

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Complete Genome Sequence of Agrobacterium sp. Strain 33MFTa1.1, Isolated from Thlaspi arvense Roots

### Permalink

<https://escholarship.org/uc/item/00h9v9b9>

### Journal

Microbiology Resource Announcements, 8(37)

### ISSN

2169-8287

### Authors

Langley, Sasha  
Eng, Thomas  
Wan, Kenneth H  
et al.

### Publication Date

2019-09-12






### DOI

10.1128/mra.00432-19

Peer reviewed



# Complete Genome Sequence of *Agrobacterium* sp. Strain 33MFTa1.1, Isolated from *Thlaspi arvense* Roots

Sasha Langley,<sup>a</sup>  Thomas Eng,<sup>a,b</sup> Kenneth H. Wan,<sup>a</sup>  Robin A. Herbert,<sup>a,b</sup> Andrew P. Klein,<sup>c,d\*</sup>  Yasuo Yoshikuni,<sup>a,e</sup> Susannah G. Tringe,<sup>a,e</sup> James B. Brown,<sup>a,f,g</sup> Susan E. Celniker,<sup>a</sup>  Jenny C. Mortimer,<sup>a,b</sup>  Aindrila Mukhopadhyay<sup>a,b</sup>

<sup>a</sup>BioSciences Area, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>b</sup>Joint BioEnergy Institute, Emeryville, California, USA

<sup>c</sup>Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

<sup>d</sup>Howard Hughes Medical Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

<sup>e</sup>U.S. DOE Joint Genome Institute, Walnut Creek, California, USA

<sup>f</sup>Centre for Computational Biology, Biosciences, University of Birmingham, Birmingham, United Kingdom

<sup>g</sup>Department of Statistics, University of California, Berkeley, California, USA

**ABSTRACT** *Agrobacterium* sp. strain 33MFTa1.1 was isolated for functional host-microbe interaction studies from the *Thlaspi arvense* root-associated microbiome. The complete genome is comprised of a circular chromosome of 2,771,937 bp, a linear chromosome of 2,068,443 bp, and a plasmid of 496,948 bp, with G+C contents of 59%, 59%, and 58%, respectively.

*Agrobacterium* is a diverse genus of soil-dwelling bacteria in the alphaproteobacterial family *Rhizobiaceae*. Many *Agrobacterium* species cause plant diseases, including *Agrobacterium tumefaciens* (crown gall disease), *Agrobacterium rhizogenes* (hairy root disease), and *Agrobacterium vitis* (lesions and tumors on grape vines). First described in 1897 (1), *Agrobacterium* has been widely studied, largely because of its ability to transform plant cells with its DNA (which is known as transfer DNA [T-DNA]). As a result, *A. tumefaciens* has become the workhorse of plant genetic engineering (2–4). Other strains of *Agrobacterium* are commensal inhabitants of plant tissue. For example, *Agrobacterium* sp. strain 33MFTa1.1 was isolated from the root endophytic compartment of *Thlaspi arvense*, a close relative of the model plant *Arabidopsis thaliana* (5), and recolonizes gnotobiotic *A. thaliana* plants without producing disease symptoms (6, 7). This report describes the complete genome sequence of *Agrobacterium* sp. 33MFTa1.1 and will facilitate plant-microbe interaction studies.

*Agrobacterium* sp. 33MFTa1.1 (NCBI taxon identifier [ID] 1279031) was obtained from Jeff Dangl. A previously published (5) draft shotgun assembly of this strain consists of 15 contigs, and we posited that long-read sequencing techniques would enable assembly at the chromosome level. Bacteria were streaked onto LB plates, single colonies were amplified, and an aliquot was used for 16S V1 and V4 PCR (8) and sequence identification (reviewed in reference 9). DNA was isolated (10), and whole-genome sequencing was performed at Lawrence Berkeley National Laboratory (LBNL) using a combination of Oxford Nanopore long-read sequencing on the MinION Mk1B (11) and Illumina paired-end 300-bp read sequencing for quality (12). Oxford Nanopore sequencing libraries were constructed from 5 to 10  $\mu$ g DNA using the Oxford Nanopore 1D native barcoding genomic DNA protocol (version NBE\_9006\_v103\_revO\_21Dec2016) and sequenced on three FLO-MIN107 R9 version flow cells. Oxford Nanopore data were demultiplexed with Porechop (13). Sequencing yielded 61,147 reads with a length of  $\geq 2,000$  bp and a filtered mean read length of 6,862 bp, totaling 419,565,480 bp ( $\sim 79$ -fold coverage). The Illumina sequencing library was constructed from 1.5  $\mu$ g DNA.

**Citation** Langley S, Eng T, Wan KH, Herbert RA, Klein AP, Yoshikuni Y, Tringe SG, Brown JB, Celniker SE, Mortimer JC, Mukhopadhyay A. 2019. Complete genome sequence of *Agrobacterium* sp. strain 33MFTa1.1, isolated from *Thlaspi arvense* roots. *Microbiol Resour Announc* 8:e00432-19. <https://doi.org/10.1128/MRA.00432-19>.

**Editor** Catherine Putonti, Loyola University Chicago

**Copyright** © 2019 Langley 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 Jenny C. Mortimer, [jcmortimer@lbl.gov](mailto:jcmortimer@lbl.gov), or Aindrila Mukhopadhyay, [amukhopadhyay@lbl.gov](mailto:amukhopadhyay@lbl.gov).

\* Present address: Andrew P. Klein, Amyris, Inc., Emeryville, California, USA.

**Received** 11 June 2019

**Accepted** 16 August 2019

**Published** 12 September 2019

The DNA was fragmented using a Diagenode Bioruptor, and libraries were constructed using the NEBNext Ultra DNA library prep kit for Illumina. Sequencing yielded 2,529,890 paired-end reads, which were trimmed using Trimmomatic (14), resulting in a filtered mean read length of 270 bp and totaling 631,387,669 bp (~128-fold coverage). Nanopore and Illumina sequencing data were used as inputs for a *de novo* hybrid assembly constructed using Unicycler version 0.4.1 with the “bold” option (15, 16). The assembly produced 3 contigs, a single circular chromosome, a single linear chromosome, and a plasmid. Annotations of protein-encoding open reading frames and noncoding RNAs (ncRNAs) were predicted with the NCBI Prokaryotic Genome Annotation Pipeline (17).

The circular chromosome annotation predicts 2,654 protein-coding genes, 63 pseudogenes, 2 rRNA operons, and 40 tRNAs, with canonical anticodon triplets that base pair with codons for amino acids. It also encodes the telomerase A (*telA*) gene that encodes the protein required to generate the covalently closed hairpin loops at the ends of linear chromosomes (3). The linear chromosome annotation predicts 1,800 protein-coding genes, 69 pseudogenes, 2 rRNA operons, and 14 tRNAs. In addition, the genome includes a single 496,948-bp plasmid, p\_JBx\_073812, which contains candidate genes for plasmid replication initiation proteins (*repA*, *repB*, and *repC*) and for conjugative transfer (*traA*, *traB*, *traC*, *traD*, *traF*, *traG*, *traH*, and *traM*). It also carries genes for arsenic resistance (*arsH* and *ACR3*) and arsenate metabolism (two copies of the arsenate reductase gene *arsC*). A comparison of the new assembly with the previous 15-contig assembly (NCBI taxon ID 1279031) using the Joint Genome Institute (JGI) microbial species identifier (MiSI) genome-wide average nucleotide identity, alignment fraction (gANI, AF) calculator (<https://ani.jgi-psf.org/html/calc.php>) reveals high similarity, as expected, with gANI values of 100 and AF values of 0.99 (previous assembly→new assembly) and 0.98 (new assembly→previous assembly) (18).

To identify gene clusters of interest for further research, we analyzed the genome with the antibiotics and Secondary Metabolite Analysis SHell (antiSMASH) version 4.2.0 (19) tool. A total of 41 clusters and putative clusters were identified. These included a type I polyketide synthase cluster, a terpene cluster, and a nonribosomal peptide synthetase cluster. Of the remainder, 5 were putative fatty acid clusters, 5 were putative saccharide clusters, and 28 were putative clusters of unknown type, as identified by the ClusterFinder algorithm (19).

**Data availability.** The complete circular and linear chromosomes and plasmid sequences described here are deposited in GenBank under the accession numbers [CP036358](#) (circular chromosome), [CP036359](#) (linear chromosome), and [CP036360](#) (plasmid). The SRA accession number is [PRJNA523206](#).

## ACKNOWLEDGMENTS

We thank B. Booth for data management. The strain was originally isolated by the lab of Jeffery Dangl at UNC Chapel Hill.

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy contract DE-AC02-05CH11231. S.G.T. and Y.Y. are supported through the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported by the Office of Science of the U.S. Department of Energy under contract number DE-AC02-05CH11231. Part of this work was funded through the DOE Joint BioEnergy Institute (<http://www.jbei.org>), supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

## REFERENCES

1. Kado CI. 2014. Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by *Agrobacterium tumefaciens*. *Front Microbiol* 5:340. <https://doi.org/10.3389/fmicb.2014.00340>.
2. Nester EW. 2014. *Agrobacterium*: nature's genetic engineer. *Front Plant Sci* 5:730. <https://doi.org/10.3389/fpls.2014.00730>.
3. Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M, Quorollo B, Goldman BS, Cao Y, Askenazi M, Halling C, Mullin L, Houmiel K, Gordon

- J, Vaudin M, Iartchouk O, Epp A, Liu F, Wollam C, Allinger M, Doughty D, Scott C, Lappas C, Markelz B, Flanagan C, Crowell C, Gurson J, Lomo C, Sear C, Strub G, Cielo C, Slater S. 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* 294:2323–2328. <https://doi.org/10.1126/science.1066803>.
4. Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF, Woo L, Chen Y, Paulsen IT, Eisen JA, Karp PD, Bovee D, Chapman P, Clendenning J, Deatherage G, Gillet W, Grant C, Kutayavin T, Levy R, Li MJ, McClelland E, Palmieri A, Raymond C, Rouse G, Saenphimmachak C, Wu Z, Romero P, Gordon D, Zhang S, Yoo H, Tao Y, Biddle P, Jung M, Krespan W, Perry M, Gordon-Kamm B, Liao L, Kim S, Hendrick C, Zhao ZY, Dolan M, Chumley F, Tingey SV, Tomb JF, Gordon MP, Olson MV, Nester EW. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* 294:2317–2323. <https://doi.org/10.1126/science.1066804>.
  5. Levy A, Salas Gonzalez I, Mittelviehhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, Rangel Alvarez B, Lundberg DS, Lu T-Y, Lebeis S, Jin Z, McDonald M, Klein AP, Feltcher ME, Rio TG, Grant SR, Doty SL, Ley RE, Zhao B, Venturi V, Pelletier DA, Vorholt JA, Tringe SG, Woyke T, Dangl JL. 2018. Genomic features of bacterial adaptation to plants. *Nat Genet* 50:138–150. <https://doi.org/10.1038/s41588-017-0012-9>.
  6. Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, Malfatti S, Glavina del Rio T, Jones CD, Tringe SG, Dangl JL. 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* 349:860–864. <https://doi.org/10.1126/science.aaa8764>.
  7. Castrillo G, Teixeira P, Paredes SH, Law TF, de Lorenzo L, Feltcher ME, Finkel OM, Breakfield NW, Mieczkowski P, Jones CD, Paz-Ares J, Dangl JL. 2017. Root microbiota drive direct integration of phosphate stress and immunity. *Nature* 543:513–518. <https://doi.org/10.1038/nature21417>.
  8. Yu Z, Morrison M. 2004. Comparisons of different hypervariable regions of *rrs* genes for use in fingerprinting of microbial communities by PCR-denaturing gradient gel electrophoresis. *Appl Environ Microbiol* 70:4800–4806. <https://doi.org/10.1128/AEM.70.8.4800-4806.2004>.
  9. Slatko BE, Kieleczawa J, Ju J, Gardner AF, Hendrickson CL, Ausubel FM. 2011. “First generation” automated DNA sequencing technology. *Curr Protoc Mol Biol* Chapter 7:Unit7.2. <https://doi.org/10.1002/0471142727.mb0702s96>.
  10. Wilson K. 2001. Preparation of genomic DNA from bacteria. *Curr Protoc Mol Biol* Chapter 2:Unit 2.4. <https://doi.org/10.1002/0471142727.mb0204s56>.
  11. Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 17:239. <https://doi.org/10.1186/s13059-016-1103-0>.
  12. Risse J, Thomson M, Patrick S, Blakely G, Koutsovoulos G, Blaxter M, Watson M. 2015. A single chromosome assembly of *Bacteroides fragilis* strain BE1 from Illumina and MinION nanopore sequencing data. *Giga-science* 4:60. <https://doi.org/10.1186/s13742-015-0101-6>.
  13. Wick R. 2017. Porechop: adapter trimmer for Oxford Nanopore reads. <https://github.com/rwick/Porechop>.
  14. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
  15. 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>.
  16. 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>.
  17. 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>.
  18. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 43:6761–6771. <https://doi.org/10.1093/nar/gkv657>.
  19. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43:W237–W243. <https://doi.org/10.1093/nar/gkv437>.