Draft genomes of the fungal pathogen *Phellinus noxius* in Hong Kong

Karen Sze Wing Tsang[1], Regent Yau Ching Lam[1,2], Hoi Shan Kwan[1]*

[1] School of Life Sciences, The Chinese University of Hong Kong, Hong Kong SAR

[2] Muni Arborist Limited, Hong Kong SAR

*Corresponding author: hskwan@eservices.cuhk.edu.hk

Abstract

The fungal pathogen *Phellinus noxius* is the underlying cause of brown root rot, a disease with causing tree mortality globally, causing extensive damage in urban areas and crop plants. This disease currently has no cure, and despite the global epidemic, little is known about the pathogenesis and virulence of this pathogen.

Using Ion Torrent PGM, Illumina MiSeq and PacBio RSII sequencing platforms with various genome assembly methods, we produced the draft genome sequences of four *P. noxius* strains isolated from infected trees in Hong Kong to further understand the pathogen and identify the mechanisms behind the aggressive nature and virulence of this fungus. The resulting genomes ranged from 30.8Mb to 31.8Mb in size, and of the four sequences, the YTM97 strain was chosen to produce a high-quality Hong Kong strain genome sequence, resulting in a 31Mb final assembly with 457 scaffolds, an N50 length of 275,889 bp and 96.2% genome completeness. RNA-seq of YTM97 using Illumina HiSeq400 was performed for improved gene prediction. AUGUSTUS and Genemark-ES prediction programs predicted 9,887 protein-coding genes which were annotated using GO and Pfam databases. The encoded carbohydrate active enzymes revealed large numbers of lignolytic enzymes present, comparable to those of other white-rot plant pathogens. In addition, *P. noxius* also possessed larger numbers of cellulose, xylan and hemicellulose degrading enzymes than other plant pathogens. Searches for virulence genes was also performed using PHI-Base and DFVF databases revealing a host of virulence-related genes and effectors. The combination of non-specific host range, unique carbohydrate active enzyme profile and large amount of putative virulence genes could explain the reasons behind the aggressive nature and increased virulence of this plant pathogen.

The draft genome sequences presented here will provide references for strains found in Hong Kong. Together with emerging research, this information could be used for genetic diversity and epidemiology research on a global scale as well as expediting our efforts towards discovering the mechanisms of pathogenicity of this devastating pathogen.

Background

Phellinus noxius is a pathogenic soil-dwelling white-rot basidiomycete fungus belonging to the Hymenochaetaceae family. It causes the devastating Brown Root Rot (BRR) disease prevalent in tropical and subtropical countries and has a broad host range of over 260 species globally [1]. This fungal pathogen has become a growing concern in many countries in recent years as it would lead to swift deterioration of the health of the plant host within a year if left untreated. The pathogen is extremely infectious, and can be spread easily over short distances through root-to-root contact or contact with infected soil or wood pieces [2, 3], or through air-borne basidiospores for long distance dispersal [4]. It targets the roots and water transport system of the host plants, causing root mortality and compromising tree stability.

The pathogen has the ability to survive in decayed root tissue in soil for over 10 years, whilst its basidiospores can survive up to 4.5 months in soil, making it difficult to control and eradicate [5]. The aggressive nature and high infectivity of *P. noxius* have

made it a global concern as it is particularly prevalent in urban settings [6-12]. *P. noxius* also has a large impact on the production of commercial crop plants such the avocados in Australia [13] and plantation conifers in Japan [10]. In Hong Kong, occurrence of *P. noxius* infections has increased exponentially over the past several years, and has become subject to public scrutiny after several collapse incidents of infected trees injuring members of the public [14-15]. Although there are several currently available methods for controlling the pathogen [16-17], there is no established cure to date.

In this study, we describe draft genome sequences of four *P. noxius* strains isolated in Hong Kong to facilitate understanding of the pathogen by investigating key factors behind its virulence and pathogenicity. Isolate YTM97 was sequenced using Ion Torrent PGM, Illumina MiSeq and PacBio RSII and hybrid assembled to generate a high quality genome sequence. Illumina HiSeq reads of the same strain were used to assemble the transcriptome to improve genome annotation. The other three strains were assembled using 3x200 bp paired-end Illumina MiSeq reads using trusted scaffolds from the previous assembly. *P. noxius* genomes generated in this study represent references for local strains in Hong Kong and allow further comparisons to global strains.

Methods

Isolation of fungal strains

P. noxius strains YTM97, YTM65, SSP14, and S39 were isolated from roots of infected trees in Hong Kong (Table S1) on 2% malt extract agar amended with streptomycin, gallic acid, benomyl, and dichloran [18] and incubated at 28°C for 10 days. Pure isolates were then cultured and maintained on 2% potato dextrose agar (PDA) before DNA extraction.

Genome sequencing and assembly

Isolate YTM97 was sequenced using Ion Torrent PGM, Illumina MiSeq (2x300 bp paired-ends) and PacBio RSII (four PacBio SMRT cells with an insert size of 20kb). The other three *P. noxius* strains were sequenced with the Illumina MiSeq platform. Illumina and PacBio sequencing was performed at the Beijing Genome Institute (BGI) of Hong Kong and Ion Torrent sequencing was performed at the Core Facilities of The Chinese University of Hong Kong.

For the isolates YTM65, SSP14 and S39, *de novo* assembly of the Illumina MiSeq raw reads was produced using the Newbler Assembler v2.8 [19]. For YTM97, FLASh v1.2.11 [20] was used to correct the Illumina reads, which were then used to correct the PacBio reads using the program proovread [21] to offset the high sequencing error rate of PacBio[1]. SPAdes v3.9.1 [22] was used for the hybrid assembly. For the other three isolates, raw Illumina MiSeq reads were *de novo* assembled using the Newbler Assembler v2.8 [19]. The quality and completeness of the assembly was assessed using the Benchmarking Universal Single-Copy Orthologs [BUSCO] v2 software [23] and assembly statistics were calculated with QUAST v4.5 [24].

Transcriptome sequencing and assembly

RNA-seq was performed on the YTM97 strain with the Illumina HiSeq4000 (PE101) platform at BGI Hong Kong. The isolate was grown on PDA for 10 days before total RNA extraction using the RNeasy Mini Kit (Qiagen). Raw RNA reads were filtered for quality and the remaining high quality reads were mapped against the genome assembly using Tophat 2.0.14 [25] and assembled into transcripts using Cufflinks 2.2.1 [26].

Genome annotation

Genome annotation was performed *de novo* using GeneMark-ES [27] and AUGUSTUS v3.2.3 [28]. Predicted proteins were annotated using BLASTp and imported into Blast2GO v4.1 [29] for gene ontology (GO) assignment. Eukaryotic orthologous groups (KOGs) were assigned by RPSBLAST v2.2.15 using the KOG database [30] on the WebMGA server [31]. Carbohydrate active enzymes (CAZymes) were predicted using family classification from the CAZy database on the dbCAN web server [32] and compared to those in nine other white rot fungi. Orthologous gene clusters were assigned using OrthoVenn [33]. RepeatMasker v4.0.7 [34] was used to find repeats in the assembled genome using cross_match [35] for hits. Protein-based RepeatMasking [34] was also used to provide a second reference for known repeats and RepeatScout [36] was used to predict *de novo* unknown repeats. Secreted proteins were identified using SignalP v4.1 [37] and ProtComp v.9.0 [38]. Putative virulence factors were predicted using the Database of Fungal Virulence Factors (DFVF) [39] and Pathogen Host Interaction Database (PHI-base) [40] using a blast cutoff value of $1e^{-10}$.

Phylogenetic analysis

The phylogenetic placement of *P. noxius* was inferred using LSU sequences of closely related members in the Hymenochaetaceae family retrieved from GenBank [Larsson et al 2006]. The sequences were clustered using ClustalW [41] and quality-filtered using Gblocks v0.91 [42]. Maximum composite likelihood trees were constructed using MEGA v6.06 [43] based on the Kimura 2-paramater model. Statistical confidences on the inferred relationships were assessed by 1000 bootstrap replicates. *Oxysporus corticola, Exidiopsis calcea* and *Protodontia piceicola* were included as outgroups.

Results and discussion

A high-quality genome of *P. noxius* from Hong Kong

Four *P. noxius* strains were isolated from infected trees around Hong Kong (Table S1). The YTM97 strain was sequenced using Ion Torrent PGM, Illumina MiSeq and PacBio RSII. Hybrid and non-hybrid *de novo* genome assemblies were performed. The PacBio RSII and Illumina MiSeq hybrid genome assembly provided the most complete assembly with the least amount of contigs. The final assembly yielded an estimated genome size of 31,684,099 bp with an N50 value of 275,889 bp (Table 1). The GC content was 41.58% and there were over 457 contigs. The coverage was estimated to be approximately 40 fold.

Illumina HiSeq4000 (PE101) RNA-seq of YTM97 yielded 74,968,996 raw reads, with a Q20 score of 98.31%. The assembled transcriptome was used to predict a total of 9,887 protein-coding genes. Genome assembly and annotation comprehensiveness recovered 279 complete BUSCOs (96.2%) suggesting the assembly is of high quality (Table S2).

Table 1. Genome assembly statistics for *P. noxius* YTM97

	Ion Torrent	Illumina MiSeq	Ion Torrent + Illumina MiSeq	PacBio RSII + Illumina MiSeq
Total length (bp)	31,387,566	31,026,813	31,304,704	31,684,099
# Contigs	1,903	635	596	457
GC (%)	41.60	41.61	41.60	41.58
N50 (bp)	78,615	130,025	202,553	275,889
N75 (bp)	36,966	76,360	108,247	164,239
Max contig length (bp)	406,953	430,359	931,211	1,232,758
# Ns per 100 kbp	0	0	0	0
Estimated fold coverage	35x	50x	45x	40x
Total protein-coding genes	9,957	9,489	9,598	9,887

The other three *P. noxius* strains were sequenced using Illumina MiSeq (2x300 bp paired end) and assembled using Newbler v2.8. The final assembly sizes varied between 30.8 – 31.9 Mb, encoding between 10,911 – 11,101 protein-coding genes (Table 2). N50 values and estimated genome completeness of these assemblies were considerably lower than those of the high quality assembly of YTM97, indicating that these assemblies were only partially complete. A pan-genome analysis for orthologous gene clusters using OrthoVenn for the *P. noxius* strains indicated that the core genome consisted of 6,744 orthologous gene clusters, whilst the pan-genome consisted of 9,399 orthologous gene clusters (Figure 1).

Table 2. Genome assembly statistics for all four *P. noxius* strains

Strain	YTM97	YTM65	SSP14	S39
Total length (bp)	31,684,099	30,898,337	31,989,945	31,622,185
# Contigs	457	5,240	4,719	4,466
GC (%)	41.58	41.52	41.51	41.50
N50 (bp)	275,889	8,236	10,463	11,000
N75 (bp)	164,239	4,285	5,225	5,522
Max contig length (bp)	1,232,758	165,448	157,720	160,989
# Ns per 100 kbp	0	0	0	0
Estimated fold coverage	40x	50x	46x	48x
Total protein-coding genes	9,887	10,911	11,101	10,962
Estimated genome completeness	96.2 %	69.6%	81.1%	75.8%

Genomic features

Out of the 9,887 predicted protein-coding genes, at least one GO term was assigned to 8,690 genes (xx%). The most represented functional categories were primary and cellular metabolic processes, mainly located intracellularly (Figure S1). KOG analysis revealed that the majority of proteins were assigned to general function, signal transduction, posttranslational modification, protein turnover, transcription, as well as intracellular trafficking, secretion and vesicular transport (Figure S2). The distributions of CDS regions, transposable elements, secreted proteins, and GC skew of the 50 largest contigs, representing 21 Mb of the final 31 Mb assembly, of the high-quality genome were visualised using Circa (Figure 2). SignalP and ProtComp were used to identify 290 potential secreted signal peptides using stringent parameters that eliminated peptides with more than one transmembrane domain and glycosylphosphatidylinositol anchors. The majority of these peptides (48%) were found to be secreted extracellularly, consistent with the degradative nature of this pathogen. The distribution of these secreted proteins was uneven in the genome (Figure 2), with large stretches of contigs containing no secreted protein-coding genes.

Repeats and transposable elements (TE)

RepeatScout was used for *de novo* repeat detection before running through RepeatMasker for improved accuracy and identified a total of 1,077,456 bp of repetitive elements which constituted for 3.38% of the whole genome (Table S3). The majority of the repeats belonged to simple repeats (1.62%) and LTR elements (1.26%), with only five elements showing no homology to any existing repetitive sequences. The distribution of transposable elements in the *P. noxius* genome was uneven (Figure 2), with some contigs containing only one element. The retrotransposon content in Basidiomycota genomes ranges between 3,108 copies in *Postia placenta* [44] to 47 copies in *Cryptococcus neoformans* [45]. A total of 546 LTR elements were detected in *P. noxius*, which is slightly lower than the average of 607 found in Basidiomycota [46]. However, in regard to the density per Mb, *P. noxius* has one of the highest LTR ratios due to its compact genome (Table S4). The majority of LTRs in *P. noxius* belonged to the Gypsy/DIRS1 category.

Low levels of local and global genetic diversity of *P. noxius*

The phylogenetic relationships between *P. noxius* isolates and its closest relatives were determined using the maximum composite likelihood method with a bootstrap consensus inferred from 1000 replicates. The analysis included LSU sequences deposited in GenBank from 46 members of the Hymenochaetaceae family, with *Oxysporus corticola, Exidiopsis calcea* and *Protodontia piceicola* as outgroups. The resulting phylogenetic tree (Figure 3 and Table S5)showed well-defined clustering of different genera in concordant to previous findings [47], with the exception of the *Phellinus* genus, which formed multiple separate clades scattered throughout the tree. This indicates that the classification of the *Phellinus* genus may warrant revision or at least reconsideration. *P. noxius* clustered with the rest of the Hymenochaetaceae family with a 98% bootstrap value, but formed its own clade separated from the other members of the *Phellinus* genus.

To identify the level of genetic diversity among the Hong Kong *P. noxius* strains, single-nucleotide polymorphisms (SNPs) were called by mapping the reads of the three strains to the reference YTM97 strain. Strains YTM65, SSP14 and S39 showed similar levels of genetic variation, with 0.66, 0.63 and 0.66 SNPs per kb, respectively (Table S6). Further comparative analysis based on a total of 5,965 single-copy genes revealed the highest similarity between YTM97 and S39 (98.5%) and the lowest similarity between SSP14 and YTM65 (98.1%), suggesting that there is none or very little genetic variation between locations (Table S7). This could further suggest that *P. noxius* in these locations may have originated from the same source.

To resolve the phylogenetic relationship between *P. noxius* strains on a global scale, internal transcribed spacer (ITS) sequences from 20 strains were obtained from GenBank. Maximum composite likelihood analysis revealed that the majority of strains formed one clade (Figure 4 and Table S8) with 99% similarity, suggesting that the pathogen has little genetic divergence on a local and global scale. Most strains from the same countries were grouped together with the exception of those from Singapore, Gabon, Malaysia and Australia, suggesting that strains from these countries do not originate from the same sources.

CAZyme analysis

CAZymes play a large role in the pathogenicity and virulence of white-rot pathogens so the degradation capabilities of *P. noxius* were compared against nine known white-rot fungi here. Basidiomycete white-rot fungi were used in this comparison, with the exception of *Fusarium graminea.* The plant pathogens *Phanerochaete carnosa, Stereum hirsutum, Fomitiporia mediterranea, Trametes versicolor*, and *Fusarium graminea* were included due to their similar pathogenicity to *P. noxius*.

The *P. noxius* genomes encoded 400 to 477 CAZymes with a larger number of carbohydrate esterases (CE) compared to the other fungi (Figure S3). It possessed an average number of auxiliary activities enzymes (AA), which is characteristic of white-rot fungi. However, the distribution of enzymes in each AA class differed from the other plant pathogens. YTM65, SSP14 and S39 possessed 10 to 11 copies of AA1 class enzymes, which encode for laccases, ferroxidases and laccase-like multicopper oxidases used for lignolytic activities. However, only one copy was present in YTM97 (Figure 5 and Table S9), suggesting that lignin was not a major substrate for this strain and that *P. noxius* possessed the ability to efficiently utilize other carbohydrate sources. AA4 and AA7 were both present in all *P. noxius* strains but mostly absent in other plant pathogens and white-rot fungi, suggesting an increased efficiency in processing aromatic compounds, detoxification and biotransformation of lignocellulosic compounds [48]. In addition, higher numbers of CE1, CE10, GT48, GH105, and GH109 enzymes in *P. noxius* suggest that the fungus utilizes xylans, cellulose, hemicellulose, and pectin substrates efficiently in addition to lignin. The ability to degrade a wide range of substrates revealed by the CAZyme profiles and the non-specific host range of *P. noxius* together might partially explain the aggressive nature and virulence of this pathogen.

Pathogenicity and virulence genes

P. noxius is known for its virulent nature toward a large number of hosts but to date, no studies have been performed to identify the mechanisms or genes that are involved in its pathogenesis. PHI-base is a collection of genes that have been experimentally proven to participate in pathogen and host interactions in different species. With an E-value threshold of $\leq 1e^{-10}$, 1,968 PHI genes were identified in YTM97. A total of 234 genes were annotated as vital for its pathogenicity, mutations or deletions of which could result in a loss of pathogenicity phenotype. Uniprot GO assignment revealed that many of these genes are involved in protein transport, autophagy, peroxisome biogenesis and the growth of appresoria and its ability to penetrate plant tissues.

Apart from virulence factors related to host interactions, a general search for putative virulence factors was performed using the DFVF with an E-value threshold of $\leq 1e^{-10}$. Around 3,482 putative virulence factors were predicted and transcriptional repressor Tup1, P21-rho-binding domain, DEAD/DEAH box helicase, AAA, 1,3-beta glucan synthase, cytochrome p450, cellulase, WD40, chitin synthases 1 and 2, and beta glucan synthesis-associated protein (SKN1) were in high abundance. Disruption of the *tup1* gene causes a reduction in the mating and filamentation capacity [49]; DEAD/DEAH box helicases are crucial in the signaling pathways for host and pathogen interactions [50]; beta glucan synthases and chitin synthases are vital for the biogenesis of cell walls; and P450s have now been linked to increased virulence in several pathogens [51-53]. The combination of all these virulence factors could explain the aggressive nature of *P. noxius*.

Conclusion

The current understanding of the pathogenicity and mechanisms of infection of this pathogen are limited. Here, we present the genome sequences of four strains of *P. noxius*, the cause of the devastating brown root rot disease, posing threats to a wide range of tree hosts and crop plants globally. We were able to reveal the large repertoire of CAZymes present, possibly explaining the ability of this pathogen to kill a tree in a short amount of time and its ability to propegate and spread quickly. It has an unspecific host range and ability to degrade carbohydrates quickly, in addition to a large quantity of putative virulence factors, making it difficult to develop prevention methods.

Comparisons between the Hong Kong strains revealed the low genetic diversity present with very similar genome sequences, suggesting that these strains originated from the same source. With LSU phylogenetic analysis of the Hymenochaetaceae family, we found *P. noxius* clustered in a distinct clade away from other family members. The *Phellinus* genus is not clearly resolved and formed multiple clades throughout the phylogenetic tree, unlike other family members who formed distinct clusters within its own genera. Location based ITS analysis also revealed no significant clustering with strains from different countries, further suggesting low genetic diversity present. However, information on the Hymenochaetaceae family is limited, with only 2 fully sequenced genome sequences available making full comparative genomics difficult.

In summary, the genome sequences and information reported in this study are an important reference point for Hong Kong strains and is a valuable resource for researchers to perform genetic diversity and epidemiology research on a global scale. This information along with emerging research and the future availability of sequences from other countries could help track the spread of brown root rot disease as well as expediting our efforts towards discovering the mechanisms of pathogenicity of this devastating pathogen.

Availability of data and materials

The datasets generated and analysed during the current study are available under BioProject ID PRJNA415882 (http://www.ncbi.nlm.nih.gov/bioproject/415882)

Funding

This project was jointly funded by the Environmental and Conservation Fund and Woo Wheelock Green Fund (ECF Project 03/2014).

References

1. Farr DF, Rossman AY. Fungal Databases. U.S. National Fungus Collections, ARS, USDA. 2017. https://nt.ars-grin.gov/fungaldatabases. Accessed 20 Jun 2017.
2. Ann PJ, Chang TT, Ko WH. *Phellinus noxius* brown root rot of fruit and ornamental trees in Taiwan. Plant Disease. 2002;86:820-826.
3. Akiba M, Ota Y, Tsai IJ, Hattori T, Sahashi N, Kikuchi T. Genetic differentiation and spatial structure of *Phellinus noxius,* the causal agent of Brown Root Rot of woody plants in Japan. PLos One. 2015;10:e0141792.
4. Chung CL, Huang SY, Huang YC, Tzean SS, Ann PJ , Tsai JN, Yang CC, Le HH, Huang TW, Huang HY et al. The genetic structure of *Phellinus noxius* and dissemination pattern of brown root rot disease in Taiwan. Plos One. 2015;10:e0139445.
5. Chang TT. Survival of *Phellinus noxius* in soil and in the roots of dead host plants. Phytopathology. 1996;86:272–276.
6. Sahashi N, Akisa M, Ishihara M, Abe Y, Morita S. First report of the brown root rot disease caused by *Phellinus noxius,* its distribution and newly recorded host plants in the Amami Islands, southern Japan. Forest Pathology. 2007;37:167-173.
7. Hodges CS, Tenorio JA. Root disease of Delonix regia and associated tree species in the Marina Islands caused by Phellinus noxius. Plant Disease. 1984;68:334-336.
8. Brooks FE. Brown root rot disease in American Samoa's tropical rain forests. Pacific Science. 2002;56:377-387.
9. Greening, Landscape & Tree Management Section, Development Bureau, The Government of the HKSAR of the People's Republic of China. Brown Root Rot Disease. 2016. http://www.greening.gov.hk/en/knowledge_database/brown_root.html. Accessed 26 Jun 2017.
10. Sahashi N, Akiba M, Takemoto S, Yokoi T, Ota Y, Kanzaki N. *Phellinus noxius* causes brown root rot on four important conifer species in Japan. European Journal of Plant Pathology. 2014;140(4):869-873.

11. Nicole M, Nandris D, Geiger JP, Rio B. Variability among African populations of *Rigidoporus lignosus* and *Phellinus noxius*. Forest Pathology. 1985;15:293-300.
12. Nandris D, Nicole M, Geiger JP. Variation in virulence among *Rigidoporus lignosus* and *Phellinus noxius* isolates from West Africa. Forest Pathology. 1987;17:271-281.
13. Dann K, Smith LA, Shuey LS, Begum F. Phellinus noxius: A basidiomycete fungus impacting productivity of Australian avocados. In: 7th World Avocado Congress 2011 - Conference Proceedings. 7th World Avocado Congress, Cairns, QLD, Australia 5-9 September 2011.
14. Developmental Bureau. Experts decide to remove Old and Valuable Tree on Nathan Road. 2012. http://www.devb.gov.hk/en/publications_and_press_releases/press/index_id_7298.html. Accessed 21 June 2017.
15. Tree Management Office. Guidelines on brown root rot disease. 2012. http://www.greening.gov.hk/filemanager/content/pdf/knowledge_database/Guidelines_on_Brown_Root_Rot_Disease(version_for_the_general_public)EN.pdf. Accessed 21 June 2017.
16. Schwarze FWMR, Jauss F, Spencer C, Hallam C, Schubert M. Evaluation of an antagonistic Trichoderma strain for reducing the rate of wood decomposition by the white rot fungus *Phellinus noxius*. Biological Control. 2012;61:160–168.
17. Chang TT, Chang RJ. Generation of volatile ammonia from urea fungicidal to *Phellinus noxius* in infested wood in soil under controlled conditions. Plant Pathology. 1999;48:337-344.
18. Chang TT. A selective medium for *Phellinus noxius*. European Journal of Forest Pathology. 1995;25:185-190.
19. Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW. High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics. 2012;13:32.
20. Magoc T, Salzberg S. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27:2957-2963.
21. Hackl T, Hedrich R, Schultz J, Forster F. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics. 2014;30:3004-3011.
22. Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. Journal of Computational Biology. 2012;19:455-477.
23. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210-3212.
24. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072-1075.
25. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105-1111.
26. Trapnell C, Williams B, Pertea G, Mortazavi A, Kwan G, van Baren J, Salzberg S, Wold B, Pachter L. Transcript assembly and quantification by

RNA-Seq reveals unannotated transripts and isoform switching during cell differentiation. Nature biotechnology. 2010;doi:10.1038/nbt.1621.
27. Borodovsky M, Lomsadze A. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Current Protocols in Bioinformatics. 2011;doi:10.1002/0471250953.bi0406s35.
28. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research. 2005;doi: 10.1093/nar/gki458.
29. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomic research. Bioinformatics. 2005;21:3674-3676.
30. Tatusov RL, Fedorva ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R,Mekhedov SL, Nikolskaya AN, Sridhar Rao B, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003;doi:10.1186/1471-2105-4-41.
31. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a Customizable Web Server for Fast Metagenomic Sequence Analysis. BMC Genomics. 2011;doi:10.1186/1471-2164-12-4.
32. Yin Y, Mao X, Yang JC, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Research. 2012;40:W445-451.
33. Wang Y, Coleman-Derr D, Chen Guoping, Gu YQ. Orthovenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic acids research. 2015;doi: 10.1093/nar/gkv487.
34. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015. http://www.repeatmasker.org.
35. Green P. Phrap version 1.090518. http://phrap.org.
36. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21:351-358.
37. Peterson TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions.Nature Methods. 2011;8:785-786.
38. ProtComp – predict the sub-cellular localization for animal/fungi proteins. http://www.softberry.com/berry.phtml?topic=protcompan&group=programs&subgroup=proloc. Accessed March 2017.
39. Lu T, Yao B, Zhang C. DFVF: database of fungal virulence factors. Database. 2012;bas032.
40. Winnenburg R, Baldwin TK, Urban M, Rawlings C, Kohler J, Hammond-Kosack KE. PHI-base: a new database for pathogen host interactions. 2006;34:D459-D464.
41. Larkin MA, Blackshields G, Brown NP, Channa R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and ClustalX version 2. Bioinformatics. 2007;23:2947-2948.
42. Talavera G, Castersana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology. 2007;56:564-577.
43. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. 2013;30:2725-2729.

44. Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* uspports unique mechanisms of lignocellulose conversion. Proceedings of the national academy of sciences of the United States of America. 2009;doi:10.1073/pnas.0809575106.
45. Chaturvedi V, Chaturvedi S. *Cryptococcus gattii*: a resurgent fungal pathogen. Trends in microbiology. 2011;19:564-571.
46. Muszewska A, Hoffman-Sommer M, Grynberg M. LTR retrotransposons in fungi. PLOS one. 2011;doi:10.1371/journal.pone.0029425.
47. Larsson KH, Parmasto E, Fischer M, Langer E, Nakasone KK, Redhead SA. Hymenochaetales: a molecular phylogeny for the hymenochaetoid clade. Mycologia. 2006;98:926-936.
48. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnology Biofuels. 2013;doi:10.1186/1754-6834-6-41.
49. Elias-Villalobos A, Fernandez-Alverez A, Ibeas JI. The general transcriptional repressor Tup1 is required for dimorphism and virulence in a fungal plant pathogen. PLOS Pathogens. 2011;doi:10.1371/journal.ppat.1002235.
50. Heung LJ, Del Poeta M. Unlocking the DEAD-box: a key to cryptococcal virulence? The Journal of Clinical Investigation. 2005;doi:10.1172/JCI200524508.
51. Zhang DD, Wang XY, Chen JY, Kong ZQ, Gui YJ, Li NY, Bao YM, Dai XF. Identification and characterization of a pathogenicity related gene VdCYP1 from *Verticillium dahliae.* Scientific reports. 2016;doi:10.1038/srep27979.
52. Coleman JJ, White GJ, Rodriguez-Carres M, Vanetten HD. An ABC transporter and a cytochrome P450 of *Nectria haematococca* MPVI are virulence factors on pea and are the major tolerance mechanisms to the phytoalexin pisatin. Molecular plant microbe interactions. 2011;doi:10.1094/MPMI-09-10-0198.
53. Lah L, Podobnik B, Novak M, Korosec B, Berne S, Vogelsang M, Krasevec N, Zupanec N, Stojan J, Bohlmann J, Komel R. The versatility of the fungal cytochrome P450 monooxygenase system is instrumental in xenobiotic detoxification. Molecular microbiology. 2011;doi:10.1111/j.1365-2958.2011.07772.x.
54. Suzuki H, MacDonald J, Syed K, Slamov A< Hori C, Aerts A, Henrissat B, Wiebenga A, VanKuyk PA, Barry K, Lindquist E, LaButti K, Lapidus A, Lucas S, Coutinho P, Gong Y, Samejima M, Mahadevan R, Abou-Zaid M, de Vries RP, Igarashi K, Yadav JS, Grigoriev IV, Master ER. Comparative genomics of the white rot fungi, Phanerochaete carnosa and P. chrysosporium, to elucidate the genetic basis of the distinct wood types they colonise. BMC Genomics. 2012;doi:10.1186/1471-2164-13-444.
55. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT, Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Gorecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, Kohler A, Kues U, Kumar TK, Kuo A, LaButti K, Larrondo LF, Lindquist E, Ling A, Lombard V, Lucas S, Lundell T, Martin R, McLaughlin DJ, Morgenstern I, Morin E, Murat C, Nagy LG, Nolan M, Ohm RA, Patyshakuliyeva A, Rokas A, Ruiz-Duenas FJ, Sabat G, Salamov A, Samejima M, Schmutz J, Slot JC, St John F, Stenlid J, Sun H, Sun

S, Syed K, Tsang A, Wiebenga A, Young D, Pisabarro A, Eastwood DC, Martin F, Cullen D, Grigoriev IV, Hibbett DS. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science. 2012;doi:10.1126/science.1221748.
56. Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA, Otillar R, Poliakov A, Ratnere I, Riley R, Smirnova T, Rokhsar D, Dubchak I. The Genome Portal of the Department of Energy Joint Genome Institute. 2012;40(Database issue):D26-32.
57. Ohm RA, Riley R, Salamov A, Min B, Choi IG, Grigoriev IV. Genomics of wood-degrading fungi. Fungal genetics biology. 2014;doi:10.1016/j.fgb.2014.05.001.
58. Alfaro M, Castanera R, Lavin JL, Grigoriev IV, Oguiza JA, Ramirez L, Pisabarro AG. Comparative and transcriptional analysis of the predicted secretome in the lignocellulose-degrading basidiomycete fungus Pleurotus ostreatus. Environmental Microbiology. 2016;doi:10.1111/1462-2920.13360.
59. Muraguchi H, Umezawa K, Niikura M, Yoshida M, Kozaki T, Ishii K, Sakai K, Shimizu M, Nakahori K, Sakamoto Y, Choi C, Ngan CY, Lindquist E, Lipzen A, Tritt A, Haridas S, Barry K, Grigoriev IV, Pukkila PJ. Strand-Specific RNA-Seq Analyses of Fruiting Body Development in Coprinopsis cinerea. PLoS One. 2015;doi:10.1371/journal.pone.0141586.
60. Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang YL, Decaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes HW, Mitterbauer R, Muehlbauer G, Munsterkotter M, Nelson D, O'donnell K, Ouellet T, Qi W, Quesneville H, Roncero MI, Seong KY, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC. The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. Science. 2007;317:1400-1402.

Table legends

Table 1. Genome assembly statistics for *P. noxius* YTM97
Table 2. Genome assembly statistics for all four *P. noxius* strains
Table S1. Table showing locations and host species of *P. noxius* infected trees used in this study
Table S2. Table showing the BUSCO statistics for YTM97
Table S3. RepeatMasker and RepeatModeler results for YTM97
Table S4. LTR density ratios of *P. noxius* and select Basidiomycota
Table S5. GenBank accessions for Hymenochaetaceae 28S phylogenetic analysis
Table S6. Table showing SNP density of Hong Kong *P. noxius* strains
Table S7. Pairwise calculations of genetic similarity of the four Hong Kong *P. noxius* strains using 5,965 single-copy orthologous genes
Table S8. GenBank accessions for *Phellinus noxius* location based ITS phylogenetic analysis

Figure legends

Figure 1. The pan genome orthologous gene clusters for the four Hong Kong strains of *Phellinus noxius.*
Figure 2: Circular representation of the genome assembly with the 50 largest contigs (21Mb) out of a complete 31Mb assembly. From outside to inside –1: Sizes of the 50 largest contigs; 2: CDS forward strand; 3: CDS reverse strand; 4: Transposable elements (TEs); 5: Secreted proteins; green – forward strand, red – reverse strand; 6: GC skew (10 kb sliding window).
Figure 3: Maximum composite likelihood phylogenetic analysis using LSU sequences of *P. noxius* strains and 46 members of the Hymenochaetaceae family with a bootstrap consensus inferred from 1000 replicates.
Figure 4: Maximum likelihood phylogenetic analysis using the ITS sequences of *P. noxius* strains from different countries with a bootstrap consensus inferred from 1000 replicates.
Figure 5: Heatmap of selected CAZymes found in *P. noxius* strains and other white rot Basidiomycetes
Figure S1. Graphical representation of Gene Ontology process assignment for YTM97. Showing the assignment of biological processes **(A)** and the cellular locations of these processes **(B)** using GO terms.
Figure S2. Graphical representation of KOG assigned classes for YTM97
Figure S3. Predicted CAZyme numbers in white-rot fungi

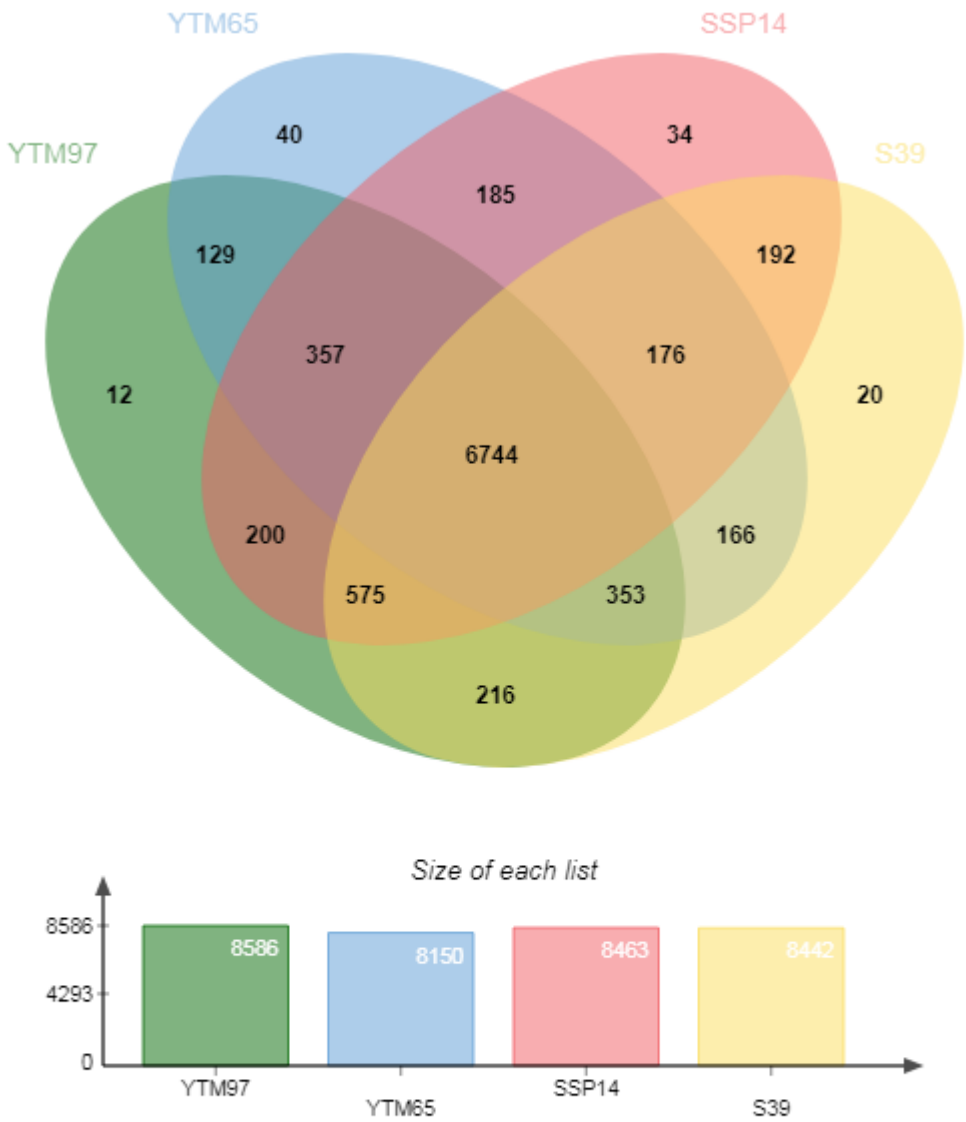

Figure 1. The pan genome orthologous gene clusters for the four Hong Kong strains of *Phellinus noxius.*

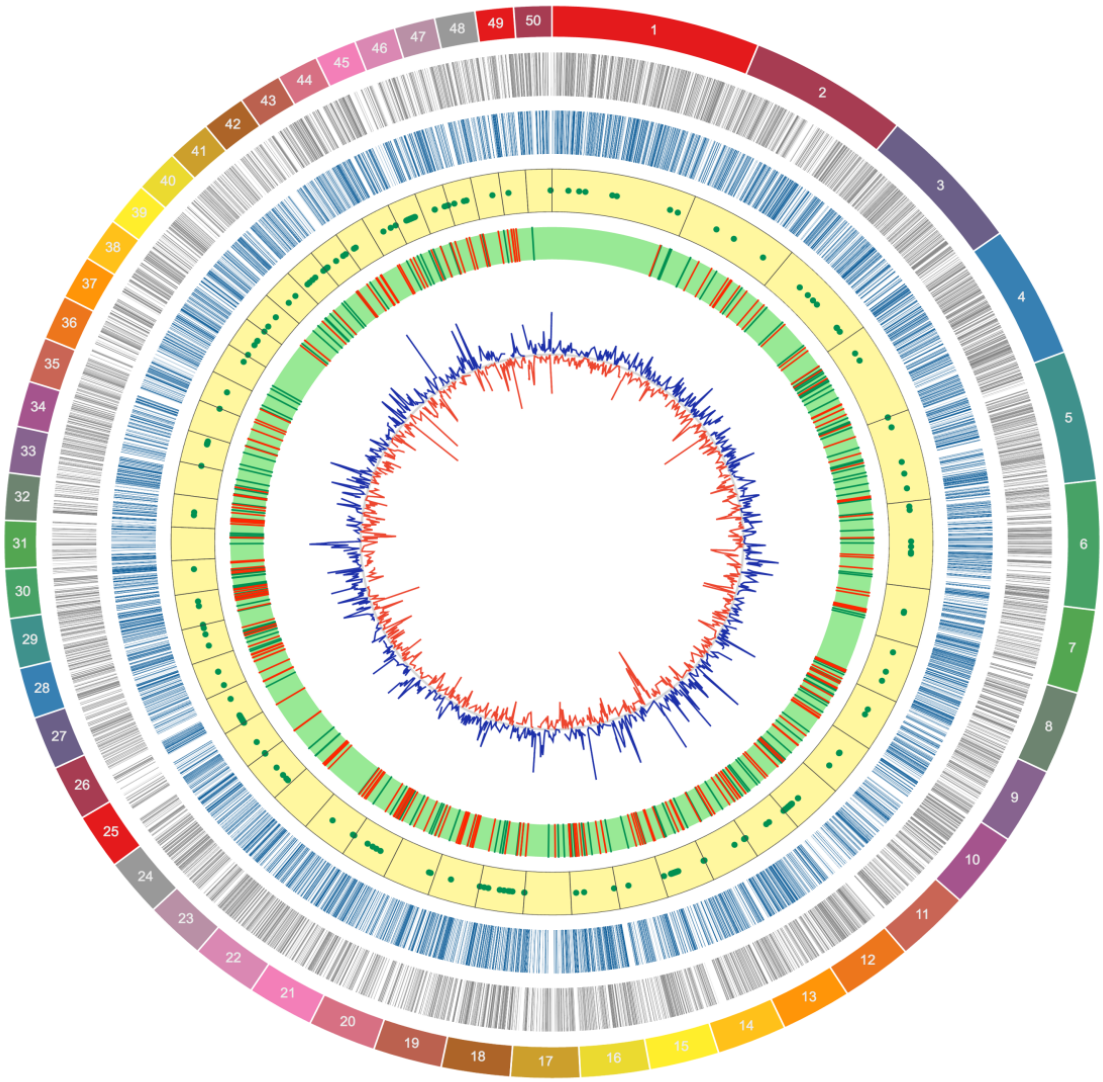

Figure 2: Circular representation of the genome assembly with the 50 largest contigs (21Mb) out of a complete 31Mb assembly. From outside to inside –1: Sizes of the 50 largest contigs; 2: CDS forward strand; 3: CDS reverse strand; 4: Transposable elements (TEs); 5: Secreted proteins; green – forward strand, red – reverse strand; 6: GC skew (10 kb sliding window)

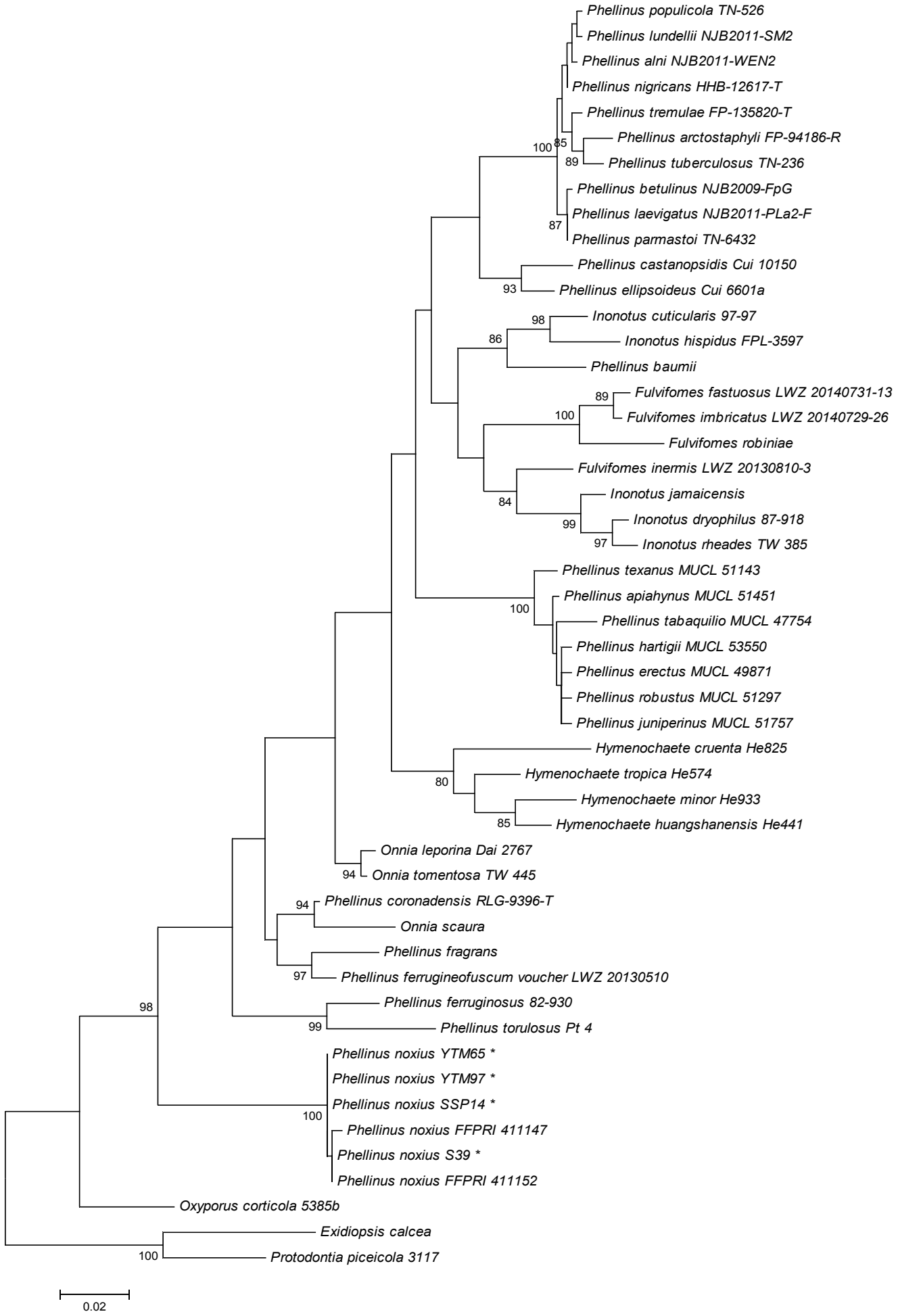

Figure 3: Maximum composite likelihood phylogenetic analysis using LSU sequences of *P. noxius* strains and 46 members of the Hymenochaetaceae family with a bootstrap consensus inferred from 1000 replicates.

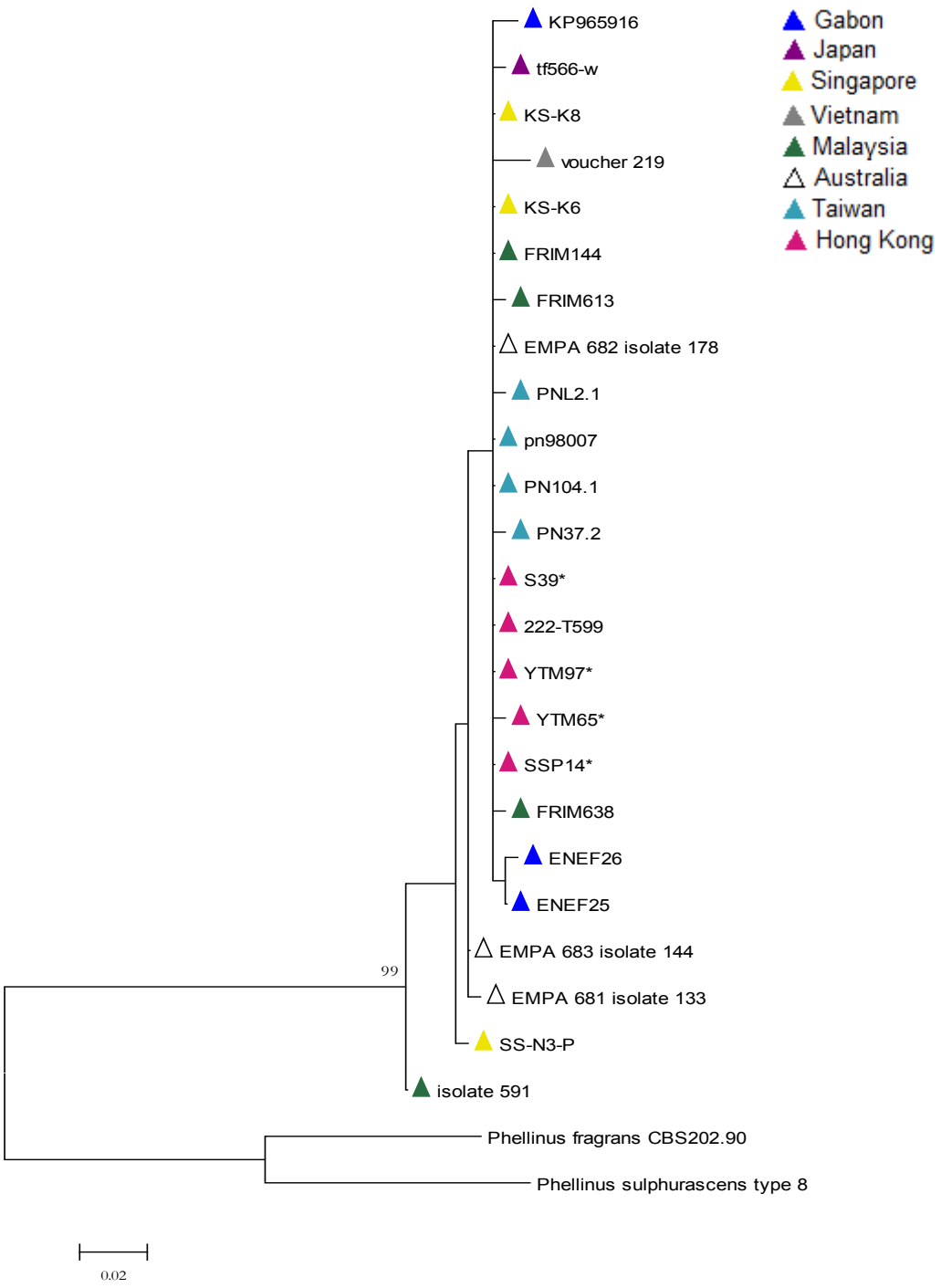

Figure 4: Maximum likelihood phylogenetic analysis using the ITS sequences of *P. noxius* strains from different countries with a bootstrap consensus inferred from 1000 replicates.

	Basidiomycetes												Asco.
White rot	x	x	x	x	x	x	x	x	x	x	x	x	x
Plant Pathogen	x	x	x	x	x	x	x	x					x
	P. noxius YTM97	*P. noxius S39*	*P. noxius SSP14*	*P. noxius YTM65*	*P. carnosa*	*S. hirsutum*	*F. mediterranea*	*T. versicolor*	*P. coccineus*	*P. chrysosporium*	*P. ostreatus*	*C. cinerea*	*F. graminearum*
AA1	1	10	11	10	10	17	12	10	7	5	7	17	14
AA2	16	14	16	16	11	7	17	26	11	15	9	1	1
AA3	9	10	10	9	37	20	27	23	26	36	29	40	20
AA4	2	2	2	1	0	0	0	0	0	0	0	0	1
AA5	4	3	4	4	6	8	4	9	7	7	11	6	5
AA6	2	2	3	3	3	0	3	1	1	4	2	3	1
AA7	13	12	13	12	0	10	0	0	0	0	2	2	4
AA8	1	1	1	1	2	1	1	2	2	2	1	7	6
AA9	17	18	16	15	11	9	13	18	16	16	21	34	13
CE1	10	3	8	8	2	1	0	3	3	4	2	4	4
CE10	45	34	41	37	0	21	0	0	1	0	0	0	0
CBM1	15	12	15	10	27	2	6	23	22	36	31	39	13
CBM50	12	2	8	6	2	3	2	1	5	12	1	19	4
GT48	11	2	3	4	2	0	2	2	2	2	2	2	1
GH105	5	5	5	5	0	2	1	0	0	0	2	0	3
GH109	7	1	5	5	0	0	0	0	0	0	0	0	0
Total	**477**	**400**	**454**	**431**	**400**	**358**	**398**	**462**	**423**	**423**	**383**	**519**	**577**

Figure 5: Heatmap of selected CAZymes found in *P. noxius* strains and other white rot Basidiomycetes