Complete Genome Sequences of *Pseudomonas alkylphenolica* Neo and *Variovorax* sp. Strain CSUSB, Obtained in Undergraduate Microbiology Courses Using a Hybrid Assembly Approach

Christopher Ne Ville,[a] Dylan Enright,[a*] Ivan Hernandez,[a*] Jeremey Dodsworth,[a] Paul Orwin[a]

[a]Department of Biology, California State University San Bernardino, San Bernardino, California, USA

ABSTRACT   Two Gram-negative bacteria with a high G+C content were isolated from soil in undergraduate microbiology classes by enriching for low nutrient growth and neonicotinoid pesticide tolerance. DNA from these isolates was purified and sequenced using a hybrid approach. Here we report the genome sequences of *Pseudomonas alkylphenolica* strain Neo and *Variovorax* sp. strain CSUSB.

Two Gram-negative bacteria were isolated from environmental soil samples by undergraduates in microbiology courses. *Pseudomonas alkylphenolica* strain Neo was initially isolated from garden soil using discs impregnated with the imidacloprid-containing pesticide Bayer Advanced fruit, citrus, and vegetable insect control (0.235% imidacloprid; ~9.2 mM). Colonies growing closest to the discs were subsequently tested for growth in the presence of higher concentrations of the pesticide, and the isolate sequenced was selected based on growth in a 40% (vol/vol) pesticide challenge (~3.67 mM imidacloprid). Initial characterization of *P. alkylphenolica* was based on phenotypic analysis and partial 16S rRNA gene sequencing. *Variovorax* sp. strain CSUSB was isolated in an experiment identifying rhizosphere bacteria that grew in the diluted agar medium (0.1% Trypticase soy broth [TSB] solidified with 1.5% agar). Isolates from this medium were subsequently tested for antibiotic resistance. *Variovorax* sp. CSUSB was isolated from the roots of *Helianthus annuus*, similar to a previous isolation of *Variovorax paradoxus* EPS (1).

DNA was extracted from cultures of the isolates grown at room temperature in 0.5% yeast extract (Fisher Scientific) broth using the high-molecular-weight DNA protocol outlined for *Escherichia coli* (https://www.protocols.io/view/ultra-long-read-sequencing-protocol-for-rad004-mrxc57n) (2). These cultures were grown from single colonies picked from plates grown directly from −80°C glycerol stocks, which were made from the initial isolated cultures. Quality assurance and quantitation were performed using the NanoDrop 1 spectrophotometer (Thermo Fisher). Long reads were obtained by preparing libraries using the rapid barcoding kit (catalog number SQK-RBK004) and sequencing these libraries using an MIN-106 flowcell (R9.4.1) in an Oxford Nanopore MinION sequencer. Two separate sequencing runs were performed using barcoded libraries derived from the same genomic sample in separate flow cells, and the data were combined for assembly after processing. Barcodes RB03 and RB06 were used for *Variovorax* sp. CSUSB, and barcodes RB05 and RB04 were used for *P. alkylphenolica* Neo. For all subsequent data-processing steps, the default parameters were used unless otherwise noted. MinION reads were basecalled in Guppy v2.3.1 using the Flipflop model and demultiplexed in Deepbinner v0.2.0 (3). Barcodes and adapters were removed with Porechop v0.2.4 (4). A total of 657,284 reads were obtained for *Variovorax* sp. CSUSB (average read length, 7,561.13 ± 9,721.0 bp; 885.56× coverage), and 35,176

reads were obtained for *P. alkylphenolica* (average read length, 9,573 ± 10,371.88 bp; 60.0× coverage). For *P. alkylphenolica,* the same genomic DNA preparation was used to generate a 250- to 300-bp library with the Nextera DNA Flex LPK kit, which was sequenced in the Illumina iSeq platform (2 × 150 bp). Short read data for *Variovorax* sp. CSUSB were obtained using the FastDNA spin kit for soil (Solon, OH) and sequenced on the Illumina MiSeq platform (500- to 600-bp library fragment size using the Nextera XT kit, 2 × 250 reads). A total of 3,800,551 reads were obtained from the *P. alkylphenolica* library (average read length, 131 bp; 177.43× coverage), and 1,582,102 reads were obtained from the *Variovorax* sp. CSUSB library (average read length, 195.69 ± 40.61 bp; 55.54× coverage). FastQC v0.11.8 was used for quality assessment of these data (5), and trimming was performed in Trimmomatic v0.38.0 (6). Assemblies of *Variovorax* sp. CSUSB and *P. alkylphenolica* Neo were created using a hybrid approach in Unicycler v0.4.8.0 (7) on the North America Galaxy hub (http://usegalaxy.org) (8).

The *Variovorax* sp. CSUSB strain genomic DNA assembled into a single circular contig with 5,574,400 bp. Initial annotation with RASTtk (http://rast.nmpdr.org) (9) identified 5,228 coding sequences and 58 RNAs, with a G+C content of 65.7%. The *P. alkylphenolica* strain assembly generated a single circular contig with 5,612,010 bp, with a G+C content of 61.2%, and a similar annotation identified 5,079 coding sequences and 89 RNAs. The assemblies uploaded to NCBI were annotated using the Prokaryotic Genome Annotation Pipeline (PGAP) by NCBI (10). Using PHASTER (https://phaster.ca/) (11), one complete prophage was identified in the *Variovorax* sp. CSUSB assembly along with two incomplete phage elements, and three complete phage elements were identified in *P. alkylphenolica* Neo.

There is substantial diversity in the genome structure in the genus *Variovorax,* with both single-chromosome and multiple-replicon structures (1, 12). The reported CSUSB strain genome represents one of the smallest *Variovorax* sp. assemblies so far. *Variovorax paradoxus* NBRC 15149 was identified by 16S rRNA sequence identity as the closest relative (>99%), and the two-way average nucleotide identity (ANI) was calculated (http://enve-omics.ce.gatech.edu/ani/) (13) as 88%. This was the highest level seen in pairwise ANI for this isolate against all sequenced *Variovorax* strains, which leads us to the conservative naming of this isolate as *Variovorax* sp. (14). Analysis of *P. alkylphenolica* Neo revealed high levels of identity by both ANI (two-way ANI, 96.76%) and 16S homology (99.93%) to the sequenced type strain KL28 (15).

Data availability. The assemblies and sequence data have been uploaded to the NCBI database. *Pseudomonas alkylphenolica* Neo can be found at BioProject number PRJNA593854, BioSample number SAMN13494299, and assembly number CP046621. *Variovorax* sp. CSUSB can be found at numbers PRJNA593854, SAMN13494298, and CP046622. Read data can be found at numbers SRR10662237 to SRR10662240, including demultiplexed fastQ files with barcodes removed for the MinION runs and paired fastQ files for the Illumina iSeq and MiSeq runs.

REFERENCES

1. Han J-I, Spain JC, Leadbetter JR, Ovchinnikova G, Goodwin LA, Han CS, Woyke T, Davenport KW, Orwin PM. 2013. Genome of the root-associated plant growth-promoting bacterium Variovorax paradoxus strain EPS. Genome Announc 1:e00843-13. https://doi.org/10.1128/genomeA.00843-13.
2. Quick J, Loman NJ. 2018. DNA extraction strategies for Nanopore sequencing, p 91–105. *In* Nanopore sequencing. World Scientific, Singapore, Singapore.
3. Wick RR, Judd LM, Holt KE. 2018. Deepbinner: demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. PLoS Comput Biol 14:e1006583. https://doi.org/10.1371/journal.pcbi.1006583.
4. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 3:e000132. https://doi.org/10.1099/mgen.0.000132.
5. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Institute, Cambridge, United Kingdom.
6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

7. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

8. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379.

9. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365.

10. 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.

11. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.

12. Han J-I, Choi H-K, Lee S-W, Orwin PM, Kim J, Laroe SL, Kim T-G, O'Neil J, Leadbetter JR, Lee SY, Hur C-G, Spain JC, Ovchinnikova G, Goodwin L, Han C. 2011. Complete genome sequence of the metabolically versatile plant growth-promoting endophyte Variovorax paradoxus S110. J Bacteriol 193:1183–1190. https://doi.org/10.1128/JB.00925-10.

13. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr 4:e1900v1. https://doi.org/10.7287/peerj.preprints.1900v1.

14. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.

15. Mulet M, Sánchez D, Lalucat J, Lee K, García-Valdés E. 2015. Pseudomonas alkylphenolica sp. nov., a bacterial species able to form special aerial structures when grown on p-cresol. Int J Syst Evol Microbiol 65:4013–4018. https://doi.org/10.1099/ijsem.0.000529.