

# Complete Genome Sequence of *Arthrobacter* sp. ATCC 21022, a Host for Bacteriophage Discovery

Daniel A. Russell, Graham F. Hatfull

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

We report the complete genome sequence of *Arthrobacter* sp. ATCC 21022, a strain maintained by ATCC and a commonly used host for bacteriophage isolation and genomic analysis. The strain is prophage-free and CRISPR-free but codes for two predicted restriction-modification systems.

Received 4 February 2016 Accepted 6 February 2016 Published 24 March 2016

**Citation** Russell DA, Hatfull GF. 2016. Complete genome sequence of *Arthrobacter* sp. ATCC 21022, a host for bacteriophage discovery. *Genome Announc* 4(2):e00168-16. doi:10.1128/genomeA.00168-16.

**Copyright** © 2016 Russell and Hatfull. 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 Graham F. Hatfull, gfh@pitt.edu.

*Arthrobacter* spp. are common soil bacteria classified within the family *Micrococcaceae* in the order *Actinomycetales*, and several species have potential bioremediation uses (1). Relatively few *Arthrobacter* genomes have been sequenced (12 complete, 46 whole-genome shotgun [WGS]) relative to other common soil bacterial genera, such as *Mycobacterium* (225 complete, 4,435 WGS), *Bacillus* (202 complete, 693 WGS), and *Streptomyces* (44 complete, 599 WGS), as of 2 February 2016 (2).

Here, we report the complete genome sequence of *Arthrobacter* sp. ATCC 21022, an ATCC-maintained strain. This strain is currently used for bacteriophage isolation in the Howard Hughes Medical Institute (HHMI) Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEAPHAGES) program (3, 4), and since 2012 more than 300 individual phages have been isolated, of which more than 100 have been completely sequenced (<http://phagesdb.org/hosts/strains/3>).

Genomic DNA was extracted from a culture of ATCC 21022 using the Promega Wizard Genomic DNA purification kit and was sequenced by Ion Torrent on a 316 chip using 200-bp reagents. This yielded ~2.6 million reads, which were assembled into 112 contigs using Newbler version 2.6. To enhance consensus quality, an additional ~7.5 million 150-bp reads were generated using an Illumina MiSeq and aligned to the existing assembly for a total average coverage of ~350×. Where possible, contigs were joined manually using Consed version 22 (5), reducing the number of contigs to 37. These contigs were scaffolded using BLAST similarity to *Arthrobacter aurescens* TC1 (6), and contigs were then joined by performing PCR and Sanger-sequencing the products.

The genome of *Arthrobacter* sp. ATCC 21022 is a single circular molecule of 4,434,904 bp with a G+C content of 63.41%, which was annotated using the NCBI's Prokaryotic Genome Annotation Pipeline. For bioinformatic purposes the genome was linearized with position 1 corresponding to the first base of the *dnaA* gene. The closest relative is *Arthrobacter* sp. Rue61a (7), with query coverage under standard BLASTn parameters of 76%. *Arthrobacter* sp. ATCC 21022 contains four rRNA operons and 53 tRNA genes and has 3,895 predicted protein-coding genes. We identified at least four different transposons, each of which is present in three

or more locations in the genome. No CRISPR elements were identified using CRISPRFinder (8), and no intact integrated prophages were predicted using PHAST (9), although phage-related genes were identified (AUT26\_17235—AUT26\_17265) that are likely components of a type VI secretion system. ATCC 21022 also encodes two components of predicted restriction-modification systems (AUT26\_04760 and AUT26\_11520), two toxin-antitoxin systems, and a putative abortive infection protein (AUT26\_01375), each of which could influence the profile of phages capable of infecting this host.

**Nucleotide sequence accession numbers.** The complete *Arthrobacter* sp. ATCC 21022 sequence and annotation are available in GenBank with the accession number [CP014196](https://ncbi.nlm.nih.gov/GenBank/CP014196). Sequencing reads have been deposited in the Sequencing Read Archive (SRA) with accession number SRP068921.

## ACKNOWLEDGMENTS

We thank Carlos Guerrero for bacterial growth and DNA isolation. This work was supported by grant 54308198 from the Howard Hughes Medical Institute.

## FUNDING INFORMATION

This work, including the efforts of Graham F. Hatfull, was funded by Howard Hughes Medical Institute (HHMI) (54308198).

## REFERENCES

- 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. <http://dx.doi.org/10.1023/A:1008309026751>.
- Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJ, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 42:D581–D591. <http://dx.doi.org/10.1093/nar/gkt1099>.
- Hatfull GF. 2015. Innovations in undergraduate science education: going Viral. *J Virol* 89:8111–8113. <http://dx.doi.org/10.1128/JVI.03003-14>.
- Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SC, Findley AM, Gissendan-

- ner CR, Golebiewska UP, Guild N, Hartzog GA, Grillo WH, Hollowell GP, Hughes LE, Johnson A, King RA, Lewis LO, Li W, Rosenzweig F, Rubin MR, Saha MS, Sandoz J, Shaffer CD, Taylor B, Temple L, Vazquez E, Ware VC, Barker LP, Bradley KW, Jacobs-Sera D, Pope WH, Russell DA, Cresawn SG, Lopatto D, Bailey CP, Hatfull GF. 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. *MBio* 5:e01051-13. <http://dx.doi.org/10.1128/mBio.01051-13>.
5. Gordon D, Green P. 2013. *Coned*: a graphical editor for next-generation sequencing. *Bioinformatics* 29:2936–2937. <http://dx.doi.org/10.1093/bioinformatics/btt515>.
  6. 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. <http://dx.doi.org/10.1371/journal.pgen.0020214>.
  7. Niewerth H, Schuldes J, Parschat K, Kiefer P, Vorholt JA, Daniel R, Fetzner S. 2012. Complete genome sequence and metabolic potential of the quinaldine-degrading bacterium *Arthrobacter* sp. Rue61a. *BMC Genomics* 13:534. <http://dx.doi.org/10.1186/1471-2164-13-534>.
  8. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57. <http://dx.doi.org/10.1093/nar/gkm360>.
  9. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res* 39:W347–W352. <http://dx.doi.org/10.1093/nar/gkr485>.