




# Complete Genome of *Aeromonas encheleia* Strain SOD01 Isolated from an Urban Freshwater Stream

Jade Arruda,<sup>a</sup> Taylor Brysgel,<sup>a</sup> Elizabeth Capestro,<sup>a</sup> Isabelle Flaherty,<sup>a</sup> Grace Foltz,<sup>a</sup> Abbey Gardner,<sup>a</sup> Michaela Hunt,<sup>a</sup> Jiana Ingrassia,<sup>a</sup> Sean O'Donnell,<sup>a</sup> Gregory Pappas,<sup>a</sup> Emily Peterson,<sup>a</sup> Sarah Ramsaran,<sup>a</sup> Marcelo Rocha,<sup>a</sup>  Laura E. Williams<sup>a</sup>

<sup>a</sup>Department of Biology, Providence College, Providence, Rhode Island, USA

**ABSTRACT** We isolated *Aeromonas encheleia* strain SOD01 from an urban freshwater stream in Providence, RI. *De novo* assembly of PacBio RSII data followed by polishing with Illumina MiSeq data generated a complete 4,450,115 bp genome with 61.8% GC content. PGAP annotation predicted 3,877 protein-coding genes, 127 tRNA, and 31 rRNA.

To isolate prey bacteria for studies of predatory *Bdellovibrio* (1), we soaked a sterile swab in water collected from an urban stream in Providence, RI (41.835 N, 71.443 W) then swabbed trypticase soy agar. After incubating plates at 28°C, we picked a colony and conducted three rounds of streaking. To establish freezer stock, we combined 0.5 mL 50% glycerol with 0.5 mL overnight culture grown in trypticase soy broth (TSB) at 28°C.

Using the Wizard Genomic DNA Purification kit (Promega, Madison, WI), we obtained genomic DNA from overnight cultures grown from freezer stock in TSB at 28°C. We extracted DNA from separate cultures for long- and short-read libraries. For long reads, the University of Maryland Institute for Genome Sciences sheared genomic DNA using g-TUBE (Covaris) at 3400 rpm, and then size selected it on a Blue Pippin instrument (Sage Scientific) with an 11,000 bp cutoff. The library was prepared using SMRTbell Template Prep Kit 1.0 (PacBio) and sequenced with P6-C4 chemistry using one SMRT cell on PacBio RS II. For short reads, the University of Rhode Island Genomics and Sequencing Center sheared genomic DNA using a Covaris S220 focused ultrasonicator. The library was prepared using the PrepX DNA Library kit (TaKaRa Bio), visualized on a high-sensitivity BioAnalyzer chip (Agilent), quantified using the KAPA Illumina Quantification kit (Roche), and sequenced on an Illumina MiSeq to obtain 2 × 250 bp paired-end reads.

Unless otherwise noted, default parameters were used for all software. For *de novo* assembly of the PacBio data (138,319 subreads at N50 12,507 bp), we compared Hierarchical Genome Assembly Process v3 (HGAP3) (2) and Canu 2.2 (3). To test HGAP3, we launched an Amazon EC2 instance of SMRT Portal v2.3.0 and set the estimated genome size at 4.5 Mbp, which yielded a 4,468,532 bp contig. After using BLASTN (4) and EMBLSS 6.6.0.0 extractseq (5) to identify and trim overlap between contig ends, the closed contig was 4,447,847 bp. To test Canu, we set the estimated genome size at 4.45 Mbp, which yielded a 4,464,381 bp contig. After using extractseq to trim overlap based on information from Canu, the closed contig was 4,448,166 bp. Comparison using dnadiff (6) identified only 20 SNVs and 2,559 indel bp between HGAP3 and Canu closed contigs, demonstrating good agreement. We proceeded with Canu output and used BLASTN comparisons and extractseq to rotate the closed contig so that the first base corresponded to the *dnaA* start codon.

For polishing, we aligned raw Illumina MiSeq read 1 and read 2 datasets (5,253,695 reads each) separately to the contig using Burrows-Wheeler aligner “mem” (BWA-mem) 0.7.17 (7) with the option to report all possible alignments for each read. Based on samtools (8) analysis, 5,190,922 R1 reads and 5,087,523 R2 reads aligned to at least one location. We used Polypolish v0.5.0 (9) to remove alignments based on insert size and then correct the sequence. Polypolish reported 456.6× coverage, and dnadiff analysis identified correction of one SNV

**Editor** J. Cameron Thrash, University of Southern California

**Copyright** © 2022 Arruda 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 Laura E. Williams, [lwilla7@providence.edu](mailto:lwilla7@providence.edu).

The authors declare no conflict of interest.

**Received** 1 July 2022

**Accepted** 1 August 2022

**Published** 18 August 2022

and 1,949 indel bp, yielding a 4,450,113 bp contig. To confirm, we aligned MiSeq R1 and R2 reads to the corrected contig using POLCA within MaSuRCA 4.0.8 (10). POLCA corrected two indel bp, yielding a final genome sequence of 4,450,115 bp and 61.8% GC. Prokaryotic Genome Annotation Pipeline (PGAP) (11) annotation predicted 3,877 protein-coding genes, 127 tRNA, and 31 rRNA. Digital DNA:DNA hybridization analysis with the Type Strain Genome Server (12) classified SOD01 as *Aeromonas encheleia*.

**Data availability.** *Aeromonas* sp. strain SOD01 genome sequence was deposited in GenBank (accession no. [CP099717](https://doi.org/10.1093/genbank/CP099717)). PacBio and MiSeq reads were deposited in the SRA (BioProject accession no. [PRJNA852219](https://doi.org/10.1093/bioinformatics/PRJNA852219) and SRA accession no. [SRX15851419](https://doi.org/10.1093/bioinformatics/SRX15851419) and [SRX15851420](https://doi.org/10.1093/bioinformatics/SRX15851420), respectively).

## ACKNOWLEDGMENTS

This research was conducted as part of an undergraduate course in genomics during the spring 2020 and spring 2022 semesters at Providence College. L.E.W. was the course instructor for both semesters. Independently of the course, S.O. isolated the strain. All other authors were undergraduate students in the course and contributed equally to the project.

We thank Lisa Sadzewicz and Luke Tallon at the Institute for Genome Sciences at the University of Maryland Baltimore for PacBio sequencing services and Janet Atoyian at the Genomics and Sequencing Center at the University of Rhode Island for Illumina sequencing services.

This research was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant no. P20GM103430 and by funding from Providence College. This research is based upon work conducted at a Rhode Island NSF EPSCoR research facility, the Genomics and Sequencing Center, supported in part by the National Science Foundation EPSCoR Cooperative Agreement EPS-1004057. The funders had no role in study design, data collection, interpretation, or the decision to submit the work for publication.

We declare no conflict of interest.

## REFERENCES

- Williams LE, Cullen N, DeGiorgis JA, Martinez KJ, Mellone J, Oser M, Wang J, Zhang Y. 2019. Variation in genome content and predatory phenotypes between *Bdellovibrio* sp. NC01 isolated from soil and *B. bacteriovorus* type strain HD100. *Microbiology (Reading)* 165:1315–1330. <https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000861>.
- Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
- 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).
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16:276–277. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2).
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv <https://doi.org/10.48550/ARXIV.1303.3997>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol* 18:e1009802. <https://doi.org/10.1371/journal.pcbi.1009802>.
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>.
- 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>.
- Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>.