Draft Genome Sequence of *Advenella kashmirensis* Strain W13003, a Polycyclic Aromatic Hydrocarbon-Degrading Bacterium

Xinxin Wang,[a,b] Decai Jin,[c] Lisha Zhou,[d] Liang Wu,[b,e] Wei An,[b] Lin Zhao[a]

School of Environmental Science and Engineering, Tianjin University, Tianjin, China[a]; China Offshore Environmental Service Co. Ltd., Tianjin, China[b]; College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China[c]; Shenzhen Key Laboratory of Environmental Microbial Genomics and Application, BGI, Shenzhen, China[d]; College of Environmental Science and Engineering, Ocean University of China, Qingdao, China[e]

X.W. and D.J. made equal contributions to this article.

Advenella kashmirensis **strain W13003 is a polycyclic aromatic hydrocarbon (PAH)-degrading bacterium isolated from PAH-contaminated marine sediments. Here, we report the 4.8-Mb draft genome sequence of this strain, which will provide insights into the diversity of *A. kashmirensis* and the mechanism of PAH degradation in the marine environment.**

Address correspondence to Xinxin Wang, wangxx200899@yahoo.com.

Polycyclic aromatic hydrocarbons (PAHs) are widely distributed in marine sediments. These hydrocarbons can be degraded by many bacteria, which have attracted increasing attention (1). *Advenella kashmirensis* W13003 was isolated in PAH-contaminated sediments from Bohai Bay in China, and it can degrade PAHs, including pyrene and phenanthrene. Genomic analysis provides an opportunity to understand microbial PAH degradation (2–4). However, only one strain, i.e., *A. kashmirensis* WT001, has been sequenced (5). Here, the draft genome sequence of PAH-degrading *A. kashmirensis* W13003 is presented.

Genomic DNA was extracted using a commercial DNA isolation kit and sequenced using an Illumina HiSeq 2000 platform (San Diego, CA) with a whole-genome shotgun (WGS) strategy. The sequencing produced 9,871,420 paired-end reads with an average insert size of 300 bp, yielding about 200-fold coverage. The filtered reads were assembled, scaffolded, gap filled, and validated using SOAP*denovo* version 2.04 (6), SSPACE version 2.0 (7), Gap-Filler version 1.10 (8), and BWA version 0.7.4 (9). The final assembly consisted of 15 contigs, with the largest length being 1,071,036 bp, which were assembled into 12 scaffolds, the largest of which is 2,399,045 bp. The genome sequence was annotated using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

The genome consists of 4.8 Mb with a G+C content of 55.1%. A total of 4,507 coding sequences (CDSs), 33 pseudogenes, 1 noncoding RNA (ncRNA), 38 tRNA genes, and 1 rRNA operon were identified. Of the CDSs, 88.3% were assigned to clusters of orthologous groups (COGs), with amino acid transport and metabolism being the most abundant class, and 50.2% were annotated into 1,688 KEGG orthologous groups by using KAAS (10), involving 193 metabolic pathways. Clustered regularly interspaced short palindromic repeat (CRISPR) elements were not found using CRISPRFinder (11). The plasmid partitioning gene, *parB*, was detected on contig 7, which suggests the presence of plasmids. Eighty-nine tandem repeats were detected, as revealed by Tandem Repeats Finder version 4.07 (12). Only one incomplete prophage region was detected using PHAST (13). The IS3 and IS21 families dominate the insertion sequence (IS) elements, as revealed by IS-Finder (14). An average nucleotide identity (ANI) analysis (15) revealed that *A. kashmirensis* W13003 is phylogenetically related to *A. kashmirensis* WT001 (78.7%).

Seven genes were identified as being involved in hydrocarbon degradation, including 2 alkane 1-monooxygenase genes, 1 protocatechuate 3,4-dioxygenase gene, 1 ring-cleavage dioxygenase gene, 1 biphenyl 2,3-dioxygenase gene, and 2 terephthalate 1,2-dioxygenase genes. Moreover, 6 genes were identified as being involved in compatible solute synthesis and uptake, including 3 glycine/betaine ABC transport genes, 2 betaine-aldehyde dehydrogenase genes, and 1 ectoine synthase gene, which may enhance tolerance to salt stress in the ocean. An inorganic sulfur compound oxidation gene cluster was detected similar to that of *A. kashmirensis* WT001. Information about the genome sequence of *A. kashmirensis* W13003 will be helpful for understanding the diversity of *A. kashmirensis* and the mechanisms of PAH degradation in the marine environment.

Nucleotide sequence accession number. The draft genome sequence of *A. kashmirensis* W13003 has been deposited in GenBank under the accession no. AYXT00000000. The version described in this paper is the first version.

REFERENCES

1. **Mahmoudi N, Porter TM, Zimmerman AR, Fulthorpe RR, Kasozi GN, Silliman BR, Slater GF.** 2013. Rapid degradation of deepwater horizon

spilled oil by indigenous microbial communities in Louisiana saltmarsh sediments. Environ. Sci. Technol. **47:**13303–13312. http://dx.doi.org/10.1021/es4036072.

2. **Cui Z, Xu G, Li Q, Gao W, Zheng L.** 2013. Genome sequence of the pyrene- and fluoranthene-degrading bacterium *Cycloclasticus* sp. strain PY97M. Genome Announc. **1**(4):e00536-13. http://dx.doi.org/10.1128/genomeA.00536-13.

3. **Maeda AH, Nishi S, Ozeki Y, Ohta Y, Hatada Y, Kanaly RA.** 2013. Draft genome sequence of *Sphingobium* sp. strain KK22, a high-molecular-weight polycyclic aromatic hydrocarbon-degrading bacterium isolated from cattle pasture soil. Genome Announc. **1**(6):e00911-13. http://dx.doi.org/10.1128/genomeA.00911-13.

4. **Cui Z, Gao W, Li Q, Xu G, Zheng L.** 2013. Genome sequence of the polycyclic aromatic hydrocarbon-degrading bacterium strain *Marinobacter nanhaiticus* D15-8W^T. Genome Announc. **1**(3):e00301-13. http://dx.doi.org/10.1128/genomeA.00301-13.

5. **Ghosh W, George A, Agarwal A, Raj P, Alam M, Pyne P, Das Gupta SK.** 2011. Whole-genome shotgun sequencing of the sulfur-oxidizing chemoautotroph *Tetrathiobacter kashmirensis*. J. Bacteriol. **193:**5553–5554. http://dx.doi.org/10.1128/JB.05781-11.

6. **Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J.** 2012. SOAP*denovo*2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience **1:**18. http://dx.doi.org/10.1186/2047-217X-1-18.

7. **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.

8. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with GapFiller. Genome Biol. **13:**R56. http://dx.doi.org/10.1186/gb-2012-13-6-r56.

9. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.

10. **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.

11. **Grissa I, Vergnaud G, Pourcel C.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **35:**W52–W57. http://dx.doi.org/10.1093/nar/gkm360.

12. **Benson G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27:**573–580. http://dx.doi.org/10.1093/nar/27.2.573.

13. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: a fast phage search tool. Nucleic Acids Res. **39:**W347–W352. http://dx.doi.org/10.1093/nar/gkr485.

14. **Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M.** 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. **34:**D32–D36. http://dx.doi.org/10.1093/nar/gkj014.

15. **Richter M, Rosselló-Móra R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. **106:**19126–19131. http://dx.doi.org/10.1073/pnas.0906412106.