AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Complete Genome Sequence of *Agrobacterium tumefaciens* 1D1609

Shu-Ting Cho (卓舒婷),[a] Mindia Haryono,[a] Hsing-Hua Chang (張興驊),[a] Mary Nia M. Santos,[a] Erh-Min Lai (賴爾珉),[a] Chih-Horng Kuo (郭志鴻)[a]

[a]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

ABSTRACT *Agrobacterium tumefaciens* 1D1609 is a highly virulent strain isolated from a crown gall tumor of alfalfa (*Medicago sativa* L.). Compared to other well-characterized *A. tumefaciens* strains, such as C58 and Ach5, 1D1609 has a distinctive host range. Here, we report its complete genome sequence to facilitate future studies.

Agrobacterium tumefaciens is known for its ability to transfer a DNA segment on the tumor-inducing (Ti) plasmid into the nuclear genome of its plant hosts (1). Because of this property, *A. tumefaciens* is widely used in genetic engineering (2, 3). The strain *A. tumefaciens* 1D1609 was isolated from a crown gall on a field-grown alfalfa plant in Imperial Valley, southern California (4). Previous infection assays demonstrated that this strain has an infectivity profile distinct from those of other well-characterized strains, such as C58 and Ach5 (4, 5). Thus, comparative analysis among these *A. tumefaciens* strains could shed light on the genetic determinants of host range and infection efficiency, which would improve their biotechnological applications. To facilitate such studies, we determined the complete genome sequence of *A. tumefaciens* 1D1609 and report the results here.

The procedures for sequencing, assembly, and annotation are based on those described previously (6–8). Briefly, the Illumina MiSeq platform was used to generate 301-bp reads from one paired-end library (~550-bp insert, 11,564,340 reads) and one mate pair library (~4,100-bp insert, 8,219,766 reads). The *de novo* assembly was performed using AllPaths-LG (9). The initial draft assembly was iteratively improved using PAGIT (10). In each iteration, the Illumina reads were mapped to the assembly using Burrows-Wheeler Aligner (BWA) (11), programmatically checked using SAMtools (12), and visually inspected using Integrative Genomics Viewer (IGV) (13). The regions with repetitive sequences (e.g., rRNA gene clusters) or low coverage of Illumina reads were confirmed by PCR and Sanger sequencing. The iterative process was continued until the complete genome assembly was obtained and verified. Gene prediction was done using RNAmmer (14), tRNAscan-SE (15), Prodigal (16), and GeneMark.hmm (17). The initial annotation was based on the homologous genes in *A. tumefaciens* C58 (18–20) and Ach5 (8) as identified by OrthoMCL (21). Subsequently, manual curation was performed based on BLASTP (22) searches against the National Center for Biotechnology Information (NCBI) nonredundant protein database (23), the NCBI Conserved Domain Database (CDD) (24), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (25, 26). Finally, noncoding RNAs were annotated based on the Rfam database (27).

The complete genome sequence of *A. tumefaciens* 1D1609 consists of one circular chromosome (3,058,772 bp), one linear chromosome (2,329,227 bp), one octopine-type Ti plasmid (166,117 bp), and two other plasmids (pAt1D1609a, 243,381 bp; pAt1D1609b, 187,640 bp). The first version of annotation includes 12 rRNA genes, 53 tRNA genes, 5,630 protein-coding genes, and 15 noncoding RNAs.

Received 27 February 2018 **Accepted** 22 March 2018 **Published** 19 April 2018

Citation Cho S-T, Haryono M, Chang H-H, Santos MNM, Lai E-M, Kuo C-H. 2018. Complete genome sequence of *Agrobacterium tumefaciens* 1D1609. Genome Announc 6: e00253-18. https://doi.org/10.1128/genomeA.00253-18.

Address correspondence to Chih-Horng Kuo, chk@gate.sinica.edu.tw.

Accession number(s). The complete genome sequence of *A. tumefaciens* 1D1609 has been deposited at DDBJ/EMBL/GenBank under the accession numbers CP026924 to CP026928.

ACKNOWLEDGMENTS

We thank the Genomic Technology Core of our institute for providing the Sanger sequencing service. The Illumina paired-end sequencing service was provided by the Genomics Core Facility (Institute of Molecular Biology, Academia Sinica), and the Illumina mate pair sequencing service was provided by Yourgene Bioscience (New Taipei City, Taiwan).

Funding for this project was provided by the Institute of Plant and Microbial Biology at Academia Sinica and the Ministry of Science and Technology of Taiwan (grant MOST 105-2311-B-001-067) to C.-H.K.

The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. Nester EW. 2015. *Agrobacterium*: nature's genetic engineer. Front Plant Sci 5:730. https://doi.org/10.3389/fpls.2014.00730.
2. Hwang H-H, Gelvin SB, Lai E-M. 2015. Editorial: "*Agrobacterium* biology and its application to transgenic plant production". Front Plant Sci 6:265. https://doi.org/10.3389/fpls.2015.00265.
3. Hwang H-H, Yu M, Lai E-M. 2017. *Agrobacterium*-mediated plant transformation: biology and applications. Arabidopsis Book 15:e0186. https://doi.org/10.1199/tab.0186.
4. Palumbo JD, Phillips DA, Kado CI. 1998. Characterization of a new *Agrobacterium tumefaciens* strain from alfalfa (*Medicago sativa* L.). Arch Microbiol 169:381–386. https://doi.org/10.1007/s002030050586.
5. Hwang H-H, Wu ET, Liu S-Y, Chang S-C, Tzeng K-C, Kado CI. 2013. Characterization and host range of five tumorigenic *Agrobacterium tumefaciens* strains and possible application in plant transient transformation assays. Plant Pathol 62:1384–1397. https://doi.org/10.1111/ppa.12046.
6. Lo W-S, Chen L-L, Chung W-C, Gasparich GE, Kuo C-H. 2013. Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. BMC Genomics 14:22. https://doi.org/10.1186/1471-2164-14-22.
7. Chung W-C, Chen L-L, Lo W-S, Lin C-P, Kuo C-H. 2013. Comparative analysis of the peanut witches'-broom phytoplasma genome reveals horizontal transfer of potential mobile units and effectors. PLoS One 8:e62770. https://doi.org/10.1371/journal.pone.0062770.
8. Huang Y-Y, Cho S-T, Lo W-S, Wang Y-C, Lai E-M, Kuo C-H. 2015. Complete genome sequence of *Agrobacterium tumefaciens* Ach5. Genome Announc 3:e00570-15. https://doi.org/10.1128/genomeA.00570-15.
9. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A 108:1513–1518. https://doi.org/10.1073/pnas.1017351108.
10. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc 7:1260–1284. https://doi.org/10.1038/nprot.2012.068.
11. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
12. 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.
13. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative Genomics Viewer. Nat Biotechnol 29:24–26. https://doi.org/10.1038/nbt.1754.
14. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.
15. Lowe T, Eddy S. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964.
16. Hyatt D, Chen G-L, LoCascio P, Land M, Larimer F, Hauser L. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
17. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29:2607–2618. https://doi.org/10.1093/nar/29.12.2607.
18. 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, Kutyavin T, Levy R, Li M-J, 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 Z-Y, Dolan M, Chumley F, Tingey SV, Tomb J-F, Gordon MP, Olson MV, et al. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science 294:2317–2323. https://doi.org/10.1126/science.1066804.
19. Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M, Qurollo 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.
20. Slater S, Setubal JC, Goodner B, Houmiel K, Sun J, Kaul R, Goldman BS, Farrand SK, Almeida N, Burr T, Nester E, Rhoads DM, Kadoi R, Ostheimer T, Pride N, Sabo A, Henry E, Telepak E, Cromes L, Harkleroad A, Oliphant L, Pratt-Szegila P, Welch R, Wood D. 2013. Reconciliation of sequence data and updated annotation of the genome of *Agrobacterium tumefaciens* C58, and distribution of a linear chromosome in the genus *Agrobacterium*. Appl Environ Microbiol 79:1414–1417. https://doi.org/10.1128/AEM.03192-12.
21. Li L, Stoeckert CJ, Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189. https://doi.org/10.1101/gr.1224503.
22. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden T. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
23. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2015. GenBank. Nucleic Acids Res 43:D30–D35. https://doi.org/10.1093/nar/gku1216.
24. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng

C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45: D200–D203. https://doi.org/10.1093/nar/gkw1129.
25. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27.
26. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38:D355–D360. https://doi.org/10.1093/nar/gkp896.
27. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res 46: D335–D342. https://doi.org/10.1093/nar/gkx1038.