**BMC
Genomics**

The genome and transcriptome of the pine saprophyte *Ophiostoma piceae*, and a comparison with the bark beetle-associated pine pathogen *Grosmannia clavigera*

Sajeet Haridas[1], Ye Wang[1], Lynette Lim[1], Sepideh Massoumi Alamouti[1], Shaun Jackman[2], Rod Docking[2], Gordon Robertson[2], Inanc Birol[2], Jörg Bohlmann[3] and Colette Breuil[1*]

Abstract

Background: *Ophiostoma piceae* is a wood-staining fungus that grows in the sapwood of conifer logs and lumber. We sequenced its genome and analyzed its transcriptomes under a range of growth conditions. A comparison with the genome and transcriptomes of the mountain pine beetle-associated pathogen *Grosmannia clavigera* highlights differences between a pathogen that colonizes and kills living pine trees and a saprophyte that colonizes wood and the inner bark of dead trees.

Results: We assembled a 33 Mbp genome in 45 scaffolds, and predicted approximately 8,884 genes. The genome size and gene content were similar to those of other ascomycetes. Despite having similar ecological niches, *O. piceae* and *G. clavigera* showed no large-scale synteny. We identified *O. piceae* genes involved in the biosynthesis of melanin, which causes wood discoloration and reduces the commercial value of wood products. We also identified genes and pathways involved in growth on simple carbon sources and in sapwood, *O. piceae*'s natural substrate. Like the pathogen, the saprophyte is able to tolerate terpenes, which are a major class of pine tree defense compounds; unlike the pathogen, it cannot utilize monoterpenes as a carbon source.

Conclusions: This work makes available the second annotated genome of a softwood ophiostomatoid fungus, and suggests that *O. piceae*'s tolerance to terpenes may be due in part to these chemicals being removed from the cells by an ABC transporter that is highly induced by terpenes. The data generated will provide the research community with resources for work on host-vector-fungus interactions for wood-inhabiting, beetle-associated saprophytes and pathogens.

Keywords: *Ophiostoma piceae*, Genome, Transcriptome, Wood-staining fungus, Saprobe

Background

Pine trees and processed wood (lumber and logs) are colonized by ascomycete ophiostomatoid fungi that include pathogens and saprobes [1,2]. As they grow in the phloem and sapwood of the trees or in the sapwood of logs or lumber, most of these fungi produce a dark melanin pigment that causes a wood discoloration known as blue stain or sap stain. Ophiostomatoid sap stain fungi were first described more than 100 years ago [3,4] and have been recognized as an economic problem for forest industries worldwide. Currently, the group contains at least five genera that include *Ophiostoma* and *Grosmannia* (Figure 1). Ophiostomatoid fungi produce sticky sexual and asexual spores that are readily vectored by specific or generalist bark beetles that colonize trees or processed wood [5]. Before 1995, in Canada, *Ophiostoma* species were reported as the major cause of pine discoloration [1,6]. However, since 1995, in western Canada, the mountain pine beetle (MPB; *Dendroctonus ponderosae*) has expanded its range, and its fungal associates from the

* Correspondence: colette.breuil@ubc.ca
[1]Department of Wood Science, University of British Columbia, Vancouver, BC V6T1Z4, Canada
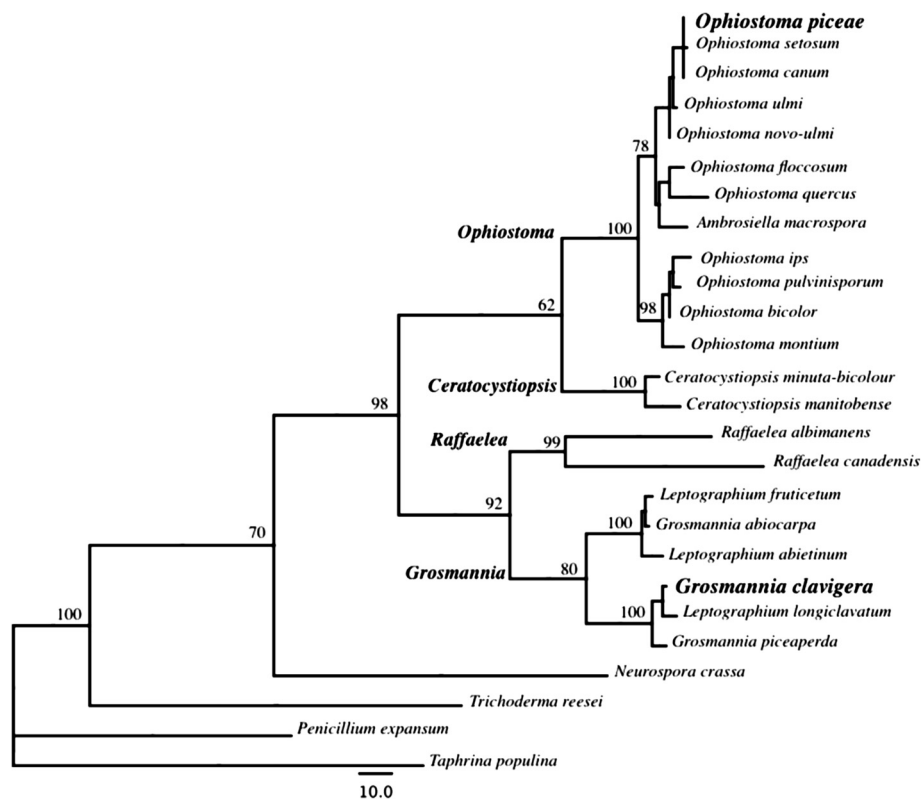Full list of author information is available at the end of the article

Figure 1 ITSs-rDNA maximum parsimony tree places *O. piceae* within the Pezizomycotina. *O. piceae* was originally described by Münch in 1907 [4]; this species and closely related species (e.g. *O. setosum, O. canum, O. novo-ulmi, O. ulmi* and *O. floccosum O. quercus*) are reported as the *O. piceae* complex. The numbers on the branches of the tree are bootstrap values based on 1000 replicates and the heuristic option [9].

genera *Grosmannia* (mainly *G. clavigera* and *Leptographium longiclavatum*) and *Ophiostoma* (*O. montium*) have become the main cause of pine wood discoloration. As well, this beetle-fungal complex has killed large areas of pine trees in western North American conifer forests [7,8]. Wood discoloration is caused by melanin, a dark pigment that is synthesized inside the fungal cell and is released as small black globules into the cell wall and outside of the cell (Figure 2).

O. piceae is a saprobe that is dispersed by generalist bark beetles [10]. This fungal species has been reported in North America, Europe and Asia [5,6,11]. In contrast to *Grosmannia* species, which penetrate deeply into the sapwood of pine logs and reach the heartwood boundary, *O. piceae* is a more superficial sap stain fungus that becomes established in the outer two to three centimetres of sapwood [1,12,13]. Species in the *O. piceae* complex have retained the attention of wood industry researchers because they cause stain in processed wood and were the most commonly isolated species of sap stain fungi in Canadian saw mills [6]. In contrast to *G. clavigera*, which is specific to pine, *O. piceae* is able to grow not only on pine, but also on wood of other conifers in Canada, including black and white spruce, balsam fir and hemlock

[6]. Because members of the *O. piceae* complex grow poorly on freshly cut pine logs and prefer the dryer environment of lumber or dead trees [1,13], their staining effects can be minimized by keeping logs frozen or saturated with water, or by prompt log processing. Green lumber can be protected by kiln drying to below 20% moisture content, or by chemical and biological treatments [14-16].

Both the pathogen *G. clavigera* and the saprophyte *O. piceae* acquire nutrients from pine species by secreting extracellular enzymes to break down large molecules like polysaccharides (e.g. hemicellulose and starch), proteins and lipids. They do not degrade wood and do not affect wood structural properties [1,16,17], so likely have limited or incomplete cellulolytic and/or lignolytic activities. However, in order to colonize conifers (e.g. lodgepole pine), fungi and their bark beetle vectors have to cope with the host's preformed and induced defense chemicals, which include terpenoid and phenolic compounds [18-20]. Pathogens like *G. clavigera* have evolved mechanisms to overcome these defences [8,21]. However, the role of such host defence compounds in cut logs and lumber, where saprophytes like *O. piceae* are generally found,
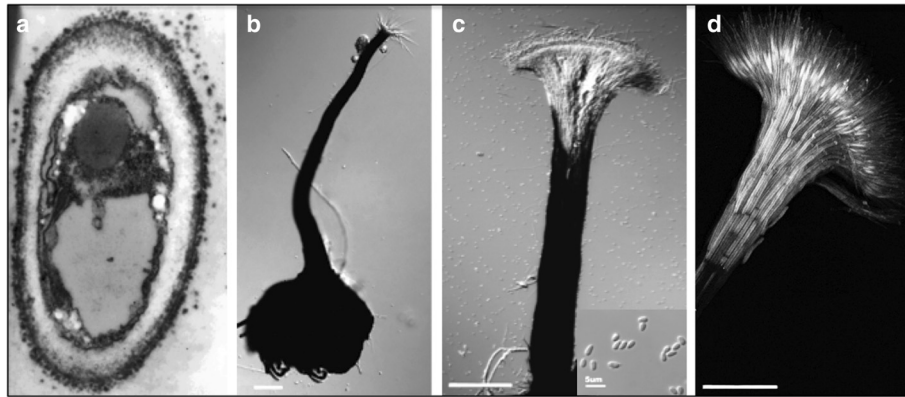
Figure 2 Micrographs of *O. piceae* mycelium, sexual and asexual structures. It grows as filamentous hyphae on solid media and as yeast like form in liquid media. *O. piceae* mycelium, perithecium and synemata are highly melanized while both asexual and sexual spores are not pigmented. The pigment accumulates as small black granules in the cell wall and the external sheath surrounding the hyphae (**a**: electron micrograph); Melanin is also present in fruiting body (perithecium) (**b**: light micrograph), as well as in the stem of the synemata (**c**: light micrograph; **d**: confocal micrograph), an aggregation of branched hyphae that produce abundant asexual spores. The fruiting bodies are easily obtained in artificial media or wood but require the pairing of two individuals with different mating types.

has not been reported. It is important to note that the composition of defence chemicals, especially terpenoids, varies with different pine genotypes across the landscape and can be affected by the environment [22]. Further, wood processing and drying affect the concentrations of chemicals in wood products, and so logs and lumber typically contain lower concentrations of the subset of terpenoids that are volatile [23,24].

In previous work we reported the annotated genome of the pine pathogen *G. clavigera*, and showed that this fungus is able to tolerate and utilize pine defence compounds, specifically terpenoids found in pine oleoresin [8,21]. Here, we report the annotated genome of the saprophyte *O. piceae*, and its gene expression responses in a range of growth conditions that include wood nutrients and host defence chemicals. We compare these results to corresponding results from *G. clavigera*, and highlight differences between a pathogen that colonizes and kills living pine trees and a saprophyte that lives on dead trees or processed wood. Neither fungus has a lignocellulolytic enzyme system that would allow it to degrade wood. Both fungi overcome terpenoid defence chemicals in their pine niches; however, only the pathogen, but not the saprophyte, can metabolize terpenoids as a carbon source. Both fungi have a similar ABC efflux transporter that is highly induced with monoterpene treatments. The functionality of the *O. piceae* transporter remains to be fully characterized.

Results
Genome assembly
We used ABySS [25] to assemble 100-nt reads from 200- and 700-nt insert Illumina HiSeq 2000 libraries, and ~300-nt reads from an 8-kb insert 454 Titanium

(Table 1, Methods). The Illumina libraries provided >100× coverage for assembly and initial scaffolding, while the 454 reads supported long-range scaffolding. Refining this assembly with two iterations of Anchor resulted in a genome that consisted of 244 scaffolds, each of which was at least 1,000 nt in length. The assembly contained 335 false gaps represented by a single lowercase 'n' (see Methods). Of these, 219 were resolved by mapping Trinity-assembled RNA-seq transcripts to the genome using exonerate est2genome [26]. The remaining 116 gaps were resolved by using exonerate to find small overlaps (<5 bp) at the ends of contigs that were joined by an 'n'.

We removed from the final assembly 187 scaffolds and contigs smaller than 10,000 bp (including gaps) that represented 1% of the assembly because they contained no genes or t-RNAs. The corrected 32.8-Mbp genome assembly consisted of 45 scaffolds. One percent of the genome consisted of 342 gaps (N's). Half of the genome was in nine scaffolds that had an N50 of approximately 1.45 Mbp, while 90% was represented in 27 scaffolds that had an N90 of approximately 0.38 Mbp. Using CEGMA [27], we identified complete copies of 233 of 248 conserved eukaryotic genes and partial copies of an additional five, which suggests that our assembly represents 94% - 96% of the *O. piceae* gene space [28]. The genome characteristics of *O. piceae*, and three other ascomycetes [8,29,30] also in

Table 1 Sequencing strategy for *O. piceae* genome

Sequencing technology	Read length (nt)	Insert length (nt)	Read pairs (Millions)
Illumina Hiseq	100	200	87.8
Illumina Hiseq	100	700	32.2
454 Titanium	318 (median)	8000	0.3

the class of the Sordariomycetes and found on dead tree and wood products are summarized in Table 2.

Genome features and annotation

We used the Maker annotation pipeline [31] to predict genes. Within the annotated genome of *O. piceae*, we identified genes and gene families for secondary metabolite processing, cytochrome P450 as well as ABC transporters. We also identified homologous *O. piceae* and *G. clavigera* proteins based on reciprocal best BLAST hits. We further characterized the MAT idiomorph that is responsible for the mating type of the sequenced strain.

Maker predicted 8,884 proteins within our acceptance criteria (see Methods), of which 8,723 were at least 100 amino acids long. Almost 65% (5,786) of the predicted proteins encoded by the gene models had a known Pfam domain. Some of the major gene families in *O. piceae* are shown in Table 3. About a third of the predicted genes (3,026) had only one exon and only 1,283 transcripts were encoded by four or more exons. In this compact genome, genes, not including their upstream and downstream untranslated regions (UTRs) represent 45% of the assembly. Almost a quarter (1,984) of the predicted gene coding sequences (CDS) was within 500 bp of their respective neighbouring CDS, and almost half (4349) were within 1,000 bp of its neighbour. Our analysis predicted that 778 CDSs encode secreted proteins.

Although *O. piceae* and *G. clavigera* share hosts, cause sap-stain in pine, and are in sister clades in the Ophiostomatales [32] (Figure 1), their genomes showed no large-scale synteny (Additional file 1). This is consistent with synteny being lost with evolutionary time between members of the class Dothideomycetes [33]. Dothideomycetes is a sister clade of the class Sordariomycetes, which include *O. piceae* and *G. clavigera*, and we anticipate that similar synteny losses have occurred

Table 2 Characteristics of the *O. piceae* (Op) genome assembly and annotation and a comparison with other related genomes

	Op	Gc[a]	Nc (10)[b]	Tr[c]
Genome size (Mbp)	32.8	30	41	33.5
Number of scaffolds	45	289	7 [d]	87
N50 (Mbp)	1.45	2	1.56	1.12
Number of ungapped contigs	388	478	956	231
Genome GC content (%)	52.8	53.4	48.25	52.7
Non-coding genome (%)	54	54.28	56	
Number of genes	8,884	8,312	9,733	9,129
Median CDS length (bp)	1,401	1,350	1,673	1,299
Exon GC content (%)	59.7	60.4		57.8

a. *G. clavigera* (Gc); b. *Neurospora crassa* Sequencing Project, Broad Institute of Harvard and MIT (NC10). c. *Trichoderma reesei* (Tr). d. Chromosome numbers.

Table 3 Major gene families in *O. piceae* (Op) and in three other ascomycetes

Gene family	Op	Gc*	Nc*	Tr*
MFS transporters	289	227	161	236
ABC transporters	34	40	36	48
ATPases	308	349	356	352
NAD binding proteins	258	254	211	301
FAD binding proteins	130	146	122	144
Cytochrome P450s	45	54	43	73
Methyltransferases	112	159	126	125
Transcription factors	115	133	106	218
Glycoside hydrolases	140	126	168	170
Glycosyl transferases	63	64	76	79

G. clavigera (Gc); *Neurospora crassa* (Nc) and *Trichoderma reesei* (Tr).

within this class. A BLAST comparison of the two predicted proteomes showed that 5,450 proteins were reciprocal best hits. These included most of the major metabolic functions. The *O. piceae* proteins with no significant homolog in the *G. clavigera* genome were overrepresented by protein kinases (GO:0004672), sequence-specific DNA binding RNA polymerase II transcription factors (GO: 0000981) and zinc ion binding proteins (GO:0008270) (Additional file 2). In addition, proteins involved in transmembrane transport (GO:0055085) were also significantly overrepresented in this group of 3,469 proteins (Additional file 2). Over 40% (1,397) of the *O. piceae* proteins with no evident homologs in *G. clavigera* were proteins of unknown function (predicted or hypothetical proteins). None of the six carboxylic ester hydrolases (GO:0052689) in the *O. piceae* genome had a homolog in the *G. clavigera* genome.

We searched for genes that may be involved in producing secondary metabolites (SMs). Such genes are typically organized as contiguous genomic clusters and can be identified by tools like SMURF [34], which uses hidden Markov models that consider genomic context and domain content. The first step in fungal SM biosynthesis is usually catalyzed by 'backbone' genes like nonribosomal peptide synthases (NRPSs), polyketide synthases (PKSs), hybrid NRPS-PKS enzymes, prenyltransferases, and terpene cyclases [34]. SMURF, which does not identify clusters containing terpene cyclases, identified thirteen backbone genes of which eleven are in SM clusters in *O. piceae* (Additional file 3), and nineteen genes in fourteen clusters in *G. clavigera* [35]. All the *O. piceae* SM genes have homologs in *G. clavigera*.

Melanin is a secondary metabolite that is produced by *O. piceae* and related species, but, as in *O. piceae*, the genes responsible for its production do not always occur in a cluster. Melanin is synthesized through the 1,8–dihydroxynaphthalene (DHN) pathway [36]. In *O. piceae*

we identified a number of genes that were similar to genes that have major roles in the DHN pathway in *Ophiostoma*, *Grosmannia* and *Ceratocystis* species [12,37,38]. These genes included a PKS (OPP_00823), two reductases (OPP_02710, OPP_00820) and a scytalone dehydratase (OPP_07153). PKS catalyze both the elongation of five ketide subunits and the cyclization of these units to form the base ring of naphthalene. The first reductase (OPP_02710) converts 1,3,6,8-hydroxynaphthalene to scytalone, while the second (OPP_00820) transforms scytalone to vermelone.

O. piceae is a heterothallic species, so requires two individuals with different mating types for sexual reproduction and production of fertile fruiting bodies. Our genome annotation identified *O. piceae*'s MAT1-2 idiomorph (OPP_06680). A truncated MAT1-1 gene was next to the MAT1-2 gene, as in *Grosmannia* and related species [8,39,40]. We have produced perithecia for *O. piceae* by mating UAMH-11346 with UAMH-11672 (Figure 2), and we have successfully amplified and sequenced the full length of the MAT1-1 loci in this latter strain; the alpha box (~978 bp) was missing in the sequenced strain.

Gene expression patterns

To identify genes that may be critical for the saprophyte *O. piceae* to grow in the presence of the nutrients and defence chemicals that are characteristic of its natural pine sapwood substrate, we profiled gene expression for the fungus growing on solid agar media supplemented with simple carbon sources (i.e. sugars and lipids), with pine sawdust, or with pine terpenes (see Methods). Mapping the RNA-seq reads to the predicted gene models identified 7,157 genes that had an abundance of at least 10 FPKM (fragments per kilobase of exon per million fragments mapped) in any of the conditions tested.

To select genes that were highly differentially regulated under different growth conditions, we required a gene to have an FPKM abundance that was at least ten times higher in a specific condition, or a related set of conditions, than in all other growth conditions. This approach identified 677 genes whose transcripts were differentially abundant in at least one growth condition (Additional file 4), and 173 genes whose transcripts were differentially abundant in only one condition (Figure 3). By manually comparing the set of 173 genes to functional information in the Gene Ontology (http://www.geneontology.org/) and to reference metabolic pathways KEGG (http://www.genome.jp), we identified pathways that were likely involved in the response of *O. piceae* to the growth conditions tested in our experiments. We added support for these pathways by manually identifying genes from the 677-gene set (Additional file 4) whose transcripts, while up-regulated, did not pass the stringent 10-fold filter used to identify the set of 173 genes.

We also assessed alternative transcript splicing across the range of growth conditions used for this study (Additional file 5). Splicing appeared not to be an important factor under these conditions.

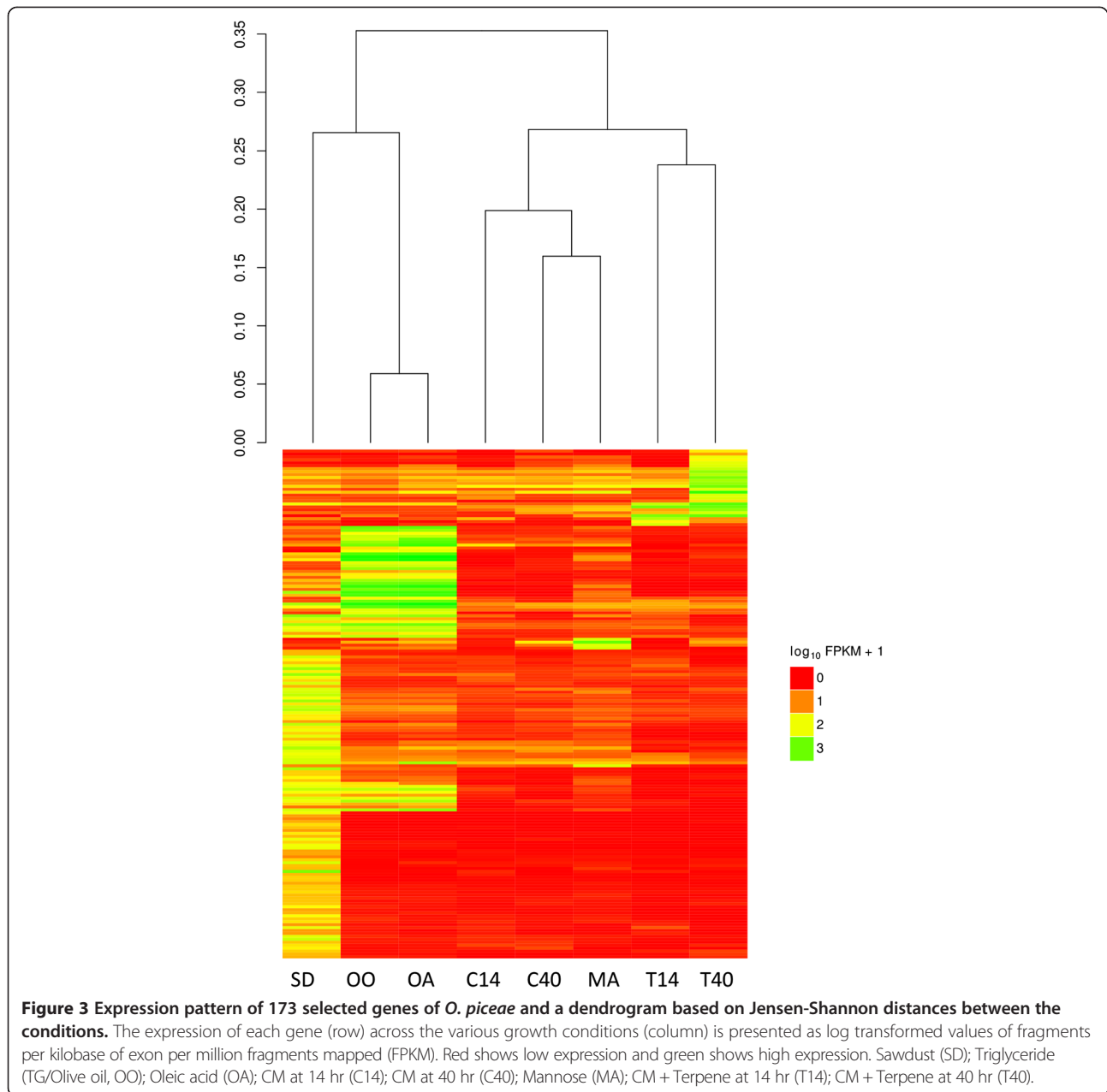### Growth on mannose, fatty acid (oleic acid) and triglycerides

Mannose is a simple monomeric epimer of glucose and can be readily utilized as a carbon source by *O. piceae*. We found five genes whose expression was at least ten times higher with mannose than in any other conditions tested (Additional file 4 shown light blue). These included two transporters, one oxidoreductase and two hypothetical proteins. Our data suggests that mannose uptake involves two transporters (OPP_03031, OPP_05665), and a simple isomerisation/ epimerization reaction by an oxidoreductase (OPP_00733) converts it into glucose. The function of two remaining up-regulated genes (OPP_02416, OPP_07274) is unknown.

We grew *O. piceae* on triglycerides and fatty acids, which are important lipid compounds in lodgepole pine sapwood, and are a major source of carbon for *O. piceae* [41]. Because most sources of triglycerides contain a small proportion of fatty acids, it was not surprising that most of the 129 genes whose transcripts were differentially abundant between these conditions were highly up-regulated in both of the conditions. Of the 25 up-regulated genes that were significantly induced only in these two conditions (shown in gold in Additional file 4), 18 were predicted to produce secreted proteins. Unexpectedly, the differentially up-regulated genes included no fungal lipases, which are necessary for the hydrolysis of triglycerides (Additional file 4). Twenty-three of the 25 up-regulated genes were predicted to be involved in the breakdown of carbohydrates and sugars; these included eight genes coding for secreted proteins in the glycoside hydrolase family and four genes for secreted proteins involved in carbohydrate and starch binding. We identified a transcription factor (OPP_02429) that showed significant up-regulation in the presence of triglycerides and oleic acid.

One of the genes differentially expressed between olive oil and oleic acid was a cytochrome P450 (OPP_02426) with a significantly higher expression with triglyceride than with fatty acid. Like its *G. clavigera* homolog (CMQ_5365; CYP630B18) and homologs in several other species including *Fusarium graminearum*, *Aspergillus niger*, *A. fumigates* and others, this gene is in close proximity to genes encoding a myo-inositol transporter, ARCA-like protein and a cytochrome P450 reductase [35].

Growth on pine sapwood, a natural substrate for *O. piceae*

O. piceae has slower growth rates than *G. clavigera* on rich media and on wood; for example, here, on MEA, *O. piceae* grows 5.8 mm/day, while in Wang et al. [21], *G. clavigera*

Figure 3 Expression pattern of 173 selected genes of *O. piceae* and a dendrogram based on Jensen-Shannon distances between the conditions. The expression of each gene (row) across the various growth conditions (column) is presented as log transformed values of fragments per kilobase of exon per million fragments mapped (FPKM). Red shows low expression and green shows high expression. Sawdust (SD); Triglyceride (TG/Olive oil, OO); Oleic acid (OA); CM at 14 hr (C14); CM at 40 hr (C40); Mannose (MA); CM + Terpene at 14 hr (T14); CM + Terpene at 40 hr (T40).

grew 14 mm/day. Of the treatments used in this study, sawdust obtained by grinding pine sapwood was the closest to the natural substrate. It contains a variety of carbon sources including mannose, triglycerides and fatty acids. In this growth condition, 366 genes were up-regulated, 91 of which were up-regulated only in the presence of sawdust (Additional file 4). The subset of 91 genes was overrepresented in GO terms for transport (GO:0005215, GO:0006810; $p < 0.0001$) (Figure 4), which may reflect the complexity of the nutrient sources used by *O. piceae*. The up-regulated transporters included several allantoate, urea, hexose, iron and sugar transporters,

and other major facilitator superfamily (MFS) transporters. As well, oxidoreductase genes that code for putative proteins involved in the modification of aromatic compounds, including phenolics, were highly up-regulated (e.g. P450s, dehydrogenases).

Among the 91 genes that were up-regulated on sawdust, 32 were found in 8 genomic clusters, each of which contained four to seven genes that may be co-regulated (Table 4). Three of the clusters contain a fungal-specific transcription factor, Zn2cys, which may be involved in primary and secondary metabolism and drug resistance [42]. Four of the clusters contain at least one of the
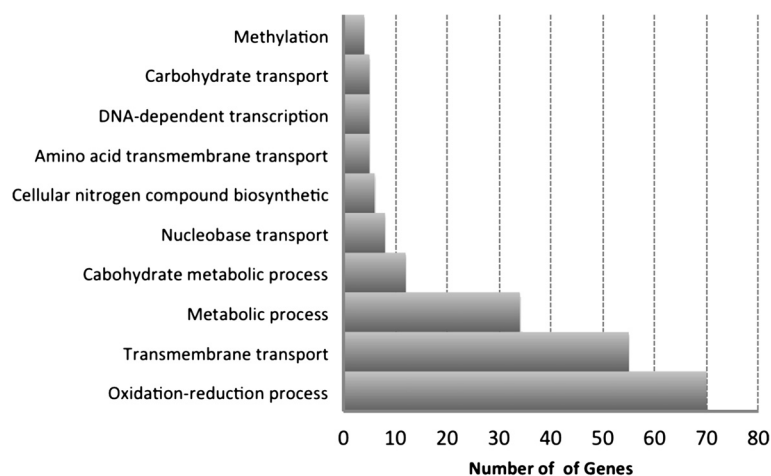
Figure 4 Functional classification of up-regulated genes of *O. piceae* grown on sawdust using Blast2go. Potential biological roles of transcribed proteins during growth experiments on different media were inferred by first identifying genes that were exclusively up-regulated (e.g. in sawdust), then associating the encoded protein sequences with biological processes using Blast2go.

following: a putative secreted salicylate dehydroxylase, an NAD-dependant epimerase, an alpha-mannosyltransferanse and an FAD–binding protein. An additional 22 genes that were up-regulated with sawdust were also up-regulated with triglycerides and oleic acid (dark blue in Additional file 4). The overall set of 113 genes (i.e. the 91 and the 22) was overrepresented in GO terms for secreted proteins (GO:0005576; p < 0.001) and carbohydrate metabolism (GO:0005975, GO:0030246; p < 0.001).

One of the above eight up-regulated genomic clusters contained seven genes that were involved in metabolizing quinic acid (OPP_8732 to OPP_8738). These include a quinate permease, two regulatory genes (an activator and a repressor), and the four genes of the quinate/shikamate catabolic pathway [43,44]. The latter four catabolic genes (OPP_08735 to OPP_08738) suggest that *O. piceae* uses quinic acid in wood as a carbon source. While this gene cluster has been reported in many fungi, we were unable to find the cluster in *G. clavigera*. To confirm that *O. piceae* can use quinic acid, while *G. clavigera* cannot, we showed that the former, but not the latter, grows on YNB media with quinic acid as the sole carbon source (Additional file 6). Finally, we identified a secreted lipase (OPP_00605) with predicted triglyceride degradation activity, whose transcript abundance was at least 50-fold higher in the mycelium grown in pine sapwood than in the control mannose.

Tolerance of pine tree defence chemicals

O. piceae did not grow when a mixture of monoterpenes (MT) (i.e. (+)-limonene, (+)-3-carene, racemic a-pinene and (−)-ß-pinene at a ratio of 5:3:1:1) was the only carbon source in YNB. However, after a month of incu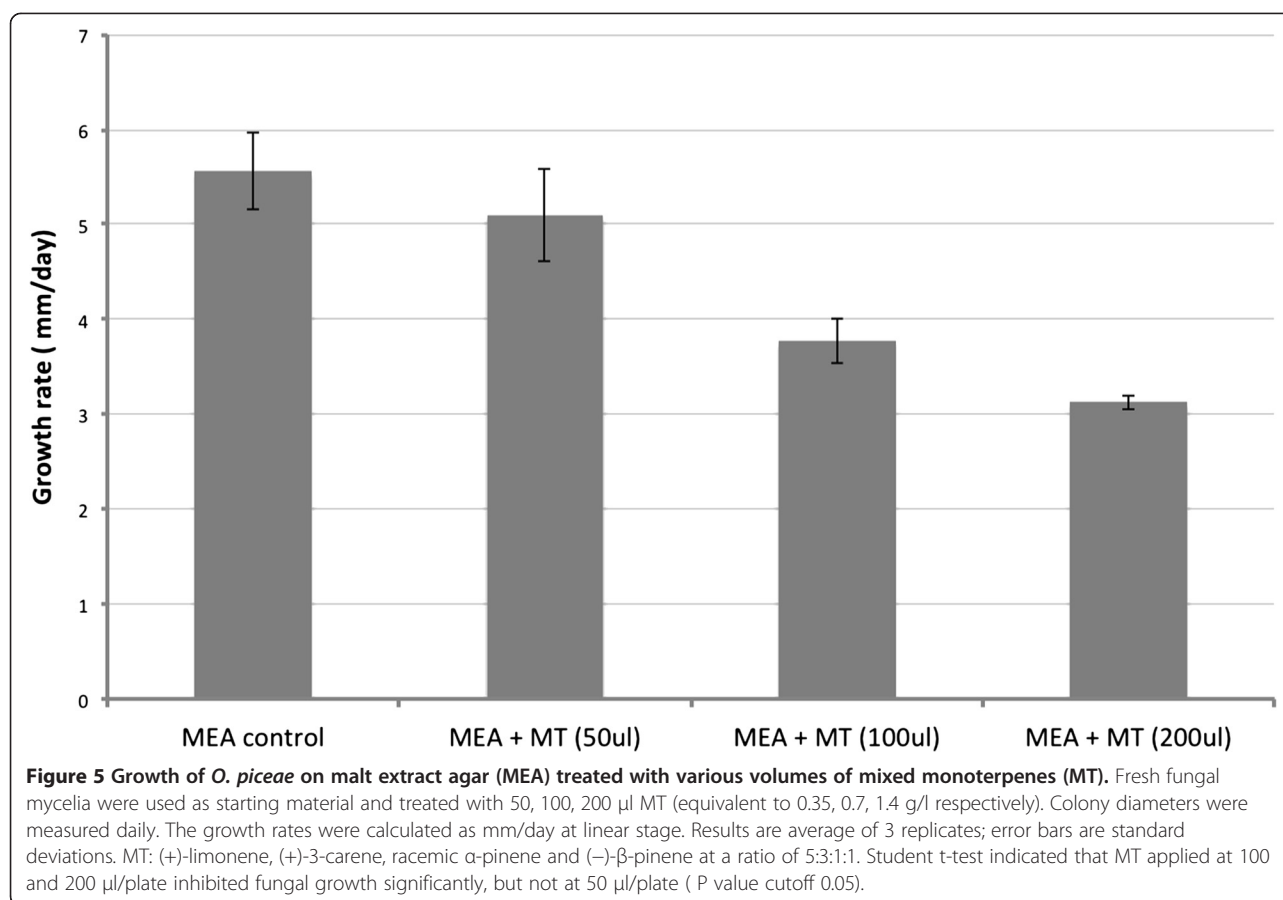bation in the presence of MT, the inocula resumed normal growth when they were transferred from YNB + MT to MEA. This suggests that *O. piceae* is able to survive in the presence of very high levels of monoterpenes. When the fungus was inoculated on MEA and treated with different amounts of MT, the growth rate was only significantly affected when at least 100 µl /plate (~ 0.7 g/l) of MT was added (Figure 5). For all MT treatments, the mycelia were more aerial and fluffy, while the asexual reproduction structures (i.e. formation of synemata) were highly inhibited (Additional file 7).

In order to identify genes involved in terpene tolerance, we grew *O. piceae* on CM and treated it with a mixture of terpenes as previously described in our studies with *G. clavigera* [8,21]. While the experiments for the two species were done at different times, we used the same conditions for both growth experiments and the same protocols for RNA extractions. We compared gene expression profiles of *O. piceae* after 14 h and 40 h treatments to profiles for untreated CM plates at the same time points. At 14 h, 295 genes were differentially abundant, 261 of which were down-regulated. While carbohydrate metabolism (GO:0005975) was associated with the down-regulated genes (p < 0.001), we were unable to identify any GO terms that were associated with the 34 up-regulated genes. After 40 h in the presence of terpenes, 264 genes were differentially abundant, 126 of which were up-regulated. While carbohydrate metabolism was still associated with down-regulated genes (p < 0.001), several transporters (OPP_06758, OPP_05515, OPP_03974, OPP_02103) were significantly up-regulated. In *G. clavigera*, which is able to utilize terpenes as a carbon source, more than 250 genes were up-regulated by at least 2-fold at 12 h and 36 h in the presence of terpenes [8]. Of the 34 *O. piceae* genes that were up-regulated

Table 4 Gene clusters up-regulated in sawdust

Gene IDs	Putative function	Secreted	Log$_2$(FC)*
Cluster 1			
OPP_08738	Inositol monophosphatase	No	3.98
OPP_08737	Catabolic 3-dehydroquinase	No	5.78
OPP_08736	3-dehydroshikimate dehydratase	No	7.20
OPP_08735	Quinate dehydrogenase	No	5.65
Cluster 2			
OPP_06948	Allantoate permease	No	8.05
OPP_06946	Sarcosine oxidase	No	5.97
OPP_06944	Fungal-specific transcription factor domain protein	No	4.50
OPP_06943	Oxoglutarate 3-dioxygenase	No	8.75
Cluster 3			
OPP_07708	Sugar transporter	No	6.06
OPP_07707	Salicylate hydroxylase (salicylate 1-monooxygenase)	Yes	7.75
OPP_07706	NAD dependent epimerase	Yes	3.62
OPP_07705	Arylacetamide deacetylase	No	4.69
Cluster 4			
OPP_08830	Amidohydrolase family protein	No	8.98
OPP_08829	Aldehyde dehydrogenase	No	7.04
OPP_08827	FAD binding domain protein	Yes	6.25
OPP_08826	Retinol dehydrogenase 13	No	7.78
OPP_08825	Cytochrome p450	No	9.25
OPP_08824	General alpha-glucoside permease	No	7.98
Cluster 5			
OPP_07998	Xaa-pro dipeptidase	No	3.34
OPP_07997	Major facilitator superfamily transporter	No	4.21
OPP_07996	Hexose transporter	No	5.56
OPP_07995	Thymine dioxygenase	No	4.95
Cluster 6			
OPP_01495	N-carbamoyl-l-amino acid hydrolase	No	9.85
OPP_01494	Gal4-like transcription factor	No	6.74
OPP_01493	Class ii aldolase adducin domain-containing protein	No	9.97
OPP_01491	Isoflavone reductase family protein	Yes	6.49
Cluster 7			
OPP_05544	Hypothetical protein	No	10.82
OPP_05543	Alpha-mannosyltransferase	Yes	8.083
OPP_05542	Ethanolamine utilization protein	No	7.16
OPP_05541	C6 zinc finger domain containing protein	No	4.99
OPP_05540	Alpha-mannosyltransferase	No	8.78
Cluster 8			
OPP_02428	Myo-inositol transporter	No	7.42
OPP_02427	Arca-like protein	No	3.81
OPP_02426	Benzoate 4-monooxygenase cytochrome p450	No	4.45
OPP_02425	NADPH-cytochrome p450 reductase	No	5.12
OPP_02424	NAD binding rossmann fold	No	4.02

* FC = fold change.

Figure 5 Growth of *O. piceae* on malt extract agar (MEA) treated with various volumes of mixed monoterpenes (MT). Fresh fungal mycelia were used as starting material and treated with 50, 100, 200 μl MT (equivalent to 0.35, 0.7, 1.4 g/l respectively). Colony diameters were measured daily. The growth rates were calculated as mm/day at linear stage. Results are average of 3 replicates; error bars are standard deviations. MT: (+)-limonene, (+)-3-carene, racemic α-pinene and (−)-β-pinene at a ratio of 5:3:1:1. Student t-test indicated that MT applied at 100 and 200 μl/plate inhibited fungal growth significantly, but not at 50 μl/plate (P value cutoff 0.05).
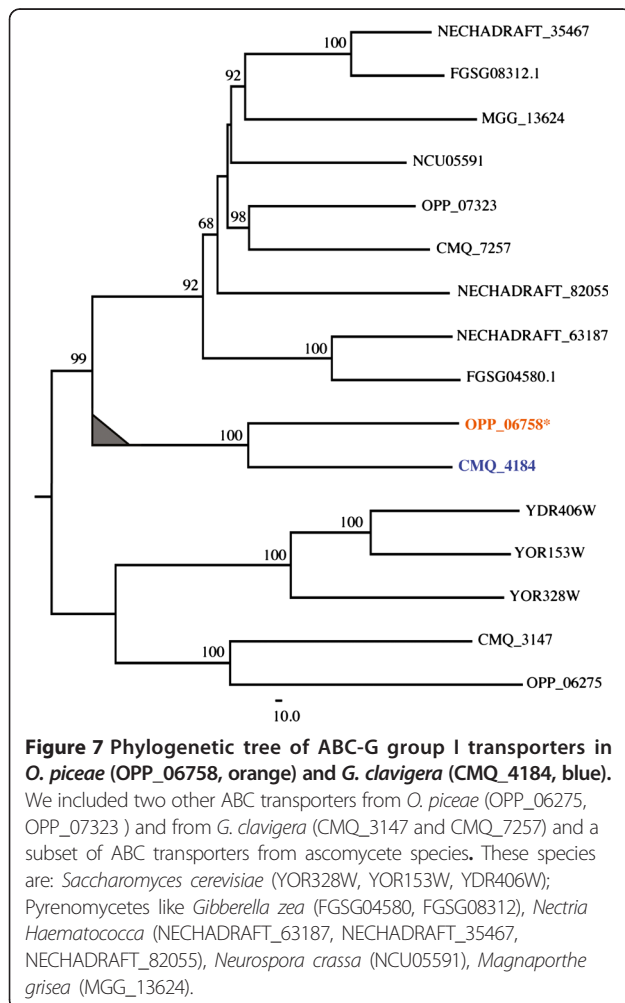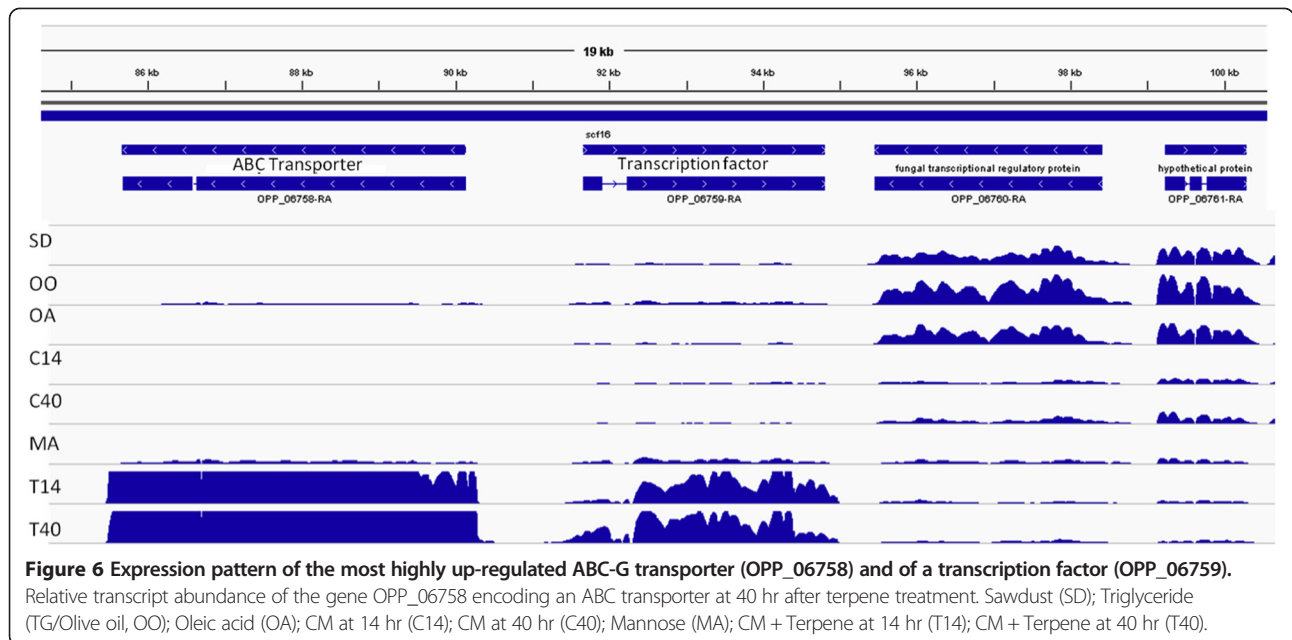
at 14 h, 26 had homologs in *G. clavigera*, and nine of these *G. clavigera* genes were up-regulated at 12 h. Similarly, of the 126 *O. piceae* genes that were up-regulated at 40 h, 75 had *G. clavigera* homologs, twenty of which were up-regulated at 36 h.

We found 26 *O. piceae* genes that were up-regulated only in the presence of terpenes, at one or both time points. Eighteen of these had *G. clavigera* homologs. The most highly up-regulated *O. piceae* gene (OPP_06758) encoded an ABC transporter that was homologous to the *G. clavigera* transporter (CMQ_4184; GcABC-G1) that confers terpene tolerance to the pathogen [21] (Figure 6). Approximately 1,500 bp upstream of the *O. piceae* ABC transporter is a gene (OPP_06759) encoding a transcription factor whose expression, like that of the transporter, was up-regulated only in the presence of terpenes (Figure 6). We found that the *O. piceae* OPP_06758 and the *G. clavigera* CMQ_4184 (GcABC-G1) ABC transporters were placed in the same clade when we constructed a phylogenetic tree for a subset of the fungal species recently analyzed by Wang et al. [21]. To our knowledge, from currently available sequence data, this clade is unique to these two fungal species (Figure 7); no similar fungal ABC transporter has been reported.

Discussion

The ecological niches of saprophytic and pathogenic wood-inhabiting filamentous fungi differ in moisture content, nutrients and defence chemicals. For such fungi to survive in and colonize their substrates and hosts, they need active transport systems that can excrete enzymes that break down complex external substrates and then import nutrients into the cells. As well, they need to modify or remove toxic host defense chemicals that have entered their cells. The wood of trees, logs and lumber has a wide range of moisture contents and a high carbon-to-nitrogen ratio [14]. *O. piceae* grows more efficiently in drier pine wood than in freshly cut logs; *G. clavigera*, which is vectored by MPB, colonizes healthy or stressed living pine trees, which have high moisture and low oxygen contents. Neither organism degrades lignocellulosic wood fibers [1,45]. *O. piceae* has to retrieve nutrients from a nutrient-poor substrate that typically contains very little nitrogen and relatively low levels of host defence chemicals. In contrast, *G. clavigera* has first to cope with high concentrations of defense chemicals produced by its pine host.

Recently, we reported the annotated genome sequence and transcriptomes of the pine pathogen *G. clavigera* [8].

Figure 6 Expression pattern of the most highly up-regulated ABC-G transporter (OPP_06758) and of a transcription factor (OPP_06759).
Relative transcript abundance of the gene OPP_06758 encoding an ABC transporter at 40 hr after terpene treatment. Sawdust (SD); Triglyceride
(TG/Olive oil, OO); Oleic acid (OA); CM at 14 hr (C14); CM at 40 hr (C40); Mannose (MA); CM + Terpene at 14 hr (T14); CM + Terpene at 40 hr (T40).

Figure 7 Phylogenetic tree of ABC-G group I transporters in
O. piceae **(OPP_06758, orange) and** ***G. clavigera*** **(CMQ_4184, blue).**
We included two other ABC transporters from *O. piceae* (OPP_06275,
OPP_07323) and from *G. clavigera* (CMQ_3147 and CMQ_7257) and a
subset of ABC transporters from ascomycete species**.** These species
are: *Saccharomyces cerevisiae* (YOR328W, YOR153W, YDR406W);
Pyrenomycetes like *Gibberella zea* (FGSG04580, FGSG08312), *Nectria
Haematococca* (NECHADRAFT_63187, NECHADRAFT_35467,
NECHADRAFT_82055), *Neurospora crassa* (NCU05591), *Magnaporthe
grisea* (MGG_13624).

Here, we report the annotated genome and transcriptomes
of the saprophyte *O. piceae*, the second pine wood-
inhabiting ophiostomatoid fungus for which a complete
genome has been sequenced. *O. piceae*'s genome size
and the number of predicted genes and proteins were
similar to those for *G. clavigera* and other sequenced
saprophytic ascomycetes in the class Sordariomycetes
(e.g. *N. crassa*, *T. reesei*). *O. piceae*'s predicted secretome is
10% larger than that of the pine-specific pathogen. Given
its more diverse range of host trees (e.g. pines, hemlock,
spruces), it is likely that the saprophyte requires more
extracellular enzymes to degrade the different chemicals
encountered in these substrates.

In both genomes we identified genes that are poten-
tially involved in the biosynthesis or processing of SMs.
In fungi, SMs are diverse and play a range of roles; some
SMs are protective, while others are virulence factors
[46]. Both *O. piceae* and *G. clavigera* produce the SM
melanin in artificial media and in their natural sub-
strates. Fungal melanin may protect cells in harsh envi-
ronments (e.g. UV radiation, extreme temperatures and
toxic compounds), and may be involved in cellular de-
velopment, differentiation and pathogenicity [36]. In all
conditions tested here, except with terpene treatments,
O. piceae mycelia and asexual structures (i.e. synemata)
were highly melanized. Scytalone dehydratase, which is
a marker gene for the DHN pathway [47], was up-
regulated in all conditions tested except with terpene
treatments, and was most highly expressed in sawdust.
Similarly, in *G. clavigera*, which is densely melanized in
its pine host, scytalone dehydratase was down-regulated
on CM with terpenes, but was up-regulated on other

media and when monoterpenes were the only carbon source. In contrast to *O. piceae*, *G. clavigera* does not produce large numbers of asexual spores when it is actively growing on these media. It is likely that melanin protects *O. piceae* from the unfavourable environmental conditions that it encounters in lumber (e.g. dessication, UV), as well as being involved in cellular development. In contrast, for *G. clavigera*, melanin may be more important in protecting the fungus from host defense chemicals.

O. piceae and *G. clavigera* can grow on a variety of simple sugars that are present in phloem or in sapwood parenchyma cells [13], and can acquire additional sugars by degrading wood hemicelluloses [13,14,45]. Both fungi grow well with mannose and maltose, and can also use starch, a stored tree nutrient [48]. For *O. piceae*, our data suggest that mannose uptake and the initial steps in its utilization are controlled by at least six genes that include two transporters. That none of the six were up-regulated with maltose suggests that maltose utilization involves an alternate pathway.

O. piceae and related species can use triglycerides and fatty acids in artificial media or wood; these lipids can account for up to 3% of the dry weight of sapwood [41]. Triglycerides are hydrolyzed by extracellular lipases into fatty acids and glycerol, which are ultimately processed through ß-oxidation and glycolysis pathways [47]. While lipase and esterase genes were present in the *O. piceae* genome and we noted that a lipase was expressed on sawdust, we were unable to detect up-regulated lipases on triglycerides. It is possible that on triglycerides the lipase was produced very early in growth, as shown by Gao and Breuil [49], who reported an optimum production of the enzyme at day 3, before the pH of the medium decreases due to the accumulation of fatty acids. Here, we collected the mycelium after seven days of growth on a solid media with triglycerides. We identified a glycerol kinase that was up-regulated for triglycerides and sawdust, which suggests that glycerol may be metabolized by the fungus. Further, we noticed that triglycerides induced a genomic cluster that contained a P450 and a reductase (described in Results). Lah *et al.* [35] reported a similar genomic cluster organization and expression pattern in *G. clavigera*, and it is likely that the clusters have similar roles. Lah *et al.* suggested that the cytochrome P450 and the reductase may be specific redox partners and may play a role in the conversion of exogenous phenolics or fatty acids.

O. piceae retrieves and metabolizes diverse nutrients that are present in low concentrations in sawdust, particularly nitrogen sources, while removing or modifying diverse toxic compounds like terpenes, and aromatic compounds that include simple phenolics. While the fungus grows more slowly on sawdust, diverse transporters were up-regulated. Some of these are involved

in acquiring nutrients like sugars and nitrogen, while others, like ABC or MFS transporters, are known to contribute to drug resistance or chemical modification or detoxification [50].

While small amounts of simple sugars are available in sapwood, *O. piceae* can retrieve additional sugars by degrading pine hemicellulose [13,45]. Fleet et al. [13] reported that mannose was the most depleted sugar in logs and lumber inoculated with *Ophiostoma* species. In our data, the genes up-regulated on sawdust also included glycoside hydrolases (e.g. two xylanases and one pectinase), which are involved in degrading hemicellulose and pectin. As well, the fungus can retrieve quinic acid through a quinate permease, and can utilize this carbon source by processing it through the quinate/shikamate pathway, which was up-regulated on sawdust. Further, in artificial media *O. piceae* can readily use inorganic or organic nitrogen. However, in pine sapwood nitrogen is found mainly as amino acids and proteins, and at very low concentrations (~0.05% of the wood dry weight) [51]. We have shown that *O. piceae* and related species have to produce proteases in order to retrieve organic nitrogen from wood [52]. In the current work, an amino acid permease, and urea and ammonium transporters were up-regulated on sawdust. Urea can be used as a source of nitrogen by many fungi, and it can be efficiently converted into ammonium by a urease enzyme [53]. However, while ammonium is present in trace amounts in pine lumber [53,54], urea has not been reported in wood.

Mono- and diterpenes are well known biocides for microorganisms, including fungi, and for beetle vectors [21,55]. Our data show that on artificial media *O. piceae* tolerates monoterpenes but does not use them as a carbon source. It is not found in living trees, which have the highest terpene concentrations. However, it is able to remain viable for extended periods in the presence of monoterpenes, and likely in the presence of diterpenes, which can account for ~0.4% of pine sapwood dry weight [41]. Here, we show that monoterpenes affected the macroscopic morphology of *O. piceae*'s mycelia, and inhibited its production of synemata and asexual spores. Further, in the saprophyte, monoterpene/diterpene treatments rapidly up-regulated expression of genes involved in oxidative and reductive processes, as well as transmembrane transport, suggesting that the fungus' primary response involves protecting itself from these chemicals. During these processes, an ABC transporter (OPP_06758), which is homologous to the *G. clavigera* efflux transporter characterized by Wang et al. [21], was highly expressed.

We have shown that this *G. clavigera* ABC-G transporter is expressed in young trees and that the transporter excretes monoterpenes [21]. As we have not yet demonstrated this function for the homologous gene in

O. piceae, at this time we can only suggest that this unique transporter may play a similar role in the saprophyte by allowing it to survive in toxic mixtures of terpenes. When *O. piceae* is treated with terpenes on rich media, there is an initial growth delay, after which the fungus resumes its growth. In this growth phase, while genes providing most of the primary protective biological functions were active, genes involved in degrading hydrophobic compounds were up-regulated. This suggests that, like *G. clavigera*, *O. piceae* may be able to modify terpenes into less toxic compounds. However, while *G. clavigera* has a gene cluster that specifically responds to terpenes and is potentially involved in metabolizing terpenes [8], in *O. piceae* we found no such gene cluster. Only five of the 30 genes in this *G. clavigera* cluster had homologs in *O. piceae*, and these five genes were dispersed through the *O. piceae* genome. In ongoing work we are characterizing *O. piceae* genes that are involved in modifying terpenes.

Conclusions

We compared the genomes of *O. piceae* and *G. clavigera*. While we found no large-scale synteny, the ecological niches of both fungi involve growing in pine wood, and both produce similar sets of diverse enzymes. Neither fungus produces a complete battery of cell-wall degrading enzymes, and neither affects the structural properties of wood. We began to clarify differences between the saprophyte and the pathogen, focusing on ABC transporters, CYP450s, genes that produce secondary metabolites like melanin, genes involved in lipid metabolism, and genes that detoxify terpenes and phenolics. *G. clavigera*, but not *O. piceae*, can use monoterpenes as a carbon source. However, both *O. piceae* and *G. clavigera* have a similar ABC-G transporter that, for both fungi, may play an important role in reducing the intracellular concentration of toxic compounds like monoterpenes. Similar specialized transporters may have evolved in other ophiostomatoid fungi that are vectored by insects and inhabit the phloem and sapwood of living or processed conifers.

Methods
Strain and growth conditions

The *O. piceae* strains used in this work, either for the genome sequencing or for the mating experiments, had been isolated from freshly sawn timber of *Pinus contorta* at Prince George in British Columbia (Canada) [6]. The strains have been deposited at the University of Alberta Microfungus Collection and Herbarium; ID: UAMH-11346 for the genome sequenced and ID: UMAH-11672 for the mating experiments or for checking the growth characteristics. For growth and maintenance, spores or plugs of fungal mycelium were inoculated and grown at room temperature on plates of MEA (1.6% Oxoid malt extract agar and 1.5% technical agar, pH 5–6). Growth and utilization of terpenes were performed as previously described by Wang et al. [21]. The only exception was for growth on sawdust where spores were inoculated and germinated on 1% MEA (Difco) for 2 days, and then transferred on sawdust plates (15% lodgepole pine sawdust, mixed with 2% granulated agar) overlaid with cellophane for one week. Growth experiments with or without terpenes or with different carbon sources were at least repeated or carried out three times. For RNA-seq analysis, fungal hyphae grown on MEA for three days were transferred to the respective treatment conditions as shown in Table 5.

Genome sequencing

DNA was extracted from fungal hyphae grown on MEA using methods described by Haridas and Gantt [56]. We used two sequencing technologies: Illumina HiSeq 2000, which generated 100 nt reads, and 454 Titanium, which generated reads with a mean length of 318 nt. The libraries for Illumina HiSeq 2000 had two different insert sizes, 200 and 700 nt, while the library for 454 had an insert size of 8000 nt. Illumina HiSeq sequencing was done at the BC Genome Sciences Centre in Vancouver, Canada and 454 sequencing was done at the Plate-forme d'Analyses Génomiques at Laval University in Québec, Canada.

Genome assembly

Reads generated by the two platforms were used with no further processing for genome assembly using ABySS v1.3.0 [25] with a kmer size of 60. This assembler filters the FASTQ sequences based on quality scores. In order to efficiently use the 454 reads for scaffolding, we used a minimum contig size (1000) and read pairs for building

Table 5 Growth conditions for RNA-seq

Medium	Carbon sources or treatment	Duration
CM	No treatment	14 h
CM	No treatment	40 h
CM	200 µl Terpene blend	14 h
CM	200 µl Terpene blend	40 h
YNB	Mannose (1% w/v)	5 days
YNB	*TG: Olive Oil (1% v/v)	5 days
YNB	Oleic acid (0.5% v/v)	5 days
Sawdust	No treatment	1 weeks

* Triglyceride: Olive oil 80% + Fatty acids 20%.
All treatment times were calculated from the initial transfer of actively growing hyphae onto the appropriate medium. CM = 0.17% yeast nitrogen base without amino acids, 1.5% granulated agar, 1% maltose, 0.1% potassium hydrogen phthalate, 0.3% asparagine. YNB = 0.67% yeast nitrogen base without amino acid and 1.5% agar with carbon sources indicated as treatment. The terpene blend was a mixture of monoterpenes and diterpenes as described by Lah et al. [35].

scaffolds (2) (SCAFFOLD_OPTIONS = '-s1000 -n2'). The assembly was scrubbed and gaps closed with Anchor (v0.3.0; www.bcgsc.ca/platform/bioinfo/software/anchor). When Abyss is unable to find overlaps between contigs where paired end data suggests that the contigs should overlap, it joins the contigs with a single lowercase 'n'. Such overlaps were resolved using transcriptome assembly (described below) or by finding small overlaps at the ends of the contigs using exonerate v2.2.0 [26]. Genome synteny was assessed by MUMmer [57].

Transcriptome assembly

RNA-seq was performed on eight RNA samples extracted from the mycelia of *O. piceae* hyphae grown under various conditions as shown in Table 5. For each RNA-seq library, we collected samples from three biological replicates, extracted RNA separately from each replicate, and pooled the samples for paired-end sequencing on an Illumina HiSeq (Canada's Michael Smith Genome Science Center, Vancouver). Multiplexed sequencing with three libraries per lane was done using the Illumina HiSeq platform to obtain 100 bp paired end reads from 250 bp fragments. Reads were analysed using fastqc (www.bioinformatics.babraham.ac.uk/projects/fastqc/) and showed read bias and in the first few bases of the reads and poor quality in the last few. Reads with minimum (Phred) quality scores less than 20 were removed and the first six and last four bases of all reads were trimmed using prinseq [58]. Processed RNA-seq reads were assembled using Trinity [28] using the jaccard_clip option to minimize fusion transcripts. The best protein coding transcripts were identified using the included scripts and aligned back to the assembled genome using exonerate v2.2.0 est2genome [26].

Genome annotation

We used the Maker annotation pipeline (v2.26) for genome annotation [31]. In addition to the trinity assembled best candidates, we also used two additional sources of evidence in Maker. The first was transcripts predicted by the Core Eukaryotic Genes Mapping Approach [27], (CEGMA) and the second was coding sequences of transcripts assembled by cufflinks [59] from RNA-seq reads mapped to the assembled genome. Within the Maker framework, we trained SNAP v2006-07-28 [60] using the Trinity assembled transcripts, gene models of *Magnaporthe grisea* for Augustus (v2.5.5) [61] and an hmm file for Genemark-ES (v2.3) [62] using an independent run. The UniProtKB/Swiss-Prot (release 2012_01) fasta file was provided as protein homology evidence and pred_flank was set to 50 to minimize fusion transcripts. Predicted genes smaller than 100 amino acids were removed unless they were at least 80 amino acids long and had transcript, protein or CEGMA evidence. Selected gene models were manually curated. Functional

identification of predicted genes was done using Blast2go (v2.5.1) [63]. tRNA's were identified using tRNAscan-SE (v1.3.1) [64]. Relative synonymous codon usage (RSCU) was calculated using a local installation of the graphical codon usage analyser [65]. Secretome predictions were made with TargetP [66] and Phobius [67]. A protein was considered to be secreted if either TargetP or Phobius suggested that it was secreted and this result was not in conflict with the other. Identification of secondary metabolism genes and clusters was done using the Secondary Metabolite Unique Regions Finder (SMURF) [34].

RNA-Seq analysis

Quality trimmed RNA-seq reads were aligned to the *O. piceae* genome using Bowtie (v0.12.7), Tophat (v2.0.4), Cufflinks (v2.0.2) as described by Trapnell et al. [59]. Because mapping the RNA-seq reads to the genome without providing fixed gene models resulted in an unacceptable number of predicted fusion transcripts, reads were mapped using the curated gene models predicted by the Maker pipeline.

Data availability

The sequences reported in this paper are being deposited in NCBI gene bank as assembly and annotations, Project NO. PRJNA182071.

Additional files

Additional file 1: Comparison of the genomes of *O. piceae* and *G. clavigera*, using MUMmer [57]. A dot plot genomic comparison showed no large scale synteny between the assembled genomes of *O. piceae* and *G. clavigera*. The genome of *O. piceae* is represented on the X-axis and that of *G. clavigera* on the Y-axis.

Additional file 2: GO term enrichment in a set of 3,469 genes of *O. piceae* which were not reciprocal best blast hits in the *G. clavigera* genome. #Test represents the number of genes annotated with the respective GO term in the set of 3,469 genes and #Ref represents the number of genes with the GO classification in the entire *O. piceae* genome. P-value is calculated for Fisher's exact test and FDR is the false discovery rate.

Additional file 3: The eleven secondary metabolism clusters in *O. piceae* predicted by SMURF [34]. The backbone gene in each cluster is highlighted in yellow.

Additional file 4: Expression patterns of the 677 significantly differentially regulated genes. Expression values measured as FPKM (fragments per kilobase of exon per million fragments mapped). Genes that showed a 10 times or greater up-regulation in one condition or a related set of conditions are colour coded as described.

Additional file 5: Alternative splicing. To assess how important alternative splicing and transcripts were, we used Tophat and Cufflinks to map the RNA-seq reads to the genome assembly using the techniques described by Trapnell et al. The results suggested that approximately 150 alternative transcripts were expressed; however, all of these appeared to be false positives. The dominant cause of these false positive predictions was that closely spaced genes with overlapping UTRs were misassembled as single contigs, and differential regulation of such genes under different growth conditions appeared as alterative isoforms. In other cases, mapping errors produced false gene calls and alternative isoforms. These results indicated that alter415native splicing is not important under the growth conditions used for this study.

> **Additional file 6: Growth of *O.piceae* and *G. clavigera* on mannose, oleic acid and quinic acid.** Fungal plugs were inoculated onto YNB plates (pH ~ 7, adjusted by KH_2PO_3-K_2HPO_3 buffer) containing a single carbon source; the plates with the fungus were incubated for 2 weeks. The growth of *O. piceae* is slower than *G. clavigera*, as shown with mannose. Control: YNB with no carbon.
>
> **Additional file 7: Growth of *O. piceae* on MEA with or without the addition of monoterpenes for a week. A**) Growth on MEA alone (arrow a: synemata and spores, arrow b: mycelium), **B**) Growth on MEA with monoterpenes, the mycelium was more aerial and fluffy (arrow c) while the production of asexual structures was highly inhibited.

Abbreviations

Mbp: Mega base pair; MPB: Mountain pine beetle; nt: Nucleotide; bp: Base pair; UTR: Untranslated region; CDS: Coding sequence; SM: Secondary metabolite; NRPS: Nonribosomal peptide synthase; PKS: Polyketide synthase; DHN: 1,8-dihydroxynaphthalene; FPKM: Fragments per kilobase of exon per million fragments mapped; MFS: Major facilitator superfamily; YNB: Yeast nitrogen base; MT: Monoterpene; MEA: Malt extract agar; CM: Complete medium; UV: Ultra violet; CEGMA: Core Eukaryotic Gene Mapping Approach; RSCU: Relative synonymous codon usage; SMURF: Secondary Metabolite Unique Regions Finder.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SJ and RD performed the initial genome assembly, and SH the final assembly. SH annotated gene models, analyzed transcriptomes, and drafted the results for the manuscript. YW participated in and finalized the transcriptome analyses. LL prepared DNA, mRNA, performed growth experiments and participated in the analyses. SMA did the phylogenetic trees and micrographs. IB supervised the sequencing of the genome and transcriptomes. GR, JB and CB wrote and completed the manuscript. CB and JB conceived and directed the project. All co-authors critically reviewed, edited and approved the manuscript.

Author details

[1]Department of Wood Science, University of British Columbia, Vancouver, BC V6T1Z4, Canada. [2]Canada's Michael Smith Genome Sciences Centre, Vancouver, BC V5Z 4E6, Canada. [3]Michael Smith Laboratories, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

References

1. Seifert K: **Sapstain of commercial lumber by species of Ophiostoma and Ceratocystis**. In *Ceratocystis and Ophiostoma: taxonomy, ecology, and pathogenicity*. Edited by Wingfield M, Seifert K, Webber J. St Paul, Minnesota: APS Press; 1993:141–151.
2. Harrington TC: **Diseases of conifers caused by species of *Ophiostoma* and *Leptographium***. In *Ceratocystis and Ophiostoma: taxonomy, ecology, and pathogenicity*. Edited by Wingfield M, Seifert K, Webber J. St Paul, Minnesota: APS Press; 1993:161–172.
3. Münch E: **Die bläufaule des nadelholzes. I-II**. *Naturwiss Z Forst-landw* 1907, **5**:531–573.
4. Upadhyay H: **Classification of the Ophiostomatoid fungi**. In *Ceratocystis and Ophiostoma: taxonomy, ecology, and pathogenicity*. Edited by Wingfield M, Seifert K, Webber J. St Paul, Minnesota: APS Press; 1993:7–13.
5. Krokene P, Solheim H: **Pathogenicity of four blue-stain fungi associated with aggressive and nonaggressive bark beetles**. *Phytopathology* 1998, **88**(1):39–44.
6. Uzunovic A, Yang D, Gagne P, Breuil C, Bernier L, Byrne A, Gignac M, Kim SH: **Fungi that cause sapstain in Canadian softwoods**. *Can J Microbiol* 1999, **45**(11):914–922.
7. Kurz WA, Dymond CC, Stinson G, Rampley GJ, Neilson ET, Carroll AL, Ebata T, Safranyik L: **Mountain pine beetle and forest carbon feedback to climate change**. *Nature* 2008, **452**(7190):987–990.
8. DiGuistini S, Wang Y, Liao NY, Taylor G, Tanguay P, Feau N, Henrissat B, Chan SK, Hesse-Orce U, Alamouti SM, Tsui CKM, Docking RT, Levasseur A, Haridas S, Robertson G, Birol I, Holt RA, Marra MA, Hamelin RC, Hirst M, Jones SJM, Bohlmann J, Breuil C: **Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen**. *Proc Natl Acad Sci USA* 2011, **108**(6):2504–2509.
9. Felsenstein J: **Confidence-limits on phylogenies - an approach using the bootstrap**. *Evolution* 1985, **39**(4):783–791.
10. Harrington TC: **Ecology and evolution of mycophagous bark beetles and their fungal partners**. In *Ecology and Evolutionary Advances in Insect-Fungal Associations*. Edited by Vega F, Blackwell M. United Kingdom: Oxford University Press; 2005:251–291.
11. Chung WH, Kim JJ, Yamaoka Y, Uzunovic A, Masuya H, Breuil C: ***Ophiostoma breviusculum* sp. *nov*. (Ophiostomatales, Ascomycota) is a new species in the *Ophiostoma piceae* complex associated with bark beetles infesting larch in Japan**. *Mycologia* 2006, **98**(5):801–814.
12. DiGuistini S, Ralph SG, Lim YW, Holt R, Jones S, Bohlmann J, Breuil C: **Generation and annotation of lodgepole pine and oleoresin-induced expressed sequences from the blue-stain fungus *Ophiostoma clavigerum*, a Mountain pine beetle-associated pathogen**. *FEMS Microbiol Lett* 2007, **267**(2):151–158.
13. Fleet C, Breuil C, Uzunovic A: **Nutrient consumption and pigmentation of deep and surface colonizing sapstaining fungi in Pinus contorta**. *Holzforschung* 2001, **55**(4):340–346.
14. Zabel R, Morrell J: **Wood stains and discolorations**. In *Wood Microbiology: decay and its prevention*. Edited by Zabel R, Morrell J. San Diego, California: Academic Press, Inc; 1992:326–343.
15. Behrendt C, Blanchette R, Farrell R: **Biological-control of blue-stain fungi in wood**. *Phytopathology* 1995, **85**(1):92–97.
16. Uzunovic A, Byrne T, Gignac M, Yang D: **Microbial discolorations**. In *Wood Discolorations and their Prevention; with an Emphasis on Bluestain*. Edited by Uzunovic A, Byrne T, Gignac M, Yang D. British Columbia: FPInnovations; 2008:16–41. Special publication AP-50 ISSN#0824-2119.
17. Scheffer T: **Microbiological degradation**. In *Wood Deterioration and its Prevention by Preservative Treatments*. Edited by Nicholas D. NY: Syracuse University Press; 1973:31–106.
18. Franceschi VR, Krokene P, Christiansen E, Krekling T: **Anatomical and chemical defenses of conifer bark against bark beetles and other pests**. *New Phytol* 2005, **167**(2):353–375.
19. Keeling CI, Bohlmann J: **Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens**. *New Phytol* 2006, **170**(4):657–675.
20. Bohlmann J: **Pine terpenoid defences in the mountain pine beetle epidemic and in other conifer pest interactions: specialized enemies are eating holes into a diverse, dynamic and durable defence system**. *Tree Physiol* 2012, **32**(8):943–945.
21. Wang Y, Lim L, DiGuistini S, Robertson G, Bohlmann J, Breuil C: **A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees**. *New Phytol* 2013, **197**(3):886–898.
22. Clark EL, Carroll AL, Huber DPW: **Differences in the constitutive terpene profile of lodgepole pine across a geographical range in British Columbia, and correlation with historical attack by mountain pine beetle**. *Can Entomol* 2010, **142**(6):557–573.
23. Turtola S, Manninen A, Holopainen J, Levula T, Raitio H, Kainulainen P: **Secondary metabolite concentrations and terpene emissions of Scots pine xylem after long-term forest fertilization RID B-1656-2008**. *J Environ Qual* 2002, **31**(5):1694–1701.
24. Eberhardt TL, Sheridan PM, Mahfouz JM: **Monoterpene persistence in the sapwood and heartwood of longleaf pine stumps: assessment of differences in composition and stability under field conditions**. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* 2009, **39**(7):1357–1365.

25. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I: **ABySS: a parallel assembler for short read sequence data.** *Genome Res* 2009, **19**(6):1117–1123.

26. Slater GS, Birney E: **Automated generation of heuristics for biological sequence comparison.** *BMC Bioinforma* 2005, **6**:31.

27. Parra G, Bradnam K, Korf I: **CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes.** *Bioinformatics* 2007, **23**(9):1061–1067.

28. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotechnol* 2011, **29**(7):644–652.

29. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, *et al*: **The genome sequence of the filamentous fungus *Neurospora crassa*.** *Nature* 2003, **422**(6934):859–868.

30. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, *et al*: **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*).** *Nat Biotechnol* 2008, **26**(5):553–560.

31. Holt C, Yandell M: **MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects.** *BMC Bioinforma* 2011, **12**:491.

32. Massoumi Alamouti S, Tsui CK, Breuil C: **Multigene phylogeny of filamentous ambrosia fungi associated with ambrosia and bark beetles.** *Mycol Res* 2009, **113**(Pt 8):822–835.

33. Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F, Hesse CN, Kosti I, LaButti K, Lindquist EA, Lucas S, Salamov AA, Bradshaw RE, Ciuffetti L, Hamelin RC, Kema GHJ, Lawrence C, Scott JA, Spatafora JW, Turgeon BG, de Wit PJGM, Zhong S, Goodwin SB, Grigoriev IV: **Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen dothideomycetes fungi.** *PLoS Pathog* 2012, **8**(12):e1003037.

34. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND: **SMURF: genomic mapping of fungal secondary metabolite clusters.** *Fungal Genet Biol* 2010, **47**(9):736–741.

35. Lah L, Haridas S, Bohlmann J, Breuil C: **The cytochromes P450 of *Grosmannia clavigera*: Genome organization, phylogeny, and expression in response to pine host chemicals.** *Fungal Genet Biol* 2013, **50**:72–81.

36. Butler MJ, Day AW: **Fungal melanins: a review.** *Can J Microbiol* 1998, **44**(12):1115–1136.

37. Wang H, Breuil C: **A second reductase gene involved in melanin biosynthesis in the sap-staining fungus *Ophiostoma floccosum*.** *Mol Genet Genomics* 2002, **267**(5):557–563.

38. Loppnau P, Tanguay P, Breuil C: **Isolation and disruption of the melanin pathway polyketide synthase gene of the softwood deep stain fungus *Ceratocystis resinifera*.** *Fungal Genet Biol* 2004, **41**(1):33–41.

39. Brasier C: **The genetic system as a fungal taxonomic tool: gene flow, molecular variation and sibling species in the "*Ophiostoma piceae – Ophiostoma ulmi*" complex and its taxonomic and ecological significance.** In *Ceratocystis and Ophiostoma: taxonomy, ecology, and pathogenicity*. Edited by Wingfield M, Seifert K, Webber J. St Paul, Minnesota: APS Press; 1993:77–92.

40. Tsui CK, Diguistini S, Wang Y, Feau N, Dhillon B, Bohlmann J, Hamelin RC: **Unequal recombination and evolution of the Mating-Type (MAT) loci in the pathogenic fungus *Grosmannia clavigera* and relatives.** *G3 (Bethesda, Md.)* 2013, **3**(3):465–480.

41. Gao Y, Breuil C, Chen T: **Utilization of triglycerides, fatty acids and resin acids in lodgepole pine wood by a sapstaining fungus *O. piceae*.** *Material und Organismen* 1994, **28**:105–118.

42. MacPherson S, Larochelle M, Turcotte B: **A fungal family of transcriptional regulators: The zinc cluster proteins.** *Microbiol Mol Biol Rev* 2006, **70**(3):583–+.

43. Asch DK, Orejas M, Geever RF, Case ME: **Comparative studies of the quinic acid (qa) cluster in several *Neurospora* species with special emphasis on the qa-x-qa-2 intergenic region.** *Mol Gen Genet* 1991, **230**(3):337–344.

44. Giles NH, Geever RF, Asch DK, Avalos J, Case ME: **The Wilhelmine E. Key 1989 invitational lecture. Organization and regulation of the qa (quinic acid) genes in *Neurospora crassa* and other fungi.** *J Hered* 1991, **82**(1):1–7.

45. Schirp A, Farrell R, Kreber B, Singh A: **Advances in understanding the ability of sapstaining fungi to produce cell wall-degrading enzymes.** *Wood Fiber Sci* 2003, **35**(3):434–444.

46. Yu J, Keller N: **Regulation of secondary metabolism in filamentous fungi.** *Annu Rev Phytopathol* 2005, **43**:437–458.

47. Wang Y, DiGuistini S, Wang TCT, Bohlmann J, Breuil C: ***Agrobacterium*-meditated gene disruption using split-marker in *Grosmannia clavigera*, a mountain pine beetle associated pathogen.** *Curr Genet* 2010, **56**(3):297–307.

48. Fischer C, Holl W: **Food Reserves of Scots Pine (*Pinus-Sylvestris* L) .2. Seasonal-Changes and Radial-Distribution of Carbohydrate and Fat Reserves in Pine Wood.** *Trees-Structure and Function* 1992, **6**(3):147–155.

49. Gao Y, Breuil C: **Extracellular Lipase Production by a Sapwood-Staining Fungus, Ophiostoma-Piceae.** *World J Microbiol Biotechnol* 1995, **11**(6):638–642.

50. Coleman JJ, Mylonakis E: **Efflux in fungi: La piece de resistance.** *PLoS Pathog* 2009, **5**(6):e1000486.

51. Abraham L, Hoffman B, Gao Y, Breuil C: **Action of *Ophiostoma piceae* proteinase and lipase on wood nutrients.** *Can J Microbiol* 1998, **44**(7):698–701.

52. Breuil C, Yagodnik C, ABRAHAM L: **Staining fungi growing in softwood produce proteinases and aminopeptidases.** *Material Und Organismen* 1995, **29**(1):15–25.

53. Abreu C, Sanguinetti M, Amillis S, Ramon A: **UreA, the major urea/H+ symporter in *Aspergillus nidulans*.** *Fungal Genet Biol* 2010, **47**(12):1023–1033.

54. Abraham L, Breuil C, Bradshaw D, Morris P, Byrne T: **Proteinases as potential targets for new generation anti-sapstain chemicals.** *For Prod J* 1997, **47**(9):57–62.

55. Raffa K, Smalley E: **Interaction of Pre-Attack and Induced Monoterpene Concentrations in Host Conifer Defense Against Bark Beetle Fungal Complexes.** *Oecologia* 1995, **102**(3):285–295.

56. Haridas S, Gantt JS: **The mitochondrial genome of the wood-degrading basidiomycete *Trametes cingulata*.** *FEMS Microbiol Lett* 2010, **308**(1):29–34.

57. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes.** *Genome Biol* 2004, **5**(2):R12.

58. Schmieder R, Edwards R: **Quality control and preprocessing of metagenomic datasets.** *Bioinformatics* 2011, **27**(6):863–864.

59. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc* 2012, **7**(3):562–578.

60. Korf I: **Gene finding in novel genomes.** *BMC Bioinforma* 2004, **5**:59.

61. Stanke M, Waack S: **Gene prediction with a hidden Markov model and a new intron submodel.** *Bioinformatics* 2003, **19**(Suppl 2):ii215–ii225.

62. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M: **Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training.** *Genome Res* 2008, **18**(12):1979–1990.

63. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674–3676.

64. Pavesi A, Conterio F, Bolchi A, Dieci G, Ottonello S: **Identification of new eukaryotic tRNA genes in genomic DNA databases by a multistep weight matrix analysis of transcriptional control regions.** *Nucleic Acids Res* 1994, **22**(7):1247–1256.

65. McInerney JO: **GCUA: general codon usage analysis.** *Bioinformatics* 1998, **14**(4):372–373.

66. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, **300**(4):1005–1016.

67. Kall L, Krogh A, Sonnhammer EL: **Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server.** *Nucleic Acids Res* 2007, **35**:W429–W432.