Draft Genome Sequence of *Agrobacterium rhizogenes* Strain NCPPB2659

Jose A. Valdes Franco,[a] Ray Collier,[a] Yi Wang,[a,b] Naxin Huo,[a,b] Yong Gu,[a] Roger Thilmony,[a] James G. Thomson[a]

USDA-ARS Crop Improvement and Genetics, Western Regional Research Center, Albany, California, USA[a]; Department of Plant Sciences, University of California, Davis, Davis, California, USA[b]

This work reports the draft genome sequence of *Agrobacterium rhizogenes* strain NCPPB2659 (also known as strain K599). The assembled genome contains 5,277,347 bp, composed of one circular chromosome, the pRi2659 virulence plasmid, and 17 scaffolds pertaining to the linear chromosome. The wild-type strain causes hairy root disease in dicots and has been used to make transgenic hairy root cultures and composite plants (nontransgenic shoots with transgenic roots). Disarmed variants of the strain have been used to produce stable transgenic monocot and dicot plants.

Address correspondence to Roger Thilmony, roger.thilmony@ars.usda.gov, or James G. Thomson, james.thomson@ars.usda.gov.

*A*grobacterium rhizogenes is a Gram-negative soil bacterial plant pathogen that causes hairy root disease (1) resulting from *in planta* expression of *rol* genes (2) after transfer DNA (T-DNA) movement into the host plant cells (3). The Ri T-DNA of *A. rhizogenes* strain NCPPB2659, isolated from cucumbers displaying hairy root disease, was characterized (4), and the pRi2659 virulence plasmid sequenced (5). *A. rhizogenes* strain NCPPB2659 induces hairy roots from transformation-recalcitrant plants, like soybean, with higher efficiency relative to other strains (6), and this characteristic has been used for gene function studies in hairy root cultures (7, 8) and composite plants (9). The strain has been disarmed by two independent groups (5, 10) and used to generate stable transgenic monocot (5, 11) and dicot (12) plants. To enhance the stability of constructs containing large synthetic T-DNA molecules in the wild-type and disarmed strains (such as *A. rhizogenes* 18r12v), the genome sequence data were used to design a cassette for the *TcR* insertional inactivation of *recA* (13, 14). These variant strains with deactivated *recA* remain competent for plant transformation (data not shown).

Genomic and plasmid DNA were extracted from *A. rhizogenes* strain NCPPB2659, according to the protocol of Wise et al. (15). Illumina paired-end libraries were constructed from sonicated DNA using the Kapa DNA library preparation kit (Kapa Biosystems). Fragments between 500 and 1,000 bp were selected with a polyethylene glycol (PEG)-NaCl solution method (Kapa Biosystems). A total of 3,165,594 paired-end reads were generated on an Illumina MiSeq, at a read length of 300 bp. Reads were error corrected with Quake version 0.3.5 (16) and quality filtered and trimmed with Trimmomatic version 0.32 (17), resulting in 2,546,326 high-quality reads. The initial sequence assembly was performed using SOAP*denovo*2 (18). This assembly produced 25 scaffolds (minimum, 617 bp; maximum, 1,934,317 bp; N_{50}, 1,468,096 bp). Due to the fragmented nature of the resulting assembly, a manual analysis was performed on the CLC Genomics Workbench version 8.5 (CLC bio/Qiagen) to reveal low-coverage regions and broken paired-end reads and identify collapsed or missing regions, within and between the scaffolds. These regions were then PCR amplified and characterized by sequencing. Alignment analysis of these sequences with the scaffolds allowed assembly of the complete circular chromosome. Subsequent mapping analyses validated the assembly (>97% of the reads mapped in pairs with uniquely assigned matches, excluding clearly repetitive areas showing higher-than-average coverage). This allowed the generation of a final collection of 18 sequences (N_{50}, 3,003,247 bp), with a complete circular chromosome and 17 scaffolds of proposed linear chromosome (excluding the prior published virulence plasmid sequence), composed of 5,277,347 bp, with a G+C content of 59.8% and an overall coverage of approximately 110×.

Automated annotation was performed by the NCBI Prokaryotic Genome Annotation Pipeline. *A. rhizogenes* strain NCPPB2659 contains 4,860 predicted coding sequences, 72 predicted RNA genes, and 4 complete copies of each rRNA. The 33 identified pseudogenes exhibit no obvious sequencing or assembly mistakes.

Nucleotide sequence accession numbers. The entire genome shotgun sequence data set has been deposited at DDBJ/EMBL/GenBank under accession no. LYBK00000000. Illumina raw reads are available at the NCBI Short Read Archive under accession no. SRR3537230.

REFERENCES

1. **Riker AJ, Banfield WM, Wright WH, Keitt GW, Sagen HE.** 1930. Studies on infectious hairy root of nursery apple trees. J Agric Res **41:**507–540.
2. **Serino G, Clerot D, Brevet J, Costantino P, Cardarelli M.** 1994. *rol* genes of *Agrobacterium rhizogenes* cucumopine strain: sequence, effects and pattern of expression. Plant Mol Biol **26:**415–422. http://dx.doi.org/10.1007/BF00039550.
3. **Chilton MD, Tepfer DA, Petit A, David C, Casse-Delbart F, Tempé J.** 1982. *Agrobacterium rhizogenes* inserts T-DNA into the genomes of the host plant root cells. Nature **295:**432–434. http://dx.doi.org/10.1038/295432a0.
4. **Combard A, Brevet J, Borowski D, Cam K, Tempé J.** 1987. Physical map of the T-DNA region of *Agrobacterium rhizogenes* strain NCPPB2659. Plasmid **18:**70–75. http://dx.doi.org/10.1016/0147-619X(87)90079-5.
5. **Mankin SL, Hill DS, Olhoft PM, Toren E, Wenck AR, Nea L, Xing L, Brown JA, Fu H, Ireland L, Jia H, Hillebrand H, Jones T, Song H-S.** 2007. Disarming and sequencing of *Agrobacterium rhizogenes* strain K599 (NCPPB2659) plasmid pRi2659. In Vitro Cell Dev Biol Plant **43:**521–535. http://dx.doi.org/10.1007/s11627-007-9071-4.
6. **Savka MA, Ravillion B, Noel GR, Farrand SK.** 1990. Induction of hairy roots on cultivated soybean genotypes and their use to propagate the soybean cyst nematode. Phytopathology **80:**503–508. http://dx.doi.org/10.1094/Phyto-80-503.
7. **Cho HJ, Farrand SK, Noel GR, Widholm JM.** 2000. High-efficiency induction of soybean hairy roots and propagation of the soybean cyst nematode. Planta **210:**195–204. http://dx.doi.org/10.1007/PL00008126.
8. **Xiao X, Ma F, Chen C-L, Guo W-W.** 2014. High efficient transformation of auxin reporter gene into trifoliate orange via *Agrobacterium rhizogenes*-mediated co-transformation. Plant Cell Tissue Organ Cult **118:**137–146. http://dx.doi.org/10.1007/s11240-014-0469-5.
9. **Collier R, Fuchs B, Walter N, Kevin Lutke W, Taylor CG.** 2005. *Ex vitro* composite plants: an inexpensive, rapid method for root biology. Plant J **43:**449–457. http://dx.doi.org/10.1111/j.1365-313X.2005.02454.x.
10. **Veena V, Taylor CG.** 2007. *Agrobacterium rhizogenes*: recent developments and promising applications. In Vitro Cell Dev Biol **43:**383–403. http://dx.doi.org/10.1007/s11627-007-9096-8.
11. **Collier R, Bragg J, Hernandez BT, Vogel JP, Thilmony R.** 2016. Use of *Agrobacterium rhizogenes* strain 18r12v and paromomycin selection for transformation of *Brachypodium distachyon* and *Brachypodium sylvaticum*. Front Plant Sci **7:**716. http://dx.doi.org/10.3389/fpls.2016.00716.
12. **Olhoft PM, Bernal LM, Grist LB, Hill DS, Mankin SL, Shen Y, Kalogerakis M, Wiley H, Toren E, Song H-S, Hillebrand H, Jones T.** 2007. A novel *Agrobacterium rhizogenes*-mediated transformation method of soybean [*Glycine max* (L.) Merrill] using primary-node explants from seedlings. In Vitro Cell Dev Biol Plant **43:**536–549. http://dx.doi.org/10.1007/s11627-007-9050-9.
13. **Comai L, Schilling-Cordaro C, Mergia A, Houck CM.** 1983. A new technique for genetic engineering of *Agrobacterium* Ti plasmid. Plasmid **10:**21–30. http://dx.doi.org/10.1016/0147-619X(83)90054-9.
14. **Ruvkun GB, Ausubel FM.** 1981. A general method for site-directed mutagenesis in prokaryotes. Nature **289:**85–88. http://dx.doi.org/10.1038/289085a0.
15. **Wise AA, Liu Z, Binns AN.** 2006. Nucleic acid extraction from *Agrobacterium* strains, p 67–76. *In* Wang K (ed), *Agrobacterium* protocols vol 1, 2nd ed. Humana Press, Totowa, NJ.
16. **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.
17. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:**2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.
18. **Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J.** 2012. SOAP*denovo*2: an empirically improved memory-efficient short-read *de novo* assembler. GigaScience **1:**18. http://dx.doi.org/10.1186/2047-217X-1-18.