Draft Genome Sequence of 2-Methylpyridine-, 2-Ethylpyridine-, and 2-Hydroxypyridine-Degrading *Arthrobacter* sp. Strain ATCC 49987

Nidhi Gupta,[a,b] Kelly A. Skinner,[a] Zarath M. Summers,[c] Janaka N. Edirisinghe,[a,b] José P. Faria,[a,b] Christopher W. Marshall,[a*] Anukriti Sharma,[b*] Neil R. Gottel,[b*] Jack A. Gilbert,[a,b*] Christopher S. Henry,[a,b] [ID] Edward J. O'Loughlin[a]

[a]Argonne National Laboratory, Lemont, Illinois, USA
[b]University of Chicago, Chicago, Illinois, USA
[c]ExxonMobil Research and Engineering Company, Annandale, New Jersey, USA

ABSTRACT Here, we report the draft genome sequence of *Arthrobacter* sp. strain ATCC 49987, consisting of three contigs with a total length of 4.4 Mbp. Based on the genome sequence, we suggest reclassification of *Arthrobacter* sp. strain ATCC 49987 as *Pseudarthrobacter* sp. strain ATCC 49987.

*A*rthrobacter spp. are ubiquitous members of soil microbial communities that play important roles in biogeochemical cycling and bioremediation, and many have been studied for their ability to grow under different environmental stresses to degrade persistent environmental pollutants (1, 2). O'Loughlin et al. isolated a 2-methylpyridine-degrading bacterium (*Arthrobacter* sp. strain ATCC 49987) from subsurface sediment from an aquifer located at an industrial site that had been contaminated with pyridine and pyridine derivatives (3). In addition to 2-methylpyridine, ATCC 49987 utilizes 2-ethylpyridine and 2-hydroxypyridine as primary nitrogen, carbon, and energy sources (3). The genome sequence reported here will help us understand the genetic basis of the biotransformation of these compounds.

ATCC 49987 was obtained from the American Type Culture Collection (ATCC) and grown in tryptic soy broth for 16 h at 30°C. DNA was extracted using the Qiagen DNeasy PowerSoil genomic DNA extraction kit (product number 12888-50) as per the manufacturer's instructions. DNA libraries were prepared using the Oxford Nanopore LSK109 kit for GridION X5 sequencing and the Nextera XT paired-end kit (Illumina) for Illumina MiSeq sequencing. For the GridION reads, base calling was carried out using Guppy v3.2.6, followed by adaptor removal and demultiplexing using Porechop v0.2.3. Paired-end reads (10,004,660 paired-end reads with an average length of 151 bp) along with single-end reads (331,392 single-end reads with an average length of 11,799 bp) were uploaded into the U.S. Department of Energy KBase system (https://kbase.us/n/64079/39) for further analysis (4). All bioinformatic tools were used with default parameters unless stated otherwise. FastQC v0.11.5 was used to assess read quality (5), and sequences were assembled using the MaSuRCA assembler v3.2.9 (6). The draft assembly had three contigs, with the largest consisting of 4,169,727 bp and the other two containing 65,535 bp and 167,082 bp. The total size of the resulting draft genome was 4.4 Mbp, with a GC content of 66.62% and genome coverage of 1,231$\times$. The quality of assembly was checked using CheckM v1.0.18 (7), and the assembly was determined to be 99.7% complete, with 1.17% contamination. Annotation was completed using Prokka v1.12 (8), revealing 4,153 total genes, 66 noncoding RNAs, and 77 noncoding repeats (https://narrative.kbase.us/#dataview/64079/42/1). ATCC 49987 is reported to utilize 2-, 3-, and 4-hydroxybenzoate, gentisic acid, protocatechuic acid, and catechol in addition to pyridine derivatives (2-methylpyridine, 2-ethylpyridine, and 2-hydroxy-

Address correspondence to Edward J. O'Loughlin, oloughlin@anl.gov.

* Present address: Christopher W. Marshall, Marquette University, Milwaukee, Wisconsin, USA; Anukriti Sharma, University of California, San Diego, San Diego, California, USA; Neil R. Gottel, University of California, San Diego, San Diego, California, USA; Jack A. Gilbert, University of California, San Diego, San Diego, California, USA.
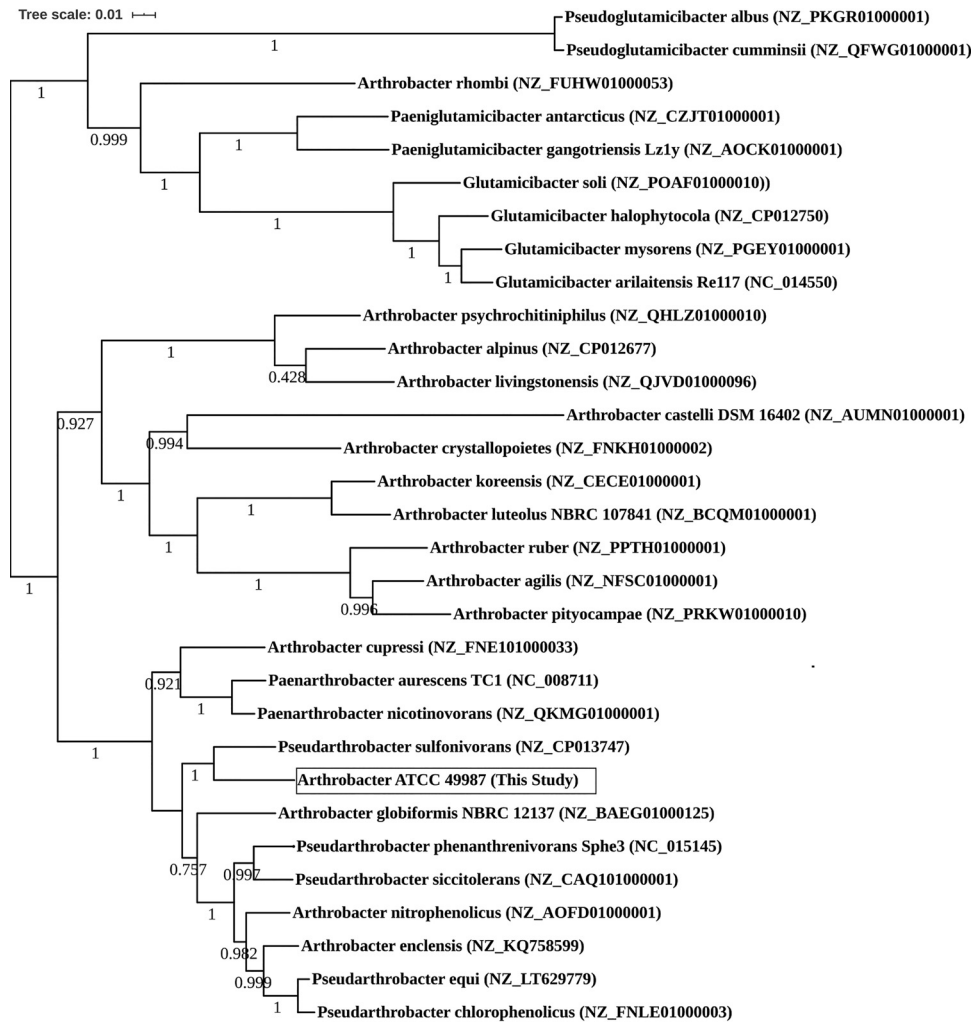
FIG 1 Phylogeny of ATCC 49987 in comparison with different species of *Arthrobacter*, *Paenarthrobacter*, *Glutam-icibacter*, *Pseudarthrobacter*, *Pseudoglutamicibacter*, and *Paeniglutamicibacter* (RefSeq accession numbers in paren-theses) selected from public KBase genomes. The whole genomes were aligned and the approximately maximum likelihood tree of ATCC 49987 was constructed using the default settings of FastTree v2.1.10 (10). Local support values for the tree nodes were computed using the Shimodaira-Hasegawa test in FastTree v2.1.10 and are represented on the tree.

pyridine) (3), suggesting that it has broad potential for the biotransformation of a range of monoaromatic compounds. In-depth analysis of the genome of ATCC 49987 will further our understanding of the underlying mechanisms, thus guiding genetic mod-ification of pathways to enhance its biotransformation capabilities.

At the time of isolation, ATCC 49987 was identified as a member of the genus *Arthrobacter* based on its physiological and morphological characteristics (3). However, certain species of the genus *Arthrobacter* have been reclassified into five novel genera, namely, *Pseudoglutamicibacter*, *Pseudarthrobacter*, *Glutamicibacter*, *Paeniglutamicibac-ter*, and *Paenarthrobacter* (9). To assess the relatedness of the ATCC 49987 genome, a tree was generated in KBase using the application "insert set of genomes into species tree." The application first ran RPS-BLAST v0.3.3 with 49 universal conserved marker genes. The marker genes were selected based on the Clusters of Orthologous Groups (COGs) from NCBI (https://github.com/kbaseapps/SpeciesTreeBuilder/tree/master/data/cogs) against the ATCC 49987 genome. The RPS-BLAST alignments of marker genes generated using the draft genome and 30 closely related RefSeq genomes were concatenated, and then a tree was constructed using FastTree v2.1.10 with default settings (10) (Fig. 1). *Arthrobacter* sp. strain ATCC 49987 is most closely related to

Pseudarthrobacter sulfonivorans; therefore, we suggest reclassification of *Arthrobacter* sp. strain ATCC 49987 as *Pseudarthrobacter* sp. strain ATCC 49987.

Data availability. The whole-genome sequences have been deposited in GenBank under accession numbers JAABNS010000001 through JAABNS010000003, and the SRA accession numbers are SRR10884004 and SRR10884005. The assembly with annotation can be accessed at NZ_JAABNS010000001, NZ_JAABNS010000002, and NZ_JAABNS010000003.

REFERENCES

1. Mongodin EF, Shapir N, Daugherty SC, DeBoy RT, Emerson JB, Shvartzbeyn A, Radune D, Vamathevan J, Riggs F, Grinberg V, Khouri H, Wackett LP, Nelson KE, Sadowsky MJ. 2006. Secrets of soil survival revealed by the genome sequence of *Arthrobacter aurescens* TC1. PLoS Genet 2:e214. https://doi.org/10.1371/journal.pgen.0020214.

2. Guo X, Xie C, Wang L, Li Q, Wang Y. 2019. Biodegradation of persistent environmental pollutants by *Arthrobacter* sp. Environ Sci Pollut Res Int 26:8429–8443. https://doi.org/10.1007/s11356-019-04358-0.

3. O'Loughlin EJ, Sims GK, Traina SJ. 1999. Biodegradation of 2-methyl, 2-ethyl, and 2-hydroxypyridine by an *Arthrobacter* sp. isolated from subsurface sediment. Biodegradation 10:93–104. https://doi.org/10.1023/a:1008309026751.

4. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163.

5. Brown J, Pirrung M, McCue LA. 2017. FQC Dashboard: integrates FastQC results into a Web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics 33:3137–3139. https://doi.org/10.1093/bioinformatics/btx373.

6. 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.

7. 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. https://doi.org/10.1101/gr.186072.114.

8. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

9. Busse HJ. 2016. Review of the taxonomy of the genus *Arthrobacter*, emendation of the genus *Arthrobacter sensu lato*, proposal to reclassify selected species of the genus *Arthrobacter* in the novel genera *Glutamicibacter* gen. nov., *Paeniglutamicibacter* gen. nov., *Pseudoglutamicibacter* gen. nov., *Paenarthrobacter* gen. nov. and *Pseudarthrobacter* gen. nov., and emended description of *Arthrobacter roseus*. Int J Syst Evol Microbiol 66:9–37. https://doi.org/10.1099/ijsem.0.000702.

10. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.