Data in brief 29 (2020) 105169

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on the genome analysis of the wood-rotting fungus *Steccherinum ochraceum* LE-BIN 3174

Konstantin Moiseenko [a, *], Olga Glazunova [a], Natalia Shakhova [b], Olga Savinova [a], Daria Vasina [a], Tatiana Tyazhelova [c], Nadezhda Psurtseva [b], Tatiana Fedorova [a]

[a] *A.N. Bach Institute of Biochemistry, Research Center of Biotechnology, Russian Academy of Sciences, Leninsky Ave. 33/2, Moscow, 119071, Russian Federation*
[b] *Komarov Botanical Institute of the Russian Academy of Sciences, Professor Popov St. 2, St. Petersburg, 197376, Russian Federation*
[c] *N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkin St. 3, Moscow, 117809, Russian Federation*

ARTICLE INFO

Article history:
Received 30 November 2019
Received in revised form 30 December 2019
Accepted 15 January 2020
Available online 22 January 2020

Keywords:
Steccherinum ochraceum
Draft genome sequence
White-rot
Wood decay

ABSTRACT

In the present article, we report data on the whole-genome sequencing of wood-rotting (white-rot) fungus *Steccherinum ochraceum* LE-BIN 3174. The *S. ochraceum* LE-BIN 3174 genome consists of 770 scaffolds (N50 = 62,812 bp) with the total length of assembly ~35 Mb. The structural annotation of the genome resulted in the prediction of 12,441 gene models, among which 181 were models of tRNA-coding genes, and 12,260 – protein-coding genes. The protein-coding genes were annotated with different databases (Pfam, InterPro, eggNOG, dbCAN, and MEROPS). The whole genome sequence and functional annotation provide an important information for the deep investigation of biochemical processes that take place during the late stages of wood decomposition by Basidiomycetes. The Whole Genome project of *S. ochraceum* LE-BIN 3174 had been deposited at DDBJ/ENA/GenBank under the accession RWJN00000000. The version described in this work is version RWJN00000000.1. For further interpretation of the data provided in this article, please refer to the research article "Fungal Adaptation to the Advanced Stages of

* Corresponding author.
E-mail address: mr.moiseenko@gmail.com (K. Moiseenko).

https://doi.org/10.1016/j.dib.2020.105169

Wood Decomposition: Insights from the *Steccherinum ochraceum*" [1].

Specifications Table

Subject	Biology
Specific subject area	Microbiology, Mycology, Genomics.
Type of data	Genome sequence data.
How data were acquired	Shotgun method using Illumina HiSeq 2500 with paired end runs.
Data format	Raw and analyzed data.
Parameters for data collection	The mycelium derived from field-collected basidiospores was statically cultivated on glucose-peptone (GP) medium at 26–28 °C in 750-mL Erlenmeyer flasks. The mycelium was ground in liquid nitrogen, and total DNA was extracted using DNeasy Plant Mini Kit (Qiagen, US).
Description of data collection	The genome was assembled with CLC Genomics Workbench 11.0 (Qiagen, US) and annotated with Funannotate pipeline v1.5.0 (https://github.com/nextgenusfs/funannotate)
Data source location	The fungal strain of *Steccherinum ochraceum* (Pers. ex J.F. Gmel.) Gray was isolated (August 01, 2013) from basidiospores collected from a fallen dry aspen branch in the polydominant temperate deciduous broadleaf forest (Kaluzhskiye Zaseki Nature Reserve, Russia; N 53°33′28.4″; E 35°38′24.4″). The strain was deposited in the Komarov Botanical Institute Basidiomycetes Culture Collection (LE-BIN; St. Petersburg, Russia) as *S. ochraceum* LE-BIN 3174
Data accessibility	The whole genome sequence of *Steccherinum ochraceum* LE-BIN 3174 had been deposited at DDBJ/ENA/GenBank under the accession RWJN00000000. The version described in this paper is version RWJN00000000.1. The BioSample and BioProject accession numbers are SAMN10505049 and PRJNA507755, respectively. All other data are within this article.
Related research article	K.V. Moiseenko, O.A. Glazunova, N.V. Shakhova, O.S. Savinova, D.V. Vasina, T.V. Tyazhelova, N.V. Psurtseva, T.V. Fedorova, Fungal Adaptation to the Advanced Stages of Wood Decomposition: Insights from the *Steccherinum ochraceum*, Microorganisms. 7 (2019) 527. https://doi.org/10.3390/microorganisms7110527 [1].

Value of the Data

- The genome of *Steccherinum ochraceum* LE-BIN 3174 is the first genome under the family *Steccherinaceae* to be reported.
- This draft genome will accelerate functional genomics research, increase the knowledge of the biochemical process of wood degradation and create an opportunity for comparative studies with other fungi.
- The CAZyme content of this genome will provide a valuable insight into the fungal adaptation to an ecological niche of pre-degraded wood.

1. Data description

Steccherinum ochraceum is a white-rot basidiomycete with wide ecological amplitude. It occurs in different regions of Russia and throughout the world occupying different climatic zones. The obtained draft genome of *S. ochraceum* LE-BIN 3174 (DDBJ/ENA/GenBank accession/version – RWJN00000000.1) is represented by the 770 scaffolds with the total length of 35.27 Mb and of comparable quality with other previously sequenced genomes of polypore fungi [2]. The gene prediction resulted in 12,441 gene models. The general information regarding genome's assembly, structural and functional annotation is presented in Table 1. The summary of the Gene Ontology (GO) classification of the protein coding genes is illustrated in Fig. 1. The whole genome sequence of *S. ochraceum* LE-BIN 3174 showed that it harbors 361 carbohydrate-active enzymes (CAZymes). The auxiliary activity enzymes (AA), carbohydrate

Table 1
General data on the genome sequencing of *S. ochraceum* LE-BIN 3174.

Sequencing			
Sequencing technology	Illumina HiSeq 2500	Total number of paired-reads (2×)	2 × 47 868 586
Length of paired-reads (2×), bp	2 × 100	Insert size, bp	300–500
Assembly		Structural annotation	
Assembly size	35.27 (Mb)	Repeat content	1.4 (%)
Overall coverage	100×	Overall GC	52.7 (%)
Number of scaffolds	770	Number of predicted genes	12 441
Longest scaffold	464 123 (bp)	Proportion covered by genes	64.1 (%)
N50 length of scaffolds	62 812 (bp)	Number of tRNA-coding genes	181
Mean size of scaffolds	45 812 (bp)	Number of protein-coding genes	12 260
Median size of scaffolds	33 955 (bp)	Mean protein size	483 (aa)
Main functional annotation			
General-content databases		Domain-specific databases	
Pfam	6965	dbCAN	369
InterPro	8186	MEROPS	382
eggNOG	9237		
Additional functional features			
Proteins with signal peptides	1093	Proteins with transmembrane helices	2585

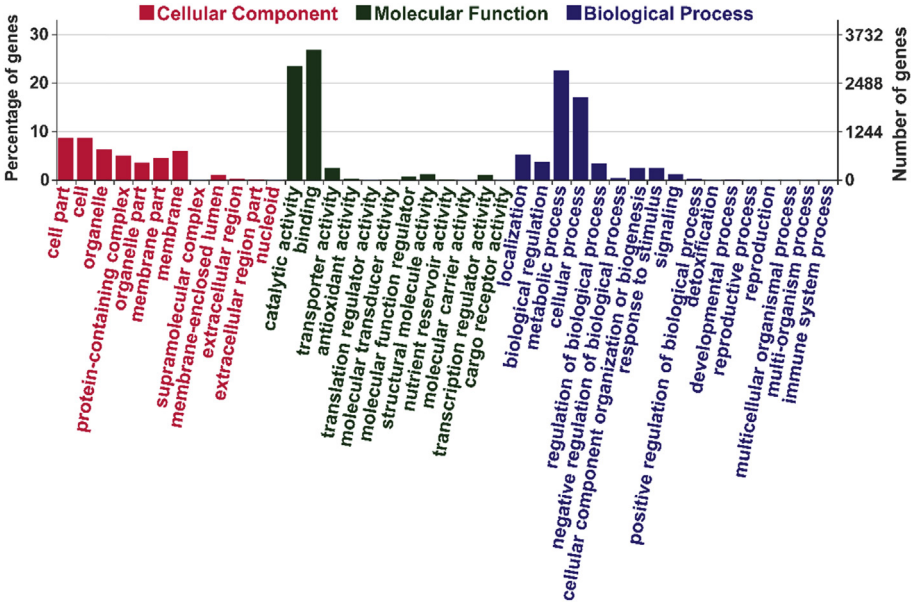

Fig. 1. The Gene Ontology (GO) functional annotation of *S. ochraceum* LE-BIN 3174.

esterase (CE), glycoside hydrolases (GH), glycosyl transferase (GT), and polysaccharide lyase (PL) superfamilies were represented by 109, 37, 151, 55, and 9 CAZymes from 9, 8, 48, 25, and 3 families, respectively. The comparison of the *S. ochraceum* CAZymes genome content with those from other lignocellulose decaying fungi belonging to different trophic groups is presented in Fig. 2 and Fig. 3.

CAZyme family	Row-wise color guide: Low								High	CAZyme family	Row-wise color guide: Low								High
	TraVer[1]	TraPub	TraHir	SteOhr	GymJun	HymRad	MycGal	CruLae	AgrPra		TraVer	TraPub	TraHir	SteOhr	GymJun	HymRad	MycGal	CruLae	AgrPra
AA1_1	7	6	7	8	3	17	28	15	8	GH1	2	2	1	2	3	0	4	3	3
AA1_2	2	1	1	1	1	7	7	1	1	GH2	5	3	4	2	2	3	8	4	2
AA1_3	1	1	1	0	0	0	2	0	1	GH3	12	10	8	8	9	12	24	10	8
AA2	27	22	18	11	14	13	61	17	16	GH5	22	20	19	17	31	34	51	24	22
AA3	2	2	3	5	1	7	18	1	2	GH6	1	1	1	1	3	3	2	4	2
AA3_1	0	0	0	0	0	2	0	0	0	GH7	4	4	4	4	4	5	12	10	5
AA3_2	16	15	21	25	22	46	97	52	24	GH9	1	1	1	1	1	1	1	1	1
AA3_3	4	3	7	14	3	15	10	3	4	GH10	6	5	5	4	5	6	8	7	3
AA3_4	1	1	0	0	0	0	0	1	0	GH11	0	0	0	0	2	0	8	1	5
AA4	0	0	0	0	0	1	5	0	0	GH12	2	2	1	1	1	1	5	5	2
AA5_1	8	6	5	4	8	19	17	6	14	GH13	6	6	7	10	6	17	15	8	9
AA6	1	1	1	4	3	6	6	4	2	GH15	4	3	3	1	2	2	3	6	2
AA7	3	2	3	16	12	19	24	10	10	GH16	19	12	14	13	11	36	27	18	14
AA8	1	1	1	1	1	1	2	1	1	GH17	1	0	0	2	2	3	1	3	2
AA9	18	17	15	18	14	33	15	34	20	GH18	20	18	18	10	12	23	35	13	16
AA10	0	0	0	0	0	0	2	0	0	GH20	5	7	5	4	3	7	4	4	2
AA12	0	0	0	0	1	0	2	1	0	GH24	0	0	0	0	1	0	3	3	2
AA13	0	0	0	0	0	0	0	0	1	GH25	1	1	1	1	0	0	1	5	1
AA14.phmm	4	4	4	2	2	8	2	3	5	GH27	0	0	0	0	0	1	1	0	0
Total_AA	95	82	87	109	85	194	298	149	109	GH28	9	7	9	5	4	9	26	5	7
CE1	1	1	1	2	2	2	6	3	3	GH29	0	0	0	0	0	3	4	1	0
CE2	1	0	0	1	1	1	2	1	2	GH30	4	4	5	1	4	3	11	3	3
CE3	0	0	0	0	0	6	8	0	0	GH31	5	5	7	3	5	11	6	6	5
CE4	2	2	0	2	7	11	20	10	7	GH32	3	1	1	0	1	0	1	0	0
CE5	0	0	0	0	2	1	10	6	5	GH35	2	1	3	2	5	4	7	5	5
CE8	2	0	2	2	3	3	8	3	5	GH37	2	2	2	2	2	2	3	2	2
CE9	1	1	1	1	0	1	1	1	1	GH38	1	1	1	1	1	1	4	1	1
CE10	35	26	15	18	24	27	52	26	31	GH39	0	0	0	0	0	0	1	0	0
CE12	0	1	0	0	0	6	3	2	3	GH43	3	3	3	2	1	16	24	8	5
CE15	2	2	2	2	3	3	3	2	2	GH44	0	0	0	1	1	1	1	1	1
CE16	7	6	9	9	5	10	16	7	5	GH45	1	1	1	1	0	5	0	0	0
Total_CE	51	39	30	37	47	71	129	61	64	GH47	5	5	6	8	8	11	19	9	8
GT1	3	2	1	1	1	2	6	3	2	GH51	2	2	1	2	1	6	3	6	6
GT2	14	15	14	14	13	23	21	14	12	GH53	1	0	1	1	1	4	2	1	1
GT3	2	1	1	1	1	0	1	1	1	GH54	0	0	0	0	0	0	8	0	0
GT4	4	4	4	4	4	5	11	4	4	GH55	2	2	2	2	2	5	7	3	2
GT8	5	4	8	4	3	7	7	3	3	GH62	0	0	0	0	0	0	3	0	0
GT15	5	5	3	5	4	13	13	8	5	GH67	0	0	0	0	0	2	1	0	0
GT17	1	1	1	1	1	2	1	1	1	GH71	5	6	6	3	3	6	13	3	1
GT20	2	2	3	2	2	3	2	2	3	GH72	1	1	1	1	1	2	6	1	1
GT21	1	1	1	1	1	2	1	1	1	GH74	1	1	1	0	1	1	2	1	1
GT22	4	3	2	2	4	5	3	3	3	GH75	0	0	0	0	0	0	1	0	0
GT24	1	1	1	0	1	1	1	1	1	GH76	0	0	1	0	2	3	5	2	1
GT25	1	0	0	1	0	2	6	0	1	GH78	3	2	0	1	1	2	9	2	2
GT28	0	0	0	1	0	0	0	0	0	GH79	11	8	8	7	8	2	14	9	9
GT31	0	0	0	0	3	0	0	4	5	GH81	0	0	0	0	0	0	1	0	0
GT32	1	1	1	1	1	1	1	1	1	GH85	1	1	0	1	1	1	1	2	1
GT33	1	1	1	1	1	1	1	1	1	GH88	1	1	1	2	1	2	1	1	1
GT35	1	1	1	1	0	1	1	1	1	GH89	1	1	1	1	0	0	0	0	0
GT39	3	3	3	3	6	4	4	3	3	GH92	4	4	3	4	2	4	10	4	4
GT48	1	1	1	2	3	2	2	2	2	GH93	0	0	0	0	0	1	3	0	0
GT49	0	0	0	0	1	0	1	0	1	GH95	1	1	1	1	1	2	3	1	1
GT50	1	1	1	1	1	0	1	1	1	GH105	3	3	2	3	4	10	6	5	2
GT57	2	2	2	2	2	2	2	2	2	GH106	0	0	0	0	0	0	2	0	0
GT58	1	1	0	1	1	2	1	1	1	GH109	1	1	0	1	0	1	1	1	0
GT59	1	1	1	1	1	0	1	1	1	GH114	0	0	0	0	0	0	3	0	0
GT66	1	1	1	1	1	1	1	1	1	GH115	2	2	2	3	2	2	2	1	1
GT69	3	3	3	2	2	5	7	2	2	GH125	1	1	1	1	2	2	6	1	1
GT76	1	1	1	1	1	1	1	1	1	GH128	3	3	5	3	3	10	15	4	5
GT90	1	1	0	1	4	3	12	5	2	GH131	3	2	3	2	2	5	2	2	2
Total_GT	61	57	55	55	63	88	109	67	62	GH133	1	1	1	1	1	1	1	1	1
PL1	0	0	0	0	0	15	8	5	0	GH134	0	0	0	0	0	0	0	1	0
PL3	0	0	0	0	0	9	0	2	0	GH135	2	2	1	1	1	1	10	3	0
PL4	1	0	1	2	0	5	1	2	2	GH145	1	1	1	1	2	1	1	1	1
PL8	1	2	2	2	1	2	4	2	3	GH146	0	1	0	0	0	0	0	0	0
PL9	0	0	0	0	0	3	0	0	0	GH152	7	2	6	2	2	7	15	3	3
PL14	6	7	6	5	4	3	5	4	5	Total_GH	204	174	179	151	174	303	477	229	185
PL20	0	0	0	0	0	0	1	0	0	[1]TraVer - Trametes versicolor, TraPub - Trametes pubescens, TraHir - Trametes hirsuta,									
PL26	0	0	0	0	0	2	0	1	0	SteOhr - Steccherinum ochraceum, GymJun - Gymnopilus junonius, HymRad - Hymenopellis radicata,									
Total_PL	8	9	9	9	5	39	19	16	10	MycGal - Mycena galopus, CruLae - Crucibulum laeve, AgrPra - Agrocybe praecox.									

Fig. 2. Families of carbohydrate-degrading enzymes (CAZymes) of *S. ochraceum* LE-BIN 3174 and other basidiomycetes.

2. Experimental design, materials, and methods

2.1. Fungal strain isolation and genetic verification

The fungal strain of *Steccherinum ochraceum* (Pers. ex J.F. Gmel.) Gray was isolated (August 01, 2013) from basidiospores collected from a fallen dry aspen branch in the polydominant temperate deciduous broadleaf forest (Kaluzhskiye Zaseki Nature Reserve, Russia; N 53°33′28.4″; E 35°38′24.4″). After morphological and genetic verifications, the strain was deposited in the Komarov Botanical Institute Basidiomycetes Culture Collection (LE-BIN; St. Petersburg, Russia) as *S. ochraceum* LE-BIN 3174.

Row-wise color guide: Low — High

Substrate		Enzyme activity	EC no.	Abbreviation	CAZyme family	*TraVer*[2]	*TraPub*	*TraHir*	*SteOhr*	*GymJun*	*HymRad*	*MycGal*	*CruLae*	*AgrPra*
Lignin		Class II peroxidase	1.11.1.13/14/16	POD	AA2	27	22	18	11	14	13	61	17	16
		Laccase	1.10.3.2	LCC	AA1_1	7	6	7	8	3	17	28	15	8
					Total	34	28	25	19	17	30	89	32	24
Cellulose		β-1,4-Endoglucanase	3.2.1.4	EG	GH3	12	10	8	8	9	12	24	10	8
					GH5	22	20	19	17	31	34	51	24	22
					GH6	1	1	1	1	3	3	2	4	2
					GH7	4	4	4	4	4	5	12	10	5
					GH9	1	1	1	1	1	1	1	1	1
					GH12	2	2	1	1	1	1	5	5	2
					GH45	1	1	1	1	0	5	0	0	0
		Cellobiohydrolase (reducing end)	3.2.1.176	CBHI	GH7	4	4	4	4	4	5	12	10	5
		Cellobiohydrolase (nonreducing end)	3.2.1.91	CBHII	GH6	1	1	1	1	3	3	2	4	2
		β-1,4-Glucosidase	3.2.1.21	BGL	GH1	2	2	1	2	3	0	4	3	3
					GH3	12	10	8	8	9	12	24	10	8
		Cellobiose dehydrogenase	1.1.99.18	CDH	AA3_1	0	0	0	0	0	2	0	0	0
					AA8	1	1	1	1	1	1	2	1	1
		Lytic polysaccharide monooxygenase	NA[3]	LPMO	AA9	18	17	15	18	14	33	15	34	20
					Total	81	74	65	67	83	117	154	116	79
Hemicelluloses	Xylan	β-1,4-Endoxylanase	3.2.1.8	XLN	GH10	6	5	5	4	5	6	8	7	3
					GH11	0	0	0	0	2	0	8	1	5
		β-1,4-Xylosidase	3.2.1.37	BXL	GH3	12	10	8	8	9	12	24	10	8
					GH43	3	3	3	2	1	16	24	8	5
					Total	21	18	16	14	17	34	64	26	21
	Galactomannan	β-1,4-Endomannanase	3.2.1.78	MAN	GH5	22	20	19	17	31	34	51	24	22
					GH26	0	0	0	0	0	0	0	0	0
		β-1,4-Mannosidase	3.2.1.25	MND	GH2	5	3	4	2	2	3	8	4	2
		β-1,4-Galactosidase	3.2.1.23	LAC	GH2	5	3	4	2	2	3	8	4	2
					GH35	2	1	3	2	5	4	7	5	5
		α-1,4-Galactosidase	3.2.1.22	AGL	GH27	0	0	0	0	0	1	1	0	0
					GH36	0	0	0	0	0	0	0	0	0
		α-Arabinofuranosidase	3.2.1.55	ABF	GH51	2	2	1	2	1	6	3	6	6
					GH54	0	0	0	0	0	0	8	0	0
					Total	36	29	31	25	41	51	86	43	37
	Xyloglucan	Xyloglucan β-1,4-endoglucanase	3.2.1.151	XEG	GH12	2	2	1	1	1	1	5	5	2
					GH74	1	1	1	0	1	1	2	1	1
		α-Arabinofuranosidase	3.2.1.55	ABF	GH51	2	2	1	2	1	6	3	6	6
					GH54	0	0	0	0	0	0	8	0	0
		α-Xylosidase	3.2.1.177	AXL	GH31	5	5	7	3	5	11	6	6	5
		α-Fucosidase	3.2.1.51	AFC	GH29	0	0	0	0	0	3	4	1	0
					GH95	1	1	1	1	1	2	3	1	1
		α-1,4-Galactosidase	3.2.1.22	AGL	GH27	0	0	0	0	0	1	1	0	0
					GH36	0	0	0	0	0	0	0	0	0
		β-1,4-Galactosidase	3.2.1.23	LAC	GH2	5	3	4	2	2	3	8	4	2
					GH35	2	1	3	2	5	4	7	5	5
					Total	18	15	18	11	16	32	47	29	22
	Arabinoxylan	Arabinoxylan arabinofuranohydrolase/arabinofuranosidase	3.2.1.55	AXH	GH62	0	0	0	0	0	0	3	0	0
		α-Glucuronidase	3.2.1.139	AGU	GH67	0	0	0	0	0	2	1	0	0
					GH115	2	2	2	3	2	2	2	1	1
		α-1,4-Galactosidase	3.2.1.22	AGL	GH27	0	0	0	0	0	1	1	0	0
					GH36	0	0	0	0	0	0	0	0	0
		β-1,4-Galactosidase	3.2.1.23	LAC	GH2	5	3	4	2	2	3	8	4	2
					GH35	2	1	3	2	5	4	7	5	5
		Acetyl xylan esterase	3.1.1.72	AXE	CE1	1	1	1	2	2	2	6	3	3
					CE5	0	0	0	0	2	1	10	6	5
		Feruloyl esterase	3.1.1.73	FAE	CE1	1	1	1	2	2	2	6	3	3
					Total	11	8	11	11	15	17	44	22	19
					Total	86	70	76	61	89	134	241	120	99
Pectin		Endopolygalacturonases	3.2.1.15	PGA	GH28	9	7	9	5	4	9	26	5	7
		Exopolygalacturonases	3.2.1.67	PGX	GH28	9	7	9	5	4	9	26	5	7
		Endorhamnogalacturonase	3.2.1.171	RHG	GH28	9	7	9	5	4	9	26	5	7
		Exorhamnogalacturonase	3.2.1.–	RHX	GH28	9	7	9	5	4	9	26	5	7
		Rhamnogalacturonan rhamnohydrolase	3.2.1.174	RGXB	GH28	9	7	9	5	4	9	26	5	7
		α-Rhamnosidase	3.2.1.40	RHA	GH78	3	2	0	1	1	2	9	2	2
		α-Arabinofuranosidase	3.2.1.55	ABF	GH51	2	2	1	2	1	6	3	6	6
					GH54	0	0	0	0	0	0	8	0	0
					GH62	0	0	0	0	0	0	3	0	0
		Endoarabinanase	3.2.1.99	ABN	GH43	3	3	3	2	1	16	24	8	5
		Exoarabinanase	3.2.1.–	ABX	GH93	0	0	0	0	0	1	3	0	0
		β-1,4-Endogalactanase	3.2.1.89	GAL	GH53	1	0	1	1	1	4	2	1	1
		Unsaturated glucuronyl hydrolase	3.2.1.–	UGH	GH88	1	1	1	2	1	2	1	1	1
		Unsaturated rhamnogalacturonan hydrolase	3.2.1.172	URH	GH105	3	3	2	3	4	10	6	5	2
		β-1,4-Xylosidase	3.2.1.37	BXL	GH3	12	10	8	8	9	12	24	10	8
					GH43	3	3	3	2	1	16	24	8	5
		β-1,4-Galactosidase	3.2.1.23	LAC	GH2	5	3	4	2	2	3	8	4	2
					GH35	2	1	3	2	5	4	7	5	5
		Pectin lyase	4.2.2.10	PEL	PL1	0	0	0	0	0	15	8	5	0
		Pectate lyase	4.2.2.2	PLY	PL1	0	0	0	0	0	15	8	5	0
					PL3	0	0	0	0	0	9	0	2	0
					PL9	0	0	0	0	0	3	0	0	0
		Rhamnogalacturonan lyase	4.2.2.23	RGL	PL4	1	0	1	2	0	5	1	2	2
					PL11	0	0	0	0	0	0	0	0	0
		Pectin methyl esterase	3.1.1.11	PME	CE8	2	0	2	2	3	3	8	3	5
		Rhamnogalacturonan acetyl esterase	3.1.1.–	RGAE	CE12	0	1	0	0	0	6	3	2	3
		Feruloyl esterase	3.1.1.73	FAE	CE1	1	1	1	2	2	2	6	3	3
					Total	84	65	75	56	51	179	286	97	85

[3]Classification is based on the on the [15]; Please note the redundancy in the classification scheme: the same CAZyme can simultaneously act on several components of lignocellulose.

[2]*TraVer* - *Trametes versicolor*, *TraPub* - *Trametes pubescens*, *TraHir* - *Trametes hirsuta*, *SteOhr* - *Stecchenirum ochraceum*, *GymJun* - *Gymnopilus junonius*, *HymRad* - *Hymenopellis radicata*, *MycGal* - *Mycena galopus*, *CruLae* - *Crucibulum laeve*, *AgrPra* - *Agrocybe praecox*.

[3]NA - not categorized by the International Union of Biochemistry and Molecular Biology (IUBMB).

Fig. 3. Families of carbohydrate-degrading enzymes (CAZymes) related to plant polysaccharide degradation in *S. ochraceum* LE-BIN 3174 and other fungal genomes.

For the genetic verification, the genomic DNA (gDNA) was extracted as described later in the "Genomic DNA Isolation, Library Preparation and Sequencing" section of this manuscript, and the sequence of ITS1-5.8S rRNA-ITS2 region was obtained using the standart primers: ITS1F 5′—CTT GGT CAT TTA GAG GAA GTA A—3′ and ITS4B 5′—CAG GAG ACT TGT ACA CGG TCC AG—3′. The PCR amplification was performed using the Encyclo PCR kit (Evrogen, Russia) under the following conditions: 1 cycle of 5 min at 95 °C; 25 cycles of 1 min at 90 °C, 1 min at 56 °C, and 1 min at 72 °C; 1 cycle of 10 min at 72 °C. Obtained PCR reaction mixture was resolved using 1,2% agarose gel. The performed PCR amplification produced the single PCR-product with approximate length of 830 bp. The obtained product was ceased from the gel and purified with QIAquick Gel Extraction Kit (Qiagen, USA), according to the manufacturer's instructions. The Sanger sequencing of the obtained fragment was performed at the Evrogen JSC (Russia, Moscow).

2.2. Genomic DNA isolation, Library Preparation and Sequencing

For gDNA extraction, *S. ochraceum* LE-BIN 3174 was statically cultivated at 26–28 °C in 750-mL Erlenmeyer flasks contained 200 mL of glucose-peptone (GP) medium (per 1 L of dH_2O): 3.0 g peptone, 10.0 g glucose, 0.6 g KH_2PO_4, 0.4 g K_2HPO_4, 0.5 g $MgSO_4$, 50 mg $MnSO_4$, 1 mg $ZnSO_4$ and 0.5 mg $FeSO_4$. The mycelium was ground in liquid nitrogen, and gDNA was extracted using DNeasy Plant Mini Kit (Qiagen, US). The quality and quantity of the isolated DNA were checked using Agilent Bioanalyzer 2100 (Agilent Technologies, US) and Qubit fluorimeter (Thermo Fisher Scientific, US).

After ultrasonic fragmentation the gDNA was prepared for sequencing using TruSeq DNA Sample Prep Kit (Illumina, US). The quality and quantity of the obtained DNA-library were checked using Agilent Bioanalyzer 2100 and StepOnePlus Real-Time PCR System (Thermo Fisher Scientific, US). The whole genome sequencing was carried out with Illumina HiSeq 2500 system (Illumina, US) using HiSeq Rapid SBS Kit v2 at the Evrogen JSC (Russia, Moscow).

2.3. Genome sequencing, assembly and annotation

The shotgun sequencing produced 2 × 47,868,586 paired-end reads (2 × 100 bp) with an insert size of 300–500 bp. The reads were further processed with CLC Genomics Workbench 11.0 (Qiagen, US) as follows: (1) adapters were removed from all reads; (2) all reads were trimmed based on their quality; (3) reads were sampled to reduce coverage to a maximum average coverage of 100 × ; (4) reads were *de novo* assembled, and resulted contigs were scaffolded.

Genome structural and functional annotations were performed using Funannotate pipeline v1.5.0 (https://github.com/nextgenusfs/funannotate).

The structural annotation step included: (1) repeat masking with the RepeatMasker package (http://www.repeatmasker.org/) using the RepBase repeats libraries [3]; (2) *ab initio* protein-coding gene prediction with self-trained GeneMark-ES [4] and AUGUSTUS [5] trained using BUSCO 2.0 [6] gene models (*Phanerochaete chrysosporium* was selected as a closely-related species); (3) *ab initio* tRNA-coding gene prediction with tRNAscan-SE [7]; (4) integration and filtering of the obtained gene models.

The functional annotation of the predicted protein-coding genes was performed with three general-content databases: the protein families database – Pfam [8], the integrative protein signature database – InterPro [9], and the orthologous groups database – eggNOG [10]. Additionally, two domain-specific databases were employed: carbohydrate-active enzyme (CAZyme) database – dbCAN [11], and peptidase database – MEROPS [12]. The prediction of transmembrane topologies and signal peptides was performed with Phobius [13] and SignalP [14], respectively.

The data on genome sequencing, assembly and annotation are presented in Table 1.

As a result of general functional prediction, 6019 genese were annotated with the GO terms. In total, 10,648 GO terms were assigned, from which 1707 were GO terms related to "Cellular component" class, 5207 – to "Molecular function" class, and 3734 – to "Biological process" class (Fig. 1).

2.4. The peculiarities of the S. ochraceum LE-BIN 3174 CAZymes genome content

Based on the sequenced genome, the CAZymes repertoire of *S. ochraceum*LE-BIN 3174 was inferred and compared with those of the 8 fungi belonging to the different ecological niches and trophic groups. From the Polyporales order: *Trametes versicolor*, *Trametes pubescens*, and *Trametes hirsuta* (all are primary colonizers on *lignum*). From the Agaricales order: *Gymnopilus junonius* (secondary colonizer on *lignum*), *Hymenopellis radicata* (deep root mushroom, *lignum*), *Mycena galopus* (saprotroph on *folia dejecta*), *Crucibulum laeve* (saprotroph on *stramentum*), and *Agrocybe praecox* (saprotroph on *humus*).

Comparison of the total CAZymes contentis present in Fig. 2.

Comparison of the content of CAZymes acting on different polymeric components of lignocellulose [15] is presented in Fig. 3. Please note, that the numbers do not add up properly due to the redundancy in the classification scheme that was advanced to reflect different enzymatic activities possessed by fungi rather than different CAZymes, since the same CAZyme can simultaneously act on several components of lignocellulose.

Acknowledgments

We would like to thank everyone who contributed directly or indirectly to this study, especially Vasil R. Sultanov for his excellent bioinformatical assistance.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2020.105169.

References

[1] K.V. Moiseenko, O.A. Glazunova, N.V. Shakhova, O.S. Savinova, D.V. Vasina, T.V. Tyazhelova, N.V. Psurtseva, T.V. Fedorova, Fungal adaptation to the advanced stages of wood decomposition: Insights from the *Steccherinum ochraceum*, Microorganisms 7 (2019) 527, https://doi.org/10.3390/microorganisms7110527.

[2] H. Nordberg, M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, I. Shabalov, T. Smirnova, I.V. Grigoriev, I. Dubchak, The genome portal of the department of energy joint genome Institute: 2014 updates, Nucleic Acids Res. 42 (2014) D26–D31, https://doi.org/10.1093/nar/gkt1069.

[3] W. Bao, K.K. Kojima, O. Kohany, Repbase Update, a database of repetitive elements in eukaryotic genomes, Mobile DNA 6 (2015) 11, https://doi.org/10.1186/s13100-015-0041-9.

[4] J. Besemer, A. Lomsadze, M. Borodovsky, GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions, Nucleic Acids Res. 29 (2001) 2607–2618. http://www.ncbi.nlm.nih.gov/pubmed/11410670.

[5] M. Stanke, R. Steinkamp, S. Waack, B. Morgenstern, AUGUSTUS: a web server for gene finding in eukaryotes, Nucleic Acids Res. 32 (2004) W309–W312, https://doi.org/10.1093/nar/gkh379.

[6] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (2015) 3210–3212, https://doi.org/10.1093/bioinformatics/btv351.

[7] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964. http://www.ncbi.nlm.nih.gov/pubmed/9023104.

[8] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: towards a more sustainable future, Nucleic Acids Res. 44 (2016) D279–D285, https://doi.org/10.1093/nar/gkv1344.

[9] S. Hunter, R. Apweiler, T.K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R.D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A.F. Quinn, J.D. Selengut, C.J.A. Sigrist, M. Thimma, P.D. Thomas, F. Valentin, D. Wilson, C.H. Wu, C. Yeats, InterPro: the integrative protein signature database, Nucleic Acids Res. 37 (2009) D211–D215, https://doi.org/10.1093/nar/gkn785.

[10] J. Huerta-Cepas, D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M.C. Walter, T. Rattei, D.R. Mende, S. Sunagawa, M. Kuhn, L.J. Jensen, C. von Mering, P. Bork, eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences, Nucleic Acids Res. 44 (2016) D286–D293, https://doi.org/10.1093/nar/gkv1248.

[11] Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, Y. Xu, dbCAN: a web resource for automated carbohydrate-active enzyme annotation, Nucleic Acids Res. 40 (2012) W445–W451, https://doi.org/10.1093/nar/gks479.
[12] N.D. Rawlings, A.J. Barrett, A. Bateman, MEROPS: the database of proteolytic enzymes, their substrates and inhibitors, Nucleic Acids Res. 40 (2012) D343–D350, https://doi.org/10.1093/nar/gkr987.
[13] L. Käll, A. Krogh, E.L.L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server, Nucleic Acids Res. 35 (2007) W429–W432, https://doi.org/10.1093/nar/gkm256.
[14] H. Nielsen, Predicting secretory proteins with SignalP, in: D. Kihara (Ed.), Protein Funct. Predict. Methods Mol. Biol. v.1611, Humana Press Inc., New York, 2017, pp. 59–73, https://doi.org/10.1007/978-1-4939-7015-5_6.
[15] J. Rytioja, K. Hildén, J. Yuzon, A. Hatakka, R.P. de Vries, M.R. Mäkelä, Plant-polysaccharide-degrading enzymes from basidiomycetes, Microbiol. Mol. Biol. Rev. 78 (2014) 614–649, https://doi.org/10.1128/MMBR.00035-14.