Complete Genome Sequence of *Pseudomonas* sp. Strain phDV1, an Isolate Capable of Efficient Degradation of Aromatic Hydrocarbons

Hao Xie,[a] Giannis Valsamidis,[b] Eirini Mathioudaki,[b] Georgios Tsiotis[b]

[a]Department of Molecular Membrane Biology, Max Planck Institute of Biophysics, Frankfurt am Main, Germany
[b]Department of Chemistry, University of Crete, Heraklion, Greece

ABSTRACT *Pseudomonas* sp. strain phDV1 is a Gram-negative bacterium capable of degrading aromatic hydrocarbons. Here, we present the complete genome sequence of this strain, which consists of 4,727,682 bp, with a 62.3% G+C content and 4,574 genes. Multiple genes responsible for the degradation of aromatics are present in this strain.

*P*seudomonas sp. strain phDV1 was isolated from a petroleum-contaminated site in Denmark (1, 2). *Pseudomonas* sp. phDV1 was shown, through gas chromatography-mass spectrometry (GC-MS) and proteomics analysis, to efficiently metabolize phenol, toluene, *o*-cresol, naphthalene, and 1,2,3-trimethylbenzene (1–4). The genome of *Pseudomonas* sp. phDV1 was chosen for sequencing due to its ability to degrade aromatics and grow in harsh hydrocarbon-containing environments.

A single colony of *Pseudomonas* sp. phDV1 was picked and cultured in lysogeny broth at 32°C. Genomic DNA was extracted from the cultures using the Gentra Puregene Yeast/Bact. kit (Qiagen, Germany). A PacBio 10-kb sequencing library was constructed using the PacBio template prep kit following the manufacturer's protocols (Pacific Biosciences, USA). The genome of *Pseudomonas* sp. phDV1 was sequenced using the PacBio RS II system on single-molecule real-time (SMRT) cells using PacBio P6-C4 chemistry at Beijing Novogene Technology Co. Ltd. (China). The raw reads were filtered in SMRT Portal (version 3.2.0), with 0.75 as the minimum read quality and 500 bp as the minimum read length. The clean data (983,266,888 bp, 98,763 reads, 208× coverage) were *de novo* assembled using the Hierarchical Genome Assembly Process (HGAP) (5) in SMRT Portal (version 3.2.0) to generate one contig without gaps. The genome of *Pseudomonas* sp. phDV1 consists of a circular chromosome, which is 4,727,682 bp in length, with a G+C content of 62.3%. No plasmid was found in this strain. The genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (6), which predicted 4,574 genes, including 4,081 coding sequences (CDSs), 411 pseudogenes, 4 copies of each of the rRNA genes (5S, 16S, and 23S), 65 tRNAs, and 5 noncoding RNAs.

For functional annotation, the predicted protein-coding sequences were searched against the Clusters of Orthologous Groups (COG) of proteins (7), Gene Ontology (GO) (8), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (9). According to the annotation results, multiple genes involved in hydrocarbon degradation were found in the genome. Among them, the operon *dmpKLMNOPQBCDEFGHI* encoding proteins involved in phenol degradation resemble the canonical genetic organization of *Pseudomonas* sp. strain CF600 type (10). The *tmoABCDEF* cluster encoding a multicomponent toluene 4-monooxygenase, involved in the degradation of toluene and various cresol compounds, is located 6 kb upstream of the *dmp* operon. In addition, homogentisate 1,2-dioxygenase, 2-alkenal reductase, P450, xylene monooxygenase,

protocatechuate 3,4-dioxygenase, and 4-carboxymuconolactone decarboxylase were present as well. The presence of these enzymes likely explains the aromatic degradation capacities of *Pseudomonas* sp. phDV1.

The 16S rRNA sequence analysis using the genomic-based 16S rRNA database (GRD; http://metasystems.riken.jp/grd/) revealed that *Pseudomonas* sp. phDV1 shares 99.93% and 99.41% similarity with *Pseudomonas pseudoalcaligenes* CECT5344 and *Pseudomonas mendocina* NK-01, respectively. At the genome level, the OrthoANIu algorithm was used to calculate the average nucleotide identity (ANI) value (11), and the Genome-to-Genome Distance Calculator (GGDC) was employed to estimate the *in silico* DNA-DNA hybridization (isDDH) value (12). The ANI and isDDH values between *Pseudomonas* sp. phDV1 and *Pseudomonas pseudoalcaligenes* CECT5344 were 96.52% and 70.30%, respectively, while both values were substantially lower (88.64% for ANI and 36.70% for isDDH) when *Pseudomonas* sp. phDV1 was compared to *Pseudomonas mendocina* NK-01. Based on these results, we suggest the transfer of *Pseudomonas* sp. phDV1 to the species *Pseudomonas pseudoalcaligenes* as strain phDV1.

Data availability. The PacBio sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRR8212849. The whole-genome sequence has been deposited at GenBank under the accession number CP031606. The version described in this paper is the first version, CP031606.1.

REFERENCES

1. Polymenakou PN, Stephanou EG. 2005. Effect of temperature and additional carbon sources on phenol degradation by an indigenous soil pseudomonad. Biodegradation 16:403–413. https://doi.org/10.1007/s10532-004-3333-1.
2. Tsirogianni I, Aivaliotis M, Karas M, Tsiotis G. 2004. Mass spectrometric mapping of the enzymes involved in the phenol degradation of an indigenous soil pseudomonad. Biochim Biophys Acta 1700:117–123. https://doi.org/10.1016/j.bbapap.2004.04.003.
3. Tsirogianni E, Aivaliotis M, Papasotiriou DG, Karas M, Tsiotis G. 2006. Identification of inducible protein complexes in the phenol degrader Pseudomonas sp. strain phDV1 by blue native gel electrophoresis and mass spectrometry. Amino Acids 30:63–72. https://doi.org/10.1007/s00726-005-0219-4.
4. Tsirogianni E, Aivaliotis M, Karas M, Tsiotis G. 2005. Detection and characterisation of catechol 2,3-dioxygenase in an indigenous soil pseudomonad by MALDI-TOF MS using a column separation. Biodegradation 16:181–186. https://doi.org/10.1007/s10532-004-4885-9.
5. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
6. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
7. 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.
8. 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. The Gene Ontology Consortium. Nat Genet 25:25–29. https://doi.org/10.1038/75556.
9. 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.
10. Powlowski J, Shingler V. 1994. Genetics and biochemistry of phenol degradation by Pseudomonas sp. CF600. Biodegradation 5:219–236.
11. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110:1281–1286. https://doi.org/10.1007/s10482-017-0844-4.
12. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60. https://doi.org/10.1186/1471-2105-14-60.