## **Supplementary Information**

Genome sequencing and analysis of *Talaromyces pinophilus* provide insights into biotechnological applications

Cheng-Xi Li<sup>†</sup>, Shuai Zhao<sup>†</sup>, Ting Zhang, Liang Xian, Lu-Sheng Liao, Jun-Liang Liu, Jia-Xun Feng<sup>\*</sup>

State Key Laboratory for Conservation and Utilization of Subtropical Agrobioresources, College of Life Science and Technology, Guangxi University, 100 Daxue Road, Nanning, Guangxi 530004, People's Republic of China.

\*Corresponding author. E-mail address: jiaxunfeng@sohu.com (J.-X. Feng) \*These authors contributed equally to this work.

## Supplementary files provided with this submission:

**Figure S1.** Venn diagram showing the unique and shared proteins in *T. pinophilus* 1-95 annotated using the Non-redundant (NR), UniProt, Kyoto Encyclopaedia of Genes and Genomes (KEGG), the Clusters of Orthologous Groups of proteins (COG) and Gene Ontology (GO) databases.

**Figure S2.** Functional annotation of proteins as determined using the Gene Ontology (GO) database, and comparative analysis of the identified proteins in *T. pinophilus* 1-95 and *T. cellulolyticus* Y-94.

**Figure S3.** Functional annotation of the secreted proteins of *T. pinophilus* 1-95 using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database.

 Table S1. Comparison analysis of genome features among *T. pinophilus* 1-95 and

 another 10 filamentous fungi producing biomass degrading enzymes.

 Table S2. Assessment analysis of *T. pinophilus* 1-95 genome by software CEGMA and

 BUSCO.

 Table S3. Comparative analysis of homologous proteins among *T. pinophilus* 1-95 and

 another 10 filamentous fungi producing biomass degrading enzymes.

**Table S4.** Paired-end (PE) libraries with average insert sizes 500 bp mapped the whole genome sequence of *Talaromyces cellulolyticus* Y-94.

**Table S5.** Comparison of plant cell wall degrading enzymes (CWDEs) among T.pinophilus 1-95, T. cellulolyticus Y-94 and Penicillium oxalicum HP7-1.

 Table S6. Comparison of starch degrading enzymes among *T. pinophilus* 1-95 and

 another 10 filamentous fungi producing biomass degrading enzymes.

**Table S7.** The predicted secondary metabolism gene clusters in *T. pinophilus* 1-95 and the comparison analysis among with another 10 filamentous fungi producing biomass degrading enzymes.

**Dataset S1.** Orthologs used for the construction of phylogenetic tree among *Talaromyces pinophilus* 1-95 and another 10 filamentous fungi producing biomass degrading enzymes.

Dataset S2. The putative transporters in *Talaromyces pinophilus* 1-95.

Dataset S3. The secretome of Talaromyces pinophilus 1-95 and function annotation



**Supplementary Figure S1.** Venn diagram showing the unique and shared proteins in *T. pinophilus* 1-95 annotated using the Non-redundant (NR), UniProt, Kyoto Encyclopaedia of Genes and Genomes (KEGG), the Clusters of Orthologous Groups of proteins (COG) and Gene Ontology (GO) databases.



**Supplementary Figure S2.** Functional annotation of proteins as determined using the Gene Ontology (GO) database, and comparative analysis of the identified proteins in *T. pinophilus* 1-95 and *T. cellulolyticus* Y-94.



Supplementary Figure S3. Functional annotation of the secreted proteins of T. pinophilus 1-

95 using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database.

**Supplementary Table S1.** Comparison analysis of genome features among *T. pinophilus* 1-95 and another 10 filamentous fungi producing biomass degrading enzymes.

Strains	Genome size (Mb)	Protein Number*	GC (%)	Data Resource
T. pinophilu 1-95	36.51	13472	50.60	This study
Aspergillus nidulans FGSC A4	30.24	9541	50.32	GCF_000149205.1
<i>A. oryzae</i> 3.042	36.57	11321	48.32	GCA_000269785.2
A. niger CBS 513.88	34.00	10503	50.35	GCF_000002855.3
Penicillium chrysogenum P2niaD18	32.52	11090	48.95	GCA_000710275.1
P. oxalicum HP7-1	30.65	9834	50.65	GCA_001723175.1
T. marneffei ATCC 18224	28.64	10002	46.67	GCF_000001985.1
T. stipitatus ATCC 10500	35.68	12442	46.07	GCF_000003125.1
T. cellulolyticus Y-94	36.40	10867	46.63	GCA_000829775.1
Trichoderma reesei QM6a	33.9	9120	52.82	GCF_000167675.1
Neurospora crassa OR74A	41.10	9,907	48.23	GCF_000182925.2

\*indicates that the number of genes is obtained through Perl script parsing GBK format.

	CEGMA				BUSCO						
Strain	Genome assembly		Gene Set		Gen	Genome Assembly		Gene Set			
	(248 genes)		(437 genes)		(1438 genes)		(1438 genes)				
Strain	Complete Partial Matching	Overlan	Doroont	Complete	Partial	Matching	Complete	Partial	Matching		
	matching	matching	percent	overlap		matching	matching	percent	matching	matching	percent
	genes	genes	(%)	genes	(%)	genes	genes	(%)	genes	genes	(%)
T. pinophilus 1-95	238	5	97.98	427	97.71	1419	15	98.67	1431	6	99.51
T. cellulolyticus Y-94	233	7	96.77	383	88.45	1411	18	98.12	1383	39	96.17

Supplementary Table S2. Assessment analysis of *T. pinophilus* 1-95 genome by software CEGMA and BUSCO.

Species	Aspergillus nidulans FGSC A4	<i>A. niger</i> CBS 513.88	A. oryzae 3.042	Neurospora crassa OR74a	Penicillium chrysogenum P2niaD18	P. oxalicum HP7-1	<i>T. marneffei</i> ATCC 18224	T. stipitatus ATCC 10500	T. cellulolyticus Y-94	Trichoderma reesei QM6a
Number of homologous proteins	3119	3824	3599	1407	3378	3520	8342	8098	10260	1626
Average identity (%)	72.7	73	72.7	71.33	72.66	72.43	88.23	84.29	98.28	70.922

**Supplementary Table S3.** Comparative analysis of homologous proteins among *T. pinophilus* 1-95 and another 10 filamentous fungi producing biomass degrading enzymes.

**Supplementary Table S4.** Paired-end (PE) libraries with average insert sizes 500 bp mapped the whole genome sequence of *Talaromyces cellulolyticus* Y-94.

94.2

99.71

90.62

DF933830

DF933831

DF933832

DF933833

DF933834

DF933835	1358518	1166859	85.89	83
DF933836	2056	0	0	0
DF933837	1279461	1122105	87.7	84
DF933838	2486755	2269975	91.28	86
DF933839	969434	819449	84.53	84
DF933840	3189086	2579602	80.89	81
DF933841	5308	0	0	0
DF933842	17731	0	0	0
DF933843	844840	668455	79.12	84
DF933844	9262	243	2.62	12
DF933845	7524	0	0	0
DF933846	577942	480141	83.08	81
DF933847	42575	0	0	0
DF933848	10449	0	0	0
DF933849	6082	0	0	0
DF933850	10834	0	0	0
DF933851	18859	0	0	0
DF933852	61417	161	0.26	4.7
DF933853	9072	0	0	0
DF933854	20907	274	1.31	190
DF933855	3406	0	0	0
DF933856	1913789	1661084	86.8	80
Total	36402801	32058530	88.07	83

**Supplementary Table S5.** Comparison of plant cell wall degrading enzymes (CWDEs) among *T. pinophilus* 1-95, *T. cellulolyticus* Y-94 and *Penicillium oxalicum* HP7-1.

Enzyme	EC Number	CAZy Family	T. pinophilus 1-95	T. cellulolyticus Y-94	P. oxalicum HP7-1
1. Cellulose-degrading enzymes					
			TP08389		POX04786
		GH6			
Cellobiohydrolase	3.2.1.91				
		CH7	TP09412	TCE0_004r00180	POX02490
		UII/			POX05587
			TP13457	TCE0_013f01292	POX01166
			TP04686	TCE0_017r03786	POX01896
	3.2.1.4	GH5	TP08784	TCE0_024f07747	POX02740
			TP07499	TCE0_033r09253	POX04137
					POX06147
β-1,4-endoglucanase					POX01206
		GH7	TP08514	TCE0_024r07350	POX05571
		GH12	TP10877	TCE0_042r14939	POX06983
					POX07535
		CI145	TP06957	TCE0_023r07056	POX05570
		01143	TP08136	TCE0_033r08342	
Xyloglucan-specific β-1,4-	3 2 1 151	GH12			
endoglucanase	5.2.1.151	01112	TP02720	TCE0_015f02847	POX00963
B-1 3(4)-endoglucanase	3216	GH16	TP07888	TCE0_033r08690	POX06959
	5.2.1.0	GIII0			POX00652
Lytia polyacobarida			TP03971	TCE0_044r17407	POX02308
monooxygenases	(LPMOs)	AA9			POX05968
monooxygenases					POX08897

					POX09256
			TP02423	TCE0_015r02411	POX01473
			TP04716	TCE0_017r03738	POX03062
		СШ1	TP05374	TCE0_018f05835	POX06079
		ОПІ	TP07549	TCE0_033f09178	POX08882
			TP11889	TCE0_060r18631	
			TP09603	TCE0_011f00448	POX00968
			TP02587	TCE0_015f02654	POX01646
			TP02719	TCE0_015r02845	POX03641
			TP06044	TCE0_018f04835	POX04358
			TP06188	TCE0_018r04615	POX06835
			TP05820	TCE0_018r05172	POX07963
ß glucosidasa	2 2 1 21		TP08794	TCE0_024f07765	POX08491
p-glucosidase	3.2.1.21		TP08074	TCE0_033f08430	POX08984
			TP08367	TCE0_033r08027	
		CH3	TP07983	TCE0_033r08553	
		GH5	TP07716	TCE0_033r08939	
			TP01149	TCE0_034f10525	
			TP00302	TCE0_034f11821	
			TP09042	TCE0_039f12945	
			TP12437	TCE0_041f14010	
			TP03558	TCE0_044f16774	
			TP13061	TCE0_050r18181	
			TP00267	TCE0_004r00287	
			TP03577	TCE0_038f12659	
			TP06192		

			TD07091		
			TP0/981		
			TP11082		
			TP12325		
			TP12326		
2. Hemicellulose-degrading enzy	mes				
			TP02736	TCE0_015f02871	POX04444
		CH27	TP02523	TCE0_015r02550	POX09085
		01127	TP02903	TCE0_015r03131	
a coloctoridado	2 2 1 22		TP05656	TCE0_018f05406	
a-galaciosidase	3.2.1.22	GH36	TP09250	TCE0_039f13249	
			TP06873	TCE0_047f17698	
			TP13280	TCE0_013f01052	
		GH10	TP13282		
			TP06900	TCE0_023r06982	POX05916
					POX08990
					POX00063
			TP01904	TCE0_015f01698	POX01417
			TP02919	TCE0_015r03151	POX02723
			TP09889	TCE0_022f06941	POX03430
β-1,4-endoxylanase	3.2.1.8	CII11	TP09024	TCE0_039r12923	POX06783
		GHII	TP12816	TCE0_041r13448	POX08484
			TP11225	TCE0_042f15439	
			TP09858		
			TP00436		
		CU20			POX04274
		GH30			POX06601
β-D-xylosidase	3.2.1.37	GH3	TP05314	TCE0_018r05917	POX00007

			TP03218	TCE0_044f16290	POX06571
			TP12892	TCE0_050r18425	
			TP07586		
			TP13153	TCE0_013r00881	POX04995
			TP13258	TCE0_013r01023	POX08661
		СЦ/2	TP02218	TCE0_015r02141	POX00006
		01143	TP02739	TCE0_015r02874	POX01744
			TP06762	TCE0_047f17875	
			TP09672		
			TP02524	TCE0_015r02553	POX05240
		GH43			POX06600
					POX07665
					POX07891
		GH51	TP13092	TCE0_013f00789	POX04752
			TP05618	TCE0_018r05464	POX06355
			TP10566	TCE0_042f14472	POX07441
α-L-arabinofuranosidase	3.2.1.55		TP09805	TCE0_011f00699	POX01914
			TP13333	TCE0_013f01127	
		GH54	TP02487	TCE0_015r02500	
			TP01576	TCE0_034f09917	
			TP04108	TCE0_044r17624	
			TP08906	TCE0_039f12752	POX05540
		GH62	TP01905	TCE0_015r01700	POX06599
			TP03986	TCE0_044f17429	
			TP02913		POX00923
α-1,5-L-endoarabinosidase	3.2.1.99	GH43	TP12748		POX01921
			TP13291	TCE0_013f01067	POX08140

					POX09154
			TP08661	TCE0_024f07565	POX00759
			TP01384	TCE0_034f10187	POX01961
0	2 2 1 25	CU2	TP10485	TCE0_042f14333	
p-mannosidase	3.2.1.25	GH2	TP03542	TCE0_044f16757	
			TP03563	TCE0_044r16782	
			TP06860	TCE0_047f17717	
		CHE	TP13093	TCE0_013f00791	POX01937
β-1,4-endomannanase	3.2.1.78	бнэ	TP12746		POX07285
		GH26			POX07536
	3.2.1	GH93	TP05094	TCE0_017r03210	POX00008
α-1,5-L-exoarabinanase			TP13027	TCE0_050f18245	POX01274
			TP12891	TCE0_050r18426	POX03420
			TP09758		
β-1,4-endogalactosidase	3.2.1.89	GH53	TP12454	TCE0_041r13985	POX06689
α-glucosiduronase	3.2.1.139	GH67			POX04275
			TP00303	TCE0_034f11818	
		CH30	TP00306	TCE0_034r11813	
		0129	TP11718	TCE0_060f18889	
			TP03070		
α-L-fucosidase	3.2.1.51		TP04936	TCE0_017r03422	POX03063
			TP07066	TCE0_023r07212	
		GH95	TP06680	TCE0_047r18022	
			TP07550		
			TP12218		
Polygalacturonasa	2 2 1 15	CH38	TP08437	TCE0_033f07936	POX03447
Polygalacturonase	3.2.1.15	0П20	TP03527	TCE0_044f16730	POX03730

			TP05070		POX04815
					POX07304
					POX04570
					POX06790
					POX08286
					POX05580
		GH28	TP08427	TCE0_033r07948	POX02024
Rhamnogalacturonan hydrolase	2 2 1 171		TP00731	TCE0_034f11163	POX04377
	5.2.1.1/1		TP09251	TCE0_039f13250	POX01225
			TP07985		
D-4,5-unsaturated β-glucuronyl hydrolase	3.2.1	GH88			POX06998
Unsaturated rhamnogalacturonyl	3.2.1.172	GH105	TP05718	TCE0_018r05318	POX08875
			TP08386	TCE0_033f08000	
hydrolase			TP00966	TCE0_034f10811	
			TP03027	TCE0_044r16075	
Rhamnogalacturonan	2 1 1 96	CE12	TP11874	TCE0_060f18653	POX08476
acetylesterase	5.1.1.80	CE12			POX07373
		CE1	TP02522	TCE0_015f02549	POX01218
		CEI	TP02906	TCE0_015r03136	
		CE3	TP06763	TCE0_047r17873	POX08861
		CE2	TP06741	TCE0_047r17910	
A patril willon astanaga	2 1 1 7 2		TP09379	TCE0_004r00149	
Acetyl xylan esterase	5.1.1.72		TP02707	TCE0_015r02825	
		CE2	TP00301	TCE0_034r11823	
		CES	TP03042	TCE0_043r15518	
		-	TP03145	TCE0_044r16257	
			TP03124		

			TP12334		
		CE5	TP08046	TCE0_033f08462	POX03711
			TP11188		
		CE1			POX00197
		CEI			POX09137
			TP13158	TCE0_013r00888	POX00727
			TP02893	TCE0_015r03115	POX01117
Feruloyl esterase	3.1.1.73		TP04811	TCE0_017r03600	POX00668
		Not assigned	TP10531	TCE0_042f14421	
			TP03050		
			TP03068		
			TP11085		
SUN family beta-glucosidase	3.2.1	GH132	TP08555	TCE0_024f07410	POX01158
					POX06380
3. Pectin-degrading enzymes					
		CE0	TP09742	TCE0_011f00620	POX07932
Destingstores	2 1 1 11		TP02150	TCE0_015f02043	POX08067
Peculiesterase	5.1.1.11	CEð	TP06538	TCE0_038r12590	POX00760
			TP01788	TCE0_015r01535	POX00729
Tannase	3.1.1.20	Not assigned	TP05371	TCE0_018r05839	POX08434
			TP08907	TCE0_039r12754	
Destin lyses	4 2 2 10	DI 1	TP12683	TCE0_041f13655	POX04920
recuitiyase	4.2.2.10	FL1	TP03086	TCE0_044r16137	POX09802
Poototo luoso	4222	DI 1	TP08387	TCE0_033f07999	POX07890
	4.2.2.2	LI PLI	TP13347	TCE0_013f01148	
Rhamnogalacturonan endolyase	4.2.2.23	PL4			POX05310

					POX06611
					POX08785
L-threo-3-deoxy-hexylosonate aldolase	4.1.2.54	Not assigned	TP01081	TCE0_034f10618	POX02774
			TP11196	TCE0_042r15394	POX07471
			TP13055	TCE0_050f18191	
		Not assigned	TP05823	TCE0_018f05169	POX04114
L coloctorate debudratace	4 2 1 146		TP07065	TCE0_023f07210	
L-galactonate denydratase	4.2.1.140		TP06402	TCE0_038r12406	
			TP06446	TCE0_038r12473	

**Supplementary Table S6.** Comparison of starch degrading enzymes among *T. pinophilus* 1-95 and another 10 filamentous fungi producing biomass degrading enzymes.

Protein Name	EC Number	T. pinophilus 1-95	Aspergillus nidulans FGSC A4	A. niger CBS 513.88	A. oryzae 3.042	Neurospora crassa OR74A	T. cellulolyticus Y-94	<i>T. marneffei</i> ATCC 18224	<i>T. stipitatus</i> ATCC 10500	Trichoderma reesei QM6a	Penicillium chrysogenum P2niaD18	P. oxalicum HP7-1
	3.2.1.1	TP03368	AN2018.2	ANI_1_1100184	Ao3042_01517	NCU05873	TCE0_033r09409	PMAA_058260	TSTA_013540	TRIREDRAFT_105956	EN45_014170	POX03230
		TP03580	AN3308.2	ANI_1_1284084	Ao3042_01518	NCU08131	TCE0_039f13307	PMAA_063250	TSTA_044280		EN45_045100	POX03231
		TP04014	AN3309.2	ANI_1_1524104	Ao3042_03622	NCU09486	TCE0_044f16809	PMAA_064990	TSTA_082630		EN45_073300	POX09352
		TP07411	AN3388.2	ANI_1_1810134	Ao3042_05212	NCU09805	TCE0_044f17478	PMAA_096850	TSTA_084740		EN45_090540	
α-amylase		TP09288	AN3402.2	ANI_1_260044			TCE0_044r16495		TSTA_089790		EN45_090550	
			AN4507.2	ANI_1_3270014							EN45_105210	
			AN6324.2	ANI_1_362084							EN45_107900	
				ANI_1_460094							EN45_107910	
				ANI_1_820104								
	3.2.1.3	TP04225	AN7402.2	ANI_1_820034	Ao3042_08133	NCU01517	TCE0_039r13271	PMAA_096710	TSTA_044290	TRIREDRAFT_1885	EN45_015780	POX01356
		TP07482	AN8904.2		Ao3042_10648		TCE0_039r13304	PMAA_096840	TSTA_044610		EN45_097990	POX02412
Glucoamylase		TP09267					TCE0_043f15539		TSTA_106690		EN45_105200	POX04642
		TP09287										
		TP12319										
1,4-α-glucan branching enzyme	24118	TP03995	AN2314.2	ANI 1 602124	Ao3042 08360	NCU05429	TCF0 044f17444	PMAA 068370	TSTA 088790	TRIREDRAFT 123368	EN45 072830	POX04938
α-glucosidase	3.2.1.20	TP09781	AN0941 2	ANI 1 1098184	A03042_05838	NCU02583	TCE0_011r00674	PMAA 002150	TSTA 023890	TRIREDRAFT 59578	EN45_014160	POX03452
		TP11464	AN2017.2	ANI 1 1476014	Ao3042_06557	NCU04674	TCE0_018f06215	PMAA 007280	TSTA 041120	TRIREDRAFT 82235	EN45_050490	POX03741
		TP12265	AN3516.2	ANI 1 1804024	Ao3042_07614	NCU06523	TCE0_010100215	PMAA 015950	TSTA 041820		EN45_053200	POX03889
		TP03913	AN4843 2	ANI 1 850114	Ao3042_07671	NCU07860	TCE0_034r10854	PMAA 099560	TSTA 101210		EN45_075930	POX06242
		TP00071	AN7345 2	1_000114	Ao3042_10501	NCU09281	TCE0_043r15609	1	TSTA 101220		EN45_084270	POX06751
		TP05786	AN8953.2		Ao3042_10503		TCE0_044r16449		TSTA_106040		EN45_084670	

		TP00293		Ao3042_11801	TCE0_044r17312	TSTA_106170	EN45_086140	
		TP00938			TCE0_060r19238	TSTA_112550	EN45_091760	
		TP01354					EN45_097770	
		TP03337						
		TP04013						
		TP04937						
		TP05120						

Cluster	T. pinophilus 1-95	T. cellulolyticus Y-94	Penicillium oxalicum HP7-1	Neurospora crassa OR74A	T. marneffei ATCC 18224	<i>T. stipitatus</i> ATCC 10500	Trichoderma reesei QM6a	Aspergillus nidulans FGSC A4	A. niger CBS 513.88	A. oryzae 3.042	P. chrysogenum P2niaD18
Indole	0	2	2	1	0	0	0	4	2	5	0
Indole-Nrps	0	0	0	0	1	0	0	0	0	0	1
Lantipeptide	0	0	0	1	0	0	0	0	0	0	0
Nrps	9	9	12	3	7	9	6	10	15	12	8
Nrps-Indole	0	0	1	0	0	0	0	1	0	1	0
Nrps-T1pks	2	2	3	0	3	3	4	2	8	2	3
Nrps-Terpene	0	0	0	0	0	1	0	0	0	0	0
Other	16	15	11	3	12	9	5	9	13	13	15
Phosphonate	1	1	0	0	0	0	0	0	0	0	0
Siderophore	0	0	1	0	0	0	0	0	2	1	1
T1pks	28	28	5	7	16	30	9	22	27	22	18
T1pks-Indole	1	1	0	0	1	0	0	0	0	1	0
T1pks-Nrps	1	0	1	0	1	3	0	0	2	0	0
T3pks	0	0	1	1	0	0	0	0	1	3	1
Terpene	9	7	7	3	7	7	6	7	14	11	7
Terpene-Nrps	0	0	0	0	0	0	0	0	0	1	1
Terpene-Nrps- Indole	0	0	1	0	0	0	0	0	0	0	0
Terpene-T1pks	1	1	0	0	0	0	0	0	0	0	0
Total	68	66	45	19	48	62	30	55	84	72	55

**Supplementary Table S7.** The predicted secondary metabolism gene clusters in *T. pinophilus* 1-95 and the comparison analysis among with another 10 filamentous fungi producing biomass degrading enzymes.