

# Draft Genome Sequence of a Xylanase-Producing Bacterial Strain, *Cellvibrio mixtus* J3-8

Yi-Rui Wu,<sup>a,c</sup> Bokun Lin,<sup>b,c</sup> Yuli Yu<sup>b,c</sup>

Department of Civil and Environmental Engineering, National University of Singapore, Singapore<sup>a</sup>; School of Public Health, Guangdong Medical College, Dongguan, Guangdong, People's Republic of China<sup>b</sup>; Department of Biology, Shantou University, Shantou, Guangdong, People's Republic of China<sup>c</sup>

Y.-R.W. and B.L. contributed equally to this work.

**The xylanase-producing bacterial strain *Cellvibrio mixtus* J3-8 was isolated from grassland giant snails. The draft genome of strain J3-8 comprises 5,171,890 bp in 152 contigs with a G+C content of 46.66%. This is the first genome report about this bacterial species.**

Received 1 November 2014 Accepted 7 November 2014 Published 11 December 2014

**Citation** Wu Y-R, Lin B, Yu Y. 2014. Draft genome sequence of a xylanase-producing bacterial strain, *Cellvibrio mixtus* J3-8. *Genome Announc.* 2(6):e01281-14. doi:10.1128/genomeA.01281-14.

**Copyright** © 2014 Wu et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Bokun Lin, [albertbclin@foxmail.com](mailto:albertbclin@foxmail.com).

Recent research on the bioconversion from lignocellulosic materials into use in the biofuel industry has received increasing attention (1). Different from the costly and environmental unfriendly acid or heat pretreatment, enzymatic hydrolysis is considered to be one of the important and indispensable processes for bioconversion (2). Most bacteria have been reported as potential degraders of cellulose, dextran, xylan, chitin, and starch, including members of the *Cellvibrio* genus (2, 3). *Cellvibrio mixtus* J3-8 is a newly xylanase-producing soil bacterial strain isolated from grassland giant snails. The 16S rRNA gene of strain J3-8 shows 99.5% identity to that of *C. mixtus* subsp. *mixtus* strain ACM 2601 (GenBank accession no. AF448515), for which genomic data are not available. As *Cellvibrio* spp. have been considered the gene pool for degrading various polysaccharides (e.g., cellulose, dextran, chitin, and starch) (3, 4), here, we present a draft genome sequence of *C. mixtus* J3-8.

The genome of *C. mixtus* J3-8 was sequenced by a whole-genome shotgun strategy using the high-throughput Illumina HiSeq 2000 at the Beijing Genomics Institute (Shenzhen, China). A total of 1,084,620 reads with a 500-bp insert size and totaling 542.31 Mbp were received, providing 105-fold coverage. The genome sequences were assembled *in silico* using SOAPdenovo (version 1.05) (5), resulting in 152 contigs with an  $N_{50}$  length of 176,538 bp and a total length of 5,171,890 bp for the whole genome. These contigs were assembled into 50 scaffolds with a maximum fragment length of 1,327,211 bp. Putative protein-coding sequences were identified by Glimmer (version 3.02) (6) and analyzed by BLASTp. The functions of the predicted protein-coding genes were annotated via the databases Kyoto Encyclopedia of Genes and Genomes (KEGG) (7), Clusters of Orthologous Groups (COG) (8), UniProtKB/Swiss-Prot (9), and UniProtKB/TrEMBL (10). The identification of tRNAs and rRNAs was performed by tRNAscan-SE 1.21 (11) and rRNAmmer 1.2 (12), respectively.

The draft genome has 5,171,890 bases and a G+C content of 46.66%, and it contains 32 tRNA genes and 3 rRNAs (5S, 16S, and 23S rRNAs). Among 4,916 predicted genes (~88.82% of the total

nucleotides), only 2,845 protein-coding sequences (CDSs) were identified (57.8% of the total obtained genes). A total of 2,030, 1,771, 825, and 478 proteins were annotated with the UniProtKB/TrEMBL, KEGG, COG, and UniProtKB/Swiss-Prot databases, respectively. The alignment analysis showed low identities for most genes, as 83.7% of them were lower than 90%.

In addition, there are a total of 145 CDSs involved in carbohydrate transport and metabolism, and 46 putative carbohydrate-hydrolase genes were detected in the whole genome, including those encoding cellulases, xylanases, glucanases, glucosidases, and xylosidases. The identities of these genes range only from 40.8% to 88.8%, which is significantly different from the submitted sequences in the GenBank database, even those in other *Cellvibrio* species. The relative poorly annotation percentage and the low identities of the genes indicate the uniqueness of this novel strain.

**Nucleotide sequence accession numbers.** The draft genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [ALBT000000000](https://www.ncbi.nlm.nih.gov/nuccore/ALBT000000000). The version described in this paper is the first version, ALBT01000000.

## ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (grant no. 31200077) and Guangdong Natural Science Foundation (grant no. S2012040006279).

## REFERENCES

- Bhat MK. 2000. Cellulases and related enzymes in biotechnology. *Biotechnol. Adv.* 18:355–383. [http://dx.doi.org/10.1016/S0734-9750\(00\)00041-0](http://dx.doi.org/10.1016/S0734-9750(00)00041-0).
- DeBoy RT, Mongodin EF, Fouts DE, Tailford LE, Khouri H, Emerson JB, Mohamoud Y, Watkins K, Henrissat B, Gilbert HJ, Nelson KE. 2008. Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. *J. Bacteriol.* 190:5455–5463. <http://dx.doi.org/10.1128/JB.01701-07>.
- Fontes CM, Gilbert HJ, Hazlewood GP, Clarke JH, Prates JA, McKie VA, Nagy T, Fernandes TH, Ferreira LM. 2000. A novel *Cellvibrio mixtus* family 10 xylanase that is both intracellular and expressed under non-inducing conditions. *Microbiology* 146(Pt 8):1959–1967.
- Carrard G, Koivula A, Söderlund H, Béguin P. 2000. Cellulose-binding

- domains promote hydrolysis of different sites on crystalline cellulose. *Proc. Natl. Acad. Sci. U. S. A.* 97:10342–10347. <http://dx.doi.org/10.1073/pnas.160216697>.
5. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J. 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272. <http://dx.doi.org/10.1101/gr.097261.109>.
  6. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <http://dx.doi.org/10.1093/bioinformatics/btm009>.
  7. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36:D480–D484. <http://dx.doi.org/10.1093/nar/gkm882>.
  8. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28:33–36. <http://dx.doi.org/10.1093/nar/28.1.33>.
  9. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. 2007. UniProtKB /Swiss-Prot. *Methods Mol. Biol.* 406:89–112. [http://dx.doi.org/10.1007/978-1-59745-535-0\\_4](http://dx.doi.org/10.1007/978-1-59745-535-0_4).
  10. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilboud S, Schneider M. 2003. The Swiss-Prot protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31:365–370. <http://dx.doi.org/10.1093/nar/gkg095>.
  11. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
  12. Lagesen, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.