

# Genome Sequence of the Octopine-Type *Agrobacterium tumefaciens* Strain Ach5

Christiaan V. Henkel, Amke den Dulk-Ras, Xiaorong Zhang, Paul J. J. Hooykaas

Department of Molecular and Developmental Genetics, Institute of Biology, Leiden University, Leiden, The Netherlands

We have sequenced the complete genome of the plant pathogen *Agrobacterium tumefaciens* strain LBA4213, a derivative of the wild-type strain *A. tumefaciens* Ach5 and the ancestor of *A. tumefaciens* strain LBA4404 used in genetic engineering. The genome consists of a circular chromosome and a linear chromosome, as well as a megaplasmid and a tumor-inducing plasmid.

Received 28 February 2014 Accepted 13 March 2014 Published 27 March 2014

**Citation** Henkel CV, den Dulk-Ras A, Zhang X, Hooykaas PJJ. 2014. Genome sequence of the octopine-type *Agrobacterium tumefaciens* strain Ach5. *Genome Announc.* 2(2): e00225-14. doi:10.1128/genomeA.00225-14.

**Copyright** © 2014 Henkel 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/3.0/).

Address correspondence to Paul J. J. Hooykaas, p.j.j.hooykaas@biology.leidenuniv.nl.

Soil bacteria of the species *Agrobacterium tumefaciens* are the causative agents of crown gall disease in plants. They induce plant tumors by transferring genes encoding phytohormones and opine biosynthetic enzymes from a tumor-inducing (Ti) plasmid to plant cell genomes.

*A. tumefaciens* strain Ach5 belongs to the tumor-inducing biotype I class of agrobacteria and harbors a Ti plasmid of the octopine type. *A. tumefaciens* strain LBA4213 is a Tn904-mutagenized derivative of the wild-type Ach5 (1, 2). LBA4213 is itself the immediate precursor of *A. tumefaciens* strain LBA4404, which has a major biotechnological application as the genomic host for the binary vector system employed in the genetic engineering of plants (3). *A. tumefaciens* LBA4404 differs from LBA4213 only by the deletion of a large region of the Ti plasmid.

The genome of *A. tumefaciens* strain LBA4213 was sequenced using 99-nucleotide paired-end reads on an Illumina HiSeq 2000. A total of 7,098,396 read pairs were obtained (approximately 250× coverage). Sequencing errors were removed using Quake version 0.3 (4), and the reads were assembled into 99 contigs, with an  $N_{50}$  of 585,488 bp, using Velvet version 1.2.03 (5).

Major ambiguous connections in the assembly graph were resolved using targeted PCRs. At minor ambiguities (e.g., variation within repeats), nucleotides were masked by Ns (1,365 positions in total). Full-length chromosome and plasmid sequences were reconstructed from the graph by making reasonable assumptions about the topology of individual molecules (e.g., the existence of a circular Ti plasmid). The final genome assembly is 5.63 Mb in length, has a C+G content of 58.5%, and consists of a circular chromosome of 2,773,134 bp, a linear chromosome of 2,095,074 bp, a “cryptic” megaplasmid of 556,485 bp, and a Ti plasmid of 205,997 bp. More than 99.95% of the corrected reads align to these sequences, and apart from masked sequence, every position is covered, indicating the assembly is complete.

Finished sequences were submitted to the NMPDR RAST server (version 4.0) for the annotation of protein-coding, rRNA, and tRNA genes (6). Gene predictions for the virulence genes on the Ti plasmid were validated manually using the previously annotated octopine Ti plasmid sequence (7). In total, 5,432 protein-

coding sequences were annotated, as well as 5 clusters of rRNA genes and 52 tRNA genes. Both chromosomes contain rRNA and tRNA genes, whereas the plasmids do not.

Based on its *recA* housekeeping gene variant, LBA4213 is a member of genomospecies 1 of *Agrobacterium* biotype I (8). To date, several *Agrobacterium* genome sequences have been finalized, revealing a division of the chromosomes into a linear molecule and a circular molecule unique to biotype I (9–11). Both chromosomes of LBA4213 are mostly collinear with the previously published complete chromosomes of *A. tumefaciens* strain C58 and *Agrobacterium* sp. strain H13-3. However, the megaplasmid shows little conservation between the three sequenced species, with only 12.4% and 26.6% aligning to C58 and H13-3 counterparts, respectively. Except for the Tn904 insertion encoding streptomycin resistance, the Ti plasmid is 99.8% identical to the previously published octopine-type consensus (7).

**Nucleotide sequence accession numbers.** The complete genome sequence of *A. tumefaciens* LBA4213 has been deposited in GenBank under accession no. CP007225 to CP007228.

## ACKNOWLEDGMENT

This work was supported by a grant from the Royal Netherlands Academy of Sciences.

## REFERENCES

1. Ooms G, Klapwijk PM, Poulis JA, Schilperoort RA. 1980. Characterization of Tn904 insertions in octopine Ti plasmid mutants of *Agrobacterium tumefaciens*. *J. Bacteriol.* 144:82–91.
2. Ooms G, Hooykaas PJJ, Van Veen RJM, Van Beelen P, Regensburg-Tuink TJG, 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).
3. 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>.
4. Kelley DR, Schatz MC, Salzberg SL. 2010. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol.* 11:R116. <http://dx.doi.org/10.1186/gb-2010-11-11-r116>.
5. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read

- assembly using de Bruijn graphs. *Genome Res.* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
6. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
  7. Zhu J, Oger PM, Schrammeijer B, Hooykaas PJJ, Farrand SK, Winans SC. 2000. The bases of crown gall tumorigenesis. *J. Bacteriol.* 182:3885–3895. <http://dx.doi.org/10.1128/JB.182.14.3885-3895.2000>.
  8. Costechareyre D, Rhouma A, Lavire C, Portier P, Chapulliot D, Bertolla F, Boubaker A, Dessaux Y, Nesme X. 2010. Rapid and efficient identification of *Agrobacterium* species by *recA* allele analysis. *Microb. Ecol.* 60:862–872. <http://dx.doi.org/10.1007/s00248-010-9685-7>.
  9. Slater SC, Goldman BS, Goodner B, Setubal JC, Farrand SK, Nester EW, Burr TJ, Banta L, Dickerman AW, Paulsen I, Otten L, Suen G, Welch R, Almeida NF, Arnold F, Burton OT, Du Z, Ewing A, Godsy E, Heisel S, Houmiel KL, Jhaveri J, Lu J, Miller NM, Norton S, Chen Q, Phoolcharoen W, Ohlin V, Ondrusek D, Pride N, Stricklin SL, Sun J, Wheeler C, Wilson L, Zhu H, Wood DW. 2009. Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* 191:2501–2511. <http://dx.doi.org/10.1128/JB.01779-08>.
  10. Slater S, Setubal JC, Goodner B, Houmiel K, Sun J, Kaul R, Goldman BS, Farrand SK, Almeida N, Jr, Burr T, Nester E, Rhoads DM, Kadoi R, Ostheimer T, Pride N, Sabo A, Henry E, Telepak E, Cromes L, Har-kleroad 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. <http://dx.doi.org/10.1128/AEM.03192-12>.
  11. Wibberg D, Blom J, Jaenicke S, Kollin F, Rupp O, Scharf B, Schneiker-Bekel S, Sczcepanowski R, Goesmann A, Setubal JC, Schmitt R, Pühler A, Schlüter A. 2011. Complete genome sequencing of *Agrobacterium* sp. H13-3, the former *Rhizobium lupini* H13-3, reveals a tripartite genome consisting of a circular and a linear chromosome and an accessory plasmid but lacking a tumor-inducing Ti-plasmid. *J. Biotechnol.* 155:50–62. <http://dx.doi.org/10.1016/j.jbiotec.2011.01.010>.