Jurak, E.; Suzuki, Hidemasa; van Erven, Gijs; Gandier, J. A.; Wong, P.; Chan, Kin; Ho, C. Y.; Gong, Y.; Tillier, E.; Rosso, M. N.; Kabel, M. A.; Miyauchi, S.; Master, E. R.

Dynamics of the Phanerochaete carnosa transcriptome during growth on aspen and spruce

BMC Genomics

Open Access

CrossMark

Dynamics of the *Phanerochaete carnosa* transcriptome during growth on aspen and spruce

E. Jurak[1,2†], H. Suzuki[3†], G. van Erven[4], J. A. Gandier[3], P. Wong[5], K. Chan[6], C. Y. Ho[6], Y. Gong[7], E. Tillier[5], M.-N. Rosso[9], M. A. Kabel[4], S. Miyauchi[8,9] and E. R. Master[1,3*]

Abstract

Background: The basidiomycete *Phanerochaete carnosa* is a white-rot species that has been mainly isolated from coniferous softwood. Given the particular recalcitrance of softwoods to bioconversion, we conducted a comparative transcriptomic analysis of *P. carnosa* following growth on wood powder from one softwood (spruce; *Picea glauca)* and one hardwood (aspen; *Populus tremuloides*). *P. carnosa* was grown on each substrate for over one month, and mycelia were harvested at five time points for total RNA sequencing. Residual wood powder was also analyzed for total sugar and lignin composition.

Results: Following a slightly longer lag phase of growth on spruce, radial expansion of the *P. carnosa* colony was similar on spruce and aspen. Consistent with this observation, the pattern of gene expression by *P. carnosa* on each substrate converged following the initial adaptation. On both substrates, highest transcript abundances were attributed to genes predicted to encode manganese peroxidases (MnP), along with auxiliary activities from carbohydrate-active enzyme (CAZy) families AA3 and AA5. In addition, a lytic polysaccharide monooxygenase from family AA9 was steadily expressed throughout growth on both substrates. P450 sequences from clans CPY52 and CYP64 accounted for 50% or more of the most highly expressed P450s, which were also the P450 clans that were expanded in the *P. carnosa* genome relative to other white-rot fungi.

Conclusions: The inclusion of five growth points and two wood substrates was important to revealing differences in the expression profiles of specific sequences within large glycoside hydrolase families (e.g., GH5 and GH16), and permitted co-expression analyses that identified new targets for study, including non-catalytic proteins and proteins with unknown function.

Keywords: *Phanerochaete carnosa*, Transcriptomics, Carbohydrate active enzymes, Lignocellulose conversions, Loosenins, Hydrophobins

Background

Fungi from the phylum Basidiomycota, class Agaricomycetes, include ectomycorrhizal fungi, saprotrophs, as well as efficient wood (lignocellulose) degraders. White-rot fungi of the orders Agaricales and Polyporales are especially adept wood-degraders. Accordingly, these fungi have been the focus of studies aimed at the bioconversion of major lignocellulose components, including strategies to hydrolyze cellulose and hemicelluloses to monosaccharides for fermentation to fuels and chemicals. Since the first publication of the *Phanerochaete chrysosporium* genome in 2004 [1], the number of Basidiomycota genome sequences has increased to several hundred (https://jgi.doe.gov/) [2]. Among these, *Phanerochaete carnosa* represents a white-rot that grows on both deciduous (hardwood) and coniferous (softwood) fibre, but has been almost exclusively isolated from softwoods [3]. Its genome was sequenced in 2012 [4], confirming *P. carnosa* encodes a full complement

* Correspondence: emma.master@utoronto.ca
†E. Jurak and H. Suzuki contributed equally to this work.
[1]Department of Bioproducts and Biosystems, Aalto University, Espoo, Finland
[3]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Canada
Full list of author information is available at the end of the article

of carbohydrate-active enzymes (CAZymes) for lignocellulose conversion and revealing a large contingent of predicted cytochrome P450 monooxygenases.

Coniferous trees are the predominant form of renewable biomass in the northern hemisphere; however, it is especially recalcitrant to bioprocess technologies. The recalcitrance of coniferous wood has been attributed to the higher lignin content, smaller pore size, and fewer hemicellulose-derived acetyl groups in comparison to deciduous woods [5, 6]. Challenges linked to softwood bioconversion have motivated studies that investigate gene and protein expression by white-rot fungi that grow on coniferous wood [7–15]. In general, corresponding studies show particularly high expression of lignolytic enzymes (e.g., lignin peroxidases (LiPs) and manganese peroxidases (MnPs)) and lytic polysaccharide monooxygenases (LPMOs); comparatively high expression of glycoside hydrolases (GHs) from families GH5, GH6, GH7, GH10, GH12, GH28, GH43, and GH131 have also been repeatedly reported. So far, such comparative analyses mainly consider either multiple substrates or multiple time points on a single wood species. Accordingly, time and substrate dependent influences on the expression of lignocellulose degrading activities remain unclear.

Herein, we apply a transcriptomic approach to track gene expression by *P. carnosa* over five growth points on heartwood of white spruce (*Picea glauca*) and trembling aspen (*Populus tremuloides*). Earlier transcriptomic analyses of *P. carnosa* grown on fir, pine, spruce and maple wood preparations show high transcript abundances corresponding to specific MnPs and LPMOs [15]; however, impacts of biomass conversion on resulting gene expression profiles could not be gleaned from the single time point included in that study. By evaluating the impact of both wood substrate and time on the *P. carnosa* transcriptome, we can identify specific enzymes, enzyme sub-families, and novel activities best correlated to plant biomass degradation and most critical to early versus late stages of wood decay.

Results

Growth on wood substrates

Mycelia samples were harvested at five equivalent radial distances (between 2 and 9 cm) from the center of solid-state cultivations on aspen and spruce. In this way, we could evaluate changes in the gene expression profiles of *P. carnosa* over a comparable extent of radial growth on the two wood substrates, and ensure in both cases that sufficient quantities of mycelia would be collected for RNA extraction. The resulting growth points (GP) 1–5 corresponded to 7 to 23 days of cultivation on aspen, and 13 to 30 days of cultivation on spruce. While growth was initially slower on spruce, the radial growth rate of *P. carnosa* was independent of substrate following GP1
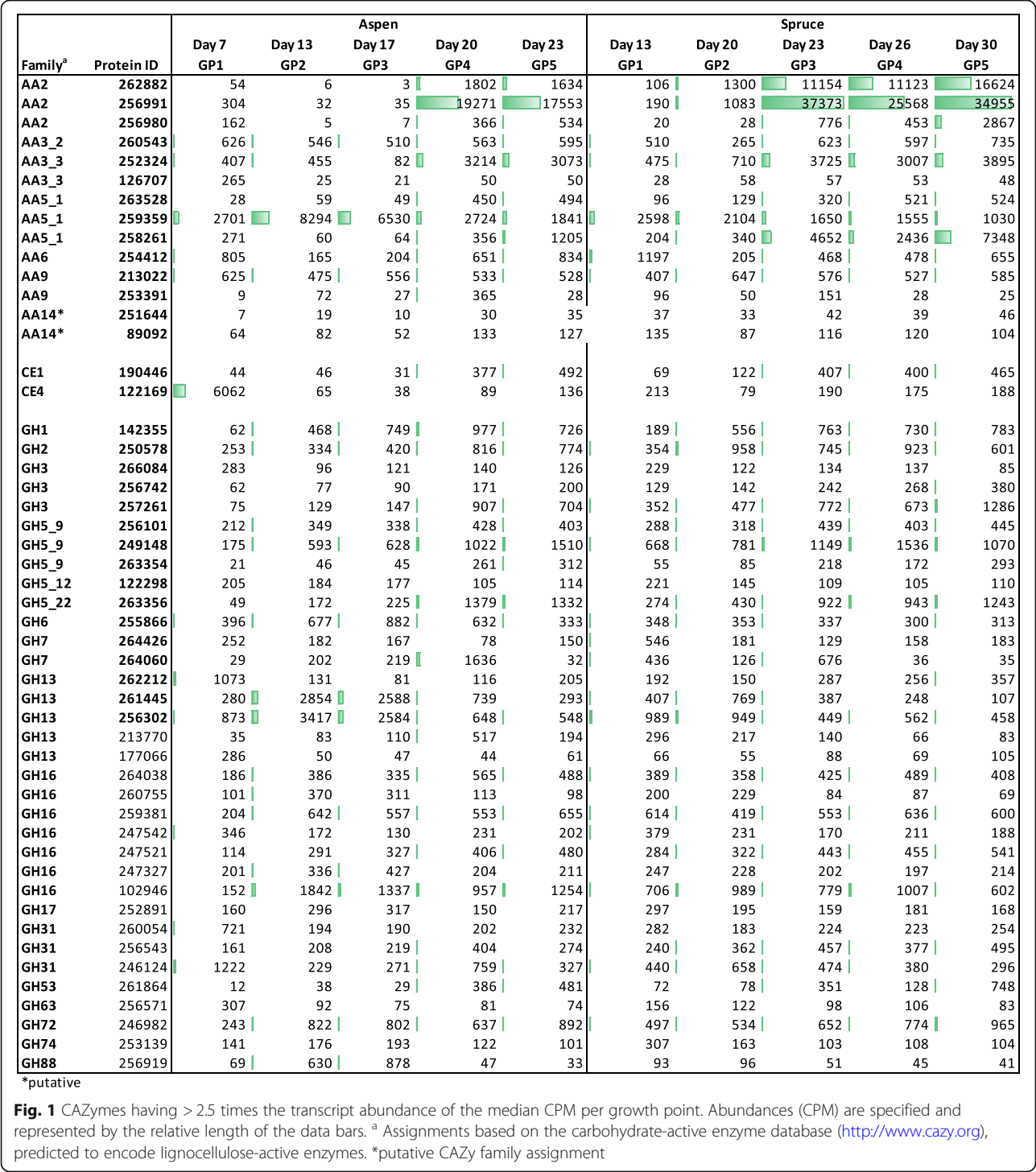
(Additional file 1). This suggests that a longer adaptation period was required to establish growth on spruce; however, following the adaptation period, *P. carnosa* grew similarly on both spruce and aspen.

Hierarchical clustering of transcriptome profiles were consistent with the growth patterns, where following the initial lag phase on spruce, similar transcriptome patterns were obtained from cultivations on spruce and aspen (Additional file 2). Notably, the relative carbohydrate composition was similar throughout growth of *P. carnosa* on both wood substrates, consistent with non-selective consumption of corresponding monosaccharides (Additional file 3). On the other hand, slight but significant loss of lignin was measured only from aspen (Additional files 4 and 5). Herein, wood samples were ball milled prior to fungal cultivation, which was expected to increase the accessibility of the wood substrates and permit comparative transcriptome analyses that reveal fungal responses to differences in wood fibre composition uncoupled from differences in wood fibre structure.

Transcriptome profiles of sequences predicted to encode lignocellulose-active CAZymes

Considering all 13,937 genes encoded by the *P. carnosa* genome [4], sequences having highest transcript abundance on both wood substrates were mainly household metabolism regulating genes, transporters, MnPs (Phaca262882, Phaca256991) and uncharacterized sequences (Additional file 6). The 246 sequences encoding carbohydrate active enzymes (http://www.cazy.org; CAZymes; Additional file 7) were considered in more detail, given they encode proteins predicted to contribute to lignocellulose conversion. This analysis uncovered a core set of CAZyme sequences present at high transcript abundance for both cultivation conditions (Fig. 1), consistent with similar extents of growth observed on both wood substrates following the initial lag phase on spruce.

Of the seven MnPs encoded by *P. carnosa*, transcript sequences corresponding to two MnPs (Phaca262882, Phaca256991) were 5 to 10 times more abundant than any other predicted CAZyme (Fig. 1). These same sequences were among the 30 most abundant transcripts expressed by *P. carnosa* during growth on maple, fir, pine, and spruce [15], confirming the biological relevance of these particular MnPs for conversion of lignin present in both deciduous and coniferous wood. In addition to MnPs, transcripts predicted to encode enzymes that provide H_2O_2 (required for MnP activity), including glyoxal oxidases (GLOX), GLOX/related copper radical oxidases (CRO) and alcohol (AOX) oxidases [16, 17] were also among the top 25 highly expressed sequences. Of these, transcript sequences encoding two AA3 alcohol oxidases (Phaca260543 and Phaca252324) and one AA5_1 oxidase (GLOX; Phaca259359) followed the MnP's in terms of

Family[a]	Protein ID	Aspen					Spruce				
		Day 7	Day 13	Day 17	Day 20	Day 23	Day 13	Day 20	Day 23	Day 26	Day 30
		GP1	GP2	GP3	GP4	GP5	GP1	GP2	GP3	GP4	GP5
AA2	262882	54	6	3	1802	1634	106	1300	11154	11123	16624
AA2	256991	304	32	35	19271	17553	190	1083	37373	25568	34955
AA2	256980	162	5	7	366	534	20	28	776	453	2867
AA3_2	260543	626	546	510	563	595	510	265	623	597	735
AA3_3	252324	407	455	82	3214	3073	475	710	3725	3007	3895
AA3_3	126707	265	25	21	50	50	28	58	57	53	48
AA5_1	263528	28	59	49	450	494	96	129	320	521	524
AA5_1	259359	2701	8294	6530	2724	1841	2598	2104	1650	1555	1030
AA5_1	258261	271	60	64	356	1205	204	340	4652	2436	7348
AA6	254412	805	165	204	651	834	1197	205	468	478	655
AA9	213022	625	475	556	533	528	407	647	576	527	585
AA9	253391	9	72	27	365	28	96	50	151	28	25
AA14*	251644	7	19	10	30	35	37	33	42	39	46
AA14*	89092	64	82	52	133	127	135	87	116	120	104
CE1	190446	44	46	31	377	492	69	122	407	400	465
CE4	122169	6062	65	38	89	136	213	79	190	175	188
GH1	142355	62	468	749	977	726	189	556	763	730	783
GH2	250578	253	334	420	816	774	354	958	745	923	601
GH3	266084	283	96	121	140	126	229	122	134	137	85
GH3	256742	62	77	90	171	200	129	142	242	268	380
GH3	257261	75	129	147	907	704	352	477	772	673	1286
GH5_9	256101	212	349	338	428	403	288	318	439	403	445
GH5_9	249148	175	593	628	1022	1510	668	781	1149	1536	1070
GH5_9	263354	21	46	45	261	312	55	85	218	172	293
GH5_12	122298	205	184	177	105	114	221	145	109	105	110
GH5_22	263356	49	172	225	1379	1332	274	430	922	943	1243
GH6	255866	396	677	882	632	333	348	353	337	300	313
GH7	264426	252	182	167	78	150	546	181	129	158	183
GH7	264060	29	202	219	1636	32	436	126	676	36	35
GH13	262212	1073	131	81	116	205	192	150	287	256	357
GH13	261445	280	2854	2588	739	293	407	769	387	248	107
GH13	256302	873	3417	2584	648	548	989	949	449	562	458
GH13	213770	35	83	110	517	194	296	217	140	66	83
GH13	177066	286	50	47	44	61	66	55	88	69	105
GH16	264038	186	386	335	565	488	389	358	425	489	408
GH16	260755	101	370	311	113	98	200	229	84	87	69
GH16	259381	204	642	557	553	655	614	419	553	636	600
GH16	247542	346	172	130	231	202	379	231	170	211	188
GH16	247521	114	291	327	406	480	284	322	443	455	541
GH16	247327	201	336	427	204	211	247	228	202	197	214
GH16	102946	152	1842	1337	957	1254	706	989	779	1007	602
GH17	252891	160	296	317	150	217	297	195	159	181	168
GH31	260054	721	194	190	202	232	282	183	224	223	254
GH31	256543	161	208	219	404	274	240	362	457	377	495
GH31	246124	1222	229	271	759	327	440	658	474	380	296
GH53	261864	12	38	29	386	481	72	78	351	128	748
GH63	256571	307	92	75	81	74	156	122	98	106	83
GH72	246982	243	822	802	637	892	497	534	652	774	965
GH74	253139	141	176	193	122	101	307	163	103	108	104
GH88	256919	69	630	878	47	33	93	96	51	45	41

*putative

Fig. 1 CAZymes having > 2.5 times the transcript abundance of the median CPM per growth point. Abundances (CPM) are specified and represented by the relative length of the data bars. [a] Assignments based on the carbohydrate-active enzyme database (http://www.cazy.org), predicted to encode lignocellulose-active enzymes. *putative CAZy family assignment

relative abundance of CAZyme sequences. Also of note, transcripts encoding two distantly related AA5_1 oxidases displayed divergent substrate-dependent expression patterns. Specifically, Phaca259359 transcript abundance was higher on aspen than spruce, whereas the reverse pattern was observed for Phaca258261. Phaca258261 is

phylogenetically related to glyoxal oxidases implicated in H_2O_2 production [16]. By contrast, Phaca259359 shares 84% sequence identity to CRO2 encoded by *P. chrysoporium*, which displays a distinct substrate preference relative to glyoxal oxidases [17], and whose biological function remains unclear.

Of the 11 family AA9 LPMOs encoded by *P. carnosa*, transcript levels corresponding to Phaca213022 were 5 to 10 times higher than the second most highly expressed AA9 sequence (Phaca253391) (Fig. 1). Moreover, Phaca213022 transcript abundance was comparatively steady over time in both aspen and spruce cultivations. The discovery of the family AA14 LPMO from the basidiomycete *Trametes coccinea* (i.e., PcAA14A) [18], prompted us to search for possible AA14 members in the *P. carnosa* genome. PcAA14A catalyzes the oxidative cleavage of xylan-coated cellulose; two potential AA14 members were identified herein, namely Phaca251644 (70.7% identity to PcAA14A) and Phaca89092 (56.8% identity to PcAA14A). Although levels were low, in both cases transcript abundances increased between the first and last growth point on aspen; by contrast, transcript abundances were steady on spruce (Fig. 1).

Abundances of transcripts predicted to encode glycoside hydrolases, carbohydrate esterases, and polysaccharide lyases were generally lower than those predicted to encode auxiliary activities. Of the 24 family GH5 sequences and 24 GH16 sequences encoded by *P. carnosa*, transcript abundances for 5 GH5s and 7 GH16s were at least 2.5 times above the median CPM value for at least one growth point (Fig. 1). Among the GH5s, three belonged to subfamily GH5_9 and one belonged to subfamily GH5_22, which are predicted to act on fungal and plant polysaccharides, respectively [19]. The transcript abundance of four GH16 sequences also increased over time, particularly during growth of *P. carnosa* on aspen (Phaca264038 Phaca259381, Phaca247521, Phaca102946). However, functional prediction for GH16 members remains complicated by the diverse activities and biological roles attributed to this CAZy family [20].

Levels of transcripts encoding the five predicted GH10 xylanases and three GH12 endoglucanases encoded by *P. carnosa*, as well as polysaccharide lyases and GH28 enzymes contributing to pectin degradation, were comparatively low and steady on both aspen and spruce (Additional file 8). By contrast, transcript abundances increased over the cultivation for sequences in families CE1, GH2, and GH3, which are known to include enzymes that target plant cell wall carbohydrates (Fig. 1; Additional file 7). Increase in transcript abundance was not observed, however, for the sole predicted GH6 cellobiohydrolase and the two most highly expressed GH7 cellobiohydrolases encoded by *P. carnosa*. Instead, corresponding transcript abundances were dependent on both time and substrate (Fig. 1).

Transcriptome profiles of sequences predicted to encode P450 monooxygenases

Cytochrome P450 monooxygenases have been implicated in the degradation of small lignin fragments and other aromatic compounds, and could thus facilitate fungal growth on wood by detoxifying lignin degradation products as well as aromatic extractives [21]. The *P. carnosa* genome comprises 266 genes predicted to encode cytochrome P450 monooxygenases, nearly twice the number encoded by *P. chrysosporium* [4].

Patterns of P450 transcript abundance were generally similar during growth of *P. carnosa* on the two wood substrates, where 50% or more of the most highly expressed P450 mainly grouped in clans CYP52 and CYP64 (Fig. 2). P450s belonging to clan CYP64 were also highly expressed in *P. coccineus* following cultivation on pine and aspen [9]. Of note, clans CYP52 and CYP64 accounted for most of the P450 sequence expansion in *P. carnosa* compared to *P. chrysosporium*. Transcript abundances were highest, however, for two sequences corresponding to clan CYP547 (Phaca260638 and Phaca259665).

Co-expression analyses

The consistency of the transcriptomic data permitted the construction of transcriptomic models using the SHIN +GO pipeline (Additional files 2, 9, 10) [22]. Resulting self-organizing maps (SOM) group genes with similar transcriptional patterns and form nodes arranged as so-called Tatami maps (Additional files 11 and 12; node number and composition listed in Additional file 8). Within a Tatami map, nodes in close proximity contain genes with relatively similar transcriptional patterns (Fig. 3).

The co-expression analyses identified groups of gene products that may operate together. For example, MnPs require a source of H_2O_2, which can be generated by family AA3 and AA5 oxidases. Clustering of the MnP (Phaca256991) and a AA3_3 oxidase (Phaca252324) within node 49, and the neighbouring positions of nodes 145, 169 and 170 (Fig. 3) that comprise the most highly expressed MnP (Phaca262882; node 145), along with a family AA3_3 oxidase (Phaca121157; node 170), and a family AA5_1 oxidase (Phaca263528; node 169), predict these specific auxiliary enzymes may act in concert to transform lignin.

Considering the profile of transcripts encoding P450 monooxygenases, co-expression analysis underscored the transition over time from sequences belonging to many P450 clans to sequences predominately from clan CYP52 and clan CYP64, which are also expanded in the *P. carnosa* genome (nodes 145, 193, and 457, Fig. 3). Moreover, nearly half of nodes including a P450 sequence also included predicted glutathione-S-transferases, which are also believed to play a role in the detoxification of compounds released during fungal growth on lignocellulosic materials [23, 24] (Additional file 8).
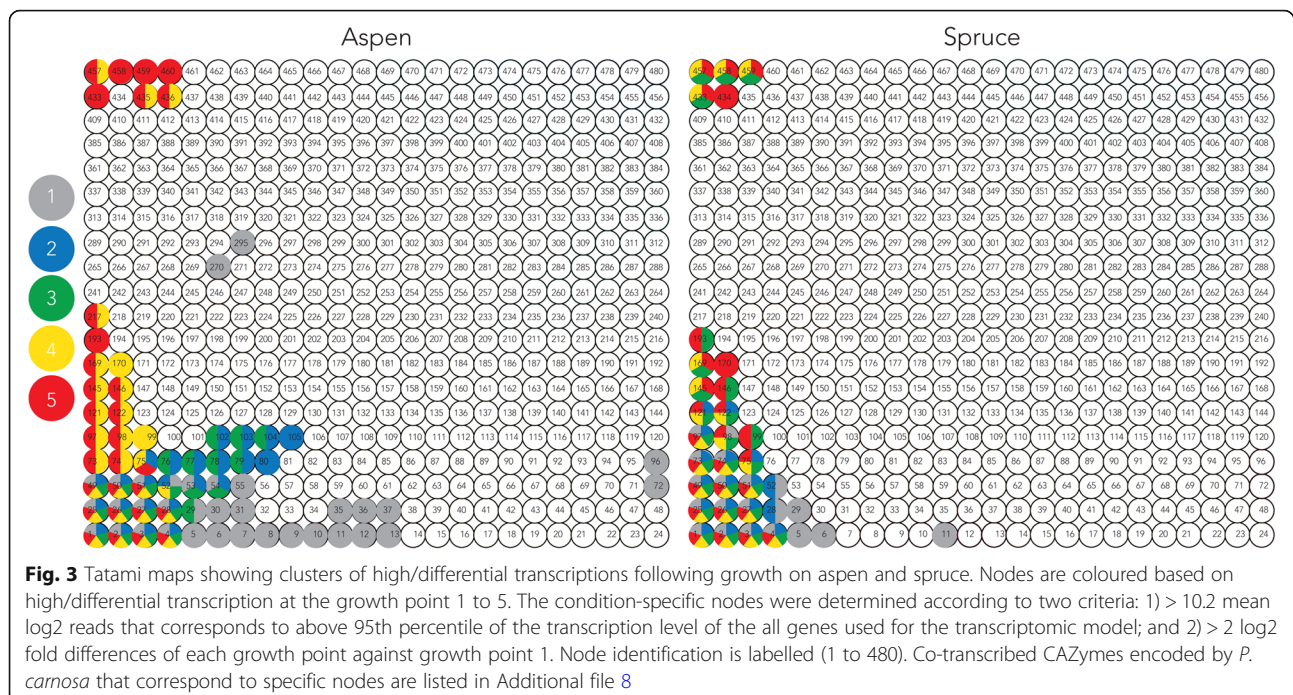
Co-expression analyses was also used herein to identify non-catalytic proteins, namely loosenins and hydrophobins, that co-express with known CAZymes and may

Jurak et al. BMC Genomics (2018) 19:815

Page 5 of 13

Family	Subfamily	Clan	Protein ID	Aspen					Spruce				
				Day 7 GP1	Day 13 GP2	Day 17 GP3	Day 20 GP4	Day 23 GP5	Day 13 GP1	Day 20 GP2	Day 23 GP3	Day 26 GP4	Day 30 GP5
CYP512	E	CYP503	246674	242	84	113	180	221	127	174	185	214	208
CYP512	G	CYP503	262885	128	157	138	265	418	69	132	367	347	634
CYP512	B	CYP503	248689	120	119	64	99	122	130	131	129	136	128
CYP512	H	CYP503	260824	98	256	274	129	93	250	359	72	53	24
CYP512	C	CYP503	248575	15	8	18	7	9	133	237	20	17	10
CYP505	D	CYP505	213993	58	74	50	55	33	814	116	44	34	32
CYP51	F	CYP51	214862	120	52	33	57	57	205	54	63	84	63
CYP63	A	CYP52	83524	1449	297	271	405	303	1356	347	281	255	291
CYP5150	A	CYP52	249950	209	187	92	120	131	2516	530	206	252	221
CYP63	A	CYP52	249548	171	350	197	591	179	540	870	363	317	213
CYP63	A	CYP52	249546	171	288	38	52	96	282	162	85	62	95
CYP5151	A	CYP52	265519	101	416	508	299	309	55	201	156	185	146
CYP63	A	CYP52	249541	92	749	332	794	107	122	713	384	209	67
CYP63	C	CYP52	263957	91	117	110	101	104	90	145	92	99	79
CYP5139	A	CYP52	250554	56	410	392	131	112	163	366	151	161	91
CYP63	B	CYP52	263397	18	325	194	175	36	107	275	152	126	42
CYP5139	A	CYP52	250529	68	255	179	269	184	149	525	249	267	169
CYP5150	B	CYP52	205063	6	186	159	10	12	43	87	25	19	11
CYP5150	B	CYP52	253633	12	67	56	222	185	59	222	237	222	252
CYP5150	A	CYP52	214286	29	27	26	158	644	57	48	190	205	318
CYP5150	A	CYP52	265429	2	7	5	148	156	6	23	123	100	199
CYP5150	B	CYP52	112917	53	106	69	119	81	164	269	101	99	63
CYP53	C	CYP53	183109	880	437	381	181	256	1494	321	256	235	340
CYP53	C	CYP53	212559	144	59	56	102	244	1748	151	261	239	398
CYP5142	B	CYP53	145790	41	118	136	33	33	95	107	32	35	24
CYP53	C	CYP53	101826	10	46	7	10	12	208	16	14	15	14
CYP53	C	CYP53	256510	19	66	28	18	20	176	38	24	24	24
CYP5141	A	CYP534	263271	40	211	238	74	93	124	108	114	118	107
CYP5137	A	CYP547	259665	592	202	208	1050	1601	1733	593	1864	1629	3029
CYP5136	NS	CYP547	262778	396	179	198	828	715	590	735	892	1054	453
CYP5136	A	CYP547	262258	32	126	120	60	19	67	60	70	44	16
CYP5137	A	CYP547	124964	21	93	120	24	34	28	49	32	30	42
CYP5136	A	CYP547	260638	28	44	51	826	945	140	279	1125	897	1090
CYP5136	A	CYP547	262256	18	59	47	218	303	1024	528	1061	1145	448
CYP61	A	CYP61	246788	181	267	261	183	200	409	149	172	223	196
CYP5148	B	CYP64	208274	7116	28	22	16	15	27	15	15	15	18
CYP5148	B	CYP64	203097	2338	58	64	26	17	31	47	28	28	30
CYP5144	C	CYP64	249435	1026	409	391	389	299	259	303	307	336	255
CYP5144	C	CYP64	252313	581	3	4	4	5	3	4	4	4	4
CYP5148	A	CYP64	254221	321	323	255	272	623	308	250	496	660	556
CYP5148	B	CYP64	198521	279	11	12	2	3	4	5	4	4	3
CYP5158	A	CYP64	212817	159	102	143	198	231	135	202	258	274	227
CYP5146	A	CYP64	256631	122	118	61	137	65	212	182	103	93	51
CYP5144	C	CYP64	256701	111	156	175	199	236	187	191	226	240	269
CYP5148	B	CYP64	265393	104	22	30	4	4	18	20	3	3	4
CYP5146	C	CYP64	258552	95	136	167	70	81	128	133	75	76	59
CYP5144	C	CYP64	256800	76	634	28	15	14	118	194	16	13	12
CYP5145	A	CYP64	205692	62	290	155	168	20	176	313	30	34	16
CYP502	B	CYP64	163592	72	179	197	381	394	151	227	293	336	391
CYP5037	B	CYP64	253611	20	172	106	28	23	25	64	23	21	20
CYP5144	A	CYP64	209389	13	154	11	12	14	40	34	21	19	16
CYP5144	A	CYP64	256793	24	71	120	198	220	76	238	331	295	306
CYP5145	A	CYP64	133539	18	71	114	118	149	67	149	160	176	152
CYP5152	A	CYP64	256887	55	73	88	427	1035	104	197	970	1057	1928
CYP5144	E	CYP64	265238	24	42	71	204	218	26	101	219	234	240
CYP5144	C	CYP64	256765	13	30	29	196	306	24	71	230	236	413
CYP5144	A	CYP64	209385	46	26	76	192	648	105	138	408	401	361
CYP5144	C	CYP64	252949	61	17	31	144	181	51	53	122	143	82
CYP5152	A	CYP64	265459	7	10	9	68	171	16	30	201	185	422
CYP5152	A	CYP64	265475	13	20	14	117	150	31	63	293	225	459
CYP5144	H	CYP64	214508	69	102	37	31	14	170	118	21	29	9
CYP5035	A	CYP67	212598	241	285	58	106	40	1037	137	132	90	63
CYP5035	A	CYP67	262568	196	121	114	70	67	808	313	76	80	70
CYP5036	A	CYP67	257741	10	18	16	62	78	26	33	134	107	183
CYP6001	NS	UA	112761	1099	713	431	278	216	693	588	180	470	252
CYP6001	NS	UA	264492	71	189	240	256	338	287	643	322	590	429

Fig. 2 P450s having > 2.5 times the transcript abundance of the median CPM per growth point. Abundances (CPM) are specified and represented by the relative length of the data bars

influence fungal growth on lignocellulosic substrates. Briefly, loosenins are single domain proteins that adopt a DPBB fold homologous to domain 1 of expansins [25, 26]. On the other hand, hydrophobins are surface active proteins secreted by filamentous fungi, which are subdivided into two classes, I and II [27, 28]. Whereas some loosenins show cellulose disruption activity [25], hydrophobin films can reverse the wettability of solid surfaces; it has also been suggested that such films could play roles in recruiting enzymes to substrates [29]. The *P. carnosa* genome is predicted to encode for twelve loosenin-like proteins (LOOL), along with one sequence that is distantly related to plant expansins (DREX) [30]. Transcripts of all thirteen of these genes were detected. Of these, transcripts encoding LOOL2 (Phaca255931) were most abundant; increasing to levels comparable to AA3 oxidases and
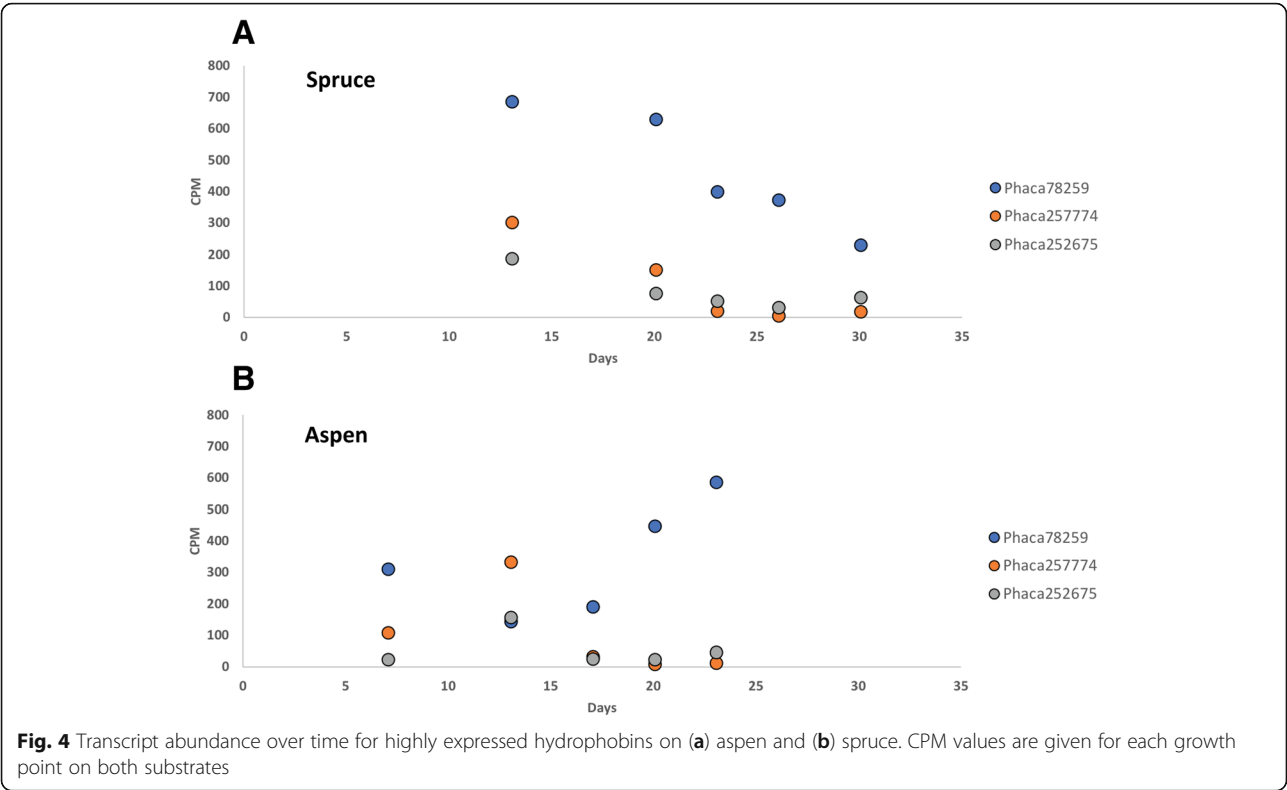
Fig. 3 Tatami maps showing clusters of high/differential transcriptions following growth on aspen and spruce. Nodes are coloured based on high/differential transcription at the growth point 1 to 5. The condition-specific nodes were determined according to two criteria: 1) > 10.2 mean log2 reads that corresponds to above 95th percentile of the transcription level of the all genes used for the transcriptomic model; and 2) > 2 log2 fold differences of each growth point against growth point 1. Node identification is labelled (1 to 480). Co-transcribed CAZymes encoded by *P. carnosa* that correspond to specific nodes are listed in Additional file 8

various glycoside hydrolases at day 13 on both spruce and aspen (CPM values of 236 and 469, respectively; Additional file 6). Co-expression analyses clustered LOOL2 with a predicted family CE9 N-acetyl-glucosamine 6-phosphate deacetylase, suggesting a role in fungal cell wall morphogenesis (Additional file 8). All 13 hydrophobin sequences predicted from the *P. carnosa* genome encode Class I proteins and were detected at the transcript level. Of these, transcript abundances for three sequences were at least 100 CPM for one or more growth points (Fig. 4; Additional file 6). In particular, the transcript abundance of Phaca78259 was up to 12 and 38 times higher than Phaca25774 and Phaca252675, reaching 590 CPM on aspen and 690 CPM on spruce. The transcript profile of Phaca78259 was also reversed on aspen versus spruce, where abundances generally increased and decreased over time, respectively (Fig. 4). Notably, the Phaca78259 transcript profile clustered into node 26 (Fig. 3), which also includes the most highly expressed LPMO (Phaca213022) along with two GH families that likely contribute to fungal cell wall modification, namely a putative β-1,3-glucanosyltransglycosylase from family GH72 and β-1,3-glucanase from family GH128 [31] (Additional file 8).

Lastly, the co-expression analyses performed herein were used to identify sequences with unknown function that co-expressed with differentially and highly expressed CAZymes. Eleven highly and differentially expressed sequences with unknown function that co-expressed with annotated CAZyme sequences were identified (Fig. 5). Of

these, three were predicted to encode a signal for secretion; moreover, Phaca259771 is predicted to encode a cupredoxin domain with the ability to bind copper. Transcript abundances for both Phaca259771 and Phaca256483 increased over time, and clustered into node 49 and 145, respectively, which also contain the most highly expressed MnPs (i.e., Phaca256991 and Phaca262882, respectively). Together, the presence of the predicted signal sequence for secretion, cupredoxin domain, and co-expression with a highly expressed MnP (Phaca256991) suggests that the protein with unknown function, Phaca259771, may in fact contribute to MnP action through, for example, H_2O_2 production.

Discussion

The wood samples used to cultivate *P. carnosa* were ball milled to increase the accessibility of the wood substrates and permit comparative transcriptome analyses that reveal fungal responses to differences in wood fibre composition uncoupled from differences in wood fibre structure. While this approach was expected to reduce the requirement for low molecular weight molecules thought to promote incipient stages of fungal growth on intact wood samples [32–34], the overall transcriptome patterns generated by *P. carnosa* during growth on spruce and aspen were similar despite differences in hemicellulose, lignin, and extractive contents in these wood substrates. Recent studies of other fungi report analogous findings. For example, *Fomitopsis pinicola* elicits similar patterns of CAZyme gene expression following growth on

Fig. 4 Transcript abundance over time for highly expressed hydrophobins on (**a**) aspen and (**b**) spruce. CPM values are given for each growth point on both substrates

aspen, pine, and spruce [35]. Instead, more significant differences were correlated to wood sample preparation (e.g., wood wafers versus wood powder) [35], underpinning the importance of 1) uncoupling fibre structure from composition when the aim is to compare impacts of compositional differences, and 2) considering the mode of fibre pretreatment when the aim is to improve enzyme formulations for biomass processing.

Transcriptomic analysis of *P. carnosa* at five growth points on two substrates uncovered expression patterns

for transcripts present at low abundance, which can be used to guide sequence selections for functional characterization. For example, transcripts encoding family GH16 sequences were grouped into those that were most abundant at initial or late stages of fungal growth, or else steadily expressed over the cultivation period (Fig. 3; Additional file 8). Functional predictions of GH16 sequences is complicated by the several activities that have been attributed to this CAZy family, including xyloglucan transglycosylase activity observed for plant GH16s, and lichenase, laminarinase,

	Aspen					Spruce				
	Day 7	Day 13	Day 17	Day 20	Day 23	Day 13	Day 20	Day 23	Day 26	Day 30
Protein ID	GP1	GP2	GP3	GP4	GP5	GP1	GP2	GP3	GP4	GP5
179284	874	1030	1227	976	1257	800	831	5641	1063	712
247092	909	638	440	793	883	2582	995	881	838	1077
265716	519	961	1106	835	1080	603	675	4183	849	646
263551*	201	2283	2484	798	955	777	1610	1099	888	1144
265294	71	707	430	189	68	2244	2823	296	506	213
259771*	251	397	371	1422	2830	808	452	1335	1707	2263
257799	69	423	305	237	353	1094	1376	402	2169	897
265297	16	726	761	274	150	1679	2778	585	1097	576
96827*	48	2159	2958	325	38	72	1232	231	69	15
256483	55	56	87	1827	2141	34	281	3019	2926	3254
262908	9	8	240	4319	245	629	2966	902	165	150

Fig. 5 Most abundant transcripts encoding proteins with unknown function that cluster with known CAZymes. Abundances (CPM) are specified and represented by the relative length of the data bars. *predicted signal peptide

Jurak *et al. BMC Genomics*   (2018) 19:815

Page 8 of 13

and agarase activities observed for microbial GH16s [20]. Diverse biological processes have also been attributed to these enzymes, including fungal cell wall synthesis [36, 37], and hydrolysis of β-glucans in endosperm cell walls of barley and other cereals [38]. Of particular note, the most abundant GH16 transcript encoded by *P. carnosa* during cultivation on wood (Phaca102946) shares less than 30% sequence identity with currently characterized fungal GH16s.

The expanded set of P450 genes in *P. carnosa* could enable its growth on heartwood and coniferous wood in general, which is typically characterized by comparatively high lignin and extractive content [4, 8]. Remarkably, over 50% of the most highly expressed P450 mainly grouped in clans CYP52 and CYP64, which also account for most of the P450 sequence expansion in *P. carnosa*. The current transcriptome analyses thus strengthen the hypothesis that both clans CYP64 and CYP52 play an important role in enabling *P. carnosa* to colonize and grow on heartwood tissue of both deciduous and coniferous sources, while at the same time, reveal the likely relevance of clan CYP547.

Co-expression analyses can identify groups of gene products that may operate together. Herein, a differentially expressed hydrophobin sequence was shown to group with known CAZymes, including sequences belonging to families AA9 and GH128. Certainly, the role that hydrophobins may play in interacting with and accessing lignocellulosic substrates remains unclear. Still, other studies are beginning to note the expression of these proteins during fungal cultivation on wood. For example, Couturier et al., [9] report 500 times higher expression of a predicted hydrophobin during *P. coccineus* cultivation on pine and aspen compared to maltose. Likewise, Kuuskeri et al., [10] found hydrophobin transcripts amongst those most upregulated in *P. radiata* during growth on wood. Co-expression analyses also identified a protein with unknown function containing a predicted cupredoxin domain (Phaca259771), which grouped with highly expressed MnP sequences. A cupredoxin containing protein of unknown function was also identified through transcriptomic analysis of *P. chrysosporium* grown on spruce [8]; however, the sequence identity to Phaca259771 is only 34%. Notably, comparisons between highly-expressed proteins with unknown function identified herein, and those highlighted in earlier transcriptome analyses of softwood-degrading basidiomycetes [8, 9, 22, 39], did not reveal a core set of related sequences.

Conclusions

Following an initial lag phase during growth on spruce compared to aspen, the transcriptome elicited by *P. carnosa* were similar on both wood substrates. For both cultivation conditions, the most abundant transcript encoded the same

MnP (Phaca256991), followed by AA3 and AA5_1 oxidases that may generate the H_2O_2 required for MnP activity. Approximately 25% of the identified P450 monooxygenases encoded by *P. carnosa* were also marked as highly expressed during growth on aspen and spruce. These mainly belonged to clans CYP52 and CYP64, which are also expanded in the *P. carnosa* genome.

Overall, transcript abundances for glycoside hydrolases and carbohydrate esterases were lower than those encoding auxiliary oxidoreducases. Of these, transcripts encoding GH2, GH5, GH6, GH7, GH16, and CE1, were among the most highly expressed sequences predicted to encode plant biomass degrading enzymes. Similar expression profiles were observed for other softwood-degrading white-rot fungi, including *Dichomitus squalens*, *Phlebia radiata*, and *Obba rivulosa*, and *Pycnoporus coccineus* [9–13]. The current study further showed that despite known differences in the compositions of spruce and aspen, *P. carnosa* produces a similar profile of CAZymes transcripts when grown on these substrates. This observation is consistent with recent studies that underscore the contribution of wood sample preparation (e.g., wood wafers versus wood powder) [35], in addition to specific tree species, age, and wood tissue on the expression of CAZymes by wood-degrading fungi.

All differentially expressed transcripts encoding carbohydrate-active enzymes belonged to the core set of plant biomass degrading enzymes previously predicted through comparative analysis of basidiomycete transcriptomes [40]. The resolution afforded by the multiple growth points included herein, however, revealed distinct expression profiles of GH families having relatively low transcript abundance yet recognized roles in plant polysaccharide conversion. These results can be used to guide the selection of *P. carnosa* sequences for functional characterization, which is especially important when considering comparatively large CAZyme families (e.g., GH5 and GH16).

The inclusion of several growth points in the current study also permitted detailed cluster analysis of co-expressed transcripts, which uncovered enzymes that may operate in concert, including MnPs and carbohydrate oxidases from families AA2 and AA3, as well as predicted P450 monooxygenases and glutathione S-transferases. Furthermore, co-expression analyses uncovered non-catalytic proteins and proteins with unknown function that could contribute to lignocellulose conversion; in particular Phaca259771, which is predicted to encode a cupredoxin domain. However, the low sequence identity of highly expressed transcripts encoding unclassified proteins from diverse lignocellulose-degrading fungi, further underscores the importance of comparable cultivation methods to expanding the core set of carbohydrate-active enzymes for lignocellulose conversion.

Methods

Fungal cultivation

Phanerochaete carnosa strain HHB-10118-sp was obtained from the U.S. Department of Agriculture (USDA) Forest Products Laboratory (Madison, WI) and maintained on YMPG agar plates (2 g yeast extract, 10 g malt extract, 2 g peptone, 10 g glucose, 2 g KH_2PO_4, 1 g $MgSO_4 \cdot 7H_2O$, and 15 g agar per 1 L in H_2O) as previously described [4]. Wood samples were obtained from New Brunswick, Canada, where a 50 cm bolt at 80 cm and 130 cm trunk heights were cut from trembling aspen (*Populus tremuloides*) and white spruce (*Picea glauca*); heartwood sections were then separated, air-dried, and then ground in a Wiley mill (Thomas scientific, NJ, USA) followed by planetary ball mill [4, 30]. Approximately 4 g of wood particles were distributed as a thin layer in Petri dishes, autoclaved, and then supplemented with 20 mL of B3 buffer [14, 30]. To maximize the reproducibility of fungal growth patterns, a 1 cm diameter agar plug taken from the growing edge of *P. carnosa* cultivated on YMPG agar plates was directly transferred to the center of each plate containing wood particles, and incubated at 27 °C under stationary conditions. As performed previously [14, 15, 30], all mycelia for 2 cm colonies, and the central 4 cm of colonies reaching 5, 6, 8, and 9 cm in diameter, were harvested and then stored at – 80 °C prior to RNA extraction and wood analyses. By sampling mycelia from the centre of the growing colony rather than the growing edge of the colony, corresponding transcriptomes were more likely to reflect responses to potentially changing substrate composition resulting from fungal growth. Moreover, this approach to fungal cultivation yielded similar and sufficient quantities of RNA for sequencing, and at the same time, ensured reproducible harvesting of biological replicates. Three replicate cultivations were prepared for each colony size (Additional file 1).

RNA extraction and sequencing

Total RNA was isolated from frozen mycelia using the Plant/Fungi Total RNA Purification Kit (Norgen Biotek) according to the manufacturer's instructions. The quality and quantity of purified RNA were monitored using a Bioanalyzer (Agilent Technologies). A portion of purified total RNA was used for first strand cDNA synthesis using RevertAid reverse transcriptase (Thermo Scientific), and the reproducibility between three biological replicates was verified by quantitative reverse-transcription PCR (qRT-PCR) for a manganese peroxidase (MnP, Phaca262882) and chitin synthase gene (Chs, Phaca257626) [15]. Two of the three biological replicates for each cultivation were then randomly selected, and total RNA from those replicates were utilized for independent RNA sequencing.

The cDNA library was prepared using TruSeq RNA Sample Prep Kit v2 (Illumina). Briefly, 1 μg of high quality total RNA was used to generate the cDNA library having an average fragment size of 350–400 bp. The quality of the barcoded library was checked using a Bioanalyzer and quantified by qPCR using KAPA SYBR FAST Universal 2X qPCR Master Mix (Kapa Biosystems) running in 7900HT Fast Real Time PCR System (Applied Biosystems) [30]. The cDNA libraries were then loaded on a flowcell for cluster generation using c-Bot and TruSeq PE Cluster Kit v3 (Illumina). Sequencing was performed using a HiSeq2000 system and the TruSeq SBS Kit v3 (pair-ended 200 cycles, Illumina). 100 bp pair-ends were generated. The real-time base call (.bcl) files were converted to fastq files using CASAVA 1.8.2 (Illumina, on CentOS 6.0 data storage and computation linux servers at the Sequencing Facility of the Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada), and then aligned to JGI *P. carnosa* gene models (v1.0, http://genome.jgi-psf.org/Phaca1/Phaca1.home.html) with the Novoalign software (Novocraft) [15, 30]). Raw sequence data were deposited to the Sequence Read Archive (SRA accession: SRP151360).

Bioinformatic analyses

Normalization of data and log2 fold differences of genes

Read counts of gene models were input into edgeR [41] for CPM (count per million) conversion and differential expression analysis; the consistency of expression values was verified by qRT-PCR and the reproducibility of CPM values between two biological replicates was verified using scatter plots [30]. For construction of transcriptomic models, the following procedures were performed. The log2 fold difference of the gene expression between time points was calculated with R package DESeq2 [42]. Genes with statistical significance were selected based on FDR (false discovery rate) adjusted p value < 0.05. Normalized read counts of the genes were also produced with DESeq2, which were subsequently log2 transformed. The consistency of normalized transcription from all biological replicates was confirmed by visualizing the distribution of read counts (Additional file 9). The expression of 27 housekeeping genes (NADH dehydrogenase and chitin synthase) under all conditions was investigated for the consistency of fungal growth (Additional file 10). A total of 11,796 genes having more than averaged five reads per condition were selected for constructing transcriptomic models.

Correlation among biological replicates

Spearman's rank correlation was calculated with normalized read counts from the biological replicates from all conditions. The estimated correlation coefficients were visualized and further examined as described below (Additional file 2).

Construction of transcriptomic models

Transcriptomic models were constructed using Self-organizing map Harbouring Informative Nodes with Gene Ontology (SHIN+GO) [22, 43]. A self-organizing map (SOM) was trained with the normalized read count of all replicates described above. The matrix of 24×20 (480) was used with a rectangular shape (four neighbouring nodes). The epoch of 1000 times more than the map size was applied (i.e. 480,000, being 480 map size times 1000). The initial radius for SOM calculation was determined using a neighbour distance function in R kohonen package [44]. The following graphic outputs (Tatami maps) were produced and investigated: 1) genome-wide transcriptomic patterns of all biological replicates, and 2) genome-wide condition-specific transcriptomic patterns (Additional files 11 and 12). Similarly-regulated condition-specific genes were determined by fulfilling either of two criteria: 1) > 10.2 log2 reads (above 95th percentile of the entire transcribed genes used for transcriptomic models), or 2) > 2 log2 transcriptional differences of each growth point against growth point 1 (Additional file 11). Functional annotation sets were integrated into the constructed model using Carbohydrate Active Enzyme database (CAZy) [45], InterPro (IPR) [46]), the Gene Ontology (GO) [47], Kyoto Encyclopedia of Genes and Genomes (KEGG) [48], and EuKaryotic Orthologous Groups (KOG) [49] (Additional file 8). IPR, GO, KEGG, KOG, SignalP were obtained from Mycocosm, JGI (https://genome.jgi.doe.gov/Phaca1/Phaca1.home.html). CAZy annotations were obtained from AFMB, CNRS-Aix-Marseille University (http://www.cazy.org). All procedures were performed with the SHIN module of SHIN+GO.

Carbohydrate content and composition

Samples were treated with 72% (*w*/w) H_2SO_4 (1 h, 30 °C) followed by hydrolysis with 1 M H_2SO_4 for 3 h at 100 °C. Hydrolysate was diluted 20 times and carbohydrate content and composition was determined by High Performance Anion Exchange chromatography (HPAEC) on a Dionex Ultimate ICS-3000 system (Thermo Scientific, Sunnyvale, CA, USA) equipped with an amperometric cell detector. Separation and quantification of monosaccharides was performed at a flow rate of 0.37 ml/min, with H_2O as the eluent: The elution profile was as follows: 0–35 min 100% H_2O; 35–42 min to 100% 0.2 M NaOH; 42–45 min to 100% H_2O.

Lignin content and composition measured using py-GC/MS (pyrolysis- gas chromatography/mass spectrometry)

Pyrolysis was performed with an EGA/PY-3030D Multi-shot pyrolyzer (Frontier Laboratories, New Ulm, MN, USA) equipped with an AS-1020E Autoshot auto-sampler as described previously by van Erven et al., (2017) [50]. The pyrolyzer was coupled to GC-MS using a Trace GC equipped with a DB-1701 fused-silica capillary column (30 m × 0.25 mm i.d. 0.25 μm film thickness) coupled to a DSQ-II mass spectrometer (both Thermo Scientific, Waltham, MA, USA). Pyrolysis, GC and MS settings were similar as previously described [51]. Samples were weighed using a XP6 excellence-plus microbalance (Mettler Toledo, Columbus, OH, USA). Pyrolysis of total biomass (70–80 μg) was performed at 500 °C for 1 min with an interface temperature of 300 °C. Pyrolysis products were injected on the column via split/splitless injection (at 250 °C) with a split ratio of 1:133 and helium was used as carrier gas with constant flow at 1.5 mL·min^{-1}. The GC oven was programmed from 70 °C (2 min) to 270 °C at 5 °C·min^{-1} and held at 270 °C for 15 min. MS detection was used with EI at 70 eV, a source temperature of 250 °C, a scan range of *m/z* 50–550 and a scan rate of 4.0 scans/sec. Compounds were identified by comparing retention time and mass spectrum with standards, the NIST library and data published by Ralph and Hatfield [52].

For qualitative identification, pyrograms were processed by AMDIS software (version 2.71, NIST, USA). For identification and deconvolution the following software settings were used: minimum match factor at 60 with multiple identifications per compound, component width at 20, adjacent peak subtraction at two, resolution at high, sensitivity at very high and shape requirements at low. Compounds identified on the basis of reference standards were annotated by evaluation of retention time (± 0.1 min), reverse search (≥ 80) and simple search (≥ 30). Peak molar area was calculated as defined by Del Río et al., [53]. Lignin content was estimated on the basis of total area of lignin-derived pyrolysis products and compared to a wheat straw reference sample with known Klason lignin content (acid-insoluble lignin + acid-soluble lignin) [51]. All samples were analyzed in triplicate.

Additional files

control sample; no fungal cultivation. Codes in brackets are used for peak annotation in Additional file 7. Since different amounts of starting material were analyzed, similar relative quantities of pyrolysis products between growth points indicates non-selective, simultaneous decay or transformation of biomass components. [a] miscellaneous [b] C_α-oxygen [c] C_β-oxygen [d] C_γ-oxygen. (DOCX 26 kb)

Additional file 5: Lignin contents estimated by py-GC-MS for aspen and white spruce control (C) and *P. carnosa* grown at growth point 1 and 5. *significantly different from control at $P \leq 0.05$. (TIFF 2495 kb)

Additional file 6: Counts per million (CPM) values for all sequences at all growth points and on each substrate (aspen heartwood – AH; white spruce heartwood – WH) is also shown. (XLSX 4817 kb)

Additional file 7: Transcript abundances (CPM values) of all annotated CAZymes and cytochrome P450s encoded by *P. carnosa*. (XLSX 586 kb)

Additional file 8: The annotations per protein IDs in 480 nodes. The nodes with high/differential transcriptions are labelled. The table also includes JGI protein IDs with following information. Log2 transformed normalized read counts of the genes averaged from the duplicates at all growth points; the log2 fold difference of the gene expression between time points with statistical significance (FDR adjusted *p* value < 0.05); functional annotation information on Carbohydrate Active Enzyme database (CAZy), InterPro (IPR), the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and EuKaryotic Orthologous Groups (KOG), and SignalP for prediction of signal peptides. (XLSX 2752 kb)

Additional file 9: The distribution and density of normalized log2 transformed read counts of 11,796 genes from 10 conditions with 2 replicates each. AH/WH: Aspen/Spruce. 1_#/2_#/3_#/4_#/5_#: Growth points and followed by replicate IDs. (TIFF 1084 kb)

Additional file 10: The normalized log2 transformed read count of chitin synthase (11 genes) and NADH dehydrogenase (16 genes). AH/WH: Aspen/Spruce. 1_#/2_#/3_#/4_#/5_#: Growth points and followed by replicate IDs. (TIFF 1395 kb)

Additional file 11: Tatami maps showing the transcriptomic patterns of 20 replicates. AH/WH: Aspen/Spruce. 1_#/2_#/3_#/4_#/5_#: Growth points and followed by replicate IDs. The log2 read count of the replicates was overlaid onto the trained SOM. The vertical bar indicates the transcription levels. (TIFF 8620 kb)

Additional file 12: Condition-wise Tatami maps showing the averaged transcriptomic patterns from aspen/ spruce at five growth points. The averaged log2 read count of replicates grown in each condition was overlaid onto the trained SOM, representing the dynamics of genome-wide transcriptions corresponding to the conditions. (TIFF 6328 kb)

Abbreviations
AA: Auxiliary activities; AOX: Alcohol oxidases; CAZy database: Carbohydrate Active Enzyme database; CAZyme: Carbohydrate-active enzyme; CPM: Count per million; CRO: GLOX/related copper radical oxidases; DREX: Distantly related to plant expansins; GH: Glycoside hydrolase; GLOX: Glyoxal oxidases; GO: Gene Ontology; GP: Growth points; HPAEC: High Performance Anion Exchange chromatography; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: EuKaryotic Orthologous Groups; LiP: Lignin peroxidase; LOOL: Loosenin-like proteins; LPMO: Lytic polysaccharide monooxygenase; MnP: Manganese peroxidase; py-GC/MS: Pyrolysis- Gas chromatography/Mass spectrometry; SOM: Self-organizing map

Acknowledgements
We thank the CAZy team at Aix-Marseille University for detailed CAZyme annotations.

Availability of data and materials
All data generated or analyzed during this study are included in this published article [and its additional files].

Authors' contributions
EJ analysed the transcriptome data and wrote the manuscript. HS designed and performed the experiments and interpreted the transcriptome data. Gv.E and MAK analysed the residual wood samples. JAG analyse transcriptome profiles of hydrophobin and loosenin sequences. YG, PW and ET assisted with manual transcriptome sequence annotations. KC and CYH performed the transcriptome sequencing, assembly, and automated annotation. MR and SM applied the SHIN+GO cluster analysis to generate the tatami maps. ERM conceived and coordinated the study. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Bioproducts and Biosystems, Aalto University, Espoo, Finland. [2]Department of Aquatic Biotechnology and Bioproduct Engineering, Groningen, The Netherlands. [3]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Canada. [4]Wageningen University, Laboratory of Food Chemistry, Bornse Weilanden 9, 6708, WG, Wageningen, The Netherlands. [5]Department of Medical Biophysics, University of Toronto, Toronto, Canada. [6]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada. [7]Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Canada. [8]Laboratory of Excellence ARBRE, INRA, Nancy, Lorraine, France. [9]Aix-Marseille Université, INRA, UMR1163, Biodiversité et Biotechnologie Fongiques, Marseille, France.

References
1. Martinez D, Larrondo LF, Putnam N, Gelpke MDS, Huang K, Chapman J, et al. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. Nat Biotechnol. 2004;22:695.
2. Ohm RA, Riley R, Salamov A, Min B, Choi I-G, Grigoriev IV. Genomics of wood-degrading fungi. Fungal Genet Biol. 2014;72:82–90.
3. Burdsall HH Jr. A contribution to the taxonomy of the genus Phanerochaete (Corticiaceae, Aphyllophorales): J. Cramer; 1985.
4. Suzuki H, MacDonald J, Syed K, Salamov A, Hori C, Aerts A, et al. Comparative genomics of the white-rot fungi, *Phanerochaete carnosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. BMC Genomics. 2012;13:444.
5. Palonen H, Thomsen AB, Tenkanen M, Schmidt AS, Viikari L. Evaluation of wet oxidation pretreatment for enzymatic hydrolysis of softwood. Appl Biochem Biotechnol. 2004;117:1–17.
6. Zhu XJ Pan JY. Woody biomass pretreatment for cellulosic ethanol production: technology and energy consumption evaluation. Bioresour Technol. 2010;101:4992–5002.
7. Vanden Wymelenberg A, Gaskell J, Mozuch M, Splinter BonDurant S, Sabat G, Ralph J, et al. Significant alteration of gene expression in wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium* by plant species. Appl Environ Microbiol. 2011;77:4499–507.
8. Korripally P, Hunt CG, Houtman CJ, Jones DC, Kitin PJ, Cullen D, et al. Regulation of gene expression during the onset of ligninolytic oxidation by *Phanerochaete chrysosporium* on spruce wood. Appl Environ Microbiol. 2015;81:7802–12.

9.   Couturier M, Navarro D, Chevret D, Henrissat B, Piumi F, Ruiz-Dueñas FJ, et al. Enhanced degradation of softwood versus hardwood by the white-rot fungus *Pycnoporus coccineus*. Biotechnol Biofuels. 2015;8:216.

10. Kuuskeri J, Häkkinen M, Laine P, Smolander O-P, Tamene F, Miettinen S, et al. Time-scale dynamics of proteome and transcriptome of the white-rot fungus *Phlebia radiata*: growth on spruce wood and decay effect on lignocellulose. Biotechnol Biofuels. 2016;9:192.

11. Rytioja J, Hildén K, Hatakka A, Mäkelä MR. Transcriptional analysis of selected cellulose-acting enzymes encoding genes of the white-rot fungus *Dichomitus squalens* on spruce wood and microcrystalline cellulose. Fungal Genet Biol. 2014;72:91–8.

12. Rytioja J, Hildén K, Di Falco M, Zhou M, Aguilar-Pontes MV, Sietiö O, et al. The molecular response of the white-rot fungus *Dichomitus squalens* to wood and non-woody biomass as examined by transcriptome and exoproteome analyses. Environ Microbiol. 2017;19:1237–50.

13. Marinović M, Aguilar-Pontes MV, Zhou M, Miettinen O, de Vries RP, Mäkelä MR, et al. Temporal transcriptome analysis of the white-rot fungus *Obba rivulosa* shows expression of a constitutive set of plant cell wall degradation targeted genes during growth on solid spruce wood. Fungal Genet Biol. 2017;112:47–54.

14. MacDonald J, Master ER. Time-dependent profiles of transcripts encoding lignocellulose-modifying enzymes of the white rot fungus *Phanerochaete carnosa* grown on multiple wood substrates. Appl Environ Microbiol. 2012;78:1596–600.

15. MacDonald J, Doering M, Canam T, Gong Y, Guttman DS, Campbell MM, et al. Transcriptomic responses of the softwood-degrading white-rot fungus *Phanerochaete carnosa* during growth on coniferous and deciduous wood. Appl Environ Microbiol. 2011;77:3211–8.

16. Kersten P, Cullen D. Copper radical oxidases and related extracellular oxidoreductases of wood-decay Agaricomycetes. Fungal Genet Biol. 2014;72:124–30.

17. Vanden Wymelenberg A, Sabat G, Mozuch M, Kersten PJ, Cullen D, Blanchette RA. Structure, organization, and transcriptional regulation of a family of copper radical oxidase genes in the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. Appl Environ Microbiol. 2006;72:4871–7.

18. Couturier M, Ladevèze S, Sulzenbacher G, Ciano L, Fanuel M, Moreau C, et al. Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. Nat Chem Biol Nature. 2018;14:306.

19. Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol. 2012;12:186.

20. Behar H, Graham SW, Brumer H. Comprehensive cross-genome survey and phylogeny of glycoside hydrolase family 16 members reveals the evolutionary origin of EG16 and XTH proteins in plant lineages. Plant J. 2018;95(6):1114–28.

21. Ichinose H. Cytochrome P450 of wood-rotting basidiomycetes and biotechnological applications. Biotechnol Appl Biochem. 2013;60:71–81.

22. Miyauchi S, Navarro D, Grigoriev IV, Lipzen A, Riley R, Chevret D, et al. Visual comparative omics of fungi for plant biomass deconstruction. Front Microbiol. 2016;7:1335.

23. Morel M, Meux E, Mathieu Y, Thuillier A, Chibani K, Harvengt L, et al. Xenomic networks variability and adaptation traits in wood decaying fungi. Microb Biotechnol. 2013;6:248–63.

24. Mathieu Y, Prosper P, Favier F, Harvengt L, Didierjean C, Jacquot J-P, et al. Diversification of fungal specific class a glutathionetransferases in saprotrophic fungi. PLoS One. 2013;8:e80298.

25. Quiroz-Castañeda RE, Martínez-Anaya C, Cuervo-Soto LI, Segovia L, Folch-Mallol JL. Loosenin, a novel protein with cellulose-disrupting activity from *Bjerkandera adusta*. Microb Cell Factories. 2011;10:8.

26. Cosgrove DJ. Loosening of plant cell walls by expansins. Nature. 2000;407:321.

27. Linder MB, Szilvay GR, Nakari-Setälä T, Penttilä ME. Hydrophobins: the protein-amphiphiles of filamentous fungi. FEMS Microbiol Rev. 2005;29:877–96.

28. Wösten HAB. Hydrophobins: multipurpose proteins. Annu Rev Microbiol. 2001;55:625–46.

29. Tanaka T, Nakayama M, Takahashi T, Nanatani K, Yamagata Y, Abe K. Analysis of the ionic interaction between the hydrophobin RodA and two cutinases of *Aspergillus nidulans* obtained via an *Aspergillus oryzae* expression system. Appl Microbiol Biotechnol. 2017;101:2343–56.

30. Suzuki H, Vuong TV, Gong Y, Chan K, Ho CY, Master ER, et al. Sequence diversity and gene expression analyses of expansin-related proteins in the white-rot basidiomycete, *Phanerochaete carnosa*. Fungal Genet Biol. 2014;72:115–23.

31. Aimanianda V, Simenel C, Garnaud C, Clavaud C, Tada R, Barbin L, et al. The dual activity responsible for the elongation and branching of β-(1,3)-glucan in the fungal cell wall. MBio. 2017;8:e00619–7.

32. Blanchette RA, Krueger EW, Haight JE, Aktar M, Akin DE. Cell wall alterations in loblolly pine wood decayed by the white-rot fungus, *Ceriporiopsis subvermispora*. J Biotechnol. 1997;53:203–13.

33. Tanaka H, Itakura S, Enoki A. Hydroxyl radical generation by an extracellular low-molecular weight substance and phenol oxidase activity during wood degradation by the white-rot basidiomycete *Phanerochaete chrysosporium*. J Biotechnol. 1999;75:57–70.

34. Tanaka H, Yoshida G, Baba Y, Matsumura K, Wasada H, Murata J, et al. Characterization of a hydroxyl-radical producing glycoprotein and its presumptive genes from the white-rot basidiomycete *Phanerochaete chrysosporium*. J Biotechnol. 2007;128:500–11.

35. Wu B, Gaskell J, Held BW, Toapanta C, Vuong T, Ahrendt S, et al. Substrate-specific differential gene expression and RNA editing in the brown rot fungus *Fomitopsis pinicola*. Appl Environ Microbiol. 2018;84(16).

36. Mouyna I, Aimanianda V, Hartl L, Prevost MC, Sismeiro O, Dillies MA, et al. GH16 and GH81 family β-(1,3)-glucanases in *Aspergillus fumigatus* are essential for conidial cell wall morphogenesis. Cell Microbiol. 2016;18(9):1285–93.

37. Roemer T, Bussey H. Yeast beta-glucan synthesis: KRE6 encodes a predicted type II membrane protein required for glucan synthesis in vivo and for glucan synthase activity in vitro. Proc Natl Acad Sci U S A. 1991;88(24):11295–9.

38. You S, Tu T, Zhang L, Wang Y, Huang H, Ma R, et al. Improvement of the thermostability and catalytic efficiency of a highly active β-glucanase from *Talaromyces leycettanus* JCM12802 by optimizing residual charge-charge interactions. Biotechnol Biofuels. 2016;9:124.

39. Hori C, Gaskell J, Igarashi K, Kersten P, Mozuch M, Samejima M, et al. Temporal alterations in the secretome of the selective ligninolytic fungus *Ceriporiopsis subvermispora* during growth on aspen wood reveal this organism's strategy for degrading lignocellulose. Appl Environ Microbiol. 2014;80:2062–70.

40. Peng M, Aguilar-Pontes MV, Hainaut M, Henrissat B, Hildén K, Mäkelä MR, et al. Comparative analysis of basidiomycete transcriptomes reveals a core set of expressed genes encoding plant biomass degrading enzymes. Fungal Genet Biol. 2018;112:40–6.

41. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.

42. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

43. Miyauchi S, Navarro D, Grisel S, Chevret D, Berrin J-G, Rosso M-N. The integrative omics of white-rot fungus *Pycnoporus coccineus* reveals co-regulated CAZymes for orchestrated lignocellulose breakdown. PLoS One. 2017;12:e0175528.

44. Wehrens R, Buydens LMC. Self- and super-organizing maps in R: the kohonen package. J Stat Softw. 2007;21:1–9.

45. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels. 2013;6:41.

46. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 2017;45:D190–9.

47. Consortium TGO. Gene ontology consortium: going forward. Nucleic Acids Res. 2015;43:D1049–56.

48. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999;27:29–34.

49. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003;4:41.

50. van Erven G, de Visser R, Merkx DWH, Strolenberg W, de Gijsel P, Gruppen H, et al. Quantification of lignin and its structural features in plant biomass using 13C lignin as internal standard for pyrolysis-GC-SIM-MS. Anal Chem. 2017;89(20):10907–16.

51. Jurak E, Punt AM, Arts W, Kabel MA, Gruppen H. Fate of carbohydrates and lignin during composting and mycelium growth of *Agaricus bisporus* on wheat straw based compost. PLoS One. 2015;10:e0138909.

52. Ralph J, Hatfield RD. Pyrolysis GC-MS characterization of forage materials. J Agric Food Chem. 1991;39:1426.

53. del Río JC, Rencoret J, Prinsen P, Martínez ÁT, Ralph J, Gutiérrez A. Structural characterization of wheat straw lignin as revealed by analytical pyrolysis, 2D-NMR, and reductive cleavage methods. J Agric Food Chem. 2012;60: 5922–35.