

Genome Sequence of *Pandoraea* sp. ISTKB, a Lignin-Degrading Betaproteobacterium, Isolated from Rhizospheric Soil

Madan Kumar,^a Rajesh Kumar Gazara,^b Sandhya Verma,^b Manish Kumar,^a Praveen Kumar Verma,^b Indu Shekhar Thakur^a

School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, India^a; National Institute of Plant Genome Research, New Delhi, India^b

R.K.G. and S.V. contributed equally to this article.

We report here the genome sequence of *Pandoraea* sp. ISTKB, a betaproteobacterium isolated from rhizospheric soil in the backwaters of Alappuzha, Kerala, India. The strain is alkalotolerant and grows on medium containing lignin as a sole carbon source. Genes and pathways related to lignin degradation were complemented by genomic analysis.

Received 9 September 2016 Accepted 12 September 2016 Published 3 November 2016

Citation Kumar M, Gazara RK, Verma S, Kumar M, Verma PK, Thakur IS. 2016. Genome sequence of *Pandoraea* sp. ISTKB, a lignin-degrading betaproteobacterium, isolated from rhizospheric soil. *Genome Announc* 4(6):e01240-16. doi:10.1128/genomeA.01240-16.

Copyright © 2016 Kumar et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Praveen Kumar Verma, pkv@nipgr.ac.in, or Indu Shekhar Thakur, isthakur@hotmail.com.

Lignin is recalcitrant to degradation due to its complex and heterogeneous structure (1). Although industry based on plant polysaccharides will add to the already large amount of lignin generation (2), sustainable biorefinery can be possible if polysaccharides and lignin are utilized in improved ways. Fungi are well-known lignin degraders, but in recent years the focus has been on bacterial degradation, since bacteria are more adapted to extreme environments and are easy to manipulate at the genetic level (3, 4). *Pandoraea* sp. ISTKB is an alkalotolerant strain, and the degradation and decolorization potential of kraft lignin and various dyes have been investigated with this strain (5). Pretreatment of sugarcane bagasse by *Pandoraea* sp. ISTKB under submerged and solid state conditions was also studied in detail (6). Insight into genomic analysis might help us to understand the novel enzymes and pathways responsible for lignin degradation and biovalorization.

The draft genome of *Pandoraea* sp. ISTKB was sequenced using the Illumina MiSeq platform, which generated 2,139,250 paired-end reads of 151-bp length. Raw reads were filtered using the NGS toolkit version 2.3.1 to obtain 1,597,718 high-quality paired-end reads. The genome assembly was performed using Velvet version 1/2/10 (7), SOAPdenovo (8), and gsAssembler using a *k*-mer value of 57 for primary assembly. Scaffolding for the primary assembled contigs was performed using SSPACE (9) to generate 115 scaffolds, with an N_{50} scaffold size of 132,761 bp (maximum scaffold length: 545,170 bp; minimum scaffold length: 258 bp). Genes were identified using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). The functional annotation of the genome sequence was also performed using Pfam (10). Pathway analysis was carried out using the KEGG Automatic Annotation Server (KAAS) (11).

The total genome size of *Pandoraea* sp. ISTKB is 6.37 Mb, and a coverage of 65 \times was achieved. The strain had a G+C content of 62.05% and a total of 5,356 protein-coding genes were predicted. The bacterium contained 54 tRNAs, six rRNAs (two copies each for 5S, 16S, and 23S), and four ncRNAs. Additionally, 154 pseudogenes were also identified, out of which

12 were frame-shifted pseudogenes. With PGAP, 68.59% of the predicted proteins were annotated. In addition, Pfam annotation was assigned to 4,603 genes (85.94%), and the KAAS tool predicted 1,351 genes (25.22%) to be involved in different pathways.

The draft genome revealed the presence of putative genes responsible for the degradation of lignin and lignin-derived aromatic compounds. The lignin-degrading enzymes identified were DyP-type peroxidases, peroxidases, multicopper oxidases, esterases, coniferyl-alcohol dehydrogenase, coniferyl-aldehyde dehydrogenase, etherases, methyltransferases, and vanillate O-demethylase oxidoreductase. Genes responsible for the catabolism of aromatic compounds identified in the genome assembly were 4-hydroxybenzoate 3-monooxygenase, vanillate monooxygenase ferredoxin subunit, salicylate hydroxylase, protocatechuate 3,4-dioxygenase, gentisate 1,2-dioxygenase, catechol 2,3-dioxygenase, extradiol ring-cleavage dioxygenase, catechol 1,2-dioxygenase, phthalate 4,5-dioxygenase, protocatechuate 4,5-dioxygenase, hydroxyquinol 1,2-dioxygenase, 3-phenylpropionate dioxygenase, and various other oxidoreductases. Additionally, genes for oxidative stress and redox signaling, such as glutathione peroxidase, glutathione synthase, glutathione reductase, thioredoxin, thioredoxin reductase, catalases, and superoxide dismutases, were also present. These findings indicate that *Pandoraea* sp. ISTKB could potentially have an application in lignocellulosic biomass valorization.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number MAOS00000000. The version described in this paper is the first version, MAOS01000000.

ACKNOWLEDGMENTS

Madan Kumar is grateful to the Council of Scientific and Industrial Research (CSIR), New Delhi, for Senior Research Fellowship. P.K.V. thanks the National Institute of Plant Genome Research, New Delhi, for financial support.

FUNDING INFORMATION

We have not received grants from funding agencies. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. Brown ME, Chang MC. 2014. Exploring bacterial lignin degradation. *Curr Opin Chem Biol* 19:1–7. <http://dx.doi.org/10.1016/j.cbpa.2013.11.015>.
2. Fuchs G, Boll M, Heider J. 2011. Microbial degradation of aromatic compounds—from one strategy to four. *Nat Rev Microbiol* 9:803–816. <http://dx.doi.org/10.1038/nrmicro2652>.
3. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M, Langan P, Naskar AK, Saddler JN, Tschaplinski TJ, Tuskan GA, Wyman CE. 2014. Lignin valorization: improving lignin processing in the biorefinery. *Science* 344:1246843. <http://dx.doi.org/10.1126/science.1246843>.
4. Beckham GT, Johnson CW, Karp EM, Salvachúa D, Vardon DR. 2016. Opportunities and challenges in biological lignin valorization. *Curr Opin Biotechnol* 42:40–53. <http://dx.doi.org/10.1016/j.copbio.2016.02.030>.
5. Kumar M, Singh J, Singh MK, Singhal A, Thakur IS. 2015. Investigating the degradation process of kraft lignin by β -proteobacterium, *Pandoraea* sp. ISTKB. *Environ Sci Pollut Res Int* 22:15690–15702. <http://dx.doi.org/10.1007/s11356-015-4771-5>.
6. Kumar M, Singhal A, Thakur IS. 2016. Comparison of submerged and solid state pretreatment of sugarcane bagasse by *Pandoraea* sp. ISTKB: enzymatic and structural analysis. *Bioresour Technol* 203:18–25. <http://dx.doi.org/10.1016/j.biortech.2015.12.034>.
7. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
8. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam T-W, Li Y, Xu X, Wong GK, Wang J. 2014. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30:1660–1666. <http://dx.doi.org/10.1093/bioinformatics/btu077>.
9. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <http://dx.doi.org/10.1093/bioinformatics/btq683>.
10. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42:D222–D230. <http://dx.doi.org/10.1093/nar/gkt1223>.
11. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185. <http://dx.doi.org/10.1093/nar/gkm321>.