

Genome sequence of a white rot fungus *Schizopora paradoxa* KUC8140 for wood decay and mycoremediation

The Faculty of Oregon State University has made this article openly available.
Please share how this access benefits you. Your story matters.

Citation	Min, B., Park, H., Jang, Y., Kim, J. J., Kim, K. H., Pangilinan, J., ... & Choi, I. G. (2015). Genome sequence of a white rot fungus <i>Schizopora paradoxa</i> KUC8140 for wood decay and mycoremediation. <i>Journal of Biotechnology</i> , 211, 42-43. doi:10.1016/j.jbiotec.2015.06.426
DOI	10.1016/j.jbiotec.2015.06.426
Publisher	Elsevier
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsfuse



Genome announcement

Genome sequence of a white rot fungus *Schizopora paradoxa* KUC8140 for wood decay and mycoremediation

Byoungnam Min^{a,1}, Hongjae Park^{a,1}, Yeongseon Jang^{b,2}, Jae-Jin Kim^b,
Kyoung Heon Kim^a, Jasmyn Pangilinan^c, Anna Lipzen^c, Robert Riley^c, Igor V. Grigoriev^c,
Joseph W. Spatafora^d, In-Geol Choi^{a,*}

^a Department of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Anam-Dong, Seongbuk-Gu, Seoul 136-713, Republic of Korea

^b Division of Environmental Science and Ecological Engineering, College of Life Sciences and Biotechnology, Korea University, Anam-Dong, Seongbuk-Gu, Seoul 136-713, Republic of Korea

^c US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, United States

^d Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331-2902, United States

ARTICLE INFO

Article history:

Received 9 June 2015

Accepted 23 June 2015

Available online 15 July 2015

Keywords:

Schizopora paradoxa KUC8140

White rot

Wood decaying fungus

Mycoremediation

ABSTRACT

Schizopora paradoxa KUC8140 is a white rot wood degrader commonly found in Korea. Tolerance to heavy metals and polycyclic aromatic hydrocarbons and dye decolorization activity make this strain a potential candidate for mycoremediation. We report the genome sequence of *S. paradoxa* KUC8140 containing 44.4 Mbp. Based on *ab initio* gene prediction, homology search and RNA-seq, total 17,098 gene models were annotated. We identified 17 lignin-modifying peroxidases and other 377 carbohydrate-active enzymes for modeling lignocellulose deconstruction and mycoremediation.

© 2015 Elsevier B.V. All rights reserved.

Soil fungi play a crucial role in deconstruction and recycling of land plant biomass. Two major wood-decay fungal groups are known: white rot and brown rot, which were categorized based on their degradation patterns. *Schizopora paradoxa* is a white rot fungus belonging to family Schizoporaceae in Basidiomycota. General habitat of *S. paradoxa* is a progressed decayed wood in all kinds of deciduous forests and in less fertile ones with deciduous trees mixed in conifer communities. *S. paradoxa* KUC8140 strain was related to most wood-decay in oak wood forests in Korea. The strain showed high tolerance to heavy metals and polycyclic aromatic hydrocarbons and dye decolorizing activity (Lee et al., 2014) that are promising traits for bioremediation. To better understand the genetic basis for wood decay and mycoremediation processes, we sequenced the first *S. paradoxa* genome as part of the 1000 Fungal Genomes Project at the US Department of Energy Joint Genome Institute (JGI) (<http://jgi.doe.gov/fungi>).

High-molecular-weight genomic DNA was extracted using a fungal DNA isolation CTAB protocol (Fulton et al., 1995). The genome of *S. paradoxa* KUC8140 was sequenced using a single 270 bp insert size of overlapped paired-end library on the Illumina HiSeq system producing over 52.4 million 150 bp reads. Quality control procedures removed artifacts, DNA spike-in, and ambiguous or low quality reads. Processed reads were subsequently assembled together with Velvet (Zerbino and Birney, 2008). The resulting assembly was used to simulate 3 kbp long mate-pair reads. In order to improve assembly quality, the virtual 3 kbp mate-pair reads were combined with the original Illumina reads by the assembly program, AllPathsLG, version R44008 (Gnerre et al., 2011). The final assembly was 44.4 Mbp presenting 1374 contigs and 1291 scaffolds. The sequencing read coverage depth of the assembly was 133.3-fold Table 1.

The genome showed 49.4% of G+C content, which was distinct from *Fomitiporia mediterranea* (40.8%) that belongs to same order, Hymenochaetales, (Floudas et al., 2012). A small portion (253 Kb, 0.6%) of repeat regions was estimated in this genome while *F. mediterranea* has a relatively high portion of the regions (23.5 Mb, 37.1%). The genome was annotated with the JGI annotation pipeline. The 17,098 protein coding genes were predicted on combination of transcriptomics data, protein homology, and model-based *ab initio* gene prediction methods. Of these, 9559

* Corresponding author.

E-mail address: igchoi@korea.ac.kr (I.-G. Choi).

¹ These authors contributed equally to this work.

² Present address: Division of Wood Chemistry & Microbiology, Korea Forest Research Institute, Seoul 130-712, Republic of Korea.

Table 1
S. paradoxa KUC8140 genome features.

Assembly statistics	
Genome assembly size (Mbp)	44.41
Sequencing read coverage depth (fold)	133.3
Number of contigs	1374
Number of scaffolds	1291
Scaffold N50 (Mbp)	0.12
Number of gaps	51
Annotation statistics	
Number of gene models	17,098
Average gene length (bp)	1735
Average exon length (bp)	246
Average intron length (bp)	68
Average protein length (aa)	413
Number of exons per gene	5.78

(56%), 7117 (42%) and 11,814 (69%) genes were supported by alignment with KEGG genes (Kanehisa and Goto, 2000), Pfam domains (Finn et al., 2010), and NCBI non-redundant proteins respectively with E-value threshold $1e-05$ for BLAST search. Most predicted genes were validated by RNA-Seq data with half supported over the entire gene length.

Carbohydrate-active enzymes (CAZymes) (Lombard et al., 2014) were predicted by the CAZymes Analysis Toolkit (CAT) software with E-value threshold $1e-02$. The 394 CAZyme-coding genes were identified in *S. paradoxa* genome. Predicted CAZymes were classified into many functional families encompassing 178 glycoside hydrolases, 52 carbohydrate esterases, 55 glycosyl transferases, 6 polysaccharide lyases, 58 carbohydrate-binding modules, and 66 auxiliary activity (AA) families. The AA2 family includes class II lignin-modifying peroxidases such as manganese peroxidase (EC 1.11.1.13), versatile peroxidase (EC 1.11.1.16) and lignin peroxidase (EC 1.11.1.14). The abundance of AA2 family members has been a genomic feature in most white rot fungi (Kirk and Farrell, 1987). *S. paradoxa* KUC8140 genome has 17 AA2 family members, consistent with its lignin degradation activity.

Accession number and availability of the genome sequence and the strain

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LBNM00000000. The version described in this paper is version LBNM01000000. The genome assembly and functional annotation are available at the DOE JGI Genome Portal MycoCosm (Grigoriev et al., 2014) (<http://genome.jgi.doe.gov/Schpa1/Schpa1.home.html>). The strain has been deposited at Korean Agricultural Culture Collection

(KACC) with the accession number 43375 (<http://www.genebank.go.kr/>). The strain is available from KACC or Korea University Collection (KUC8140) on request.

Acknowledgements

This work is supported by Korea University and the Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01044003) Rural Development Administration, Republic of Korea. The work conducted by Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

- Finn, R.D., Mistry, J., Tate, J., Coghill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., 2010. The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222.
- Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martinez, A.T., Otilar, R., Spatafora, J.W., Yadav, J.S., Aerts, A., Benoit, I., Boyd, A., Carlson, A., Copeland, A., Coutinho, P.M., de Vries, R.P., Ferreira, P., Findley, K., Foster, B., Gaskell, J., Glotzer, D., Gorecki, P., Heitman, J., Hesse, C., Hori, C., Igarashi, K., Jurgens, J.A., Kallen, N., Kersten, P., Kohler, A., Kues, U., Kumar, T.K., Kuo, A., LaButti, K., Larrondo, L.F., Lindquist, E., Ling, A., Lombard, V., Lucas, S., Lundell, T., Martin, R., McLaughlin, D.J., Morgenstern, I., Morin, E., Murat, C., Nagy, L.G., Nolan, M., Ohm, R.A., Patyshakuliyeva, A., Rokas, A., Ruiz-Duenas, F.J., Sabat, G., Salamov, A., Samejima, M., Schmutz, J., Slot, J.C., St John, F., Stenlid, J., Sun, H., Sun, S., Syed, K., Tsang, A., Wiebenga, A., Young, D., Pisabarro, A., Eastwood, D.C., Martin, F., Cullen, D., Grigoriev, I.V., Hibbett, D.S., 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**, 1715–1719.
- Fulton, T.M., Chunwongse, J., Tanksley, S.D., 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol. Biol. Rep.* **13**, 207–209.
- Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., Berlin, A.M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E.S., Jaffe, D.B., 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1513–1518.
- Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otilar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**, D699–D704.
- Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30.
- Kirk, T.K., Farrell, R.L., 1987. Enzymatic combustion: the microbial degradation of lignin. *Annu. Rev. Microbiol.* **41**, 465–505.
- Lee, H., Jang, Y., Choi, Y.S., Kim, M.J., Lee, J., Lee, H., Hong, J.H., Lee, Y.M., Kim, G.H., Kim, J.J., 2014. Biotechnological procedures to select white rot fungi for the degradation of PAHs. *J. Microbiol. Methods* **97**, 56–62.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–495.
- Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829.