



Complete Genome Sequences of Seven Strains of *Pseudomonas* spp. Isolated from Boreal Forest Soil in Interior Alaska

Tracie Haan,^a Taylor J. Seitz,^a Alexis Francisco,^a Kaytlin Gliner,^a Alexander Gloger,^a Anna Kardash,^a Nana Matsui,^a Emily Reast,^a Kaelen Rosander,^a Carmen Sonnek,^a Rebecca Wellman,^a  Devin M. Drown^{a,b}

^aDepartment of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, Alaska, USA

^bInstitute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, Alaska, USA

ABSTRACT Here, we describe the complete genome assemblies of seven *Pseudomonas* sp. isolates collected from a boreal forest soil on the University of Alaska Fairbanks campus. Using the VolTRAX v2 multiplex library preparation for Nanopore sequencing and Illumina reads for polishing, we assembled complete genome sequences for each of the isolates.

We collected soil from the University of Alaska Fairbanks campus, Alaska (64.859°N, 147.855°W). We sampled this site as part of a multiweek workshop for undergraduates with the goal of isolating, screening, and characterizing subarctic soil bacteria with antibiotic activity against ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) pathogen relatives. The ESKAPE pathogens are the leading cause of nosocomial infections worldwide and a major public health threat due to their ability to rapidly develop multidrug resistance to the safest antibiotics used clinically (1). Discovering novel antibiotics that act on the ESKAPE pathogens is therefore critical in tackling the global antibiotic resistance crisis. Although *Pseudomonas* spp. found in soils have been previously established as prolific antibiotic producers (2), the isolates described here originated from subarctic soils that are not well characterized and therefore may be a segue into the discovery of novel antibiotic biosynthesis with activity against ESKAPE pathogens.

We homogenized two 10-cm soil cores and added 1 g of soil to 9 ml tryptic soy broth (TSB). To ensure discrete colony formation, we plated 3 dilutions (1:10, 1:100, 1:1,000) onto tryptic soy agar and used the streak plate method 3 times to purify our randomly selected isolates. Here, we describe 7 isolates that we determined to be *Pseudomonas* spp. To obtain DNA for sequencing for each isolate, we inoculated liquid cultures of TSB, incubated at 22°C overnight, and used 1.8 ml as input for the DNeasy UltraClean microbial kit (Qiagen).

We used a total of 583.5 ng of DNA (range, 19.7 to 224 ng; Table 1) as input for the VolTRAX v2 (Oxford Nanopore Technologies [ONT]) to prepare a barcoded sequencing library (VSK-VMK002 workflow, cartridge ID VAB59563). We sequenced the prepared library using a MinION device (ONT) on an r9.4.1 flow cell (FLO-MIN106, flow cell ID FAK97975) for 72 h (VMK002 script). We base called the raw data using Guppy v3.4.5 (ONT) specifying the high-accuracy model (-c dna_r9.4.1_450bps_hac.cfg) and default parameters. This run generated a total of 14,168,620,548 bp in 2,963,836 reads with an N_{50} read length of 6,282 bp. We demultiplexed isolate samples with the guppy_barcode function of Guppy with parameters to discard sequences with middle adapters (-detect_mid_strand_barcodes) and trim barcodes (-trim_barcodes). We used FilTlong v0.2.0 (<https://github.com/rwick/filtlong>) to filter by length (≥ 50 bp; -min_length 50)

Citation Haan T, Seitz TJ, Francisco A, Gliner K, Gloger A, Kardash A, Matsui N, Reast E, Rosander K, Sonnek C, Wellman R, Drown DM. 2020. Complete genome sequences of seven strains of *Pseudomonas* spp. isolated from boreal forest soil in Interior Alaska. Microbiol Resour Announc 9:e00511-20. <https://doi.org/10.1128/MRA.00511-20>.

Editor David A. Baltrus, University of Arizona

Copyright © 2020 Haan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Devin M. Drown, dmdrown@alaska.edu.

Received 4 May 2020

Accepted 25 May 2020

Published 18 June 2020

TABLE 1 Summary of sequencing data statistics

Isolate	Input DNA (ng)	Barcode	ONT yield (bp)	No. of ONT reads	Avg length (bp)	Coverage (x)	Illumina yield (bp)	No. of Illumina reads	Genome size (bp)	GC%	No. of contigs	M_{50} (bp)	No. of tRNAs	No. of rRNAs	No. of CDs ^a	GenBank accession no.
ADAK2	19.7	BC08	337,234,174	84,274	4,001.64	47	445,101,405	1,739,815	7,105,260	59.6	1	7,105,260	72	19	6,548	CP052862
ADAK7	26.8	BC04	877,616,940	239,767	3,660.29	124	474,255,253	1,827,970	7,105,297	59.6	1	7,105,297	72	19	6,546	CP052861
ADAK13	75.6	BC03	1,820,660,366	357,747	5,089.24	254	389,799,732	1,488,149	7,300,098	60.95	1	7,300,098	67	16	6,704	CP052860
ADAK18	55.8	BC06	575,305,360	144,686	3,976.23	88	379,821,480	1,445,569	6,471,780	59.21	1	6,471,780	66	16	5,872	CP052859
ADAK20	69.6	BC07	951,937,436	177,124	5,374.41	153	361,587,647	1,363,699	6,003,946	60.7	1	6,003,946	68	19	5,389	CP052858
ADAK21	224.0	BC01	1,307,688,884	180,133	7,259.57	227	368,641,802	1,382,086	6,003,863	60.69	1	6,003,863	68	19	5,389	CP052857
ADAK22	112.0	BC02	1,332,557,628	272,239	4,894.81	201	308,344,076	1,155,182	6,509,129	60.71	1	6,509,129	66	16	5,989	CP052856

^aCDs, coding DNA sequences.

and quality (Q) score (≥ 10 ; $-\text{min_mean_q } 90$). We assembled the genome sequence for each isolate using Flye v2.7 (3) with default parameters specifying the estimated genome size ($-\text{genomesize} = 5\text{m}$) and Nanopore reads ($-\text{nanopore-raw}$) and subsampling for initial disjointing assembly ($-\text{asm-coverage } 100$). The Microbial Genome Sequencing Center (Pittsburgh, PA) prepared a multiplex Nextera library using the previously extracted DNA for sequencing on an Illumina NextSeq 550 to generate paired-end reads (2×150 bp). Each individual strain was separately indexed. We used the `unicycler_polish` tool of Unicycler v0.4.8 (4) for genome polishing. As input, we included the Illumina reads and the Flye assembly. In this mode, Unicycler runs multiple rounds of polishing with Pilon v1.22 (5). Table 1 contains a summary of each of our complete genome sequences.

We used PATRIC v3.6.3 (6) for initial genome annotation and to extract the 16S rRNA gene sequences for each isolate. To assign taxonomy, we used BLASTn (7) against the NCBI 16S rRNA database. For each isolate, the top five hits (ranked by bit score) included members of the *Pseudomonas* genus. Therefore, we assigned each isolate as a member of that genus. The whole-genome sequences deposited in GenBank were annotated with PGAP (8) as part of the submission pipeline.

Data availability. This genome project is indexed at GenBank under BioProject accession number [PRJNA627971](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA627971). These whole-genome sequences have been deposited in GenBank under accession numbers [CP052856](https://www.ncbi.nlm.nih.gov/nuclseq/CP052856) through [CP052862](https://www.ncbi.nlm.nih.gov/nuclseq/CP052862). Direct links are listed in Table 1. The raw sequencing data for this project can be found in the NCBI SRA under accession number [PRJNA627971](https://www.ncbi.nlm.nih.gov/sra/PRJNA627971).

ACKNOWLEDGMENTS

We thank Kate Pendleton and Ursel Schütte for logistical and technical support. The inspiration for this project came from the Tiny Earth Project.

This research was supported by a mentoring award to T.H. and T.J.S. from the University of Alaska URSA program, with additional support from the Institute of Arctic Biology and Alaska INBRE. The research reported in this publication was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103395.

REFERENCES

1. Santajit S, Indrawattana N. 2016. Mechanisms of antimicrobial resistance in ESKAPE pathogens. *Biomed Res Int* 2016:2475067. <https://doi.org/10.1155/2016/2475067>.
2. Gross H, Loper JE. 2009. Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat Prod Rep* 26:1408–1446. <https://doi.org/10.1039/b817075b>.
3. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
4. 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>.
5. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
6. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* 45:D535–D542. <https://doi.org/10.1093/nar/gkw1017>.
7. Zhang A, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7:203–214. <https://doi.org/10.1089/10665270050081478>.
8. 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>.