

Complete Genome Sequence of *Agrobacterium tumefaciens* Ach5

Ya-Yi Huang,^a Shu-Ting Cho,^a Wen-Sui Lo,^{a,b,c} Yi-Chieh Wang,^a Erh-Min Lai,^{a,b,d}  Chih-Horng Kuo^{a,b,d}

Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan^a; Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, National Chung Hsing University and Academia Sinica, Taipei, Taiwan^b; Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan^c; Biotechnology Center, National Chung Hsing University, Taichung, Taiwan^d

***Agrobacterium tumefaciens* is a phytopathogenic bacterium that causes crown gall disease. The strain Ach5 was isolated from yarrow (*Achillea ptarmica* L.) and is the wild-type progenitor of other derived strains widely used for plant transformation. Here, we report the complete genome sequence of this bacterium.**

Received 27 April 2015 Accepted 30 April 2015 Published 4 June 2015

Citation 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(3):e00570-15. doi:10.1128/genomeA.00570-15.

Copyright © 2015 Huang et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/4.0/).

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

Agrobacterium tumefaciens is a soil-dwelling bacterium that is often associated with plants. Some strains harbor a tumor-inducing (Ti) plasmid, which is required for their phytopathogenicity. During infection, the transferred DNA (T-DNA) in the Ti plasmid is integrated into the host nuclear genome. Because of this property, *A. tumefaciens* has been developed into an efficient system for plant transformation (1).

A. tumefaciens Ach5 was isolated from a crown gall on yarrow collected in Contra Costa County, CA, USA (2). It harbors an octopine-type Ti plasmid and could induce tumors on various woody fruit plants, as well as tobacco and tomato (3). Two notable strains have been derived from Ach5, first by Tn904 mutagenesis to generate LBA4213 (4), and subsequently, by deletions of the T-DNA to generate LBA4404 (5). The disarmed LBA4404 is widely used for plant transformation (6). Although draft genomes are available for LBA4213 (GenBank accession numbers CP007225 to CP007228 [7]) and LBA4404 (JMKN01000001 to JMKN01000039), the genomic information of their wild-type progenitor is still lacking. To fill this gap, we report the complete genome sequence of *A. tumefaciens* Ach5 here.

The procedures for sequencing, assembly, and annotation are based on those described previously (8–16). Briefly, the Illumina MiSeq platform was used to generate 301-bp reads from one paired-end library (~391-bp insert, 8,352,732 reads) and one mate pair library (~4,150-bp insert, 9,594,498 reads). Two parallel assembly approaches were used, both using LBA4213 (7) as the reference. For the *de novo* approach, we used ALLPATHS-LG (17) to generate an initial draft. The scaffolds were oriented according to the reference and iteratively improved by PAGIT (18). For each iteration, the raw reads were mapped using BWA (19) and manually inspected using IGV (20). For the resequencing approach, the reference was used as the starting point for iterative corrections. The raw reads were mapped using BWA and polymorphisms were checked using SAMtools (21). Gaps were manually inserted at large indel sites and closed by using PAGIT. The final assembly was validated by both approaches; eight low-coverage regions were confirmed by Sanger sequencing. Gene prediction was done using RNAmmer (22), tRNAscan-SE (23), and Prodigal

(24). The initial annotation was based on orthologs in *A. tumefaciens* C58 (25, 26) as identified by OrthoMCL (27). Subsequent manual curation was based on BLASTP (28) searches against the NCBI nonredundant database (29) and the KEGG database (30, 31). For the KEGG tool, we added representatives from *Agrobacterium* and *Rhizobium* (abbreviated identifiers: atu, ara, avi, ret, rec, rle, rlt, and rlg) to the default “prokaryotes” reference set.

The complete genome sequence of *A. tumefaciens* Ach5 has an overall G+C content of 58.5%. It consists of one circular chromosome (2,833,887 bp), one linear chromosome (2,095,752 bp), one mega plasmid (designated pAt; 544,752 bp), and one Ti plasmid (194,264 bp). The first version of annotation includes 15 rRNA genes, 56 tRNA genes, and 5,276 protein-coding genes.

Nucleotide sequence accession numbers. The complete genome sequence of *A. tumefaciens* Ach5 has been deposited at DDBJ/EMBL/GenBank under the accession numbers CP011246 to CP011249.

ACKNOWLEDGMENTS

The funding for this project was provided by the Institute of Plant and Microbial Biology, Academia Sinica, to E.-M.L. and C.-H.K.

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

REFERENCES

- Nester EW. 2014. *Agrobacterium*: nature’s genetic engineer. *Front Plant Sci* 5:730. <http://dx.doi.org/10.3389/fpls.2014.00730>.
- Archdeacon J, Bouhouche N, O’Connell F, Kado CI. 2000. A single amino acid substitution beyond the C2H2-zinc finger in Ros derepresses virulence and T-DNA genes in *Agrobacterium tumefaciens*. *FEMS Microbiol Lett* 187:175–178. <http://dx.doi.org/10.1111/j.1574-6968.2000.tb09156.x>.
- Cervera M, López MM, Navarro L, Peña L. 1998. Virulence and super-virulence of *Agrobacterium tumefaciens* in woody fruit plants. *Physiol Mol Plant Pathol* 52:67–78. <http://dx.doi.org/10.1006/pmpp.1997.0135>.
- Ooms G, Klapwijk PM, Poulsen JA, Schilperoord RA. 1980. Characteriza-

- tion of Tn904 insertions in octopine Ti plasmid mutants of *Agrobacterium tumefaciens*. *J Bacteriol* 144:82–91.
5. Ooms G, Hooykaas PJ, Van Veen RJ, Van Beelen P, Regensburg-Tuink TJ, Schilperoort RA. 1982. Octopine Ti-plasmid deletion mutants of *Agrobacterium tumefaciens* with emphasis on the right side of the T-region. *Plasmid* 7:15–29. [http://dx.doi.org/10.1016/0147-619X\(82\)90023-3](http://dx.doi.org/10.1016/0147-619X(82)90023-3).
 6. Hoekema A, Hirsch PR, Hooykaas PJJ, Schilperoort RA. 1983. A binary plant vector strategy based on separation of *vir*- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature* 303:179–180. <http://dx.doi.org/10.1038/303179a0>.
 7. Henkel CV, den Dulk-Ras A, Zhang X, Hooykaas PJ. 2014. Genome sequence of the octopine-type *Agrobacterium tumefaciens* strain Ach5. *Genome Announc* 2(2):e00225–14. <http://dx.doi.org/10.1128/genomeA.00225-14>.
 8. 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. <http://dx.doi.org/10.1371/journal.pone.0062770>.
 9. 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. <http://dx.doi.org/10.1186/1471-2164-14-22>.
 10. Ku C, Lo W-S, Chen L-L, Kuo C-H. 2013. Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in *Spiroplasma*. *Genome Biol Evol* 5:1151–1164. <http://dx.doi.org/10.1093/gbe/evt084>.
 11. Lo W-S, Ku C, Chen L-L, Chang T-H, Kuo C-H. 2013. Comparison of metabolic capacities and inference of gene content evolution in mosquito-associated *Spiroplasma diminutum* and *S. taiwanense*. *Genome Biol Evol* 5:1512–1523. <http://dx.doi.org/10.1093/gbe/evu033>.
 12. Ku C, Lo W-S, Chen L-L, Kuo C-H. 2014. Complete genome sequence of *Spiroplasma apis* B31^T (ATCC 33834), a bacterium associated with May disease of honeybees (*Apis mellifera*). *Genome Announc* 2(1):e01151–13. <http://dx.doi.org/10.1128/genomeA.01151-13>.
 13. Chang T-H, Lo W-S, Ku C, Chen L-L, Kuo C-H. 2014. Molecular evolution of the substrate utilization strategies and putative virulence factors in mosquito-associated *Spiroplasma* species. *Genome Biol Evol* 6:500–509. <http://dx.doi.org/10.1093/gbe/evu033>.
 14. Lo W-S, Chen H, Chen C-Y, Kuo C-H. 2014. Complete genome sequence of *Vibrio vulnificus* 93U204, a bacterium isolated from diseased tilapia in Taiwan. *Genome Announc* 2(5):e01005–14. <http://dx.doi.org/10.1128/genomeA.01005-14>.
 15. Su H-J, Hogenhout SA, Al-Sadi AM, Kuo C-H. 2014. Complete chloroplast genome sequence of Omani lime (*Citrus aurantiiifolia*) and comparative analysis within the rosids. *PLoS One* 9:e113049. <http://dx.doi.org/10.1371/journal.pone.0113049>.
 16. Chang H-H, Cho S-T, Canale MC, Mugford ST, Lopes JR, Hogenhout SA, Kuo C-H. 2015. Complete genome sequence of “*Candidatus* Sulcia muelleri” ML, an obligate nutritional symbiont of maize leafhopper (*Dalbulus maidis*). *Genome Announc* 3(1):e01483–14. <http://dx.doi.org/10.1128/genomeA.01483-14>.
 17. 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. <http://dx.doi.org/10.1073/pnas.1017351108>.
 18. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. 2012. A postassembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc* 7:1260–1284. <http://dx.doi.org/10.1038/nprot.2012.068>.
 19. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>.
 20. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29:24–26. <http://dx.doi.org/10.1038/nbt.1754>.
 21. 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 of Map Format and SAMtools. *Bioinformatics* 25:2078–2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>.
 22. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
 23. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
 24. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
 25. 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. <http://dx.doi.org/10.1126/science.1066803>.
 26. 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, Kutayin 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, Nester EW. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* 294:2317–2323. <http://dx.doi.org/10.1126/science.1066804>.
 27. Li L, Stoekert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <http://dx.doi.org/10.1101/gr.1224503>.
 28. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <http://dx.doi.org/10.1186/1471-2105-10-421>.
 29. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2015. GenBank. *Nucleic Acids Res* 43:D30–D35. <http://dx.doi.org/10.1093/nar/gku1216>.
 30. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28:27–30. <http://dx.doi.org/10.1093/nar/28.1.27>.
 31. 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. <http://dx.doi.org/10.1093/nar/gkp896>.