #	Peptide reference sequence	Cluster Position	Domain score	Top conserved domain hit
Cluster 1 – Scaffold 192				
	MUStwsD_GLEAN_10000751	-3	1	PGDH (cd05288)
	MUStwsD_GLEAN_10000750	-2	1	P450 (pfam00067)
	MUStwsD_GLEAN_10000749	-1	1	MFS (cd06174)
PKS→	MUStwsD_GLEAN_10000748	0	1	PKS (cd00833)
Cluster 2 – Scaffold 172				
	MUStwsD_GLEAN_10000886	-3	1	MFS (cd06174)
	MUStwsD_GLEAN_10000885	-2	0	DUF2461 (pfam09365)
	MUStwsD_GLEAN_10000884	-1	0	MPP_239FB (cd07379)
PKS→	MUStwsD_GLEAN_10000883	0	0	Macro_Appr_pase_like (cd02908)
Cluster 3 – Scaffold 132				
NRPS→	MUStwsD_GLEAN_10001307	0	0	SDR_e1 (cd05235)
	MUStwsD_GLEAN_10001306	1	0	None
	MUStwsD_GLEAN_10001305	2	0	AXO (cd01150)
	MUStwsD_GLEAN_10001304	3	0	BBE (pfam08031)
	MUStwsD_GLEAN_10001303	4	1	17beta-HSD-like_SDR_c (cd05374)
Cluster 4 – Scaffold 82				
	MUStwsD_GLEAN_10002091	-11	1	PHOX_C (cd02979)
	MUStwsD_GLEAN_10002090	-10	0	ADIP (pfam11559)
	MUStwsD_GLEAN_10002089	-9	0	None
	MUStwsD_GLEAN_10002088	-8	0	None
	MUStwsD_GLEAN_10002087	-7	0	BRLZ (smart00338)
	MUStwsD_GLEAN_10002086	-6	0	Retinol-DH_like (cd05327)
	MUStwsD_GLEAN_10002085	-5	1	AAT_like (cd00609)
	MUStwsD_GLEAN_10002084	-4	0	Abhydrolase_3 (pfam07859)
	MUStwsD_GLEAN_10002083	-3	0	Transferase (pfam02458)
	MUStwsD_GLEAN_10002082	-2	1	MFS (cd06174)
	MUStwsD_GLEAN_10002081	-1	1	MDR_like_2 (cd05289)
NRPS→	MUStwsD_GLEAN_10002080	0	0	Peptidase_C19R (cd02674)
	MUStwsD_GLEAN_10002079	1	0	BAH_fungalPHD (cd04710)
	MUStwsD_GLEAN_10002078	2	0	HATPase_c (cd00075)
	MUStwsD_GLEAN_10002077	3	0	RanBD (cd00835)
	MUStwsD_GLEAN_10002076	4	0	DPG_synthase (cd04188)
	MUStwsD_GLEAN_10002075	5	1	Salicylate_mono (TIGR03219)

Cluster - Scaffold #	Peptide reference sequence	Cluster Position	Domain score	Top conserved domain hit
Cluster 5 – Scaffold 74				
	MUStwsD_GLEAN_10002282	-5	1	Aldo_ket_red (cd06660)
	MUStwsD_GLEAN_10002281	-4	0	GMC_oxred_C (pfam05199)
	MUStwsD_GLEAN_10002280	-3	1	P450 (pfam00067)
	MUStwsD_GLEAN_10002279	-2	0	ALDH_F1-2_Ald2-like (cd07091)
	MUStwsD_GLEAN_10002278	-1	0	None
NRPS→	MUStwsD_GLEAN_10002277	0	0	SDR_e1 (cd05235)
	MUStwsD_GLEAN_10002276	1	1	MFS (cd06174)
	MUStwsD_GLEAN_10002275	2	1	P450 (pfam00067)
Cluster 6 – Scaffold 58				
	MUStwsD_GLEAN_10002754	-7	1	RhlG_SDR_c (cd08942)
	MUStwsD_GLEAN_10002753	-6	0	GRX_SH3BGR (cd03030)
	MUStwsD_GLEAN_10002752	-5	0	LMWPc (cd00115)
	MUStwsD_GLEAN_10002751	-4	0	ALDH_F6_MMSDH (cd07085)
	MUStwsD_GLEAN_10002750	-3	0	None
	MUStwsD_GLEAN_10002749	-2	0	PTZ00363 (PTZ00363)
	MUStwsD_GLEAN_10002748	-1	1	PHOX_C (cd02979)
NRPS→	MUStwsD_GLEAN_10002747	0	0	SDR_e1 (cd05235)
	MUStwsD_GLEAN_10002746	1	0	AR_SDR_e (cd05227)
	MUStwsD_GLEAN_10002745	2	1	CAD_like (cd08296)
	MUStwsD_GLEAN_10002744	3	1	Methyltransf_2 (pfam00891)
	MUStwsD_GLEAN_10002743	4	0	AP_endonuc_2 (pfam01261)
	MUStwsD_GLEAN_10002742	5	1	TRAPPC_Trs85 (pfam12739)
	MUStwsD_GLEAN_10002741	6	0	Mpv17_PMP22 (pfam04117)
	MUStwsD_GLEAN_10002740	7	0	Mito_carr (pfam00153)
	MUStwsD_GLEAN_10002739	8	1	SMC_prok_B (TIGR02168)
Cluster 7 – Scaffold 50				
	MUStwsD_GLEAN_10003056	-6	1	P450 (pfam00067)
	MUStwsD_GLEAN_10003055	-5	0	None
	MUStwsD_GLEAN_10003054	-4	0	Cellulase (pfam00150)
	MUStwsD_GLEAN_10003053	-3	1	Dnd_assoc_3 (TIGR03238)
	MUStwsD_GLEAN_10003052	-2	0	lucD (COG3486)
	MUStwsD_GLEAN_10003051	-1	0	AlcB (smart01006)
NRPS (*fso1*)→	MUStwsD_GLEAN_10003050	0	1	PRK12316 (PRK12316)
	MUStwsD_GLEAN_10003049	1	0	HAD_like (cd01427)
	MUStwsD_GLEAN_10003048	2	0	GT1_Glycogen_Phos (cd04300)
	MUStwsD_GLEAN_10003047	3	0	NPD_like (cd04730)
	MUStwsD_GLEAN_10003046	4	0	PRK06334 (PRK06334)
	MUStwsD_GLEAN_10003045	5	1	ABCC_MRP_domain2 (cd03244)
	MUStwsD_GLEAN_10003044	6	1	MFS (cd06174)
Cluster 8 – Scaffold 39				
	MUStwsD_GLEAN_10003602	-2	1	SPD_interacting (pfam08616)
	MUStwsD_GLEAN_10003601	-1	0	Pepsin_like (cd05471)
PKS→	MUStwsD_GLEAN_10003600	0	0	PKS (cd00833)
Cluster 9 – Scaffold 18				
	MUStwsD_GLEAN_10005128	-1	1	Enoyl_reductase_like (cd08249)
PKS→	MUStwsD_GLEAN_10005127	0	1	PKS (cd00833)
	MUStwsD_GLEAN_10005126	1	0	PLN02464 (PLN02464)
	MUStwsD_GLEAN_10005125	2	0	rpsS (PRK00357)
	MUStwsD_GLEAN_10005124	3	0	None

Cluster - Scaffold #	Peptide reference sequence	Cluster Position	Domain score	Top conserved domain hit
	MUStwsD_GLEAN_10005123	4	0	PTZ00096 (PTZ00096)
	MUStwsD_GLEAN_10005122	5	0	Ribosomal_P2 (cd05833)
	MUStwsD_GLEAN_10005121	6	1	Fungal_trans (smart00906)
Cluster 10 – Scaffold 2				
	MUStwsD_GLEAN_10007722	-5	0	ATP_synt_H (pfam05493)
	MUStwsD_GLEAN_10007721	-4	0	Bacter_Hen1 (TIGR04074)
	MUStwsD_GLEAN_10007720	-3	0	ispDF (PRK09382)
	MUStwsD_GLEAN_10007719	-2	0	RhoGAP_fSAC7_BAG7 (cd04396)
	MUStwsD_GLEAN_10007718	-1	0	Drf_GBD (pfam06371)
NRPS (FAS)→	MUStwsD_GLEAN_10007717	0	0	KR_fFAS_SDR_c_like (cd08950)
	MUStwsD_GLEAN_10007716	1	1	DEADc (cd00268)
	MUStwsD_GLEAN_10007715	2	0	GH16_Lam16A_gluc (cd02181)
	MUStwsD_GLEAN_10007714	3	0	ALDH_EDX86601 (cd07102)
	MUStwsD_GLEAN_10007713	4	0	None
	MUStwsD_GLEAN_10007712	5	0	Gtr1_RagA (pfam04670)
	MUStwsD_GLEAN_10007711	6	0	PolyPPase_VTC4_like (cd07751)
	MUStwsD_GLEAN_10007710	7	0	DUF3468 (pfam11951)
	MUStwsD_GLEAN_10007709	8	1	GAL4 (cd00067)
	MUStwsD_GLEAN_10007708	9	1	PH (cd00821)

Table S6 Comparison of the number of predicted cytochrome P450 and sesquiterpene synthase (STS) genes in *O. olearius* with other Basidiomycota genomes in the Fungal Cytochrome P450 Database (http://p450.riceblast.snu.ac.kr/index.php?a=view). Shown is also the number of total ORFs for each genome sequence as well as the percentage of total predicted ORFs that are putative P450 enzymes. See **Figure 6** for the illudin biosynthetic cluster.

Basidiomycota reference genome	STS	P450	ORF	P450 %
Agaricus bisporus	6	117	10,438	1.12 %
Bjerkandera adusta	11	188	15,473	1.22 %
Coprinus cinereus	6	139	13,410	1.04 %
Cryptococcus neoformans	0	5	6,475	0.08 %
Dichomitus squalens	14	175	12,290	1.42 %
Heterobasidion annosum	15	143	13,405	1.07 %
Laccaria bicolor	9	91	20,614	0.44 %
Malassezia globosa	0	6	4,286	0.14 %
Melampsora larici-populina 98AG31	0	29	16,694	0.17 %
Omphalotus olearius VT-653.13	**11**	**121**	**8,172**	**1.48 %**
Phanerochaete chrysosporium	5	145	10,048	1.44 %
Pleurotus ostreatus	15	141	11,603	1.22 %
Postia placenta	22	353	17,173	2.06 %
Puccinia graminis f. sp. tritici	0	18	20,567	0.09 %
Schizophyllum commune H4-8	3	115	13,181	0.87 %
Serpula lacrymans S7.3	8	161	14,495	1.11 %
Sporobolomyces roseus	0	7	5,536	0.13 %
Tremella mesenterica	0	8	8,313	0.10 %
Ustilago maydis FB1	0	21	6,950	0.30 %

Supplemental Figures

Figure S1 Mass spectra of identified sesquiterpenoid compounds. Compound numbers correspond to compound names discussed in the main text and shown in figures. See **Figure 2**, **Figure 3**, and **Figure 4** for individual compound occurrences. Values for the measured retention index (RI) (top) and the library values (bottom) are listed below each compound name.

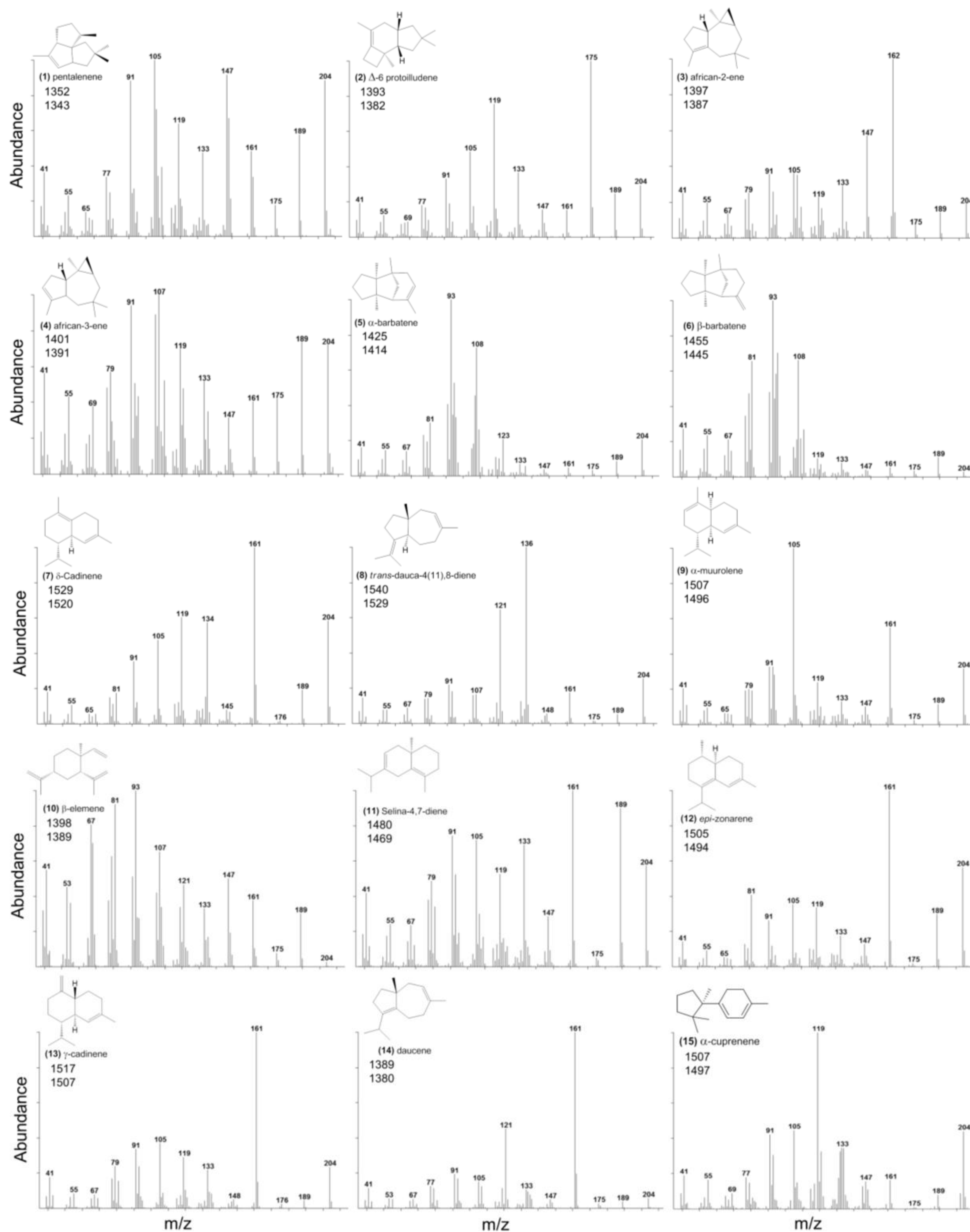


Figure S2 Time course of sesquiterpene production by *O. olearius*. Volatile metabolites produced by liquid cultures of *O. olearius* were sampled by solid phase microextraction (SPME) and analyzed by Gas Chromatography-Mass Spectrometry (GC-MS). Identified sesquiterpene compounds are numbered and structures are shown in **Figure 2** (see **Figure S1** for mass spectra for each compound). One major sesquiterpene product and many small terpene peaks could not be conclusively identified. Culture conditions between the two cultures shown here (**a** and **b**) and the culture shown in **Figure 2** were identical but resulted in significantly different relative sesquiterpene levels.

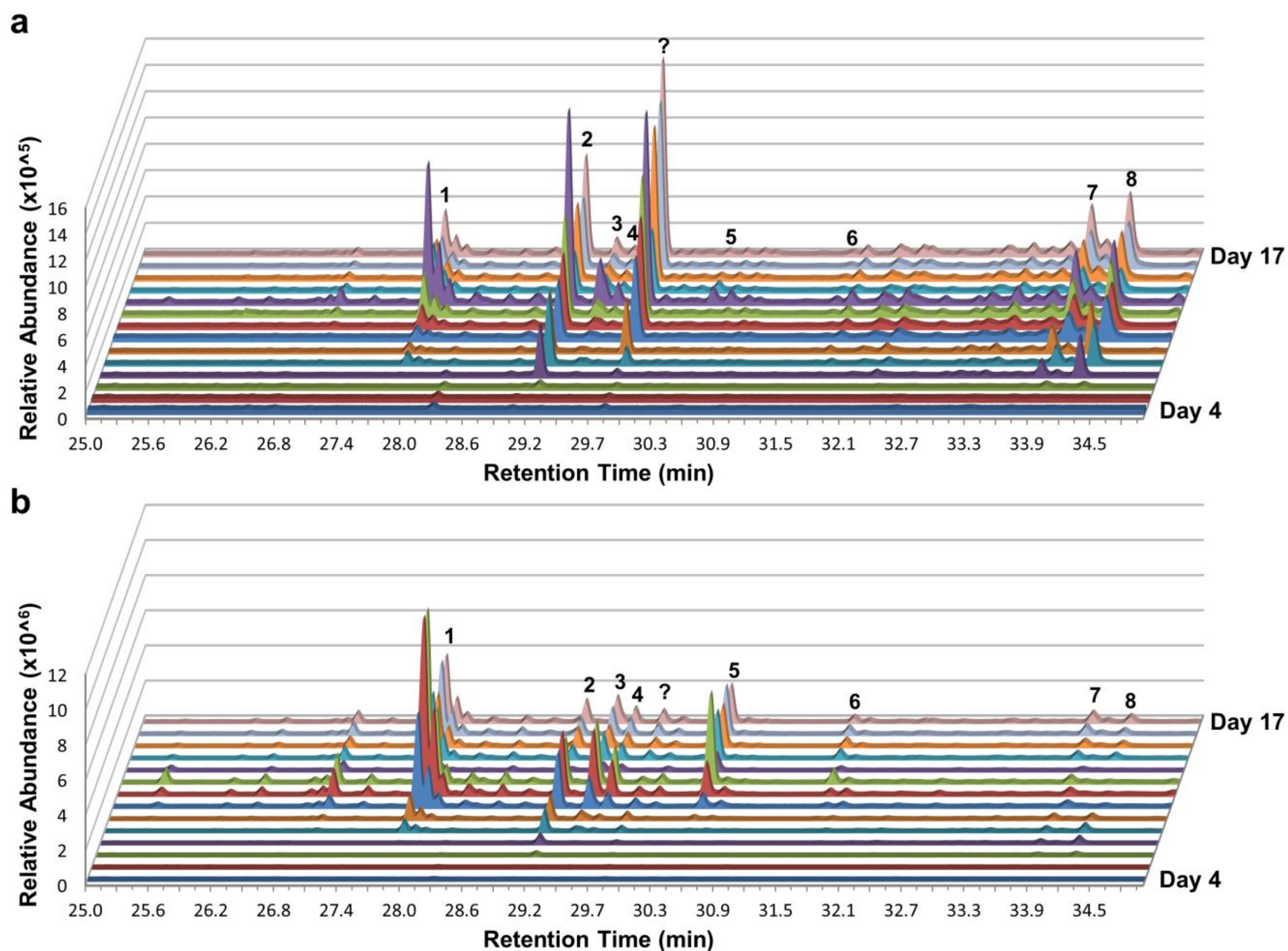
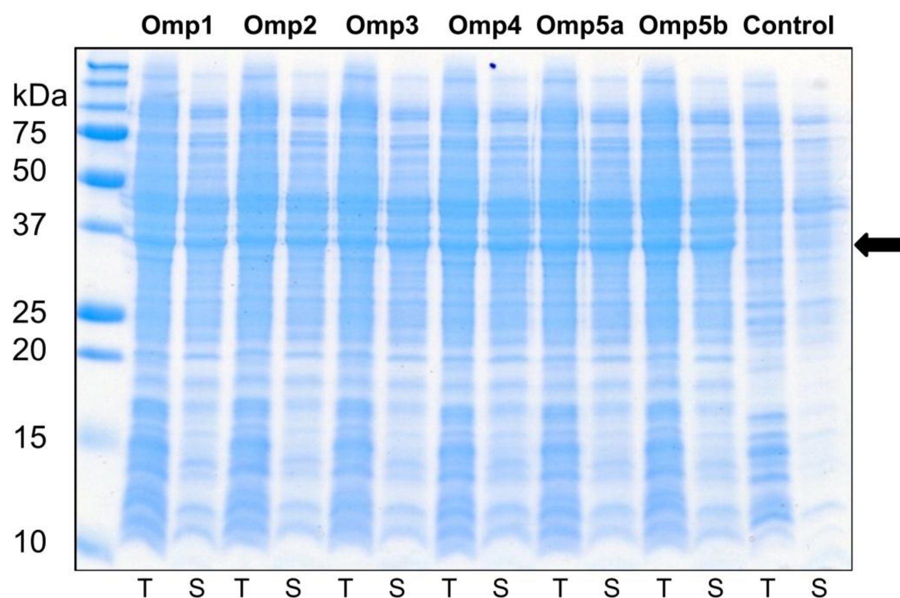


Figure S3 *O. olearius* sesquiterpene synthase (Omp1-10) expression analysis in *E. coli*. Protein levels for each Omp protein, under constitutive expression in *E. coli* strain BL21, were visualized by SDS-PAGE and Coomassie staining. Total lysate (T) or the soluble fraction (S), collected after centrifugation, was analyzed. An empty vector was used as a control. Expected sizes in kDa for the analyzed Omp proteins are 38.2 (Omp1), 38.1 (Omp2), 41.8 (Omp3), 39.1 (Omp4), 39.2 (Omp5a), 39.5 (Omp5b), 41.1 (Omp6), 39.8 (Omp7), 40.0 (Omp8), 39.5 (Omp9), and 36.6 (Omp10). See **Figure 3** for sesquiterpene production in *E. coli* by each sesquiterpene synthase.



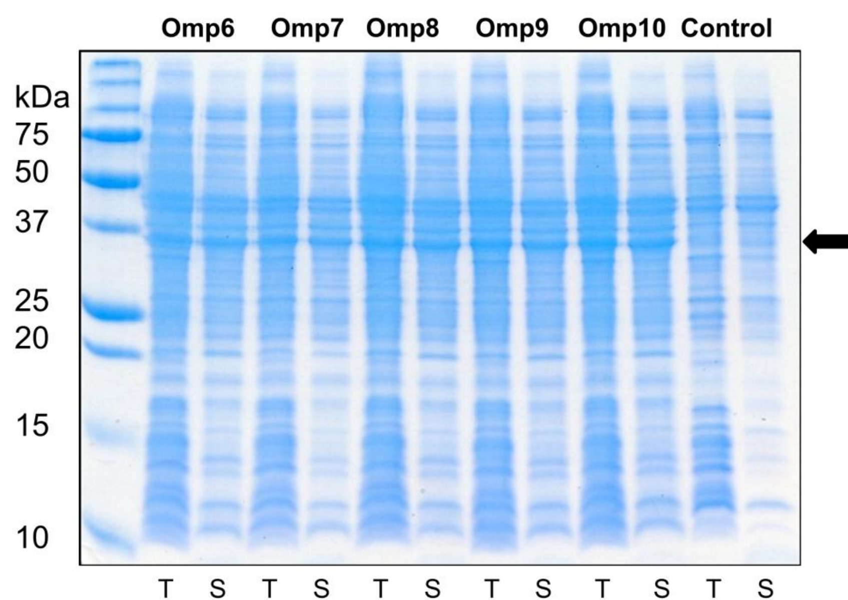


Figure S4 Expression of Omp6, Omp7 and putative modifying enzymes in *O. olearius* liquid cultures. qRT-PCR analysis of the putative Omp6 biosynthetic cluster showed that these adjacent genes displayed analogous expression profiles, suggesting that Omp6, 6-j, 6-k and 6-l are transcriptionally co-regulated. Comparable expression patterns of Omp7 and 7-b also suggest that these two genes belong to a functional biosynthetic cluster. Levels of Δ-6 protoilludene in culture headspace increased steadily over time, likely due to gradual accumulation of Δ-6 protoilludene synthase enzymes. Transcript and Δ-6 protoilludene levels were normalized in order to have each respective maximum align with a value of one. Error bars represent the standard deviation of 3 replicates. See **Table S3** for gene information and **Figure 6** for illudin cluster organization.

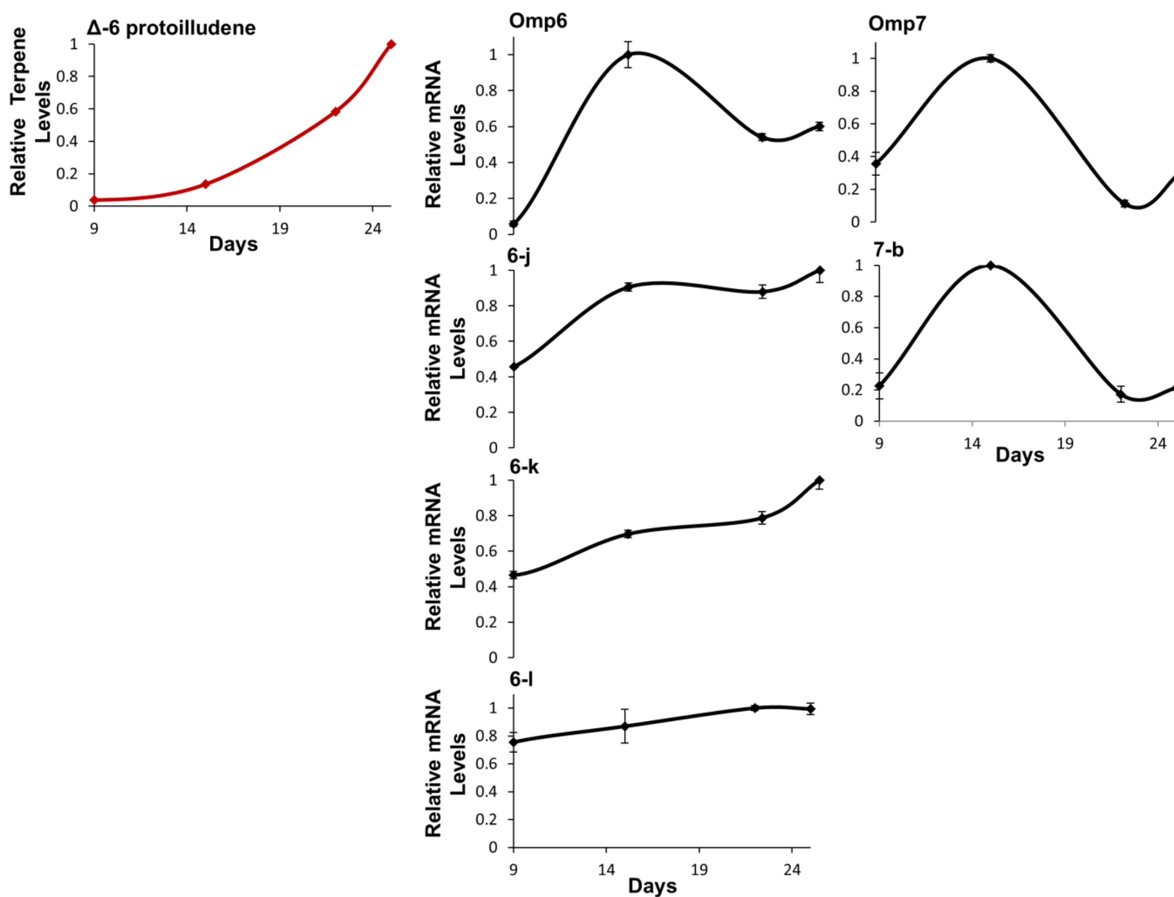
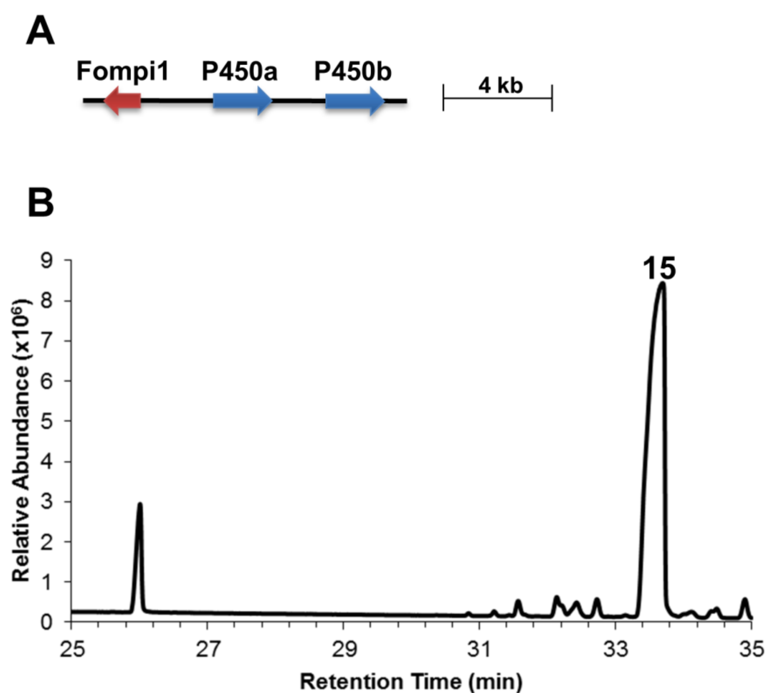


Figure S5 Identification and characterization of a new STS, Fompi1, from *Fomitopsis pinicola* strain CS-1. (a) Biosynthetic cluster prediction for the genomic region surrounding Fompi1. The cluster consists of a STS and two putative P450 enzymes. (b) GC-MS analysis of *E. coli* cultures expressing Fompi1. The major product, α-cuprenene **15**, is labeled and the structure can be found in **Figure S1**. The second largest peak, with a retention time of roughly 26 min, is the endogenous production of indole in *E. coli*, as described previously. α-cuprenene was identified by comparison with α-cuprenene produced by Cop6 from *C. cinereus* (Agger et al., 2009). See the phylogram in **Figure 7** for information on gene homology.



KEY FOR SUPPLEMENTAL DATA SETS 1-3

The following describes file contents included in the supplementary data sets 1-3.

Data S1 Gene and protein sequence predictions & Protein function annotation results.

Scaffold.fa.Glean.cds	Coding sequences for all predicted proteins
Scaffold.fa.Glean.pep	Amino acid sequences for all predicted proteins
*. kegg.list.anno	KEGG annotation result file
*. kegg.list.filter	Result file in M8 format of BLAST alignment with KEGG
KEGG_MAP/*	KEGG map of metabolic pathway
*.kog.list.anno	KOG annotation result file
*.kog.list.filter	Result file in M8 format of BLAST alignment with KOG
*.swissprot.list.anno	SwissProt annotation result file
*.swissprot.list.filter	Result file in M8 format of BLAST alignment with SwissProt
*.trembl.list.anno	Trembl annotation result file
*.trembl.list.filter	Result file in M8 format of BLAST alignment with Trembl
*.nr.list.anno	NR annotation result file
*.nr.list.filter	Result file in M8 format of BLAST alignment with NR
*.iprscan.gene.ipr	Ipr number and annotation result
*.iprscan.gene.GO	GO number and annotation result

Data S2 Complete sequence information for characterized *O. olearius* sesquiterpene synthases (STS) & Alignment and tree file of Basidiomycota sesquiterpene synthases.

OoleariusSTS_gDNA_cds_pep.fas	Coding sequences, amino acid sequences, and genomic regions for all 11 STS, including biosynthetic genes.
allsequences_clustalw.fas	FASTA alignment of all basidiomycota STS for Figure 6
allsequences_clustalw.mts	MEGA tree file used to generate Figure 6

Data S3 – Predicted P450 sequences in *O. olearius*.

O_olearius_P450.fas	FASTA file containing putative P450 enzymes

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cloning of sesquiterpene synthases

O. olearius or *F. pinicola* mRNA was extracted using the TRIzol® reagent (Invitrogen, Carlsbad, CA) following the manufacturers protocol. The extracted mRNA was used with the Superscript® III First-Strand Synthesis SuperMix kit (Invitrogen, Carlsbad, CA) to synthesize cDNA for subsequent amplification of the biosynthetic genes.

A first round of PCR with *Taq* polymerase (New England Biolabs, Ipswich, MA) was carried out with primers designed against the conserved DDxxD motif and the NSE/DTE motifs of terpene synthases in conjunction with outside gene-specific primers. The two amplification products obtained for each of the putative STS genes were then assembled by overlap-extension PCR to the predicted full-length gene sequences. PCR products were TA ligated into the pCR2.1-TOPO vector (Invitrogen, Carlsbad, CA) and transformed into *E. coli* TOP10. Cloned sequences were verified by DNA sequencing. Mutations introduced by PCR were repaired by site-directed mutagenesis using the QuikChange protocol (Stratagene, La Jolla, CA) and missing 5' or 3' sequences were added by overlap extension PCR. The chimeric Omp8 gene was synthesized by Life Technologies Corporation (previously GeneArt®). The final STS sequences were each cloned into our in-house vector pUCBB (Vick et al.) by ligation-independent cloning (Aslanidis and de Jong, 1990) for expression under the control of a constitutive *lac* promoter (Schmidt-Dannert et al., 2000) for subsequent analysis of recombinant sesquiterpene production in *E. coli* BL21.

Overexpression and purification of Omp6 and Omp7

Omp6 and Omp7 were cloned into pET32b(+) (Novagen) using *Bam*HI and *Not*I restriction sites. Overnight cultures of *E. coli* BL21 (DE3) cells harboring Omp6-pET32b and Omp7-pET32b were used to inoculate 0.5 L of LB media supplemented with 100 µg/ml ampicillin. The cultures were incubated at 30 °C with shaking at 300 rpm until an OD_{600} of 0.6 was reached. Protein overexpression was induced by addition of 1 mM IPTG and incubation at 30 °C for an additional 4 hours. Cells were harvested by centrifugation at 3000 x g for 30 min at 4 °C and the resulting pellet was resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$). The cells were lysed by sonication and the slurry was clarified by centrifugation at 10,000 x g for 30 min. The soluble thioredoxin-tagged proteins were purified metal-affinity chromatography, using BD TALON resin (Clontech, Mountain View, CA), by following the manufacturer's procedures. Cleavage of the thioredoxin tag was carried out using enterokinase (Novagen) at 4 °C overnight, and the enterokinase was removed from the reaction mixture using the agarose spin columns provided by the manufacturer. The complete cleavage of the thioredoxin tag was verified by SDS-PAGE, and the cleaved tag was separated from Omp6 and Omp7 using Ni-NTA Spin Columns (QIAGEN). The proteins were judged to be > 90 % pure by SDS-PAGE, and were subsequently used for in vitro enzymatic assays, which were performed as described previously (Lopez-Gallego et al., 2010).

Homology searches, phylogenetic tree construction and biosynthetic cluster prediction

The six fungal STS previously identified from *C. cinereus* (Agger et al., 2009) used with the NCBI BLAST (Altschul et al., 1990) software to perform homology searches of the *O. olearius* assembly. A BLAST search with Cop1-3, which share greater than 45 % amino acid sequence identity, identified a set of three putative sesquiterpene synthases (STS) from *O. olearius*, Omp1-3. A search using Cop4 returned three more putative STS, Omp4, 5a, and 5b.

Omp5a and 5b shared very high amino acid sequence identity (89%) and are adjacent to one another on the same scaffold. Two full-length putative STS sequences (Omp9 & 10) and one partial sequence missing ~100 amino acids from its N-terminal region (Omp8) were identified that were most closely related to Cop6. Two additional putative STS from *O. olearius,* Omp6 and Omp7, were identified that were most closely related to Cop5.

Following the identification of 11 STS sequences from *O. olearius*, all 17 STS protein sequences from both *C. cinereus* and *O. olearius* were used to search basidiomycete genomes hosted by the Joint Genome Institute under the Fungal Genomics Program (http://genome.jgi-psf.org/programs/fungi/index.jsf) (Martin et al., 2011).

Upon identification of putative STS amino acid sequences in other basidiomycete genomes, alignments were computed using ClustalW (Thompson et al., 2002) and phylogenetic analyses were conducted using MEGA version 5.05 (Tamura et al., 2011) using the default parameters for the Neighbor-Joining method (Saitou and Nei, 1987) with a bootstrap test of phylogeny (500 replicates) (Felsenstein, 1985). Alignments were manually refined by removing STS sequences that did not contain both of the well conserved metal-binding motifs characteristic for sesquiterpene synthases (Christianson, 2006), showed clear mistakes in splicing predictions, or were not a characteristic length (250-425 amino acids) for this class of enzyme.

For the identification putative sesquiterpene biosynthetic gene clusters, each of the scaffolds containing the 11 STS from *O. olearius* and the scaffold containing *Fompi1* were analyzed for the presence of nearby enzymes implicated in secondary metabolite biosynthesis. Gene predictions were repeated using Augustus (Stanke et al., 2004) and nearby genes were assigned a putative function based on the presence of conserved domains (CDD).

Quantitative RT-PCR

Liquid cultures of *O. olearius* VT-653.13 were inoculated and grown as described above. At different time-points, the culture headspace was assayed for sesquiterpene production, and fungal tissue was harvested. *O. olearius* mRNA and cDNA were prepared using the protocols described above. Quantitative real-time PCR (qRT-PCR) was conducted on an Applied Biosystems 7900HT Fast Real-Time PCR System, using a previously described protocol (Kurrasch et al., 2004). cDNAs (100 ng) were added to a 10 µl qRT-PCR mix containing 5 µl Power SYBR Green PCR Master Mix (Applied Biosystems) and 150 nM primers. Each reaction was performed in triplicate, and non-reverse-transcribed samples were used as controls. All primer pairs were validated by cDNA template titration, and their efficiencies were assessed using the slope of the qRT-PCR standard curve. Expression levels of the housekeeping gene beta-tubulin were used to determine changes in transcript levels of each gene (***n***-fold) by comparing mean threshold cycle values.

SUPPLEMENTAL REFERENCES

Agger, S., Lopez-Gallego, F., and Schmidt-Dannert, C. (2009). Diversity of sesquiterpene synthases in the basidiomycete *Coprinus cinereus*. Mol. Microbiol. 72, 1181-1195.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Antelo, L., Schlipp, A., Hof, C., Eisfeld, K., Berg, H., Hornbogen, T., Zocher, R., and Anke, H. (2009). The fatty acid synthase of the basidiomycete *Omphalotus olearius* is a single polypeptide. Z. Naturforsch. (C) 64, 244-250.

Aslanidis, C., and de Jong, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 18, 6069-6074.

Christianson, D.W. (2006). Structural biology and chemistry of the terpenoid cyclases. Chem. Rev. 106, 3412-3442.

Felsenstein, J. (1985). Confidence-limits on phylogenies - an approach using the bootstrap. Evolution 39, 783-791.

Guo, A.-Y., Zhu, Q.-H., Chen, X., and Luo, J.-C. (2007). GSDS: a gene structure display server. Yi chuan 29, 1023-1026.

Kurrasch, D.M., Huang, J., Wilkie, T.M., and Repa, J.J. (2004). Quantitative real-time polymerase chain reaction measurement of regulators of G-protein signaling mRNA levels in mouse tissues. Methods Enzymol 389, 3-15.

Lopez-Gallego, F., Wawrzyn, G.T., and Schmidt-Dannert, C. (2010). Selectivity of fungal sesquiterpene synthases: Role of the active site's H-1 alpha loop in catalysis. Appl. Environ. Microbiol. 76, 7723-7733.

Martin, F., Cullen, D., Hibbett, D., Pisabarro, A., Spatafora, J.W., Baker, S.E., and Grigoriev, I.V. (2011). Sequencing the fungal tree of life. New Phytol. 190, 818-821.

Saitou, N., and Nei, M. (1987). The neighbor-joining method - a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406-425.

Schmidt-Dannert, C., Umeno, D., and Arnold, F.H. (2000). Molecular breeding of carotenoid biosynthetic pathways. Nat. Biotechnol. 18, 750-753.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32, W309-W312.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-2739.

Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using ClustalW and ClustalX. Current Protocols in Bioinformatics Chapter 2, Unit 2.3.

Vick, J.E., Johnson, E.T., Choudhary, S., Bloch, S.E., Lopez-Gallego, F., Srivastava, P., Tikh, I.B., Wawrzyn, G.T., and Schmidt-Dannert, C. Optimized compatible set of BioBrick vectors for metabolic pathway engineering. Appl. Microbiol. Biotechnol. 92, 1275-1286.

Welzel, K., Eisfeld, K., Antelo, L., Anke, T., and Anke, H. (2005). Characterization of the ferrichrome A biosynthetic gene cluster in the homobasidiomycete *Omphalotus olearius*. FEMS Microbiol. Lett. 249, 157-163.