



Whole-Genome Sequence of the 1,4-Dioxane-Degrading Bacterium *Mycobacterium dioxanotrophicus* PH-06

Ya He,^a Kangfei Wei,^b Kaiwei Si,^b Jacques Mathieu,^a Mengyan Li,^c Pedro J. J. Alvarez^a

Department of Civil and Environmental Engineering, Rice University, Houston, Texas, USA^a; International Academic Support and Delivery Unit, Beijing Genomic Institute, Shenzhen, Guangzhou, China^b; Department of Chemistry and Environmental Science, New Jersey Institute of Technology, Newark, New Jersey, USA^c

ABSTRACT We report here the complete genome sequence of *Mycobacterium dioxanotrophicus* PH-06, which is capable of using 1,4-dioxane as a sole source of carbon and energy. The reported sequence will enable the elucidation of this novel metabolic pathway and the development of molecular biomarkers to assess bioremediation potential at contaminated sites.

Mycobacterium dioxanotrophicus PH-06 was isolated in South Korea from river sediment that had been contaminated with 1,4-dioxane (dioxane) for more than one decade (1). PH-06 utilizes dioxane, a groundwater contaminant of emerging concern, as its sole carbon and energy source. However, it does not harbor the well-studied monooxygenase gene cluster *thmADBC* that codes for the initiation of dioxane biodegradation in *Pseudonocardia dioxanivorans* CB1190 (2–4). Therefore, the genome sequence of PH-06 furthers our capability to discern novel dioxane biodegradation pathways and facilitates the development of biomarkers to assess dioxane bioremediation potential. Additionally, knowledge of the PH-06 genome broadens our understanding of the genus *Mycobacterium* and enables an assessment of PH-06 survival and performance in bioaugmentation applications.

PH-06 was grown in ammonium mineral salts medium (5) amended with 500 mg/L of dioxane and incubated at 30°C while shaking at 150 rpm. Cells were harvested during exponential growth, and genomic DNA was extracted using the UltraClean microbial DNA isolation kit (Mo Bio, Inc.). Whole-genome sequencing was conducted using both the PacBio RS II (Yale Center for Genome Analysis, <http://ycga.yale.edu>) and Illumina HiSeq 4000 (Beijing Genomic Institute [BGI], <http://www.genomics.cn>) platforms. The whole genome was assembled and annotated in collaboration with BGI as follows. First, prior to assembly, *k*-mer analysis was used to evaluate genome size, heterogeneity, and repeat information based on the data obtained by Illumina sequencing (6). Second, PacBio RS II reads were assembled using the RS_HGAP assembly of SMRT Analysis version 2.3.0 (<https://github.com/PacificBiosciences/SMRT-Analysis>) to obtain the main contig with a length close to the estimated genome size, and Illumina reads were used to correct and optimize the assembly results. Third, the contig's bases were corrected with Quiver, Pilon, SOAPsnp, SOAPindel (<http://soap.genomics.org.cn>), and GATK (<http://www.broadinstitute.org/gatk>). Fourth, contig circle analysis was completed by verifying overlap regions. Fifth, Glimmer (7–9), TRF (10), RNAmmer version 1.2 (11), tRNAscan-SE version 1.3.1 (12), Infernal (13), Rfam (14), and BLAST were used to predict genes, repeat sequences, rRNAs, tRNAs, and noncoding RNAs (ncRNAs). Finally, predicted genes were analyzed against the GO (15), KEGG (16–19), COG (20), NR, Swiss-Prot (21), PHI (22), VFDB (23), ARDB (24), and CAZy (25) databases to annotate gene function and identify metabolic pathways, pathogenicity, and drug resistance.

Received 15 May 2017 Accepted 31 May 2017 Published 31 August 2017

Citation He Y, Wei K, Si K, Mathieu J, Li M, Alvarez PJJ. 2017. Whole-genome sequence of the 1,4-dioxane-degrading bacterium *Mycobacterium dioxanotrophicus* PH-06. Genome Announc 5:e00625-17. <https://doi.org/10.1128/genomeA.00625-17>.

Copyright © 2017 He 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 Mengyan Li, mengyan.li@njit.edu, or Pedro J. J. Alvarez, alvarez@rice.edu.

The PH-06 genome consists of 4 contigs, including the chromosome (circular, 7.6 Mb), Plasmid_1 (circular, 156 kb), Plasmid_2 (circular, 153 kb), Plasmid_3 (linear, 106 kb), and Plasmid_4 (linear, 70 kb), and has an average G+C content of 66.46%. A total of 7,339 protein-encoding genes, 83 tRNAs, 9 rRNAs, and 4 ncRNAs are present. KEGG database analysis revealed genes encoding the complete citric acid and pentose phosphate pathways. Furthermore, 1,071 genes appear to be involved in the metabolism of xenobiotics. Pathogenicity analysis indicates that PH-06 harbors no known toxins or pathogenicity islands, suggesting it may be safe for bioaugmentation (22–24).

One gene cluster encoding putative propane monoxygenase is located on Plasmid_3. This gene cluster has high similarity to genes in (hydrocarbon-degrading) *Rhodococcus wratislaviensis* IFP2016 (89%) and *Mycobacterium chubuense* NBB4 (86%) (26, 27). Further studies are needed to determine the role of this gene cluster in dioxane biodegradation.

Accession number(s). The whole-genome sequence of *M. dioxanotrophicus* PH-06 has been deposited in GenBank under the accession numbers [CP020809](#) to [CP020813](#).

ACKNOWLEDGMENTS

This work was funded by SERDP (grant no. ER-2301).

We declare no competing financial interest.

We thank Yoon-Seok Chang for providing *M. dioxanotrophicus* PH-06.

REFERENCES

- Kim YM, Jeon JR, Murugesan K, Kim EJ, Chang YS. 2009. Biodegradation of 1,4-dioxane and transformation of related cyclic compounds by a newly isolated *Mycobacterium* sp. PH-06. *Biodegradation* 20:511–519. <https://doi.org/10.1007/s10532-008-9240-0>.
- Sales CM, Mahendra S, Grostern A, Parales RE, Goodwin LA, Woyke T, Nolan M, Lapidus A, Chertkov O, Ovchinnikova G, Sczyrba A, Alvarez-Cohen L. 2011. Genome sequence of the 1,4-dioxane-degrading *Pseudonocardia dioxanivorans* strain CB1190. *J Bacteriol* 193:4549–4550. <https://doi.org/10.1128/JB.00415-11>.
- Grostern A, Sales CM, Zhuang WQ, Erbilgin O, Alvarez-Cohen L. 2012. Glyoxylate metabolism is a key feature of the metabolic degradation of 1,4-dioxane by *Pseudonocardia dioxanivorans* strain CB1190. *Appl Environ Microbiol* 78:3298–3308. <https://doi.org/10.1128/AEM.00067-12>.
- Sales CM, Grostern A, Parales JV, Parales RE, Alvarez-Cohen L. 2013. Oxidation of the cyclic ethers 1,4-dioxane and tetrahydrofuran by a monoxygenase in two *Pseudonocardia* species. *Appl Environ Microbiol* 79:7702–7708. <https://doi.org/10.1128/AEM.02418-13>.
- Parales RE, Adamus JE, White N, May HD. 1994. Degradation of 1,4-dioxane by an actinomycete in pure culture. *Appl Environ Microbiol* 60:4527–4530.
- Liu B, Shi Y, Yuan J, Hu X, Zhang H, Li N, Li Z, Chen Y, Mu D, Fan W. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv:1308.2012*. <https://arxiv.org/abs/1308.2012>.
- Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27:4636–4641. <https://doi.org/10.1093/nar/27.23.4636>.
- Salzberg SL, Delcher AL, Kasif S, White O. 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 26:544–548. <https://doi.org/10.1093/nar/26.2.544>.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* 23:673–679. <https://doi.org/10.1093/bioinformatics/btm009>.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580. <https://doi.org/10.1093/nar/27.2.573>.
- Lagesen K, 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. <https://doi.org/10.1093/nar/gkm160>.
- 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.
- Nawrocki EP. 2014. Annotating functional RNAs in genomes using infernal. *Methods Mol Biol* 1097:163–197. https://doi.org/10.1007/978-1-62703-709-9_9.
- Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. 2009. Rfam: updates to the RNA families database. *Nucleic Acids Res* 37:D136–D140. <https://doi.org/10.1093/nar/gkn766>.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. *Nat Genet* 25:25–29. <https://doi.org/10.1038/75556>.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–D280. <https://doi.org/10.1093/nar/gkh063>.
- Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 34:D354–D357. <https://doi.org/10.1093/nar/gkj102>.
- Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 278:631–637. <https://doi.org/10.1126/science.278.5338.631>.
- Kanehisa M. 1997. A database for post-genome analysis. *Trends Genet* 13:375–376. [https://doi.org/10.1016/S0168-9525\(97\)01223-7](https://doi.org/10.1016/S0168-9525(97)01223-7).
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <https://doi.org/10.1186/1471-2105-4-41>.
- Magrane M, UniProt Consortium. 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database* 2011:bar009. <https://doi.org/10.1093/database/bar009>.
- Vargas WA, Martín JMS, Rech GE, Rivera LP, Benito EP, Díaz-Minguez JM, Thon MR, Sukno SA. 2012. Plant defense mechanisms are activated during biotrophic and necrotrophic development of *Colletotricum graminicola* in maize. *Plant Physiol* 158:1342–1358. <https://doi.org/10.1104/pp.111.190397>.
- Chen L, Xiong Z, Sun L, Yang J, Jin Q. 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence

- factors. *Nucleic Acids Res* 40:D641–D645. <https://doi.org/10.1093/nar/gkr989>.
24. Liu B, Pop M. 2009. ARDB—antibiotic resistance genes database. *Nucleic Acids Res* 37:D443–D447. <https://doi.org/10.1093/nar/gkn656>.
25. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 37:D233–D238. <https://doi.org/10.1093/nar/gkn663>.
26. Coleman NV, Yau S, Wilson NL, Nolan LM, Migocki MD, Ly MA, Crossett B, Holmes AJ. 2011. Untangling the multiple monooxygenases of *Mycobacterium chubuense* strain NBB4, a versatile hydrocarbon degrader. *Environ Microbiol Rep* 3:297–307. <https://doi.org/10.1111/j.1758-2229.2010.00225.x>.
27. Auffret MD, Yergeau E, Labbé D, Fayolle-Guichard F, Greer CW. 2015. Importance of *Rhodococcus* strains in a bacterial consortium degrading a mixture of hydrocarbons, gasoline, and diesel oil additives revealed by metatranscriptomic analysis. *Appl Microbiol Biotechnol* 99:2419–2430. <https://doi.org/10.1007/s00253-014-6159-8>.