Complete Genome Sequence of *Azoarcus* sp. DD4, a Gram-negative Propanotroph That Degrades 1,4-Dioxane and 1,1-Dichloroethylene

Daiyong Deng[1], Fei Li[1], Lei Ye[2], Mengyan Li[1,*]

1. Department of Chemistry and Environmental Science, New Jersey Institute of Technology, Newark, NJ, USA 07102

2. HaploX Biotechnology Co., Ltd, Shenzhen, Guangdong, China 518055

*Address correspondence to Dr. Mengyan Li (mengyan.li@njit.edu)

ABSTRACT

Azoarcus sp. DD4 can cometabolically degrade 1,4-dioxane and 1,1-dichloroethylene when grown with propane and other substrates. The complete genome sequence of DD4 reveals a diverse collection of bacterial monooxygenase genes which may contribute to its versatility of degrading commingled groundwater pollutants.

Azoarcus sp. DD4 is a propanotrophic bacterium isolated from an activated sludge sample collected at a municipal wastewater treatment plant in Northern New Jersey (1). Notably, DD4 presents a synchronic degradation ability on 1,4-dioxane and 1,1-dichloroethylene (1,1-DCE) via cometabolism with the induction of propane and some other substrates (1). Like other *Azoarcus* strains, DD4 is also a diazotroph that can assimilate atmospheric nitrogen (1-3). The growth and activity of DD4 can be sustained under a wide variety of aquifer-relevant conditions (1), suggesting its potential as an effective inoculum for *in situ* or *ex situ* bioaugmentation to treat the commingled contamination of 1,4-dioxane and 1,1-DCE. Therefore, the whole-genome sequence of DD4 provide insights into the genetic relevance regarding its lifestyle and degradation capabilities, which are valuable to optimize and assess its field applications.

DD4 cells were harvested at its exponential phase after grown in nitrate mineral salts (NMS) medium with propane (0.1% v/v in the headspace) as the sole carbon and energy source. Total genomic DNA of DD4 was extracted using the MagAttract HMW DNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's instruction. The extract DNA was purified with AMPure PB magnetic beads and further used for the library preparation using the combination of the SMRTbell Damage Repair Kit and Barcoded Adapter Complete Prep Kit (Pacific Biosciences, Menlo Park, CA). The genome of DD4 was sequenced using the Pacbio Sequel™ System (Pacific Biosciences, Menlo Park, CA), which generated approximately 1.69 Gbp long-read sequencing data. The average length of raw sequences for sample DD4 is estimated as 2.1 kb as the final number of raw reads is 802,558. Following the hierarchical genome-assembly process (HGAP), DD4 genome of high quality and accuracy was assembled using the RS_HGAP_Assembly.3 protocol and polished by Quiver in SMRT Portal v2.3.0 with default parameters (4). For genome component prediction, the GeneMarkS+ program (5) was employed to retrieve the related coding

genes. Seven databases were then used for the annotation of gene functions (e-value less than 1e-5, minimal alignment length percentage greater than 40%) (6), including respective GO (Gene Ontology) (7), KEGG (Kyoto Encyclopedia of Genes and Genomes) (8), COG (Clusters of Orthologous Groups) (9), NR (Non-Redundant Protein Database databases) (10), TCDB (Transporter Classification Database) (11), and Swiss-Prot and TrEMBL (12).

There exists one single circular chromosome in DD4 without circular or linear plasmids. The genome size of DD4 is 5,400,077 bp with a GC content of 66.7%. A total of 5001 putative genes are annotated covering approximately 90.1% of the genome. The genome of DD4 contains 57 tRNA genes and 4 rRNA gene clusters (5S, 16S, and 23S). Four gene clusters encoding the soluble di-iron monooxygenases (SDIMOs) (13, 14) are found on the chromosome. Based on the phylogenetic analysis of the amino acid sequences of their hydroxylase alpha subunits, these four SDIMOs are categorized as a group-1 phenol hydroxylase, a group-2 toluene monooxygenase, a group-3 butane monooxygenase, and a group-5 propane monooxygenase. In addition, genes encoding a copper membrane particulate monooxygenase (15) and a cytochrome P450 CYP153 alkane hydroxylase are identified. One or more of these monooxygenases in DD4 may be responsible for initiating the oxidation of propane, 1,4-dioxane, 1,1-DCE, and other environmental pollutants (1, 16-19).

Accession number(s). The whole-genome sequence of *Azoarcus* sp. DD4 has been deposited in the GenBank with the accession number CP022958 (https://www.ncbi.nlm.nih.gov/nuccore/CP022958). The raw reads have been deposited in the SRA with the accession number PRJNA398544 (https://www.ncbi.nlm.nih.gov/sra/PRJNA398544).

ACKNOWLEDGMENTS

This work was funded by National Science Foundation (NSF, CAREER CBET-1846945), United States Geological Survey (USGS) State Water Resources Research Act Program (2018NJ400B), and the start-up fund from the Department of Chemistry and Environmental Science at NJIT. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

We declare no competing financial interest.

References

1. Deng D, Li F, Wu C, Li M. 2018. Synchronic biotransformation of 1,4-dioxane and 1,1-dichloroethylene by a gram-negative propanotroph *Azoarcus* sp. DD4. Environmental Science & Technology Letters doi:10.1021/acs.estlett.8b00312.
2. Krause A, Ramakumar A, Bartels D, Battistoni F, Bekel T, Boch J, Böhm M, Friedrich F, Hurek T, Krause L. 2006. Complete genome of the mutualistic, N_2-fixing grass endophyte *Azoarcus* sp. strain BH72. Nature biotechnology 24.
3. Faoro H, Rene Menegazzo R, Battistoni F, Gyaneshwar P, do Amaral FP, Taulé C, Rausch S, Gonçalves Galvão P, de los Santos C, Mitra S. 2017. The oil-contaminated soil diazotroph *Azoarcus* olearius DQS-4^T is genetically and phenotypically similar to the model grass endophyte *Azoarcus* sp. BH72. Environmental microbiology reports 9:223-238.
4. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature methods 10:563.
5. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic acids research 29:2607-2618.
6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of molecular biology 215:403-410.
7. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. 2000. Gene Ontology: tool for the unification of biology. Nature genetics 25:25.
8. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2011. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research 40:D109-D114.
9. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN. 2003. The COG database: an updated version includes eukaryotes. BMC bioinformatics 4:41.
10. Li W, Jaroszewski L, Godzik A. 2002. Tolerating some redundancy significantly speeds up clustering of large protein databases. Bioinformatics 18:77-82.
11. Saier Jr MH, Reddy VS, Tamang DG, Västermark Å. 2013. The transporter classification database. Nucleic acids research 42:D251-D258.
12. Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'donovan C, Phan I. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic acids research 31:365-370.
13. Leahy JG, Batchelor PJ, Morcomb SM. 2003. Evolution of the soluble diiron monooxygenases. Fems Microbiology Reviews 27:449-479.
14. Notomista E, Lahm A, Di Donato A, Tramontano A. 2003. Evolution of bacterial and archaeal multicomponent monooxygenases. J Mol Evol 56:435-45.
15. Coleman NV, Le NB, Ly MA, Ogawa HE, McCarl V, Wilson NL, Holmes AJ. 2012. Hydrocarbon monooxygenase in *Mycobacterium*: recombinant expression of a member of the ammonia monooxygenase superfamily. ISME J 6:171-82.
16. Deng D, Li F, Li M. 2018. A novel propane monooxygenase initiating degradation of 1,4-dioxane by *Mycobacterium dioxanotrophicus* PH-06. Environmental Science & Technology Letters 5:86-91.

17. Li M, Mathieu J, Yang Y, Fiorenza S, Deng Y, He Z, Zhou J, Alvarez PJ. 2013. Widespread distribution of soluble di-iron monooxygenase (SDIMO) genes in Arctic groundwater impacted by 1, 4-dioxane. Environmental Science & Technology 47:9950-9958.
18. Sales CM, Grostern A, Parales JV, Parales RE, Alvarez-Cohen L. 2013. Oxidation of the cyclic ethers 1,4-dioxane and tetrahydrofuran by a monooxygenase in two *Pseudonocardia* species. Applied and Environmental Microbiology 79:7702-7708.
19. Mahendra S, Alvarez-Cohen L. 2006. Kinetics of 1,4-dioxane biodegradation by monooxygenase-expressing bacteria. Environmental Science & Technology 40:5435-5442.