

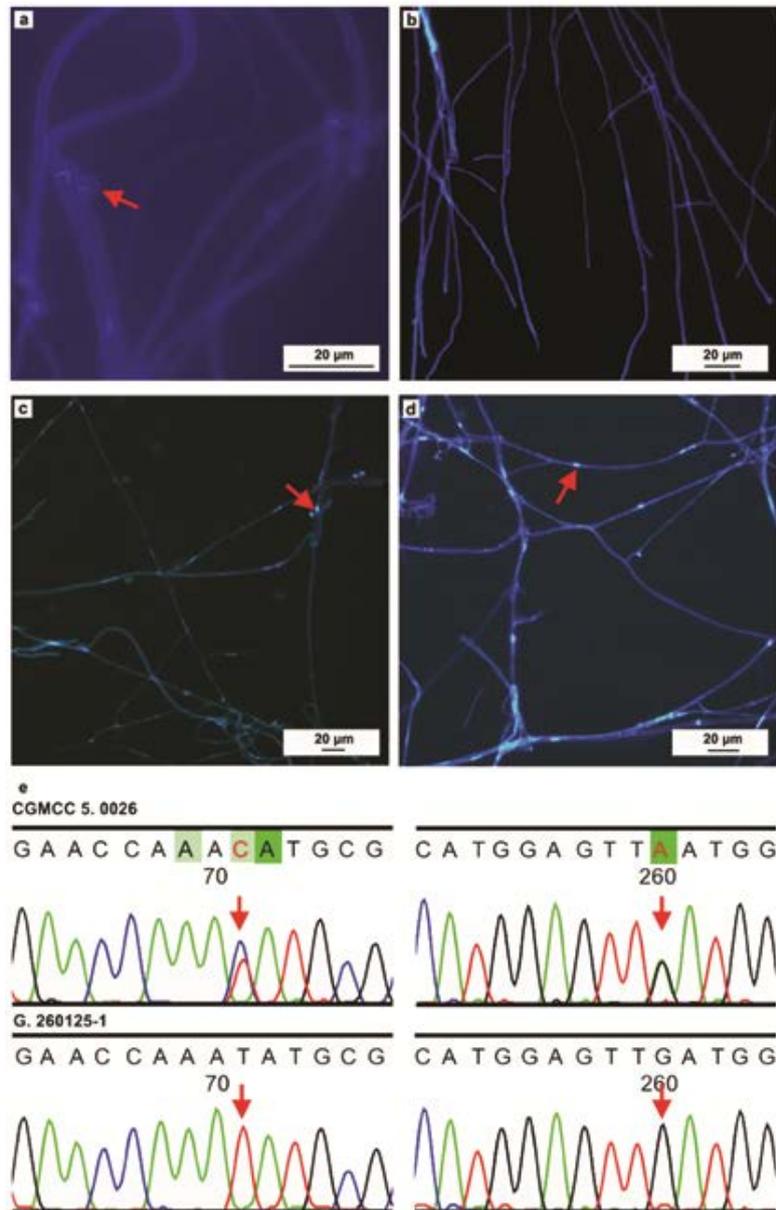
Supplementary Information

Genome sequence of the model medicinal mushroom *Ganoderma lucidum*

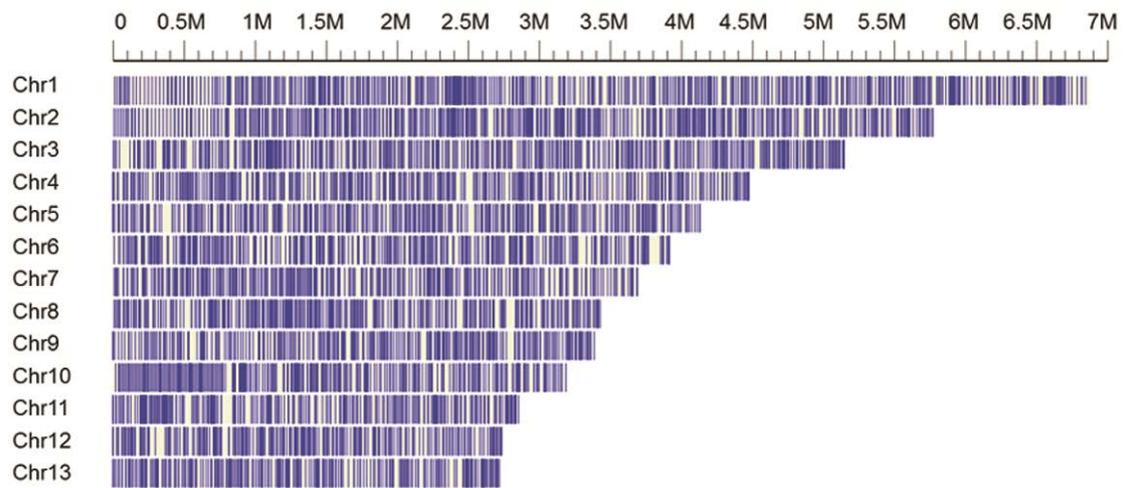
Shilin Chen^{1,*}, Jiang Xu^{1,*}, Chang Liu^{1,*}, Yingjie Zhu¹, David R. Nelson², Shiguo Zhou³, Chunfang Li¹, Lizhi Wang¹, Xu Guo¹, Yongzhen Sun¹, Hongmei Luo¹, Ying Li¹, Jingyuan Song¹, Bernard Henrissat⁴, Anthony Levasseur⁵, Jun Qian¹, Jianqin Li¹, Xiang Luo¹, Linchun Shi¹, Liu He¹, Li Xiang¹, Xiaolan Xu¹, Yunyun Niu¹, Qiushi Li¹, Mira V. Han⁶, Haixia Yan¹, Jin Zhang¹, Haimei Chen¹, Aiping Lv⁷, Zhen Wang¹, Mingzhu Liu¹, David C. Schwartz³ & Chao Sun¹

¹ Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China. ² Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. ³ Laboratory for Molecular and Computational Genomics, Department of Chemistry, Laboratory of Genetics, UW Biotechnology Center, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA. ⁴ Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, CNRS UMR 7257, 163 Avenue de Luminy, 13288 Marseille Cedex 09, France. ⁵ INRA, UMR1163 de Biotechnologie des Champignons Filamenteux, Aix-Marseille Université, ESIL, 163 Avenue de Luminy, CP 925, 13288 Marseille Cedex 09, France. ⁶ National Evolutionary Synthesis Center, 2024 W. Main Street A200, Durham, North Carolina 27705, USA. ⁷ China Academy of Chinese Medical Sciences, Beijing 100700, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S. C. (email: slchen@implad.ac.cn) or to C. S. (email: sunchao_1972@hotmail.com)

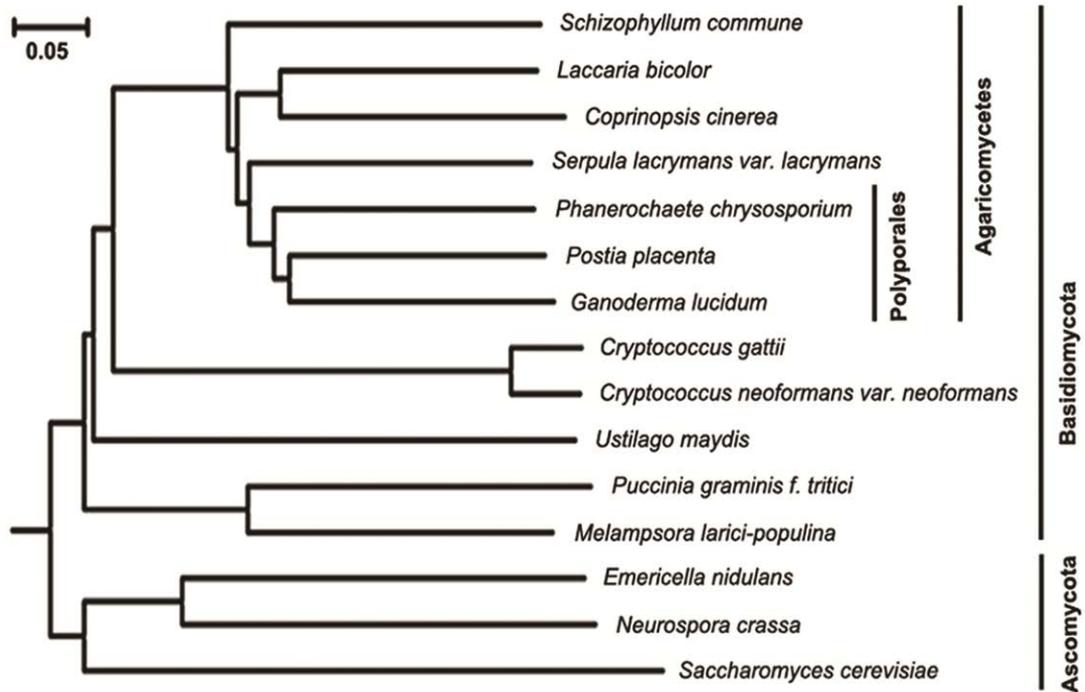
SUPPLEMENTARY FIGURES



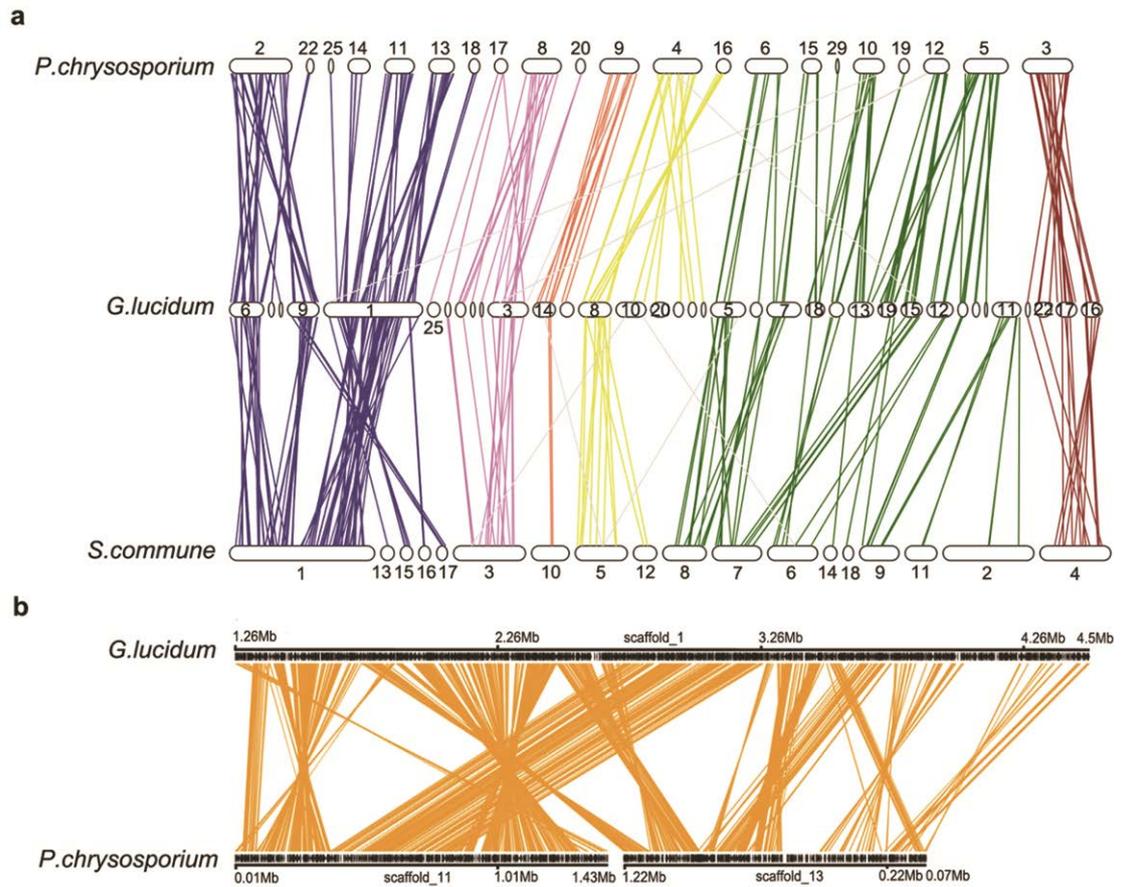
Supplementary Figure S1. Monokaryotic mycelia verified by Calcofluor White staining, DAPI staining and SNP-PCR. (a) Dikaryotic mycelia from the CGMCC5.0026 strain stained by Calcofluor White. The formation of a morphological structure known as the clamp connection is shown. (b) Monokaryotic mycelia from the G.260125-1 strain stained by Calcofluor White, and no clamp connection was found. (c) Dikaryotic mycelia from the CGMCC5.0026 strain stained by DAPI, and the two nuclei are located in close proximity within a single cell. (d) Monokaryotic mycelia from the G.260125-1 strain stained by DAPI, demonstrating a single nucleus per cell. (e) Mycelia from the dikaryotic strain CGMCC5.0026 and the monokaryotic strain G.260125-1 were analyzed by SNP-PCR, and two SNPs are shown.



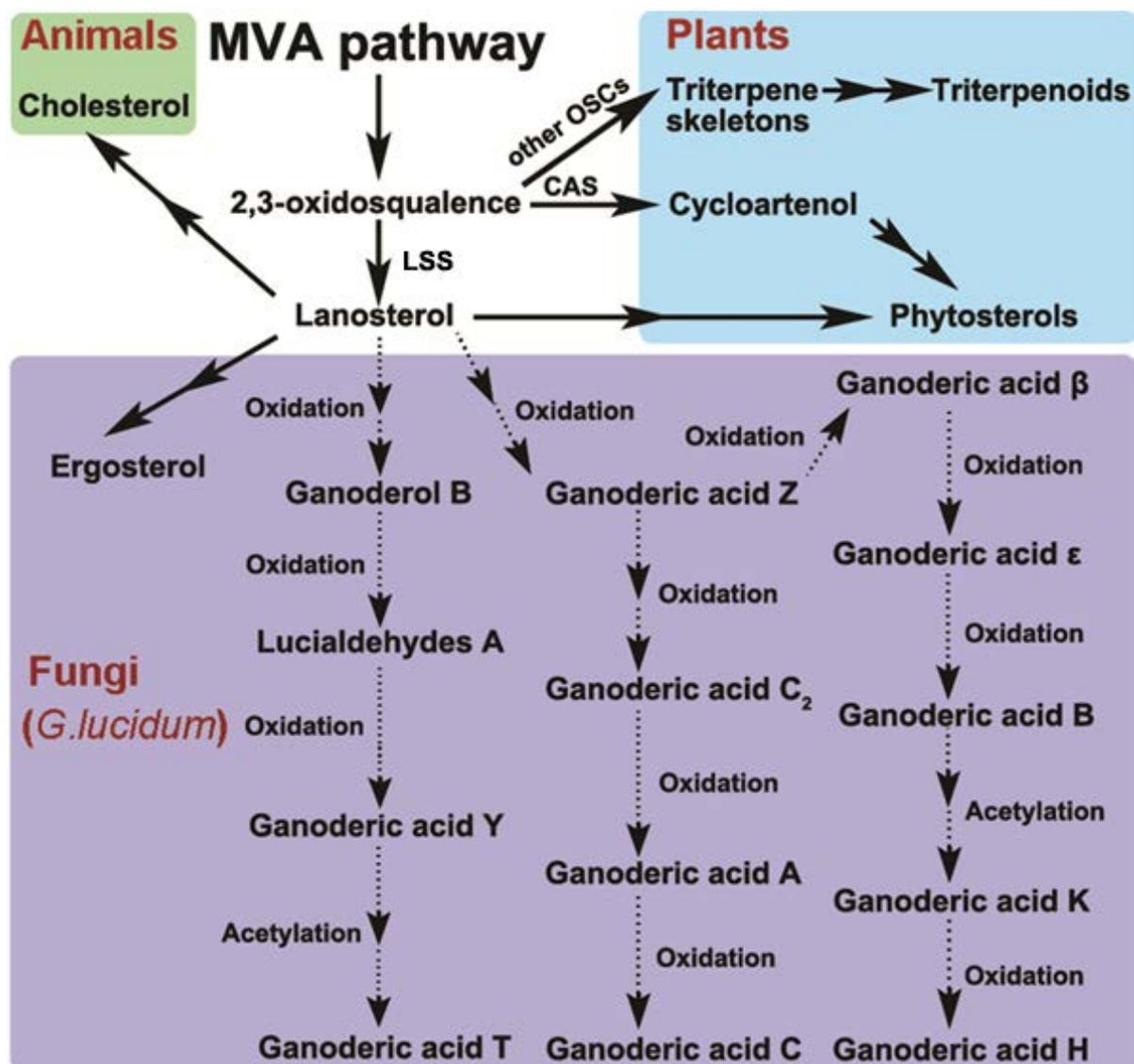
Supplementary Figure S2. Optical map of *G. lucidum* consisting of thirteen chromosomes. Each vertical line represents a *SpeI* restriction site. Chromosomes are ranked by size.



Supplementary Figure S3. Phylogenetic analysis among fungi in basidiomycetes based on 296 single copy orthologous genes. A neighbor-joining phylogenetic tree (*P-distance* model) of the fungal species was constructed using PAUP(4b10), and a bootstrap analysis with 1,000 replications was performed to evaluate the stability of the phylogenetic tree.

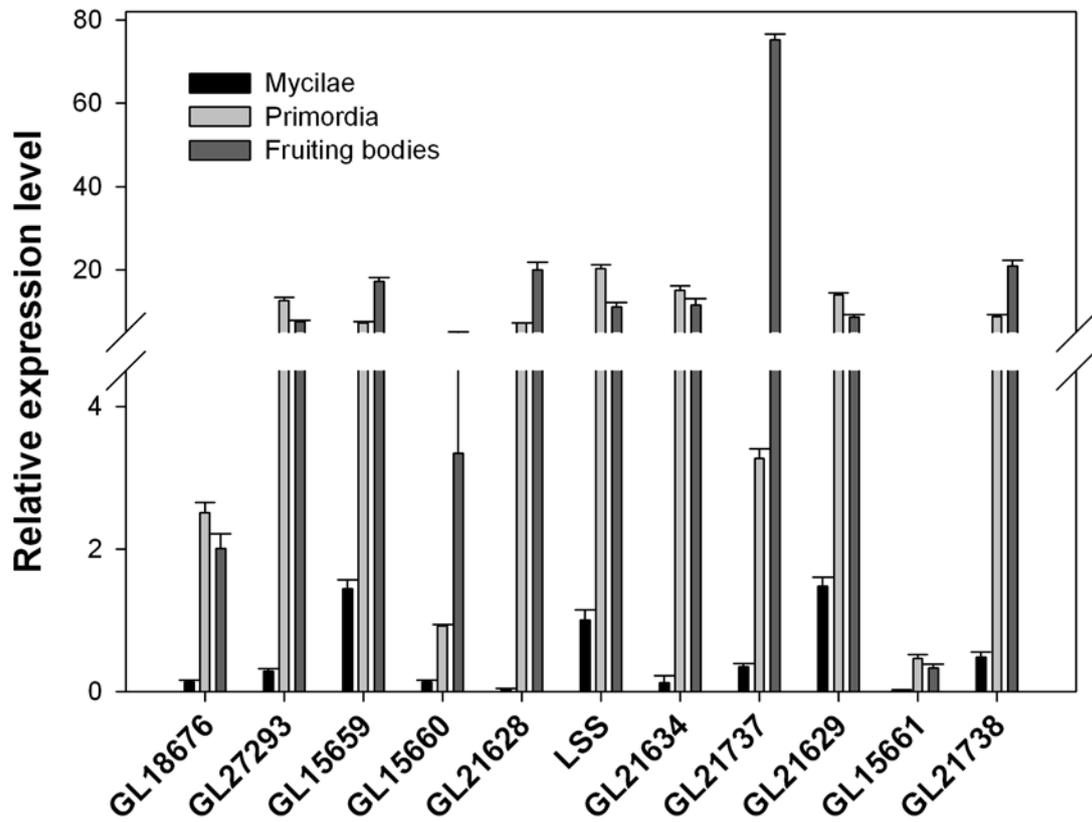


Supplementary Figure S4. Synteny analysis of *G. lucidum* and its close relatives. (a) Whole genome synteny analysis of *G. lucidum*, *P. chrysosporium* and *S. commune*. **(b)** Synteny identification between *G. lucidum* and *P. chrysosporium* at scaffold level.

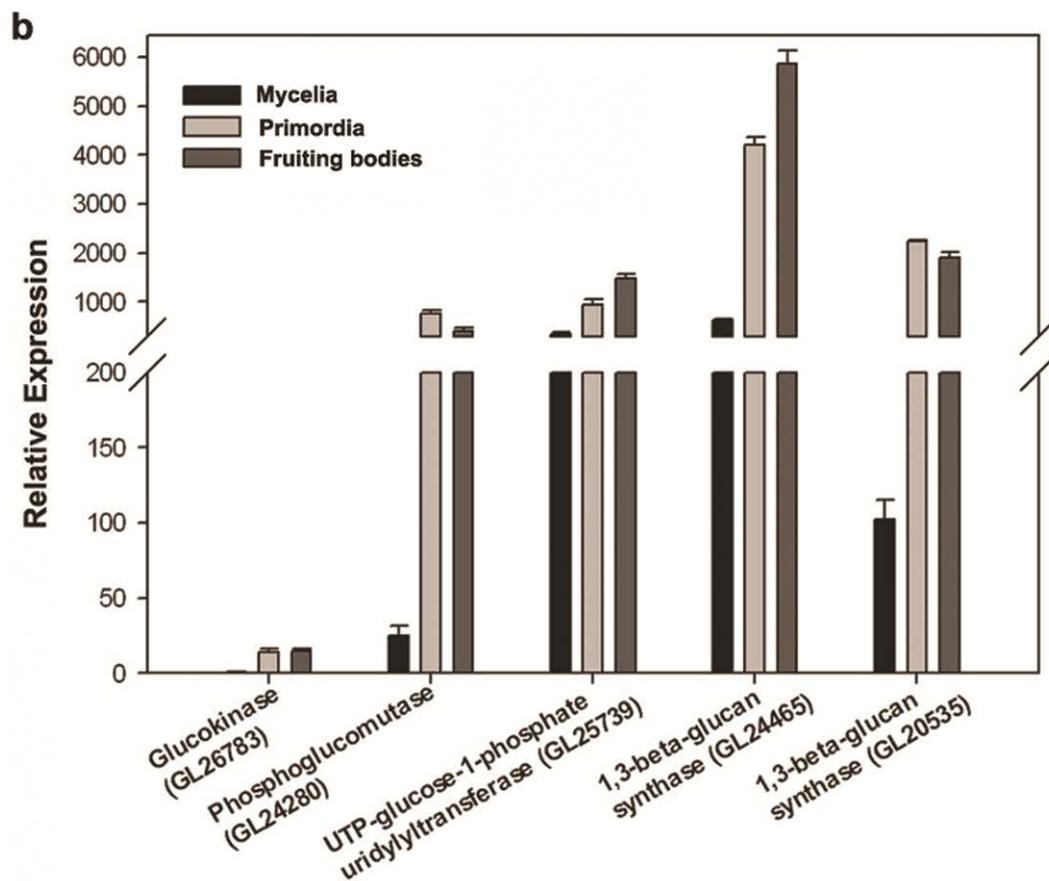
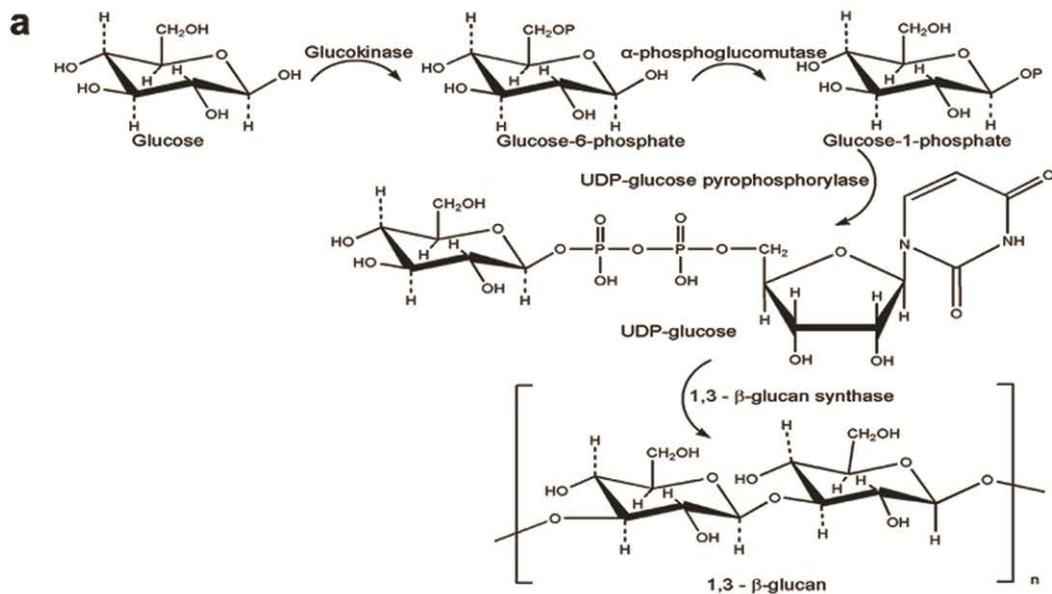


Supplementary Figure S5. Triterpenoid biosynthetic pathway in *G. lucidum*. Lanosterol is synthesized through the MVA pathway. In animals, lanosterol is metabolized to generate cholesterol (shaded green square on the left). In plants, 2,3-oxidosqualene and lanosterol are metabolized to produce triterpenes and phytosterols (shaded light-blue square on the right). In *G. lucidum*, multiple triterpenoid compounds are synthesized through oxidation reactions (shaded in purple square at the bottom). Each single arrow represents a single-step chemical reaction. Double arrows represent multiple-step chemical reactions. Previously known pathways are indicated by solid arrows. Novel pathways proposed for *G. lucidum* are indicated by dotted arrows. LSS: lanosterol synthase; CAS: cycloartenol synthase; OSC: 2,3-oxidosqualene cyclase.

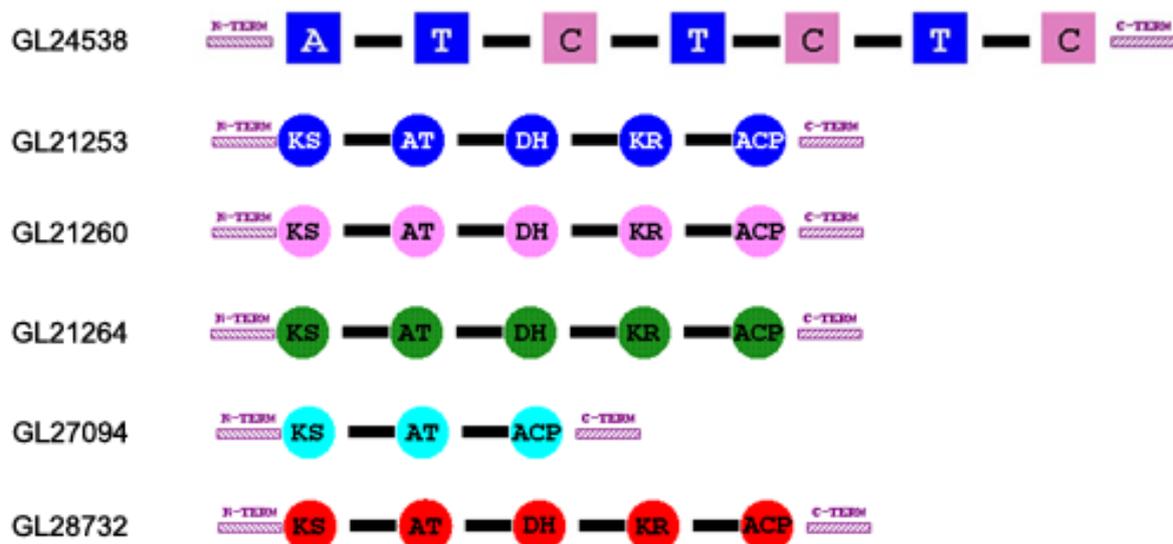
Supplementary Figure S6. A phylogenetic analysis of CYP450s from *G. lucidum*, *P. placenta* and *P. chrysosporium*. A total of 78 CYP450s co-expressed with lanosterol synthases (LSS) from 21 families in *G. lucidum* are shaded in yellow. The minimal evolution tree was generated by heuristic search using the Close-Neighbour-Interchange (CNI) algorithm in MEGA (version 5.05). A bootstrap value based on 1,000 replications was set and displayed on the tree branches with a value larger than 50.



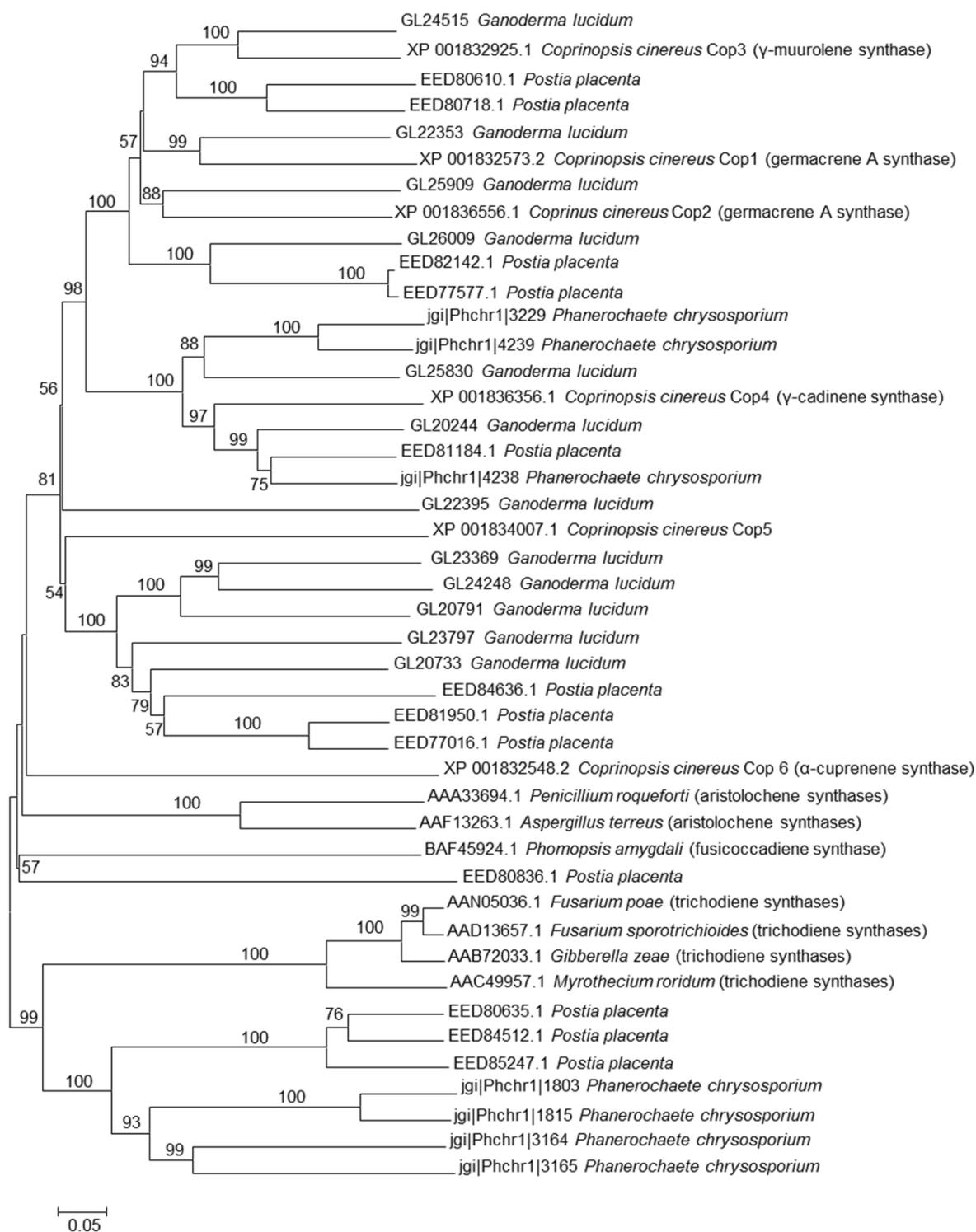
Supplementary Figure S7. Real-time PCR analysis of ten genes in close proximity to *LSS* on chromosome 6.



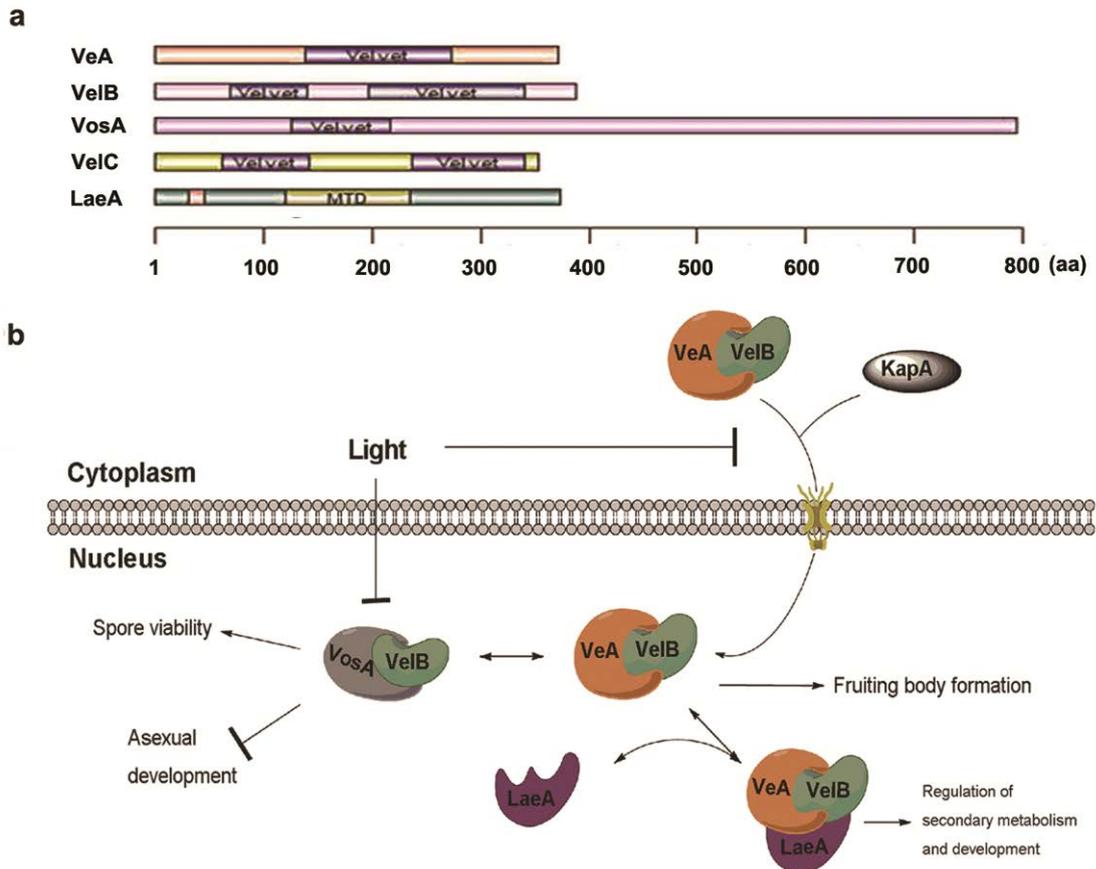
Supplementary Figure S8. Genes involved in polysaccharide biosynthesis of in *G. lucidum*. (a) The polysaccharide biosynthetic pathway; (b) Expression analysis of genes involved in polysaccharide biosynthesis across three developmental stages (mycelia, primordia and fruiting bodies). The X axis shows the gene names.



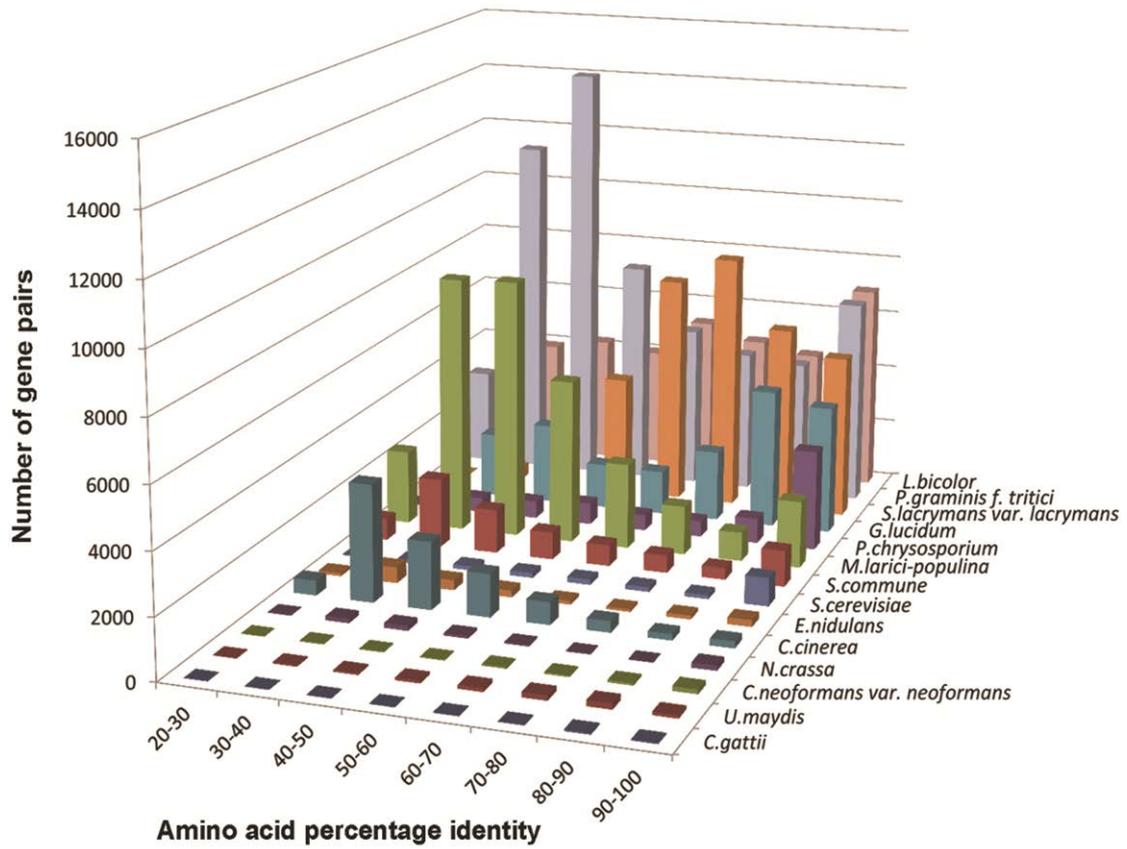
Supplementary Figure S9. The domain organisation of predicted non-ribosomal peptide synthetase (GL24538) and polyketide synthase (GL21253, GL21260, GL21264, GL27094, GL28732) in the *G. lucidum* genome. A: adenylation domain; T: thiolation domain or peptidyl carrier protein; C: condensation domain; KS: ketosynthase; AT: acyltransferase; DH: dehydratase; KR: keto reductase; ACP: acyl carrier protein.



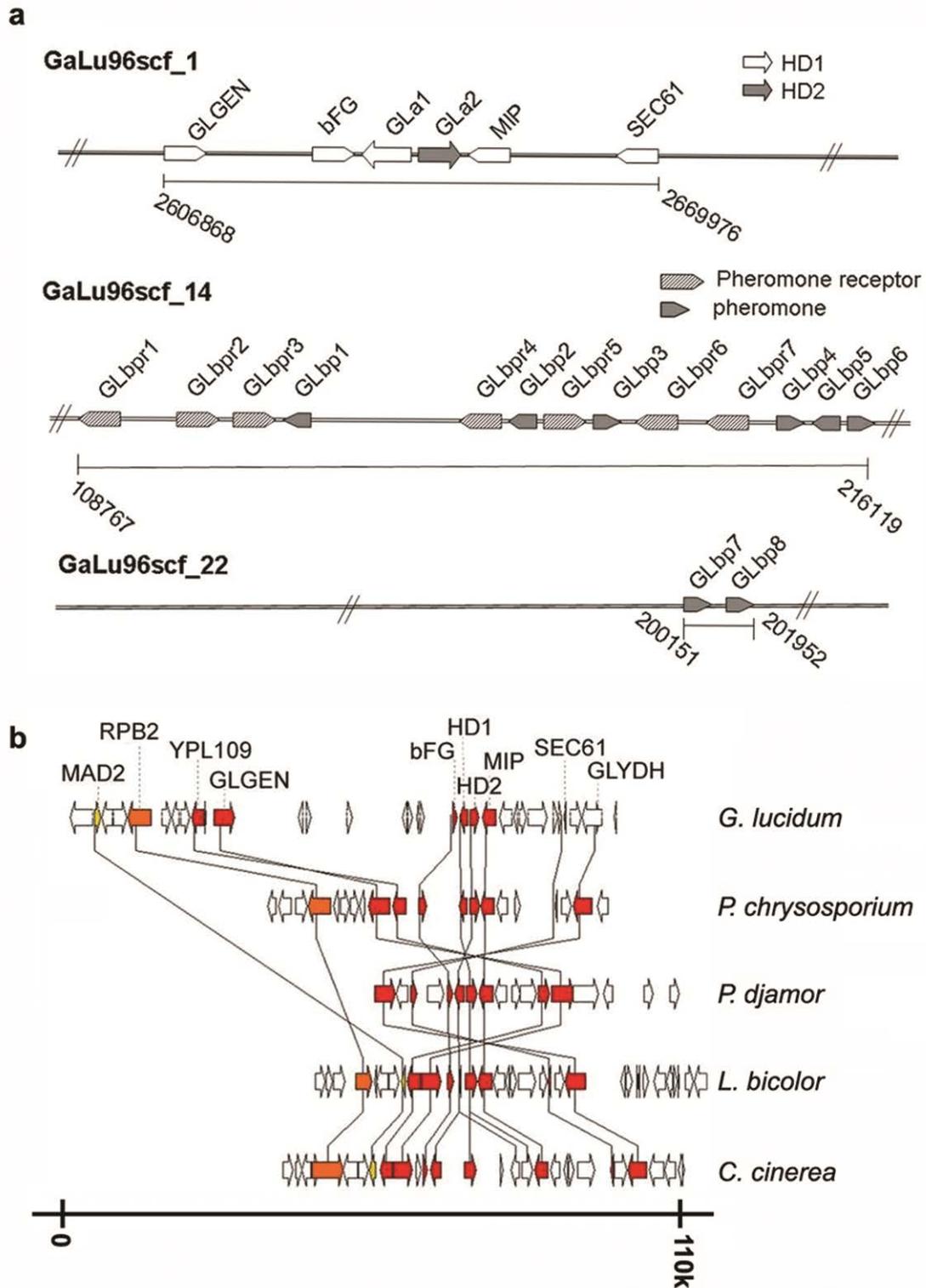
Supplementary Figure S10. An unrooted neighbour-joining tree of terpene synthase homologues from *G. lucidum* and other fungi. A Poisson model was used to account for the variation of substitution rates across amino acid sites using MEGA (version 5.05). A bootstrap value based on 1000 replications was set and displayed on the tree branches with a value larger than 50.



Supplementary Figure S11. The domain architecture and putative mechanism of velvet family proteins in *G. lucidum*. (a) Domain architecture of candidate velvet family proteins and a putative LaeA gene identified in *G. lucidum*. Velvet: Velvet domain; SAM_binding_site: S-adenosyl methionine-binding site; MTD: methyltransferase domain. (b) Proposed model showing the velvet family proteins and their coordination of secondary metabolism and development in *G. lucidum*. Velvet family includes four members, VeA, VelB, VosA, and VelC, which share a common velvet domain. Velvet family proteins interact with LaeA protein and regulate secondary metabolism and fungal development. The VeA–VelB heterodimer enters the nucleus with the help of the α -importin protein KapA in the dark and stimulates sexual fruiting body formation. VelB also forms homodimers VelB–VelB and VosA–VelB. VosA–VelB promotes spore viability and inhibits asexual development, which can be reversed by light and activation of LaeA. LaeA interacts with the VeA–VelB dimer to form a trimeric complex that coordinates secondary metabolism and development.



Supplementary Figure S12. Gene duplication analysis of 14 fungal genomes. Proteins from each genome were subjected to pair-wise intra-genomic comparisons using BLASTP (E-value < 1e-25). The pairs of proteins were then binned based on their similarity score.



Supplementary Figure S13. Analysis of the *matA* and *matB* gene loci. (a) Organization of the *matA* and *matB* loci of *G. lucidum*. (b) Schematic comparison of the *matA* locus regions in *G. lucidum*, *P. chrysosporium*, *P. djamor*, *L. bicolor*, and *C. cinerea*. Arrows represent individual genes. The color gradient indicates the number of species in which the gene is present, from yellow (present in only one species) to red (present in all species).

SUPPLEMENTARY TABLES

Supplementary Table S1. Sequencing statistics.

Library No.	Sequencing Platform	Insert size (kb)	No. Reads	Sizes (M)	Coverage (X)
1	Roche FLX Titanium	0	2,467,698	1,032.23	23.84
2	Roche FLX Titanium	3	1,084,053	372.53	8.60
3	Roche FLX Titanium	8	623,696	210.79	4.87
4	Roche FLX Titanium	20	465,569	158.31	3.66
5	Illumina GA II	0.5	84,223,832	6,822.13	157.58
6	Illumina GA II	2	79,077,412	6,405.27	147.95
7	Illumina GA II	5	49,883,208	4,040.54	93.33
Total			217,825,468	19,041.81	439.84

Supplementary Table S2. Assembly statistics.

Scaffold total number	82
Scaffold size (bp)	43,292,570
Scaffold mean length (bp)	527,958
Scaffold maximum length (bp)	4,834,011
Scaffold minimum length (bp)	2,064
Scaffold N50 ^a	11
Scaffold L50 (bp) ^a	1,387,710
Scaffold N90 ^b	34
Scaffold L90 (bp) ^b	356,137
Contig total number	194
Contig size (bp)	43,121,057
Contig mean length (bp)	222,273
Contig maximum length (bp)	2,085,703
Contig minimum length (bp)	728
Contig N50 ^a	19
Contig L50 (bp) ^a	649,708
Contig N90 ^b	79
Contig L90 (bp) ^b	135,338

^aN50/N90: number of scaffolds or contigs that collectively covered at least 50%/90% of the assembly; ^bL50/L90: length of the shortest scaffold or contig among those that collectively covered 50%/90% of the assembly.

Supplementary Table S3. Chromosome size determined by optical mapping.

ID	Length (kb)	No. of Fragments
Chr1	6,863.413	670
Chr2	5,768.626	619
Chr3	5,146.144	517
Chr4	4,477.195	442
Chr5	4,130.16	408
Chr6	3,918.09	407
Chr7	3,690.919	360
Chr8	3,426.668	369
Chr9	3,388.745	332
Chr10	3,185.851	361
Chr11	2,856.383	294
Chr12	2,737.476	287
Chr13	2,722.906	275

Supplementary Table S4. Gene model statistics.

Genes Total	16,113
Average Gene length (bp)	1,556
Gene sizes (bp)	25,078,722
Transcripts Total	16,495
Average Transcripts length (bp)	1,250
Exons per transcript	4.7
Proteins Total	16,495
Average Protein Length (aa)	396

Supplementary Table S5. Classification of repeated sequences in *G. lucidum*.

Element	Total number	Length (bp)	% assembled genome
LTR elements	3729	2346389	5.42
Copia	730	578616	1.34
Gypsy	2733	1698699	3.92
LINE	35	31701	0.07
SINE	0	0	0
DNA elements	2452	723332	1.67
Mariner/Tc1	19	19116	0.04

Helitron	8	13380	0.03
MITE	92	26545	0.06
Unclassified	359	261890	0.60
Total interspersed repeats		3363312	7.77
Satellites	1	75	0
Simple repeats	1545	75218	0.17
Low complexity	1671	90860	0.21
Total repeats		3529465	8.15

Supplementary Table S6. Gene models supported by hits from the corresponding public databases.

Total models	16,113
Nt	4,103 (25.5% ^a)
Nr	10,921 (67.8%)
KEGG	10,952 (68.0%)
Swissprot	6,369 (39.5%)
COG	4,934 (30.6%)
KOG	6,141 (38.1%)
Pfam	8,682 (53.9%)
GO	5,598 (34.7%)
InterPro	7,768 (48.2%)
Total	11,179 (69.4%)

^a The numbers in the parentheses indicate the percentages of the gene models with significant hits (E-value < 1e-5) in the databases.

Supplementary Table S7. Protein families and protein domains with significant expansion and compression in the branch leading to *G. lucidum*.

Protein Families			
PFAM ID	Annotation	Min <i>p</i>-value	Size
PF00026	Asp	0.000000	28→53 (25)
PF06985	HET	0.000000	2→103 (101)
PF07727	RVT_2	0.000000	11→30 (19)
PF08386	Abhydrolase_4	0.000000	3→14 (11)
PF01185	Hydrophobin	0.000001	11→25 (14)
PF07249	Cerato-platanin	0.000004	4→13 (9)
PF01828	Peptidase_A4	0.000007	3→11 (8)
PF07690	MFS_1	0.000007	96→126 (30)
PF04140	ICMT	0.000038	8→18 (10)
PF01764	Lipase_3	0.000041	7→16 (9)
PF07250	Glyoxal_oxid_	0.000082	4→11 (7)

	N		
PF04616	Glyco_hydro_4 3	0.000156	3→9 (6)
PF01753	zf-MYND	0.000246	7→15 (8)
PF00171	Aldedh	0.000251	16→27 (11)
PF01425	Amidase	0.000251	8→16 (8)
PF00141	Peroxidase	0.000353	4→10 (6)
Protein Domains			
PFAM ID	Annotation	Min <i>p</i>-value	Size
PF00067.16	P450	0.000000	126→237 (111)
PF00069.19	Pkinase	0.000001	130→171 (41)
PF00082.16	Peptidase_S8	0.000227	6→14 (8)
PF00135.22	COesterase	0.000153	19→32 (13)
PF00385.18	Chromo	0.000000	16→43 (27)
PF00450.16	Peptidase_S10	0.000069	11→22 (11)
PF00665.20	Rve	0.000000	8→39 (31)
PF00732.13	GMC_oxred_N	0.000005	23→41 (18)
PF03171.14	2OG-FeII_Oxy	0.000016	8→19 (11)
PF05199.7	GMC_oxred_C	0.000015	21→37 (16)

Supplementary Table S8. Genes involved in upstream section of triterpenoid biosynthetic pathway in *G. lucidum*.

Gene ID	Name of gene	Abbr.	E-value	Accession Number	Organism
GL23502	acetyl-CoA acetyltransferase	AACT	0	XP_001831920	<i>C. cinerea</i>
GL26574	acetyl-CoA acetyltransferase	AACT	1.00E-105	Q6NU46	<i>X. laevis</i>
GL24922	3-hydroxy-3-methyl glutaryl-CoA synthase	HMGS	1.00E-104	P17425	<i>R. norvegicus</i>
GL24088	3-hydroxy-3-methyl glutaryl-CoA reductase	HMGR	0	ABY84849	<i>G. lucidum</i>
GL17879	mevalonate kinase	MK	7.00E-90	O94350	<i>S. pombe</i>
GL17808	phosphomevalonate kinase	PMK	1.00E-122	EFI28147	<i>C. cinerea</i>
GL25304	mevalonate pyrophosphate decarboxylase	MVD	0	AEB00647	<i>G. lucidum</i>
GL29704	isopentenyl	IDI	1.00E-113	XP_0018287	<i>C. cinerea</i>

	diphosphate isomerase			02	
GL22068	farnesyl diphosphate synthase	FPS	1.00E-170	ACB37021	<i>G. lucidum</i>
GL25499	farnesyl diphosphate synthase	FPS	0	ACB37021	<i>G. lucidum</i>
GL21690	squalene synthase	SQS	0	ABF57213	<i>G. lucidum</i>
GL23376	squalene monooxygenase	SE	1.00E-115	XP_001840890	<i>C. cinerea</i>
GL18675	2,3-oxidosqualene-lanosterol cyclase	LSS	0	ADD60470	<i>G. lucidum</i>

Supplementary Table S9. Genes involved in polysaccharide biosynthesis and its regulation.

Gene ID	Gene name	E-value	Accession Number	Organisms
GL26783	Glucokinase	2.00E-72	Q92407	<i>A. niger</i>
GL24280	phosphoglucomutase	0	XP_001875148.1	<i>L. bicolor</i> <i>S238N-H82</i>
GL25739	UTP-glucose-1-phosphate uridylyltransferase	0	XP_001830101.1	<i>C. cinerea</i> <i>okayama7#130</i>
GL20535	1,3-beta-glucan synthase	0	O93927	<i>C. neoformans</i>
GL24465	1,3-beta-glucan synthase	0	O93927	<i>C. neoformans</i>
GL25075	Beta-glucan synthesis-associated protein KRE6	1.00E-88	P32486	<i>S. cerevisiae</i>
GL29980	Beta-glucan synthesis-associated protein KRE6	8.00E-99	P32486	<i>S. cerevisiae</i>
GL31622	Beta-glucan synthesis-associated protein KRE6	2.00E-88	P32486	<i>S. cerevisiae</i>
GL18643	Beta-glucan synthesis-associated protein SKN1	4.00E-87	P87024	<i>C. albicans</i>
GL21597	Beta-glucan synthesis-associated protein SKN1	7.00E-95	P87024	<i>C. albicans</i>
GL23415	Beta-glucan synthesis-associated protein SKN1	5.00E-92	P87024	<i>C. albicans</i>
GL22144	Beta-glucan	9.00E-74	P33336	<i>S. cerevisiae</i>

	synthesis-associated protein SKN1			
GL30704	Calnexin	2.00E-53	XM_0029121 64.1	<i>C. cinerea</i> <i>okayama7#130</i>
GL22742	GTPase-activating protein BEM2/IPL2	2.00E-13	P39960	<i>S. cerevisiae</i>
GL21985	GTP-binding protein rho1	3.00E-90	Q09914	<i>S. pombe</i>
GL23194	GTP-binding protein rho1	2.00E-52	Q09914	<i>S. pombe</i>
GL23196	GTP-binding protein rho1	2.00E-52	Q09914	<i>S. pombe</i>
GL26079	Protein cwh43	1.00E-15 2	Q9HDZ2	<i>S. pombe</i>
GL21836	Protein zds1	6.00E-08	O14100	<i>S. pombe</i>
GL25902	RHO1 GDP-GTP exchange protein 1	4.00E-15	P53046	<i>S. cerevisiae</i>
GL22624	RHO1 GDP-GTP exchange protein 2	8.00E-12	P51862	<i>S. cerevisiae</i>
GL23023	RHO1 GDP-GTP exchange protein 2	1.00E-12 9	P51862	<i>S. cerevisiae</i>
GL27304	RHO1 GDP-GTP exchange protein 2	3.00E-51	P51862	<i>S. cerevisiae</i>
GL26033	Rho-GTPase-activating protein BAG7	3.00E-25	Q12128	<i>S. cerevisiae</i>
GL21668	Rho-GTPase-activating protein LRG1	2.00E-50	P35688	<i>S. cerevisiae</i>
GL23239	ROT1 protein	2.00E-57	XP_00183581 5.1	<i>C. cinerea</i> <i>okayama7#130</i>

Supplementary Table S10. Assignment of *G. lucidum* genes to CAZy families.

CAZy	No. of Members	Gene_ID
CBM1	2	GL16848 GL21061
CBM1-CE1	1	GL24404
CBM1-CE16	1	GL17156
CBM1-GH10	3	GL20229 GL20392 GL26036
CBM1-GH27	1	GL25150
CBM1-GH27- CBM1-GH27	1	GL23450
CBM1-GH5	2	GL25283 GL29421
CBM1-GH6	1	GL24712
CBM12	1	GL16814
CBM13	7	GL15828 GL21988 GL21989 GL22047 GL22051 GL22052 GL22063
CBM13-CBM13	1	GL22099

CBM18-GH16	1	GL22445
CBM18-GH18	1	GL23035
CBM21	2	GL16344 GL30108
CBM48	2	GL18175 GL19454
CBM48-GH13	1	GL31461
CBM5	2	GL22148 GL23163
CBM50	2	GL15261 GL31063
CBM50-CBM50- CBM50-CBM50- CBM50-CBM50	1	GL16442
CE1	2	GL23410 GL29422
CE12	1	GL23909
CE15	2	GL25044 GL28882
CE16	16	GL15069 GL16742 GL17326 GL20550 GL20604 GL22017 GL22713 GL22714 GL22747 GL23497 GL23891 GL24754 GL25520 GL26151 GL27674 GL29068
CE4	3	GL27817 GL29877 GL31433
CE8	3	GL23684 GL29440 GL30634
CE9	1	GL24032
EXPN	14	GL15421 GL18494 GL18502 GL20163 GL21189 GL21245 GL22361 GL23584 GL24541 GL25701 GL28117 GL29801 GL30313 GL30757
GH1	3	GL20553 GL24039 GL30154
GH10	4	GL16247 GL17195 GL18655 GL27713
GH105	1	GL25389
GH115	2	GL16486 GL23132
GH115-GH115	1	GL21283
GH12	3	GL15948 GL16577 GL23260
GH125	1	GL26577
GH13	6	GL15361 GL18711 GL24887 GL24914 GL26000 GL27531
GH13-CBM20	1	GL31059
GH13-GT5	1	GL24554
GH15	1	GL23287
GH15-CBM20	2	GL23580 GL23600
GH16	33	GL15467 GL15468 GL18643 GL21208 GL21289 GL21331 GL21597 GL22144 GL23085 GL23128 GL23415 GL23462 GL23531 GL23820 GL23822 GL24931 GL25004 GL25008 GL25009 GL25075 GL25120 GL25601 GL27846 GL28634 GL29318 GL29641 GL30175 GL30185 GL30186 GL30437 GL30439 GL30548 GL31622
GH16-GH16	1	GL29980
GH17	1	GL25813
GH17-GH17	1	GL25603

GH18	32	GL15508 GL15547 GL15633 GL15744 GL17045 GL17093 GL17641 GL18544 GL19744 GL20390 GL20391 GL21171 GL21294 GL21821 GL21822 GL22147 GL24065 GL24125 GL24153 GL24162 GL24182 GL24217 GL24235 GL25033 GL25627 GL26628 GL26750 GL26751 GL29007 GL29216 GL29255 GL31157
GH18-CBM5	6	GL22188 GL24376 GL25816 GL25941 GL26040 GL28316
GH18-CBM5- CBM5	1	GL29885
GH2	3	GL23228 GL26515 GL29551
GH20	6	GL18102 GL24346 GL24347 GL24372 GL28725 GL29544
GH23	1	GL25256
GH25	2	GL15238 GL18370
GH27	3	GL21024 GL27011 GL30909
GH28	13	GL20947 GL21083 GL22113 GL22234 GL22335 GL22710 GL23205 GL23206 GL23208 GL25232 GL27403 GL27405 GL29527
GH3	12	GL15526 GL19093 GL19708 GL20743 GL21370 GL22586 GL22886 GL24911 GL27550 GL27561 GL29912 GL29918
GH30	2	GL23795 GL23866
GH31	6	GL19735 GL22859 GL30020 GL30116 GL30117 GL30925
GH32	1	GL19397
GH35	8	GL16607 GL16615 GL23224 GL23249 GL23250 GL23288 GL23290 GL23291
GH35-GH35	1	GL23225
GH37	2	GL24361 GL26035
GH38	1	GL22570
GH43	11	GL15164 GL15780 GL20525 GL21890 GL21973 GL22781 GL23275 GL24175 GL25001 GL27397 GL29868
GH45	2	GL19184 GL23097
GH47	10	GL15199 GL20698 GL20848 GL20849 GL20850 GL21617 GL21844 GL21852 GL26660 GL30115
GH5	17	GL19326 GL19331 GL19824 GL20194 GL21683 GL21738 GL22636 GL22898 GL24030 GL24196 GL24387 GL25895 GL26529 GL26922 GL28282 GL29456 GL30087
GH51	2	GL22224 GL22560
GH53	1	GL31140
GH55	3	GL21451 GL23395 GL24581
GH61	12	GL17404 GL19332 GL20770 GL22070 GL22659 GL23340 GL23362 GL23664 GL25237 GL25471 GL30301 GL31181
GH61-CBM1	1	GL25022
GH61-GH61	1	GL29277
GH7	3	GL18725 GL29727 GL30351

GH71	6	GL16222 GL20161 GL21154 GL21155 GL23952 GL31177
GH72-CBM43	1	GL29873
GH74	1	GL30540
GH76	2	GL21575 GL27991
GH78	5	GL17304 GL18961 GL25806 GL26025 GL27541
GH79	8	GL19451 GL23161 GL25669 GL26459 GL26618 GL29243 GL31446 GL31462
GH79-GH79- GH79	1	GL23706
GH85	1	GL29163
GH88	1	GL16281
GH89	1	GL22177
GH9	2	GL26499 GL26847
GH92	6	GL18249 GL20210 GL20609 GL23422 GL29257 GL29258
GH93	2	GL17353 GL24173
GH95	1	GL21099
GT1	10	GL17818 GL23138 GL24551 GL25278 GL25279 GL25442 GL25611 GL25612 GL25762 GL29876
GT15	3	GL22001 GL22007 GL24561
GT17	2	GL25872 GL26844
GT2	12	GL15273 GL18134 GL19116 GL22207 GL25613 GL26644 GL27969 GL28043 GL28060 GL30737 GL30799 GL31550
GT20	3	GL21342 GL22527 GL26018
GT21	1	GL30004
GT22	3	GL17041 GL20526 GL29468
GT24	1	GL22535
GT3	1	GL24971
GT31	1	GL30565
GT32	1	GL31413
GT33	1	GL29609
GT35	1	GL21375
GT39	3	GL21929 GL24505 GL26631
GT4	4	GL20810 GL22888 GL24108 GL24380
GT48	4	GL15036 GL20535 GL20577 GL24465
GT49	1	GL29342
GT50	1	GL24601
GT57	2	GL16923 GL31336
GT58	1	GL17348
GT59	1	GL29870
GT66	1	GL21733
GT69	2	GL19669 GL21087
GT76	2	GL16434 GL16435
GT8	6	GL16171 GL20444 GL21494 GL24498 GL29679 GL30965

GT90	1	GL20961
PL14	6	GL16333 GL16341 GL17811 GL22848 GL22849 GL24016
PL8	4	GL23979 GL30246 GL30937 GL30938

Supplementary Table S11. Comparison of CAZy and oxidoreductase gene numbers of *G. lucidum* with those of other fungi.

Species	GH*	CE	GT	PL	CBM	AA1_1 ^a	AA2	AA3_1	AA3_2 ^a	AA3_2 ^b	AA3_3	AA3_4	AA4	AA5_1	AA5_2	AA6	AA8
<i>P. graminis f. tritici</i>	158	25	83	4	8	0	0	0	0	0	0	0	0	4	0	1	0
<i>M. laricis-populina</i>	175	36	84	6	7	0	0	0	0	0	0	0	0	4	0	1	0
<i>U. maydis</i>	100	13	58	1	7	0	0	0	0	1	0	0	0	3	1	1	0
<i>G. lucidum</i>	288	30	70	10	53	13	8	1	3	0	1	0	0	9	0	1	2
<i>P. chrysosporium</i>	190	16	65	4	47	0	16	1	3	0	3	1	0	7	0	4	2
<i>P. placenta</i> ^c	252	21	100	8	28	3	2	0	4	0	8	0	0	2	0	2	0
<i>L. bicolor</i>	168	17	87	7	24	8	1	0	3	0	2	0	0	9	0	2	0
<i>C. cinerea</i>	216	51	71	13	89	17	1	1	22	0	2	0	0	6	0	3	6
<i>S. commune</i>	245	39	76	16	40	2	0	1	1	1	4	1	0	2	0	4	3
<i>S. lacrymans</i> var. <i>lacrymans</i>	160	12	62	6	29	4	0	2	0	0	5	0	0	3	0	2	4
<i>C. bacillisporus</i>	83	6	66	3	11	0	0	0	0	0	0	0	0	3	0	1	0
<i>C. neoformans</i> var. <i>neoformans</i>	81	7	64	3	12	0	0	0	0	0	0	0	0	3	0	1	0
<i>S. cerevisiae</i>	46	3	68	0	12	0	0	0	0	0	0	0	0	0	0	3	0
<i>N. crassa</i>	179	22	76	4	42	0	0	2	0	1	1	0	0	1	1	1	8
<i>A. nidulans</i>	267	30	97	22	47	0	0	1	1	1	1	1	1	0	0	1	6

* Sequence-based family definitions for oxidoreductase enzyme modules as determined by Levasseur and Henrissat (unpublished): GH: glycoside hydrolases, CE: carbohydrate esterases, GT: glycosyl transferases, PL: polysaccharide lyases, CBM: carbohydrate-binding module, AA1_1: classical laccase, AA2: class-II peroxidase, AA3_1: cellobiose dehydrogenase, AA3_2^a: aryl-alcohol oxidase, AA3_2^b: glucose oxidase, AA3_3: alcohol oxidase, AA3_4: pyranose oxidase, AA4: vanillyl-alcohol oxidase, AA5_1: glyoxal oxidase, copper radical oxidase, AA5_2:

galactose oxidase, AA6: benzoquinone reductase, AA8: iron reductase domain. ^c *P. placenta* figures are for the sequenced dikaryon.

Supplementary Table S12. Classification of *G. lucidum* gene involved in lignin degradation.

Module	Protein models	Functional inference	Nb	Peptide signal	Peptide signal size
AA1_1	GL16398	candidate laccase	14	Yes	[1-21]
AA1_1	GL29486	candidate laccase		Yes	[1-21]
AA1_1	GL29253	candidate laccase		Yes	[1-23]
AA1_1	GL21497	candidate laccase		Yes	[1-21]
AA1_1	GL29490	candidate laccase		Yes	[1-22]
AA1_1	GL30788	candidate laccase		Yes	[1-21]
AA1_1	GL29234	candidate laccase		Yes	[1-24]
AA1_1	GL23477	candidate laccase		Yes	[1-20]
AA1_1	GL19134	candidate laccase		Yes	[1-24]
AA1_1	GL17426	candidate laccase		Yes	[1-21]
AA1_1	GL18428	candidate laccase		Yes	[1-22]
AA1_1	GL16401	candidate laccase		Yes	[1-24]
AA1_1	GL22987	candidate laccase		Yes	[1-17]
AA1_1	GL15267	candidate multicopper oxidase; related to laccase		No	
AA1_2	GL17009	candidate ferroxidase	1	Yes	[1-17]
AA2	GL25672	candidate manganese peroxidase or lignin peroxidase or versatile peroxidase	8	Yes	[1-21]
AA2	GL20012	candidate manganese peroxidase or lignin peroxidase or versatile peroxidase		Yes	[1-19]
AA2	GL20019	candidate manganese peroxidase or lignin peroxidase or versatile peroxidase		Yes	[1-21]
AA2	GL25168	candidate manganese peroxidase or lignin peroxidase or versatile peroxidase		Yes	[1-15]
AA2	GL20537	candidate manganese peroxidase or lignin peroxidase or versatile peroxidase		Yes	[1-18]
AA2	GL30176	candidate manganese peroxidase or lignin peroxidase or versatile peroxidase		Yes	[1-18]
AA2	GL25550	candidate manganese peroxidase or lignin peroxidase or versatile peroxidase		Yes	[1-18]
AA2	GL25427	candidate manganese peroxidase or lignin peroxidase or versatile peroxidase		Yes	[1-18]
AA3_1	GL22256	candidate cellobiose dehydrogenase	1	No	
AA3_2	GL23730	candidate aryl alcohol oxidase	4	No	
AA3_2	GL24292	candidate aryl alcohol oxidase		Yes	[1-20]
AA3_2	GL28208	candidate aryl alcohol oxidase		No	
AA3_2	GL22369	related to aryl alcohol oxidase		No	
AA3_2	GL30236	related to glucose oxidase	3	Yes	[1-23]
AA3_2	GL25583	related to glucose oxidase		No	
AA3_2	GL21702	related to glucose oxidase		Yes	[1-19]
AA3_3	GL18144	candidate alcohol oxidase	3	No	
AA3_3	GL21354	related to alcohol oxidase		No	
AA3_3	GL17761	related to alcohol oxidase		No	
AA5_1	GL28574	candidate copper radical oxidase	9	Yes	[1-21]

AA5_1	GL21192	candidate copper radical oxidase		Yes	[1-21]
AA5_1	GL25348	candidate copper radical oxidase		No	
AA5_1	GL27858	candidate copper radical oxidase		No	
AA5_1	GL20734	candidate glyoxal oxidase		Yes	[1-17]
AA5_1	GL26939	candidate glyoxal oxidase		Yes	[1-21]
AA5_1	GL20654	candidate glyoxal oxidase		Yes	[1-18]
AA5_1	GL20662	candidate glyoxal oxidase		Yes	[1-18]
AA5_1	GL26948	candidate glyoxal oxidase		Yes	[1-18]
AA6	GL22248	candidate benzoquinone reductase	1	No	
AA8	GL22256	candidate iron reductase domain	2	No	
AA8	GL24596	candidate iron reductase domain			

Supplementary Table S13. Comparison of CYP families in *G. lucidum*, *P. placenta*, and *P. chrysosporium*.

CYP family	<i>G.lucidum</i>	<i>P. placenta</i>	<i>P. chrysosporium</i>
CYP51	2	1	1
CYP53	1	7	1
CYP61	1	1	1
CYP63	6	5	7
CYP502	1	4	1
CYP505	4	2	7
CYP512	22	14	14
CYP537	1	2	0
CYP642	1	0	0
CYP5027	0	9	0
CYP5035	16	3	13
CYP5036	0	0	5
CYP5037	6	13	5
CYP5065	1	0	0
CYP5136	7	0	5
CYP5137	1	6	2
CYP5138	1	1	1
CYP5139	7	8	1
CYP5140	1	1	1
CYP5141	2	4	7
CYP5142	0	0	7
CYP5143	0	0	2
CYP5144	3	3	34
CYP5145	0	0	3
CYP5146	0	0	6
CYP5147	0	0	6
CYP5148	2	1	2
CYP5149	0	1	1
CYP5150	36	25	6
CYP5151	1	1	1
CYP5152	1	2	2
CYP5153	0	0	1

CYP5154	0	0	1
CYP5155	0	0	1
CYP5156	1	1	1
CYP5157	0	0	1
CYP5158	1	2	1
CYP5339	0	2	0
CYP5340	2	1	0
CYP5341	2	3	0
CYP5342	0	1	0
CYP5343	0	1	0
CYP5344	0	3	0
CYP5345	0	1	0
CYP5346	0	1	0
CYP5347	1	2	0
CYP5348	3	34	0
CYP5349	1	2	0
CYP5350	0	11	0
CYP5351	1	1	0
CYP5352	0	1	0
CYP5353	0	1	0
CYP5354	0	2	0
CYP5355	0	1	0
CYP5356	0	1	0
CYP5357	2	0	0
CYP5358	1	0	0
CYP5359	47	0	0
CYP5360	1	0	0
CYP5361	1	0	0
CYP5362	1	0	0
CYP5363	1	0	0
CYP5364	3	0	0
CYP5365	1	0	0
CYP5366	1	0	0
CYP6005	2	0	0
Total	197	186	148

Supplementary Table S14. Proteins involved in pheromone-related MAPK pathways.

Pathway	<i>G. lucidum</i>	<i>S. cerevisiae</i>	<i>U. maydis</i>	<i>S. commune</i>
Pheromone	GLbp	mfa	mfa1/2	Bap/Bbp
Receptor	GLbpr	Ste3	pra1/2	Bar/Bbr
G protein	GL25376/GL29474/?	Gpa1/Ste4/18	Um2517/Um00703/?	Gpa1/?/?
PaK	GL21839	Ste20	Um04583	Ste20
MAPKKK ^a	GL28553/GL21789	Ste11-Ste50	Kpp4-Ubc2	Ste11-Ste50
MAPKK ^b	GL23765	Ste7	Fuz7	Ste7
MAPK ^c	GL20693	Fus3/Kss1	Kpp2	Fus3
Transcription factor	GL15449	Ste12	prf1	Ste12

^aMAPKKK: Mitogen-activated protein kinase kinase kinase; ^bMAPKK: Mitogen-activated

protein kinase kinase; ^cMAPK: Mitogen-activated protein kinase.

SUPPLEMENTARY NOTES

1. Origin of the monokaryotic strain *Ganoderma lucidum* G.260125-1

Ganoderma lucidum G.260125-1 was derived from the dikaryotic mycelia of *G. lucidum* CGMCC5.0026 deposited in the China General Microbiological Culture Collection Center (Beijing, China, <http://www.cgmcc.net/>) using the protoplast monokaryogenesis method⁶¹. In contrast to the dikaryotic strain CGMCC5.0026, no clamp connections were observed in the filaments of G.260125-1 (Supplementary Fig. S1a and b). Moreover, in the dikaryotic strain CGMCC5.0026, each hyphal cell had two nuclei, whereas in the strain G.260125-1, each hyphal cell had only one nucleus (Supplementary Fig. S1c and d). SNP analysis also supported the classification of strain G.260125-1 as a monokaryon. Eight regions containing SNPs in the dikaryotic strain CGMCC5.0026 were randomly selected. These SNPs were then confirmed by sequencing the PCR products that were amplified from the template genomic DNA prepared from the dikaryotic strain CGMCC5.0026. However, no SNPs were found when the sequenced PCR products were amplified from genomic DNA of the strain G.260125-1 (Supplementary Fig. S1e). This result indicated that only one set of chromosomes is present in the G.260125-1 strain.

2. Repetitive DNA in the genome of *G. lucidum*

Approximately 8.15% of the assembled genome of *G. lucidum* consists of repetitive sequences. The identified repeats include long terminal repeat (LTR) elements (5.42%), DNA transposons (1.67%), simple repeats (0.17%), and low-complexity sequences (0.21%). The remaining repeats, which constitute 0.60% of the genome, are unclassified. Transposable elements (TEs) were distributed throughout the genome (Supplementary Table S5). LTRs were found on all the scaffolds with lengths over 40 kb, except for scaffolds 47 and 58. For example, fifteen *cop*ia-class retroelements that almost perfectly matched the consensus sequences were located on scaffolds 1 (2,302,881-2,309,423; 4,796,732-4803,281), 2

(2,032,718-2,039,339), 3 (472,456-479,005), 7 (167,176-173,713), 9 (340,402-346,947), 14 (533,145-539,692), 16 (342,694-349,237), 17 (452,058-458,607), 18 (321,863-328,416), 21 (583,930-590,477), 27 (374,709-381,260), 28 (387,050-393,604), 33 (174,249-179,616), and 35 (248,344-253,716). Similarly, 9 *gypsy*-class retroelements were located on scaffolds 3 (1,556,666-1,563,637), 4 (1,840,503-1,847,591), 5 (1,267,762-1,274,742; 1,336,662-1,343,645), 9 (410,562-417,535), 11 (970,773-977,757), 26 (100,694-112,352), 35 (228,923-240,686), and 39 (69,516-74,563).

3. Comparison of fungal genomes

A total of 20,059 Ortholog Cluster Groups (OCGs) were constructed (Supplementary Data S1); of these, 7182 OCGs contained 11,552 *G. lucidum* proteins. The OCGs that were most highly represented in the *G. lucidum* genome were OCG10000 and OCG10026, which contain 186 and 107 proteins, respectively. A total of 5,714 OCGs contain only one *G. lucidum* protein, and 5,838 proteins have at least one paralog.

To compare the number of proteins belonging to various protein families across the 15 fungal species, we have compared proteins from these 15 fungal species (Supplementary Data S2 and S3) with the Pfam Database⁶². Entries for Pfam families and Pfam domains were examined separately and there are as many as 1,262 protein domains and 2,219 protein families with an E-value < 1e-5 found from these comparisons. The most abundant protein families across the 15 fungal species evaluated included PF07690.10, MFS_1 (1,512 proteins), PF00153.21, Mito_carr (1,261 proteins), and PF00271.25, Helicase_C (844 proteins). In *G. lucidum*, HET domain (PF06985.5) proteins were found to be the most abundant except MFS_1 (PF07690.10), and there are 103 proteins that contain HET domain. Similarly, when exploring the Pfam domains, the most abundant Pfam domains across the 15 fungal species included PF00069.19 (Pkinase, 1,914 domains) and PF00067.16 (P450, 1,234 domains). These two domains were also the most abundant ones in *G. lucidum*.

As many proteins have more than one protein domain, we examined the domain organization (DO) of the proteins from *G. lucidum* as well as from other fungi (Supplementary Data S12). A total of 7,440 DOs were found; of these, 467 DOs were found in all 15 species, and 2,614 DOs were found only in some of the species. The most abundant DOs (with at least two domains) were DO11932 (PF00270.23:DEAD; PF00271.25:Helicase_C) and DO12712 (PF00172.12:Zn_clus; PF04082.12:Fungal_trans). In total, 299 DOs were unique to *G. lucidum*.

We identified blocks of genes that were syntenic between *G. lucidum* and three other species in Agaricomycetes (*Phanerochaete chrysosporium*¹¹, *Schizophyllum commune*¹² and *Postia placenta*⁶³) using MCscan⁶⁴(version 0.8). A total of 250 syntenic blocks of co-linear genes were found between *G. lucidum* and *P. chrysosporium*, and these blocks contain 3,008 genes in the *G. lucidum* genome and 2,986 genes in the *P. chrysosporium* genome (Supplementary Fig. S4). On average, each syntenic block in *G. lucidum* genome is about 24.8 kb in size and contains 12 genes. We also detected 201 collinear blocks common to the *G. lucidum* and *S. commune* genomes. These blocks contain 1,958 genes in the *G. lucidum* genome and 1,963 genes in the *S. commune* genome. On average, each block contains 9.92 genes; only 52 blocks have more than ten genes. However, only 79 syntenic blocks were found between *G. lucidum* and *P. placenta*, which may be ascribed to the incomplete sequence assembly of *P. placenta* Genome. Comparison of *G. lucidum* itself revealed 107 syntenic blocks, in which the segment pairs may have arisen from genome duplication⁶⁵.

4. Polysaccharide biosynthesis

Polysaccharides are another group of active components found in *G. lucidum*. Among these polysaccharides, the water-soluble 1,3- β - and 1,6- β -glucans are the most active ones for immunomodulatory compounds^{2,17}. UDP-glucose is the precursor of these glucans, and its biosynthesis involves glucokinase, α -phosphoglucomutase, and UDP-glucose-1-phosphate

uridylyltransferase, three enzymes that are all encoded by single-copy genes in *G. lucidum* (Supplementary Fig. S8a). However, 1,3- β -glucan synthase, which is involved in 1,3- β -glucan biosynthesis, is encoded by two genes. Analysis of gene expression indicated that the relative expression levels of these genes were higher in primordia and fruiting bodies than in mycelia, evidencing stage-dependent expression patterns (Supplementary Fig. S8b). In addition, genes encoding the β -glucan biosynthesis-associated proteins SKN1 and KRE6 were also discovered in the *G. lucidum* genome. These two proteins play key roles in the biosynthesis of 1,6- β -glucans. Notably, a protein family (PF03935) containing the SKN1 domain was expanded in the *G. lucidum* genome compared with *P. chrysosporium* and *P. placenta*. Finally, we have also identified many genes (Supplementary Table S9) that are similar to those in *Saccharomyces cerevisiae*, which regulate 1,3- β -glucan and/or 1,6- β -glucan biosynthesis, and play important roles in regulating the polysaccharide content in the cell wall^{18,19}.

5. Wood digestion

A total of 417 *G. lucidum* genes could be assigned to carbohydrate-active enzymes (CAZymes) families as defined by the CAZy database³⁹ (www.cazy.org). This total covers 288 glycoside hydrolase domains, 70 glycosyltransferase domains, 30 carbohydrate esterase domains, 10 polysaccharide lyase domains and 53 carbohydrate-binding modules (Supplementary Table S10), making this fungus one of the richest basidiomycetes examined to date (Supplementary Table S11). In particular, the genome encodes candidate enzymes for the digestion of the three major classes of plant cell wall polysaccharides: cellulose (CAZy families GH6, GH7, GH12, GH45 and GH61), hemicellulose (CAZy families GH10, GH74, GH95, GH115) and pectin (GH28, GH35, GH43, GH51, GH53, GH78 and GH93). Interestingly, although *G. lucidum* is the richest known basidiomycete in terms of genes encoding enzymes for pectin digestion, its strategy for pectin breakdown relies solely on

hydrolytic enzymes; the genome does not encode any pectin/pectate lyase (Supplementary Data S11).

The breakdown of the lignin barrier is necessary to expose the structural polysaccharides of plant cell walls in wood. Unlike the hydrolysis of polysaccharides, lignin digestion is considered an “enzymatic combustion” reaction involving several oxidoreductases such as laccases, ligninolytic peroxidases and peroxide-generating oxidases⁴⁰. In terms of lignin digestion, wood-decaying basidiomycetes are traditionally divided into two major categories: white-rot (WR) and brown-rot (BR) fungi. WR fungi make use of an array of extracellular hydrolases and oxidative enzymes that simultaneously depolymerise carbohydrates and a large proportion of the lignin. In contrast, BR decay involves an initial non-enzymatic attack on the wood cell wall, which generates extracellular hydroxyl radicals *via* the Fenton reaction^{30,63}. Thus, BR fungi depolymerise cellulose chemically and subsequently modify lignin extensively, leaving a brown lignin-enriched polymeric residue.

The annotation of the candidate ligninolytic enzymes encoded in the *G. lucidum* genome revealed a set of 36 ligninolytic oxidoreductases, including 13 laccases, eight class-II peroxidases, one cellobiose dehydrogenase, three aryl-alcohol oxidases, nine copper-radical oxidases, one benzoquinone reductase and one alcohol oxidase (Supplementary Table S12). The lignocellulolytic enzyme complement of *G. lucidum* is clearly different from that of the model BR fungus *P. placenta*. In particular, the cellulose-targeting families GH6, GH7, GH45, GH74, and CBM1, as well as lignin peroxidase (LiP), manganese peroxidase (MnP), versatile peroxidase (VP) and cellobiose dehydrogenases, are present in *G. lucidum* and absent in *P. placenta*. Only two low-redox-potential peroxidases lacking the Mn(II)-oxidation site and the exposed tryptophan responsible for high-redox-potential substrate oxidation were reported in *P. placenta*⁶³. Recent genome comparisons revealed that BR fungi have the fewest hydrolytic and oxidoreductase-encoding genes; they have

undergone contractions in multiple gene families involved in cellulose and lignin digestion³⁰. Instead of losing these gene families and evolving to become a BR wood decayer, *G. lucidum* has expanded specific gene families usually found in WR fungi. Interestingly, the ligninolytic enzyme arsenal of *G. lucidum* sets this basidiomycete apart from model WR fungi such as *P. chrysosporium* and *S. commune* (Supplementary Table S11). Indeed, *P. chrysosporium* encodes 16 class-II peroxidases but no classical laccase, whereas *S. commune* encodes two laccases but lacks genes encoding peroxidases. In contrast, *G. lucidum* shows a large and complete set of ligninolytic peroxidases together with laccases and a cellobiose dehydrogenase, suggesting that this fungus may exploit different strategies to breakdown lignin, including oxidation by hydrogen peroxide in a reaction catalysed by class-II peroxidases. In addition, *G. lucidum* laccases were able to degrade recalcitrant lignin compounds in the presence of redox mediators or generate lignocellulose-degrading hydroxyl radical *via* Fenton chemistry⁴¹. In agreement with the presence of candidate class-II peroxidases, several peroxide-generating oxidases were identified in the *G. lucidum* genome, particularly in the copper-radical oxidase family. In summary, the distribution of its lignocellulolytic gene families classifies *G. lucidum* as a particularly versatile WR fungus equipped with a remarkable enzymatic arsenal able to degrade the entire wood structure.

6. Gene duplication and genome duplication

To identify possible duplication events in *G. lucidum*, we performed a self-vs-self sequence similarity search using the predicted proteins. At an E-value of $< 1e-25$, we found 4,872 genes in the *G. lucidum* genome that appear to have arisen from duplication events⁶⁶. A duplicated block is further defined as a region that contains duplicated genes located within a window of 50 kb. In total, 884 duplicated blocks containing 664 paralogous groups were identified. The total lengths of these blocks span 44% of the *G. lucidum* genome. The average block size is 22 kb, and the mean number of duplicated genes is 2.75. According to the GO

annotation, 17.2 % of the duplicated genes are involved in metabolic processes, 8.6% of the duplicated genes are involved in transporter activity, and 3% of the duplicated genes are involved in transcriptional regulation activity. Importantly, 107 genes in the CYP superfamily, 27 genes in carbohydrate-active enzyme families, and 33 genes for lignin-digestion enzyme families are also probably originated via gene duplication. We also compared similar gene pairs of 13 other fungal genomes. *G. lucidum* has abundant highly similar gene pairs (Supplementary Fig. S12).

Genome-wide sequence duplications were detected using BLASTN with an identity of 90% for the aligned regions. After removing the self-matched regions, 1,422 duplicated pairs were identified on all 13 chromosomes. Moreover, 132 pairs are located on the same chromosomes, and the other 1,290 pairs are located on different chromosomes. The total length of these regions is 1,582,796 bp, covering 3.7% of the assembled genome (Fig. 1d). The mean length of these duplicated fragments is approximately 7 kb. The largest fragment, with a size of 22,496 bp, is located on chromosome 1.

7. Analysis of the *matA* and *matB* gene loci

We searched the *G. lucidum* genome for the *matA* locus and identified the genes (*GLa1* and *GLa2*) encoding the HD1 and HD2 proteins (Supplementary Fig. S13). These two genes are adjacent to one another on chromosome 3 and are arranged in opposite orientations, as observed in other basidiomycete species^{67,68}. Compared with Mitochondrial intermediate peptidase (*MIP*) and beta flanking-gene (*bFG*), most published HD1-encoding genes in the Agaricales order are positioned tail-to-tail and adjacent to *MIP*. However, in *G. lucidum*, the HD2-encoding gene is located tail-to-tail adjacent to *MIP*, which is similar to what has been observed in the polypore fungi *P. chrysosporium*¹¹ and *P. placenta*⁶³.

In addition, 11 sequences were identified as mating type B (*matB*) locus genes that encode pheromones and pheromone receptors. Seven STE3-like pheromone receptor genes (*GLbpr1*

to *GLbpr7*) were clustered with six pheromone genes (*GLbp1* to *GLbp6*) within a region of approximately 100 kb on chromosome 10. There are two pheromone candidate genes (*GLbp7* and *GLbp8*) on chromosome 6, and the *GLbpr* proteins show high degrees of similarity to their orthologs in *P. placenta*⁶³, *C. cinerea*⁶⁹, and *S. commune*¹² (E-value < 1e-60 and greater than 40% sequence identity). All five *GLbpr* genes encode proteins with seven transmembrane helices, a structural feature that is a hallmark of STE3-like receptors (as predicted by the HMMTOP and TMHMM programs). Mating type-specific pheromone genes (*GLbp1* to *GLbp8*) encode proteins with CXXA motifs at their C-termini.

We then searched for proteins involved in pheromone-related MAPK pathways using BLASTP. The results revealed that the pheromone response pathway in *G. lucidum* is highly conserved compared with those in *S. cerevisiae*⁷⁰, *Ustilago. maydis*⁷¹, and *S. commune*¹². With the exception of Ste18, the gamma subunit of the STE3-type G-protein-coupled pheromone receptor, all putative proteins in this pathway were found within the genome of *G. lucidum* (Supplementary Table S14).

SUPPLEMENTARY METHODS

1. Production and regeneration of *G. lucidum* protoplasts

Mycelia were inoculated into 100 ml PDB media (200 g potato, 20 g dextrose, 1 L water) in 250 ml flasks. The mycelia were cultivated in the dark at 28°C under continuous shaking at 50 rpm. The cultures were harvested by filtering through sterile gauze. Approximately 5 g of mycelia were placed in a 50 ml tube containing 20 ml of filter-sterilized 1.5% lywallzyme (Guangdong Institute of Microbiology, China) in 0.6 M mannitol. The mixture was incubated at 30°C for approximately 2 h with occasional agitation. Hyphae fragments were removed after 2 h of digestion. MYG agar medium (2% malt extract, 0.2% yeast extract, 2% dextrose, 1.5% agar) was used for *G. lucidum* protoplast regeneration. The protoplast suspension was diluted to 1×10^4 cells/ml, and 50 μ l of the diluted protoplast suspension was added to a MYG medium plate and inoculated at 25°C in the dark.

2. Selection and validation of Monokaryotic strain

After 10 days incubation, the protoplasts grew into visible colonies, and each single colony was transferred onto a new plate. To observe the clamp connection and count the number of nuclei in each cell, hyphae from each colony were stained by Calcofluor White (Sigma, USA) and DAPI (Sigma, USA) to visualize septa and nuclei, respectively. PCR was used to detect SNPs from each colony to determine the number of nuclei present in each hyphal cell. To identify SNPs in the dikaryotic strain CGMCC5.0026, its genome was shotgun sequenced using Roche 454 GS FLX to generate six million bases in a single run. After assembly with CABOG (Celera assembler with the best overlap graph)⁴⁵, SNPs were detected using the AMOS package^{72,73}. The primer list can be found in Supplementary Data S13. The trace files from the ABI 3730xl sequencer (Life Technologies, USA) were transferred to CodonCode Aligner (version 3.7.1) software for further analysis.

3. Genome sequencing

3.1 Genome sequencing using the Roche 454 GS FLX platform

Approximately 500 ng of genomic DNA was used to construct each shotgun genomic library, according to the GS FLX Titanium Rapid Library Preparation Kit manual (Roche, USA). DNA was nebulized for 1 min and was then end-repaired using T4 DNA polymerase and T4 polynucleotide kinase (Roche, USA). Adaptors were blunt-end ligated to fragment ends using T4 DNA ligase. AMPure beads (Beckman Coulter, USA) were used to remove small DNA fragments and to collect DNA fragments between 600 bp and 900 bp in length.

Roche 454 paired-end libraries were constructed for scaffolding. Approximately 50 µg of double-stranded genomic DNA was randomly sheared to average sizes of 3, 8, or 20 kb using a HydroShear apparatus (Digilab, USA). The shearing parameters were adjusted to produce DNA fragments of a specific length, and circularization adaptors containing a loxP site and a biotin moiety were blunt-ligated to end-repaired fragments. Fragments of particular size were extracted from excised gel slices via electroelution, and a fill-in reaction was conducted using Bst DNA polymerase (New England Biolabs, USA) to remove nicks present at 3' junctions. Filled-in DNA was circularized using the Cre recombinase (New England Biolabs, USA), and these circularized molecules were randomly fragmented using a nebulizer to an average size of approximately 450 to 500 bp. Fragment ends were polished using the T4 DNA polymerase and T4 polynucleotide kinase. Library fragments with an internal biotin moiety were isolated using Dynal M-270 magnetic streptavidin beads (Life Technologies, USA), and amplification adaptors were ligated to the library fragment ends. Adapted paired-end molecules were amplified using 15 cycles of PCR. PCR products were purified with calibrated AMPure beads (Beckman Coulter, USA) to accurately select DNA of the desired size. Single-stranded DNA libraries were isolated using the biotin-streptavidin interaction followed by alkaline treatment. The library fragments were validated using a Bioanalyzer 2100 (Agilent, USA) and were quantitated using the Quant-iT RiboGreen reagent (Life

Technologies, USA) prior to emulsion PCR.

Using the emulsion PCR, the DNA molecules in the shotgun library were amplified with the GS FLX Titanium LV emPCR package (Roche, USA), according to the manufacturer's recommendations. After amplification, the beads bound to amplified molecules were collected, and the emulsion oil was removed by washing, according to the manufacturer's protocol. Beads that were bound to enough copies of the clonally amplified library fragments were selected using the specified enrichment procedure and were subsequently counted with a Multisizer 3 Coulter Counter (Beckman Coulter, USA) prior to sequencing.

After the emulsion PCR enrichment, the selected beads were loaded into wells of a Titanium Series PicoTiterPlate device by centrifugation. The Roche 454 sequencing was performed according to the manufacturer's instruction manual. The image analysis, signal processing, and base calling were performed using Newbler 2.3 software (Roche, USA).

3.2 Genome sequencing using the Illumina platform

Approximately 1 µg of DNA was used to construct a 500-bp DNA library according to the protocol provided by Illumina Corporation. Genomic DNA was fragmented by nebulization using compressed nitrogen gas. The DNA ends were then polished, and an "A" base was added to the end of each fragment. Next, DNA adaptors with a single "T" base at the 3' end were ligated onto the DNA fragments. The ligated products were then purified from the excised gel fragments via electroelution following electrophoresis in a 2% agarose gel. The fragments were amplified by PCR to ensure the presence of sufficient amounts of template for sequencing. Approximately 10 to 30 µg of genomic DNA was used to construct long (≥ 2 kb), mate-paired libraries using the manufacturer's mate-paired library kit (Illumina, USA). The specific biotin-labeled dNTPs provided with the kit were employed for polishing. Fragments in desired sizes (2kb and 5kb) were extracted from the excised gel slices via electroelution. The selected DNA fragments were circularized, and the linear DNA fragments

were removed by DNA exonuclease digestion. The circularized DNA was then refragmented and purified using biotin/streptavidin-coated magnetic beads. The ends were polished, and an “A” base and adaptors were added. The collected fragments were amplified by PCR. All the libraries were sequenced on an Illumina Genome Analyzer II according to the manufacturer’s protocol (Illumina, USA).

4. Genome assembly

A combined total of 1.9×10^{10} bases were produced by sequencing using both the Roche 454 GS FLX and Illumina platforms. Approximately 9.3% of all sequences were derived from the Roche 454 GS FLX platform (Supplementary Table S1). The assembly was performed using a combination of the CABOG (version 6.1)⁴⁵ and SSPACE (version 1.1)⁴⁶ assemblers. The Roche 454 sequencing data were used for the primary assembly, and the Illumina data were used for scaffolding, gap filling, and error correction. First, low-quality and duplicate reads from the Roche 454 SFF files were filtered, and reads shorter than 75 bp were marked for deletion. The average insert size of each Roche 454 paired-end library was estimated using the Roche GS Assembler 2.3. Then, the Roche 454 SFF files were converted into an FRG format using the sffToCA embedded in CABOG. For the Roche 454 paired-end library, each read with a Titanium linker sequence was divided into mated reads. The primary assembly was produced using the CABOG pipeline (version 6.1) for the filtered Roche 454 data, with the following parameters: merSize=22, overlapper=mer, and unitigger=bog. The Illumina mate-pair and paired-end libraries were then used to construct the scaffolds. After filtering for low-quality, duplicate reads and contaminants, the primary assemblies were scaffolded using SSPACE (version 1.1)⁴⁶. The Illumina reads with 500 bp inserts were subjected to error correction and gap filling using NESONI (version 0.49) and SOAP GapCloser⁴⁷ respectively. A total of 528 indels and 514 substitutions were corrected, and the 402 gaps were filled. The final assembled sequence consisted of 194 contigs spanning a total of 43.3 Mb, which were

organized into 82 scaffolds with N/L50 being 11/1.4 Mb respectively (Supplementary Table S2).

5. Construction of Fosmid library

High-molecular-weight DNA was isolated from young mycelia using the CTAB DNA isolation method, as described⁷⁴. Genomic DNA was randomly sheared to approximately 40 kb in size and the sheared molecules were size-selected using Pulse Field Gel Electrophoresis (Bio-Rad, USA). The sheared DNA was end-repaired and ligated into the pCC2FOS vector (Epicentre, USA). The ligations were packaged using MaxPlax lambda packaging extracts (Epicentre, USA), and the lambda phages carrying foreign DNA were used to infect EPI300-T1^R *E. coli* cells (Epicentre, USA). Fosmid clones were selected on LB-chloramphenicol (12.5 µg/ml) plates. A total of 8,448 clones were selected and transferred to 384 well plates containing LB freezing medium and were then stored at -80°C.

6. Shotgun sequencing and assembly of fosmid clones

Two fosmid clones, B22A2 and B22J20, were used to construct the shotgun libraries. The fosmid DNA was sheared randomly using a Digilab HydroShear DNA Shearing apparatus (Digilab, USA). The DNA was end-repaired and size-selected for inserts ranging from 1 to 2 kb in size. The processed DNA was then ligated into the pMD-20 T-vector (Takara, Japan), and the ligation products were transformed into DH5α *E. coli* cells (Tiangen, P.R.China) and plated on LB plates containing 50 µg/ml ampicillin. Single clones were selected and amplified, and DNA was extracted and sequenced using the Sanger method. Reads were assembled into 37,181bp and 32,113 bp contigs using the CodonCode Aligner version 3.7.1.

7. Validation for genome assembly

To evaluate the local assembly accuracy of the scaffolds, two fosmid clones were aligned to 82 scaffolds independently using BLASTN⁷⁵. Approximately 99.8% of the bases in B22A2 mapped to scaffold_7, and 99.96% of the bases of B22J20 mapped to scaffold_43. We did

not observe any obvious assembly errors within the aligned regions.

An optical map of *G. lucidum* was constructed for the ordering of the scaffolds and an assessment of the genome assembly. The optical maps and *in silico* maps of the sequence scaffolds were aligned based on restriction map patterns that contain restriction recognition sites and the ordered restriction fragment sizes in kilobases⁴⁸. The optical map and the sequence scaffolds were highly congruent, as more than 86% of the assembled scaffolds aligned with the whole genome optical map.

8. Transcriptome sequencing

8.1 Transcriptome sequencing using the Roche 454 GS FLX platform

EST sequencing was performed on material obtained from the three developmental stages of *G. lucidum*, including the mycelia, primordia, and fruiting bodies. Approximately 2 µg of total RNA from each of these stages was isolated using the miRNeasy kit (Qiagen, Germany), according to the manufacturer's protocol. An additional DNaseI digestion step was performed to ensure that the samples were not contaminated with genomic DNA. RNA was converted into cDNA using a modified SMART cDNA synthesis protocol (Clontech, USA). Because long poly(A/T) tails in cDNA can lead to low-quality sequencing reads from the GS FLX system, a modified poly(T) primer with a *BsgI* site introduced between the adaptor and the poly(T) (5'-AAG CAG TGG TAT CAA CGC AGT GCA GT(20)VN-3') was designed to overcome this limitation. For cDNA synthesis, this poly(T) primer was used in combination with the Clontech SMART IV primer. Double-stranded (ds) cDNA was purified using the PureLink™ PCR Purification kit (Life Technologies, USA) with HC buffer to remove cDNAs smaller than 300 bp in size. The cDNA was then treated overnight with *BsgI* (New England Biolabs, USA) at a concentration of 3 units/µg cDNA. This restriction enzyme cuts within the poly(A) tail to reduce its length to four bases, which greatly increases the quantity and quality of the Roche 454 reads. Digested cDNA was recovered using the QIAquick PCR

Purification kit (Qiagen, Germany). The cDNA samples were sequenced using the Roche 454 GS FLX platform (Roche, USA).

8.2 RNA-Seq analysis using the Illumina platform

Total RNA and poly A⁺ RNA were isolated from the three developmental stages of *G. lucidum* (the mycelia, primordia, and fruiting bodies) using the RNeasy Plus Mini Kit (Qiagen, Germany) and the Poly(A)Purist Kit (Life Technologies, USA), respectively. Samples were analyzed for quality using the RNA 6000 Nano II kit on a Bioanalyzer 2100 (Agilent, USA). RNA-seq was performed as recommended by the manufacturer (Illumina, USA). Poly(A) mRNA was fragmented, and first-strand cDNA was synthesized using random hexamers. After the second-strand cDNA was synthesized, the cDNA ends were polished, and an “A” base was added to each cDNA fragment end. Subsequently, DNA adaptors with a single “T” base overhang at the 3’ end (Illumina, USA) were ligated to the above products. Ligation products consisting of approximately 200 bp of cDNA were extracted from excised gel slices via electroelution, and the samples were sequenced using an Illumina HiSeq 2000.

9. Transcriptome assembly

The Roche 454 contigs were created as described below. First, we prepared samples from three different development stages of *G. lucidum*, including mycelia, primordia and fruiting bodies. These samples were subjected to sequencing using the Roche 454 GS FLX platform. The sequencing resulted in 686,899, 660,659 and 641,820 reads for the three samples, respectively. The reads were assembled separately using the Newbler assembler (version 2.5.3). During the pre-processing step, the low-quality sequences (below a Phred 20 quality value) were filtered out from the SFF files using custom scripts. We then clustered the duplicates using the cd-hit tool⁷⁶. For identical reads, only one read was retained. In these steps, approximately 5% of the sequences were filtered out. During assembly, the vector file

was included in the parameter settings, and the short sequences (<20 bp) and repeat sequences were masked out by Newbler. Other parameters used include the ‘-cdna’ option used to assemble reads and a minimum overlap of 40 bp with 90% identity. The assembly resulted in 12,420, 14,431 and 12,881 contigs from mycelia, primordia and fruiting bodies, respectively. The average length of contig transcripts was between 600 bp and 700 bp.

10. Gene prediction

Gene models for the *G. lucidum* genome were predicted using an automatic annotation pipeline MAKER (version 2.10)⁵⁰. Prior to gene prediction, repeat elements, including simple repeats, satellites, and interspersed repeats, were masked throughout the genome using RepeatMasker (version open-3.2.9) and the RepBase library (version 16.08)⁷⁷. The following three types of prediction methods were used: 1) *ab initio* methods, including the Fgenesh (version v3.1.1)⁵¹, SNAP (version 2010-07-28)⁵³, and Augustus (version 2.5)⁵⁴ programs based on full gene models of *P. chrysosporium*; 2) homology-based methods, for which the predictor was seeded with alignments against the UniProtKB database (version 2011_7), the *P. chrysosporium* proteome (Release 2.0), and the *P. placenta* proteome (JGI) using BLASTX (version 2.2.17) and Exonerate:protein2genome⁷⁸ (version 1.1.0); and 3) EST-based methods, for which the predictor was seeded with 454 contigs derived from the mycelia, primordia, and fruiting bodies. These 454 contigs were directly mapped onto the genome scaffolds with a threshold of 80% coverage and 85% identity across the entire length of the corresponding contigs using BLASTN (version 2.2.17) and Exonerate:est2genome⁷⁸ (version 1.1.0). A total of 7,068 genes were supported by 454 contigs. Alternatively, the contigs were used to predict spliced transcripts or to extend the predicted gene models into full-length genes through the addition of 5’ and/or 3’ UTRs. For the synthesis steps, information was synthesized from the previously described prediction methods using MAKER to detect the optimal models and to produce a non-redundant representative set of

genes. This technique generated a set of 16,113 predicted genes, and more than 1,600 of these were manually curated using Apollo⁷⁹ based on alignment of ESTs and proteins.

11. Gene functional annotation

All predicted gene models were functionally annotated using the BLAST (version 2.2.17) database with a significance threshold E-value of 1e-5, and these predicted models were compared with a series of protein and nucleotide databases, including the NCBI nucleotide (Nt), the nonredundant set (Nr), the UniProt/Swissprot, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) protein databases. The gene models were also annotated using InterProScan⁸⁰ (version 4.6) and were assigned Gene Ontology (GO)⁸¹ terms. To obtain an overview of the gene functions encompassed within the *G. lucidum* genome, all predicted gene models were classified according to GO, eukaryotic orthologous groups (KOGs)⁸², and KEGG⁸³ metabolic pathways. A detailed description of the functional annotation is presented in Supplementary Table S6. Of the 16,113 predicted gene models, 11,179 (69.4%) demonstrated at least one hit in the above-mentioned databases. The genome of *G. lucidum* contains 215 transfer RNAs, as predicted by tRNAscan-SE⁸⁴ (version 1.23) and including 10 pseudogenes.

12. Identification of repeat content

The REPET⁵⁵ (release 1.4, <http://urgi.versailles.inra.fr/Tools/REPET>) package, which contains the two pipelines TEdenovo and TEannot, was used to detect and annotate transposable elements. The TEdenovo pipeline compared the genome with itself using BLASTER. Then, GROUPER, RECON, and PILER were used to generate clusters that were subsequently subjected to multiple alignments to derive a consensus sequence. Finally, these consensus sequences were classified according to known TE structural and coding features and/or known TEs in RepBase (REPET edition, <http://www.girinst.org>). After filtering out consensus sequences classified as “NoCat” because they were built from fewer than 10

sequences, the TEdenovo pipeline generated a library of 155 classified, non-redundant consensus sequences. The TEannot pipeline annotated the genome using BLASTER, RepeatMasker, and CENSOR with the consensus sequences produced by TEdenovo. Satellites, simple repeats, and low-complexity sequences were annotated using RepeatMasker (open-3.2.9).

13. Comparison of proteomes across different fungal species

Predicted *G. lucidum* proteins were clustered into orthologous groups using OrthoMCL⁸⁵ version 2.0 using default parameters. Orthologs from *C. cinerea*⁶⁹, *Cryptococcus gattii* (WM276)⁸⁶, *Cryptococcus neoformans* var. *neoformans* (JEC21)⁸⁷, *Emericella nidulans*⁸⁸, *Glucidum*, *Laccaria bicolor*⁸⁹, *Melampsora laricis-populina*⁹⁰, *Neurospora crassa*⁹¹, *P.chryso sporium*¹¹, *P.placenta*⁶³, *Puccinia graminis* f. sp. *tritici*⁹⁰, *S. cerevisiae*(S288C)⁹², *S.commune*¹², *Serpula lacrymans* var. *lacrymans*³⁰, and *U.maydis*⁷¹ were analyzed. All-versus-all BLASTP was accomplished using the NCBI BLAST software (version 2.2.20) and an E-value < 1e-5. Based on the OrthoMCL analysis, 296 single-copy orthologs were identified and subsequently used to perform the phylogenetic analyses. A neighbor-joining phylogenetic tree (*P-distance* model) of the fungal species was constructed using PAUP (version 4b10)⁹³, and a bootstrap analysis with 1,000 replications was performed to evaluate the stability of the phylogenetic tree.

The proteomes were compared at the following three levels: protein sequence, domain, and domain organization. We first built clusters of orthologs using proteins from *G. lucidum* and 14 other species. The numbers of proteins in each species were then counted, and a second method of comparison was employed, which was based on the diversity and abundance of Pfam domains, with an E-value < 1e-5. The type and number of protein domains in the proteomes of each species were counted and compared. The domain organizations (DOs) were also compared, and proteins were searched against the Pfam database^{7,94} for domains

with an E-value $< 1e-10$. A total of 3,166 DOs were found within the *G. lucidum* proteome (Supplementary Data S12), and 229 of these are unique to the *G. lucidum* genome.

14. Analysis of protein families expansion

For the family size analysis, we filtered out families that were inferred as being absent from the most recent common ancestor. We analyzed 1,798 gene families that contained a total of 50,267 genes. The CAFE⁸ software was used to estimate the rate of gene duplication and loss across the 15 fungal genomes. We began with the default model of the software, which estimates a single equal birth and death rate across the entire dataset. We also tested several more complex models that used branch-specific rates and family clustered rates. The model with a single birth and death rate was selected as the best-fit model using the AIC. The rate of gene birth and death across the 15 taxa was estimated as 0.011/unit time. This rate translates to approximately 37 genes that have been duplicated or lost per genome per unit time. Based on this model, we calculated p-value for the size variation of each gene family. Families with p-value less than 0.0001 were considered to have experienced significant expansion or contraction, whereas families with a branch-specific p-value less than 0.001 were considered to have experienced significant branch-specific change. Using these criteria, we found 196 families that have experienced significant change within a single branch. Of these, 16 protein families and 10 protein domains experienced a significant change in size in the branch leading to *G. lucidum*. These protein families are involved in both primary and secondary metabolism, wood degradation, and growth and development.

15. Real-time PCR analysis of CYP genes

Because the members of large CYP gene family have high sequence similarity, it is difficult to distinguish RNA-seq reads from identical regions of different genes. Consequently, the quantification of these regions cannot be reliably determined by counting short reads. Therefore, we designed primers that are specific for each gene and then performed RT-PCR

to determine their expression levels. Total RNA was extracted from *G. lucidum* mycelia, primordia, and fruiting bodies and was then treated with recombinant DNase I (Life Technologies, USA) at a concentration of 1.5 units/ μg of total RNA. The PrimeScript™ 1st Strand cDNA Synthesis kit (TaKaRa, Japan) was used for single-strand cDNA synthesis using 1 μg RNase-free DNase I-treated (TaKaRa, Japan) total RNA and random primers. Quantitative PCR (qPCR) was conducted at least three times for each sequence using SYBR®Premix Ex Taq™ (Perfect Real Time) (TaKaRa, Japan) and an ABI PRISM 7500 real-time PCR System (Life Technologies, USA). Each reaction contained 7.5 μl of 2x SYBR Green Master Mix Reagent (Life Technologies, USA), 1.0 μl (10 ng) of cDNA, and 200 nM of the gene-specific primer in a total volume of 15 μl . The PCR amplification program consisted of denaturation at 95°C for 30 s followed by 40 cycles of 95°C for 5 s and 60°C for 34 sec. The CYP expression data were normalized against an internal reference gene, *glyceraldehyde-3-phosphate dehydrogenase (GAPDH)*. The relative expression levels were calculated by comparing the Ct (cycle threshold) of the target gene with that of the housekeeping gene *GAPDH* using the $2^{-\Delta\Delta\text{Ct}}$ method. The CYP primers were designed using Primer3 (<http://frodo.wi.mit.edu/primer3/>).

16. Analysis of triterpenoids contents by high-performance liquid chromatography

16.1 Materials, reagents and instrument

Materials were collected from *G. lucidum* at four developmental stages (monokaryotic mycelia, dikaryotic mycelia, primordia, and fruiting bodies). HPLC-grade methanol, phosphoric acid, and acetonitrile were purchased from Fisher (Fisher, USA). Distilled water was purified using a Milli-Q system (Millipore, USA). Other chemicals and solvents were purchased from Beihua Chemical Reagent Co., Inc. (Beihua, China). A Waters HPLC system (Waters, USA), which was equipped with a 996 pumping system, a vacuum degasser, a quaternary gradient pump, a 2995 UV Detector, and Empower 2 software, was utilized.

Agilent ZorbaxAq (250 mm × 4.6 mm; 5 μm) and Agilent Extent C18 (250 mm × 4.6 mm; 5 μm) columns were obtained from the Agilent Corporation (Agilent, USA). The Waters Xbridge™ C18 column (250 mm × 4.6 mm; 5 μm) was purchased from Waters (Waters, USA).

16.2 Preparation of sample solutions

All samples were ground into a powder and were then dried at 60°C until they reached a constant weight. A weighed sample (2 g) of ground powder was extracted with 50 mL chloroform using an ultrasonic bath for 60 min at 25°C. The process was repeated three times, and the extracts were combined and dried under a stream of air at room temperature. The residue was then re-dissolved in 2 mL methanol and filtered through a 0.45-μm syringe filter prior to analysis. In addition, two methods for sample preparation were tested, and the chromatography peaks obtained from samples ground in liquid nitrogen overlapped perfectly with those from samples dried in an oven at 60°C.

16.3 Optimization of HPLC conditions

Several HPLC conditions were optimized to achieve successful separation. The selection of chromatography columns was the first step to optimize the protocol. Three types of chromatography columns, Agilent ZorbaxAq, Agilent Extent C18, and Waters Xbridge™ C18, were used to separate the analytes. The Waters Xbridge™ C18 column provided the best separation. The mobile phase was also optimized using Waters Xbridge™ C18 column. Methanol-aqueous phosphoric acid, acetonitrile-aqueous phosphoric acid and acetonitrile-methanol-aqueous phosphoric acid were tested, and the results indicated that the chromatographic separation was ideal when a mobile phase consisting of acetonitrile, methanol, and 3% aqueous phosphoric acid was used. Additionally, wavelengths within a scanning range between 200 nm and 400 nm were also tested, and 250 nm was found to be the optimal wavelength.

Therefore, the optimal HPLC condition is as following. Chromatographic separations were performed using the Waters Xbridge™ C18 column (250 mm × 4.6 mm; 5 μm), and the samples were eluted in a mobile phase consisting of 0.3% aqueous phosphoric acid (v/v, A), methanol (B), and acetonitrile (C). The following gradient program was applied: 25% B from 0 min to 30 min; 25% to 30% B from 30 min to 65 min; 30% to 31% B from 65 min to 80 min; 31% B from 80 min to 100 min; and 31% to 43% B from 100 min to 110 min. The flow rate was maintained at 1 mL min⁻¹, and the temperature was set to 35°C. The injection volume was 10 μL, and the DAD detector for chromatogram acquisition was set at 250 nm. UV spectra and 3D-plots were recorded between 200 and 400 nm.

SUPPLEMENTARY REFERENCES

61. Zhao, J. & Chang, S.T. Monokaryotization by protoplasting heterothallic species of edible mushrooms. *World J. Microbiol. Biotechnol.* **9**, 538-543 (1993).
62. Finn, R.D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **38**, D211-D222 (2010).
63. Martinez, D. *et al.* Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc. Natl. Acad. Sci. USA* **106**, 1954-1959 (2009).
64. Tang, H. *et al.* Synteny and collinearity in plant genomes. *Science* **320**, 486-488 (2008).
65. Xu, X. *et al.* Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189-195 (2011).
66. Cannon, S., Mitra, A., Baumgarten, A., Young, N. & May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biology* **4**, 10 (2004).
67. Casselton, L.A. & Olesnický, N.S. Molecular genetics of mating recognition in basidiomycete fungi. *Microbiol. Mol. Biol. Rev.* **62**, 55-70 (1998).
68. James, T.Y., Kues, U., Rehner, S.A. & Vilgalys, R. Evolution of the gene encoding mitochondrial intermediate peptidase and its cosegregation with the A mating-type locus of mushroom fungi. *Fungal Genet. Biol.* **41**, 381-390 (2004).
69. Stajich, J.E. *et al.* Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc. Natl.*

- Acad. Sci. USA* **107**, 11889-11894 (2010).
70. Park, H.-O. & Bi, E. Central Roles of Small GTPases in the Development of Cell Polarity in Yeast and Beyond. *Microbiol. Mol. Biol. Rev.* **71**, 48-96 (2007).
 71. Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97-101 (2006).
 72. Phillippy, A.M., Schatz, M.C. & Pop, M. Genome assembly forensics: finding the elusive mis-assembly. *Genome Biol.* **9**, R55 (2008).
 73. Schatz, M.C., Phillippy, A.M., Shneiderman, B. & Salzberg, S.L. Hawkeye: an interactive visual analytics tool for genome assemblies. *Genome Biol.* **8**, R34 (2007).
 74. Allen, G.C., Flores-Vergara, M.A., Krasynanski, S., Kumar, S. & Thompson, W.F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **1**, 2320-2325 (2006).
 75. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
 76. Niu, B. Fu, L. Sun, S. & Li, W. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* **11**, 187 (2010).
 77. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462-467 (2005).
 78. Slater, G. S. & E. Birney. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
 79. Lewis, S.E. *et al.* Apollo: a sequence annotation editor. *Genome Biol.* **3**, research0082 (2002).
 80. Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**, W116-W120 (2005).
 81. Harris, M.A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258-D261 (2004).
 82. Koonin, E.V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7 (2004).
 83. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277-D280 (2004).
 84. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* **25**, 0955-0964 (1997).
 85. Li, L., Stoeckert, C.J., & Roos, D.S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178-2189 (2003).

86. D'Souza, C.A. *et al.* Genome Variation in *Cryptococcus gattii*, an Emerging Pathogen of Immunocompetent Hosts. *mBio* **2**, e00342-10 (2011).
87. Loftus, B.J. *et al.* The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**, 1321-1324 (2005).
88. Galagan, J.E. *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**, 1105-1115 (2005).
89. Martin, F. *et al.* The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**, 88-92 (2008).
90. Duplessis, S. *et al.* Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci. USA* **108**, 9166-9171 (2011).
91. Galagan, J.E. *et al.* The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**, 859-868 (2003).
92. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546-567 (1996).
93. Posada, D. *Current Protocols in Bioinformatics* (John Wiley & Sons, Ltd. New York, USA, 2003).
94. Bateman, A. *et al.* Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* **27**, 260-262 (1999).