Metagenome-Assembled Genome Sequence of *Pseudomonas stutzeri* Strain CO183 Isolated from a Coalbed Methane Well

Daniel E. Ross,[a,b] Djuna Gulliver[a]

National Energy Technology Laboratory, Pittsburgh, Pennsylvania, USA[a]; AECOM, Pittsburgh, Pennsylvania, USA[b]

A near-complete *Pseudomonas stutzeri* draft genome was extracted from a coalbed metagenome. The draft genome described herein provides insight into the functional pathways encoded by this bacterium and its potential role in coalbed methane environments.

Address correspondence to Djuna Gulliver, gullived@netl.doe.gov.

The ever-increasing demand for affordable domestic energy supplies necessitates alternative methods to maximize utilization of these supplies. One alternative method, microbial-enhanced coalbed methane, harnesses resident microbial populations for *in situ* natural gas production from unmineable coal seams (1). The *Pseudomonas stutzeri* strain CO183 draft genome was extracted from the assembled metagenome of coalbed methane cuttings. The sub-bituminous coal samples were retrieved from the Basal Coal Zone at a depth of 730 m. DNA extraction was performed by An and coworkers, as described previously (2). Illumina paired-end read data used in this study was obtained from the Hydrocarbon Metagenome Project database (http://hmp.ucalgary.ca).

Metagenome reads were assembled with metaSPAdes (3), binned with VizBin (4), and the most complete genome bin was assigned to the genus *Pseudomonas*. Specifically, metagenome contigs were mapped to multiple *P. stutzeri* strains (5), and a total of 866 contigs (4,740,859 bp) were mapped to *P. stutzeri* RCH2 (E-value cutoff = $1e^{-20}$). Using 833 *Pseudomonas*-specific marker genes in CheckM (6), the resultant genome bin was 92.84% complete and had 13.64% contamination. The mapped *Pseudomonas*-assigned contigs and the raw metagenome reads were used as input for paired-read iterative contig extension (PRICE) (7), and after 30 cycles the genome bin improved significantly, resulting in a bin of 117 contigs (4,522,918 bp) at 99.27% completeness and 0.4% contamination. The longest contig was 280,110 bp, the N_{50} was 74,098, and the GC content was 62.97%. The final draft genome bin had an average nucleotide identity of 97.59% with *P. stutzeri* RCH2.

The genome was annotated with Rapid Annotations using Subsystem Technology (RAST) version 2.0 using the RASTtk pipeline (8, 9) and the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAP) (http://www.ncbi.nlm.nih.gov/genome/annotation_prok). RAST estimated 4,396 coding sequences and 64 RNAs. PGAP estimated 4,429 total genes (4,066 coding genes), 68 RNA genes (four rRNAs, 60 tRNAs, and four ncRNAs), and 115 pseudogenes. Initial analysis suggests that the genome encodes for a complete tricarboxylic acid cycle, heavy-metal resistance (copper,

chromium, cadmium, zinc, arsenic), and denitrification pathways. Future comparative analyses of *Pseudomonas stutzeri* strain CO183 with other *Pseudomonas* spp. found in coalbed methane environments will provide insight into the ecology of these strains, how they evolved, and their contribution to biogenic coal degradation.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number MCNJ00000000. The version described in this paper is the first version, MCNJ01000000.

ACKNOWLEDGMENTS

FUNDING INFORMATION

REFERENCES

1. **Park SY, Liang Y**. 2016. Biogenic methane production from coal: a review on recent research and development on microbially enhanced coalbed methane (MECBM). Fuel **166:**258–267. http://dx.doi.org/10.1016/j.fuel.2015.10.121.
2. **An D, Caffrey SM, Soh J, Agrawal A, Brown D, Budwill K, Dong X, Dunfield PF, Foght J, Gieg LM, Hallam SJ, Hanson NW, He Z, Jack TR,**

Klassen J, Konwar KM, Kuatsjah E, Li C, Larter S, Leopatra V. 2013. Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. Environ Sci Technol **47:** 10708–10717. http://dx.doi.org/10.1021/es4020184.

3. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single cell sequencing. J Comput Biol **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

4. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, Coronado S, van der Maaten L, Vlassis N, Wilmes P. 2015. VizBin: an application for reference-independent visualization and human-augmented binning of metagenomic data. Microbiome **3:**1. http://dx.doi.org/10.1186/s40168-014-0066-1.

5. Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. 2011. CONTIGuator: a bacterial genome finishing tool for structural insights on draft genomes. Source Code Biol Med **6:**11. http://dx.doi.org/10.1186/1751-0473-6-11.

6. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res **25:**1043–1055. http://dx.doi.org/10.1101/gr.186072.114.

7. Ruby JG, Bellare P, DeRisi JL. 2013. PRICE: software for the targeted assembly of components of (meta) genomic sequence data. G3 (Bethesda) **3:**865–880. http://dx.doi.org/10.1534/g3.113.005967.

8. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9:**75. http://dx.doi.org/10.1186/1471-2164-9-75.

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. http://dx.doi.org/10.1038/srep08365.