Short Genome Communications

The genome of *Pleurotus eryngii* provides insights into the mechanisms of wood decay

Rui-Heng Yang¹, Yan Li¹, Ying Wáng, Jia-Ning Wan, Chen-Li Zhou, Ying Wăng, Ying-Nv Gao, Wen-Jun Mao, Li-Hua Tang, Ming Gong, Ying-Ying Wu, Da-Peng Bao*  
National Engineering Research Center of Edible Fungi, Ministry of Science and Technology (MOST), Key Laboratory of Edible Fungi Resources and Utilization (South), Ministry of Agriculture; Key Laboratory of Agricultural Genetics and Breeding of Shanghai; Institute of Edible Fungi, Shanghai Academy of Agricultural Sciences, Shanghai, PR China

**A R T I C L E  I N F O**  
Received 19 July 2016  
Received in revised form 14 September 2016  
Accepted 7 October 2016  
Available online 11 October 2016

**Keywords:**  
*Pleurotus eryngii*  
White-rot fungi  
PacBio RS II  
Carbohydrate-active enzymes (CAZymes)  
Oxidoreductases

**A B S T R A C T**

*Pleurotus eryngii* (DC.) Quél. is widely used for bioconverting lignocellulosic byproducts into biofuel and value added products. Sequencing and annotating the genome of a monokaryon strain *P. eryngii* 183 allows us to gain a better understanding of carbohydrate-active enzymes (CAZymes) and oxidoreductases for degradation of lignocellulose in white-rot fungi. The genomic data provides insights into genomic basis of degradation mechanisms of lignin and cellulose and may pave new avenues for lignocellulose bioconversion.

© 2016 Published by Elsevier B.V.

Lignocellulose, including cellulose, hemicellulose and lignin, is the most abundant organic compound in natural environments (Sánchez, 2009). Bioconversion of lignocellulosic biomass plays important roles in exploiting and utilizing biofuel and value added products and maintaining a balanced environment. *Pleurotus eryngii*, also known as the king oyster mushroom with high nutrient and medicinal value, is widely exploited for bioconverting lignocellulosic byproducts into growth substances (Yao and Lan, 2004). It has been reported as a tetrapolar heterothallic basidiomycete, whereby dikaryons produce edible fruiting bodies on a large range of lignocellulosic substrates (Kim et al., 2014). It is commonly cultivated in Europe, Middle East, North America and some parts of Asia. Especially in some developing countries, the governments make efforts to promote the production in rural communities as one of the tools for poverty alleviation and diversification of agricultural production. As a white-rot fungus, *P. eryngii* could secret various extracellular enzymes (cellulases, hemicellulases, pectinase, ligninase, protease and peptidases) for transforming all components of plant biomass, including cellulose, hemicellulose and lignin (Yao and Jin, 2004; Xie et al., 2016). However the complex compositions of CAZymes and oxidoreductases acting on lignocellulose at genome-scale in this fungus remain to be revealed.

Monokaryon strain *P. eryngii* '183' was obtained from dikaryon strain 'Xinghan' from Institute of Edible Fungi, Shanghai Academy of Agricultural Sciences using protoplast isolation method published previously (Chang et al., 1983). The strain was maintained on Potato Dextrose Agar (PDA) medium at temperature 25 °C and kept in the dark. The genome DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Courtaboeuf, France) according to the manufacturer’s specification and sequenced using two separate platforms Illumina and Pacbio. For Illumina genome sequencing, 151 bp pair-ended sequencing data of two libraries with insert sizes of 300 bp (pair-end) and 8000 bp (mate pair) were generated on HiSeq 2000 (Illumina, USA). For PacBio sequencing, genomic DNA was sheared to 10 kb using g-TUBE(Covaris USA) and converted into the proprietary SMRTbell™ library format using RS DNA Template Preparation Kit (Pacific Biosciences, Melon Park, CA). The library was sequenced on a PacBio RSII (Expression Analysis, Durham, NC, USA) platform. All the sequencings were conducted in Southgene Technology Co., Ltd (Shanghai, China).

After filtering the low quality reads (the quality scores <20 or contained 'N' in the reads) using the method previously published (Karlsson et al., 2015), 27,547,890 and 11,162,107 clean...
pair-ends reads were obtained from 30,123,660 and 15,322,772 raw pair-ends datasets generated from 300bp (paire-end) and 8000 bp (mate pair) libraries respectively. A total of 1,129,497 polymerase reads were generated from Pacbio platform, after removing adapter inserted in the long polymerase reads, resulting in 1,703,117 subreads with average length of 4166 bp. The PacBio subreads were assembled into a primary assembly using HGAP (Pacific Biosciences). The pair-ends reads were mapped the draft genome using bowtie2, a total of 29 SNPs and 3236 Indels were detected from the genome of strain ‘183’ using samtools to eliminate the errors from the PacBio. After scaffold building and gap filling using mate-pair data, genome sequences of 43.78 Mb with 49.28% G+C content spread across 166 scaffolds were obtained by assembling total approximately 475 × coverage data. The N50 size of the scaffolds was 0.5 Mb (Table 1). 14,443 gene models coding for proteins were annotated, of which 13,805 (95.58%) genes encoded proteins with homologous sequences in the NCBI nr protein databases (Table 1). KEGG analysis revealed 1733 proteins involved in different pathways. GO analysis assigned 1349 proteins into different GO terms. Altogether, 4902 proteins were assigned to different KOG classes, and 2939 protein domains were revealed in a total of 6127 proteins. Our pipeline also showed that there are 186 tRNAs and 8 rRNAs in the genomic sequences.

A total of 494 CAZyme-coding gene homologs determined using domain-based annotation dbCAN (Yin et al., 2012) with the parameters provided for fungi (E-value < 1e-17 and coverage > 45%) in the genome of P. eryngii strain ‘183’ comprise with 139 auxiliary activities (AA), 22 carbohydrate-binding modules (CBMs), 51 carbohydrate esterases (CEs), 201 glycoside hydrolases (GHS), 59 glycosyl transferases (GTs) and 22 polysaccharide lyases (PLs) (Fig. 1). The number of CAZymes was the highest than that in other fungi (Supplementary S1). The cluster conducted based on the composition of CAZymes using hclust in R packages with the method ‘average’ revealed that CAZymes in P. eryngii were different from the closely evolutionarily related P. ostreatus and the fungi with different life-styles (white-rot, brown-rot, and plant pathogens or mycorrhiza) dispersed across the tree (Fig. S1, Supplementary S2). Many more families AA1 (33 copies), AA7 (15 copies), AA8 (26 copies), AA9 (26 copies), CE10 (28 copies), GT41 (8 copies), GH7 (15 copies), GH16 (16 copies) and PL3 (14 copies) were presented in P. eryngii than that in other fungi (Supplementary S1), which were potentially involved in biotransformation of lignocellulosic compounds. Family AA1 belonging to multicomponent oxidases was considered as a characteristic of lignin-degrading white rot fungi (Kersten and Cullen, 2007). AA7 enzymes were potentially involved in the biotransformation or detoxification of lignocellulosic compounds (Levasseur et al., 2013). AA8 belonging to iron reductase domain might be involved in the non-enzymatic cellulose chain breakage and AA9 copper-dependent lytic polysaccharide monoxygenases were believed to act on cellulose directly (Levasseur et al., 2013). CE10 were reported to have a wide range of substrate specificities with activities of carboxylesterase and endo-1,4-β-xylanase (Zhao et al., 2013). 8 family GT41 genes were only possessed in P. eryngii. Higher number of copies of family GH7 might reveal that more cellobiohydrolases were participated in the degradation of crystalline cellulose (Morin et al., 2012; Choi et al., 2013; Zhao et al., 2013; Ohm et al., 2014). Pectate lyases PL1 possessed in this fungus suggested that these enzymes could degrade pectin to associate with plants of Apiaceae for weak parasitism (Zervakis et al., 2014).

In addition, there were also other two cellulase classes for the complete degradation of cellulose by β-1,4-endoglucanase of families GH5 (16 copies) and GH45 (1 copy) and β-glucosidase of families GH1 (3 copies) and GH3 (13 copies) (Supplementary S1). Families 7 GH27 (7 copies), GH43 (6 copies), GH31 (5 copies), GH35 (4 copies), GH10 (3 copies) and so on (Supplementary S1) were involved in degradation of hemicellulose (Zhao et al., 2013).

The oxidoreductases were extracted from the proteins predicted from the genome using a combination of IPR domain searches and the JGI cluster pipeline (http://genome.jgi.doe.gov/31.SAP) according to the method published previously (Floudas et al., 2012; Morin et al., 2012). The number of oxidoreductases (41 copies) for lignin degradation, which including 13 copper radical oxidases (CRO), 11 multicomponent oxidases (MCO), 6 class II peroxidases (POD), 6 heme-thiolate peroxidases (HTP), 2 oxalate decarboxylases (ODC), 2 dye-decolorizing (Dy), 1 cellulobiose dehydrogenase (CDH) and 1 quinone reductases (QRD) (Fig. 2), was much higher than that of other saprophytic fungi (Floudas et al., 2012; Morin et al., 2012). 13 copper radical oxidases (CRO) were composed of 7 glyoxal oxidases, 3 CROs3-5, 2 CROs2 and 1 CROs1. Class II peroxidases are classified into four main groups in fungi: lignin peroxidases, manganese peroxidases, versatile peroxidases and generic peroxidases, which involved in lignin degradation and also commonly distributed in white-rot fungi (Floudas et al., 2012; Morin et al., 2012). As a characteristic of lignin-degrading white rot fungi, 9 laccases, 4 manganese peroxidases, 2 versatile peroxidases have been detected in P. eryngii. In agreement with previous study, P. eryngii strain ‘183’ could not produce the enzymes with lignin peroxidases activities (Xie et al., 2016). In addition, 147 cytochrome P450s which were related to intracellular metabolism of lignin metabolites were determined. Just like other white-rot fungi, CYP5144 (23 copies) and CYP63 (19 copies) were the most dominant, which suggested that P. eryngii strain ‘183’ could oxidize multiple xenobiotic compounds or synthesize secondary metabolites (Syed et al., 2013).

The CAZymes and oxidoreductases in the genome of P. eryngii strain ‘183’ elucidated the mechanisms of bioconversion of cellulose and lignin in this fungus and increased the diversity of the enzymes related to degradation of lignocellulose. Though functions for some of the member CAZymes and oxidoreductases could
be predicted according to genomic data, the activities need to be further studied by experimental designs.

**Nucleotide sequence accession number**

The sequences of the genome have been submitted to NCBI under the accession MAZY00000000.

**Acknowledgments**

This study was founded by the Key Project of Science and Technology for Agriculture of Shanghai (No. 2015-6-2-1); Shanghai Agriculture Applied Technology Development Program (China) (2014)-7-1-6; National Engineering Research Center of Edible Fungi (16DZ2281300).

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2016.10.007.

**References**


