




Isolation and Whole-Genome Sequencing of Four Antibiotic-Producing *Pseudomonas* Strains

Audrey L. E. Armstrong,^a Kyle Cole,^a Emily Collins,^a Jarret M. Easterday,^a Lauren Feddersen,^a Brianna Fleury,^a Lauren E. Johnson,^a Ankush F. Kecht,^a Logan H. McClure,^a Joselyn M. Molinar,^a Anjali S. Patel,^a Ria Patel,^a Jackson R. Rapala,^a Arunima Salomi,^a Amari A. Stepter,^a Michal Wrobel,^a  Loralyn M. Cozy^a

^aDepartment of Biology, Illinois Wesleyan University, Bloomington, Illinois, USA

Audrey L. E. Armstrong, Kyle Cole, Emily Collins, Jarret M. Easterday, Lauren Feddersen, Brianna Fleury, Lauren E. Johnson, Ankush F. Kecht, Logan H. McClure, Joselyn M. Molinar, Anjali S. Patel, Ria Patel, Jackson R. Rapala, Arunima Salomi, Amari A. Stepter, and Michal Wrobel contributed equally to this work. Author order was determined alphabetically by last name.

ABSTRACT Here, we report the isolation, whole-genome sequencing, and annotation of four novel *Pseudomonas* isolates. We also evaluate the biosynthetic potential of each genome.

Pseudomonas is a Gram-negative, rod-shaped, polar flagellated bacterial genus with more than 140 identified species (1). Within this genus, there are many species that inhabit a diverse range of environments, resulting in their multifaceted metabolic capacities and adaptation abilities in changing environments. Here, we report the draft genomes of four novel *Pseudomonas* isolates that exhibit high antibiotic activity.

Soil samples were collected from a variety of locations (Table 1). One gram of each sample was resuspended in 10 mL of sterile saline, diluted serially, and plated on Reasoner's 2 agar (R2A) (2) hardened medium plates at 25°C for 48 h. Once diluted, 50 colonies were isolated, and individual colonies were screened for antibiotic activity by patching onto lawns of other bacteria, followed by incubation at 25°C for 48 h. The strains described here, LF19, LM20, JR33AA, and KCA11, were selected because they produced zones of inhibition against lawns of several bacterial species (Table 1). The 16S rRNA gene PCR product was sequenced from each strain using the primer set 27F (5'-AGR GTT TGA TYM TGG CTC AG-3') and 1492R (5'-GGY TAC CTT GTT ACG ACT T-3'), with 55°C annealing and 30 s of extension. Using NCBI BLAST (3), it was determined that the isolates belong to the *Pseudomonas* genus.

Axenic cultures grown on R2A were sent to the Microbial Genome Sequencing Center (Pittsburgh, PA) for DNA isolation using the Qiagen DNeasy blood and tissue kit and whole-genome sequencing (Fig. 1A). Sequence libraries were prepared with a small-volume tagmentation protocol using the Nextera DNA library preparation kit (Illumina, San Diego, CA, USA). Barcodes and adapters were attached, and libraries were amplified using the KAPA HiFi library amplification kit (4). Paired-end libraries were subsequently sequenced on the Illumina NextSeq 550 platform. FastQC v0.11.9 (5, 6) was used to verify the quality of the reads. These reads were then assembled *de novo* in PATRIC v3.6.12 (7) using the Unicycler v0.4.8 program (8). The assembly included polishing using two Pilon iterations (9) and examination using QUAST v5.0.2 (10). Read number and length information can be found in Table 1. All programs related to genome assembly were run with default parameters. Annotation was performed in PATRIC v3.6.12 using the RASTtk pipeline (11) with default bacterial parameters, using *Pseudomonas* as a taxonomic guide (Table 1).

Taxonomic relationships between the four isolates and other known *Pseudomonas* species were evaluated using the Codon Tree pipeline (12–15) in PATRIC v3.6.12 (7) in combination with average nucleotide identity (ANI) calculations using the Kostas Lab

Editor David A. Baltrus, University of Arizona

Copyright © 2022 Armstrong et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Loralyn M. Cozy, lcozy@iwu.edu.

The authors declare no conflict of interest.

Received 26 January 2022

Accepted 16 May 2022

Published 6 June 2022

TABLE 1 Sampling locations, accession numbers, sequencing features, and genomic characteristics of the *Pseudomonas* strains

Isolate	Location ^a	SRA accession no.	Assembly accession no.	GenBank accession no.	No. of reads	Read length (bp)	No. of contigs	N ₅₀ (bp)	Avg coverage (×)	Size (bp)	GC content (%)	No. of genes	No. of proteins	Antibiotic activity ^b
JR33AA	47°37'47"N, 116°46'5"W	SRR17071229	GCF_021378985.1	JAJS AW0000000000	4,285,246	146	65	324,782	230	5,407,927	62.02	6,001	5,031	Bs, Sa, Ec, Ab, Pa, Ea
KCA11	40°30'55"N, 88°59'26"W	SRR17071230	GCF_021378935.1	JAJS DJ0000000000	4,432,406	151	63	309,066	231	5,554,878	63.15	5,318	5,250	Bs, Sa, Ec, Ab, Pa, Ea
LF19	40°5'51"N, 88°19'9"W	SRR17071231	GCF_021378965.1	JAJS DK0000000000	3,550,179	146	76	247,141	177	5,811,296	62.60	5,388	5,322	Bs, Sa, Ab
LM20	40°49'6"N, 89°29'18"W	SRR17071232	GCF_021378925.1	JAJS DL0000000000	3,533,723	146	60	316,867	183	5,596,202	63.08	5,295	5,227	Bs, Sa, Pa

^a Coordinates (latitude and longitude) of the soil collection site.

^b Ab, *Acinetobacter baylyi*; Bs, *Bacillus subtilis*; Ea, *Enterobacter aerogenes*; Ec, *Escherichia coli*; Pa, *Pseudomonas aeruginosa*; Sa, *Staphylococcus aureus*.

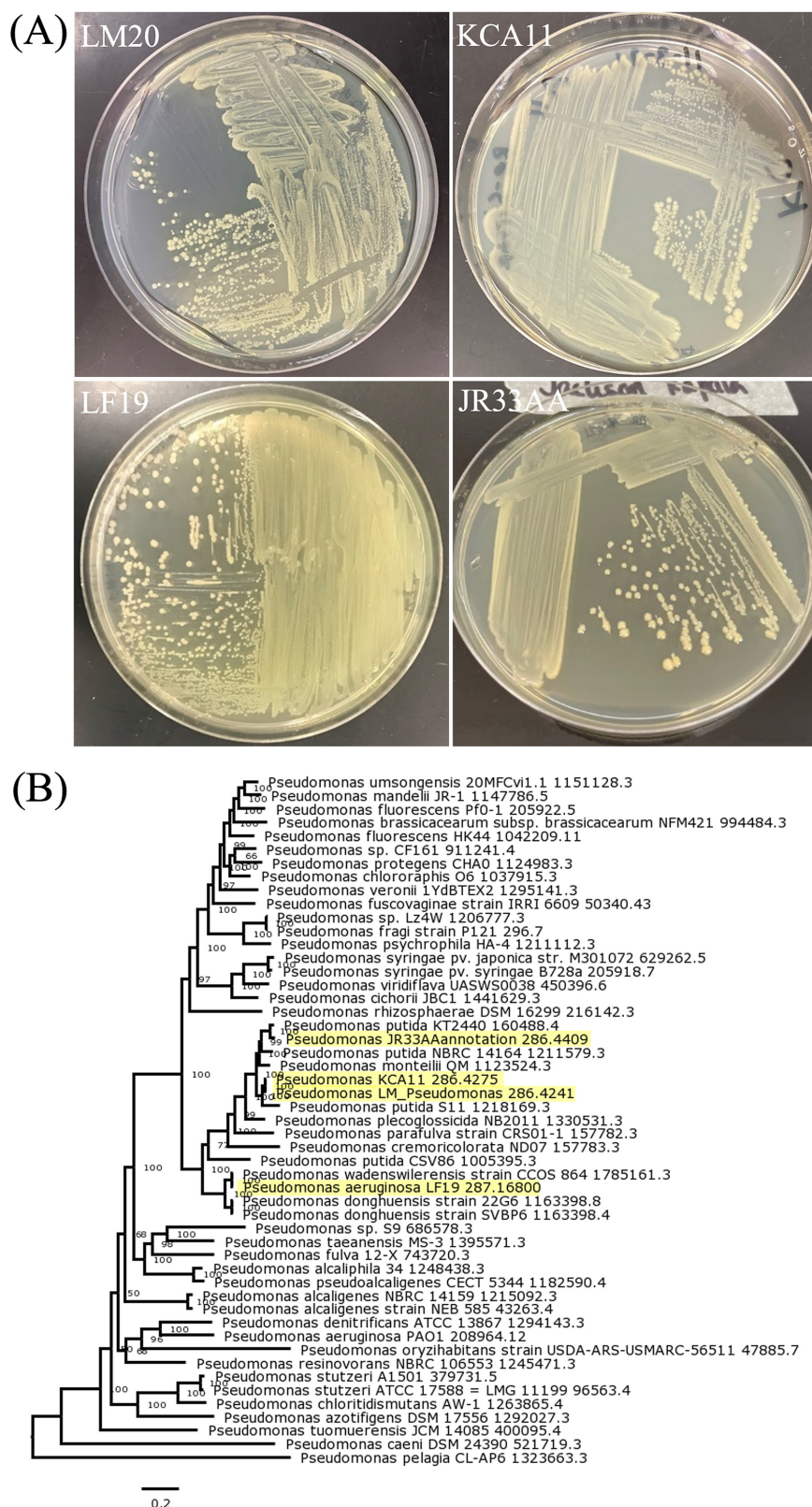


FIG 1 Colony morphology and phylogenetic relatedness. (A) R2A streak plates of each strain as indicated. (B) Phylogenetic tree generated using the Codon Tree pipeline in PATRIC with default parameters, with an input of 51 good-quality genomes. A total of 23,327 amino acids and 69,981 nucleotides were represented by 100 single-copy genes. The best tree was determined with 100 rounds of RAXML bootstrapping. Isolates LF19, LM20, JR33AA, and KCA11 are highlighted in yellow.

ANI Calculator v1.0 (16, 17), both run with default parameters. These calculations revealed that LM20 and KCA11 shared 99.5% nucleotide identity, while the remaining relationships all ranged between 82.5% and 89.3%. Additionally, isolates LM20, KCA11, and JR33AA all group with *Pseudomonas putida* strains, while LF19 is most related to *Pseudomonas wadsworthii* and *Pseudomonas donghuensis* strains (Fig. 1B).

Genome mining for antibacterial compounds using antiSMASH v6.0.1 (18), with the relaxed strictness setting, provided evidence that each of the four strains contains biosynthetic gene clusters (BGCs) potentially encoding secondary metabolites. The strain with the most predicted BGCs is LF19, with 13, and the strain with the least is LM20, with 8. KCA11 and JR33AA are both predicted to possess 10 BGCs. These strains add to the genomic data available for the *Pseudomonas* genus, supporting further investigation into its biosynthetic potential.

Data availability. This whole-genome shotgun project was deposited in GenBank under BioProject [PRJNA784595](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA784595). GenBank assembly and Sequence Read Archive (SRA) accession numbers are presented in Table 1.

ACKNOWLEDGMENTS

We thank the Tiny Earth Network (<https://tinyearth.wisc.edu>) for laboratory guidance, as well as the Biology Department and the Office of the Provost at Illinois Wesleyan University for their support of this project.

REFERENCES

1. Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW. 2011. *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* 35:652–680. <https://doi.org/10.1111/j.1574-6976.2011.00269.x>.
2. Reasoner DJ, Geldreich EE. 1985. A new medium for the enumeration and subculture of bacteria from potable water. *Appl Environ Microbiol* 49:1–7. <https://doi.org/10.1128/aem.49.1.1-7.1985>.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
4. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* 10:e0128036. <https://doi.org/10.1371/journal.pone.0128036>.
5. Andrews S, Lindenbaum P, Howard B, Ewels P. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
6. Leggett R, Ramirez-Gonzalez R, Clavijo B, Waite D, Davey R. 2013. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front Genet* 4:288. <https://doi.org/10.3389/fgene.2013.00288>.
7. Davis JJ, Wattam AR, Aziz RK, Brettn T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 48:D606–D612. <https://doi.org/10.1093/nar/gkz943>.
8. 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>.
9. 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>.
10. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
11. Brettn 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. <https://doi.org/10.1038/srep08365>.
12. Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, Shukla M, Vonstein V, Wattam AR, Yoo H. 2016. PATtyFams: protein families for the microbial genomes in the PATRIC database. *Front Microbiol* 7:118. <https://doi.org/10.3389/fmicb.2016.00118>.
13. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
14. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
15. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
16. Ciufo S, Kannan S, Sharma S, Badretdin A, Clark K, Turner S, Brover S, Schoch CL, Kimchi A, DiCuccio M. 2018. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int J Syst Evol Microbiol* 68:2386–2392. <https://doi.org/10.1099/ijsem.0.002809>.
17. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* 4:e1900v1. <https://doi.org/10.7287/peerj.preprints.1900v1>.
18. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* 49:W29–W35. <https://doi.org/10.1093/nar/gkab335>.