

Draft genome sequences of 12 *Agrobacterium* strains isolated from coastal plants in Taiwan

Hsi-Ching Yen,¹ Fan-Chen Huang,^{2,3} Pei-Ru Chien,³ Chih-Lin Wu,³ Liang-Yu Chen,³ Chuan-Wen Ho,³ Guo-Zhang M. Song,⁴ Wei-Jen Lin,⁵ Hsing-Juh Lin,³ Chieh-Chen Huang,^{3,6,7} Hau-Hsuan Hwang,^{3,6,7} Chih-Horng Kuo^{1,8}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We report draft genome sequences of 12 *Agrobacterium* strains isolated from coastal plants in Taiwan. Assemblies range from 4.95 to 5.42 Mb with >97.5% completeness. Genomic analyses reveal species identities and secretion system genes, providing resources for future research.

KEYWORDS genome, *Agrobacterium*, endophyte, Taiwan

The genus *Agrobacterium* comprises diverse species associated with soil and plants. Some strains harbor oncogenic plasmids and can be used for *Agrobacterium*-mediated transformation (AMT) (1). A major limitation of AMT is that many plants are recalcitrant to common laboratory strains. Early work demonstrated that wild-type *Agrobacterium* strains differ substantially in host range, suggesting that strain diversity could broaden AMT application to different crops (2). Guided by this rationale, we screened draft genome assemblies from our collection of endophytic bacteria isolated from coastal plants in Taiwan, identified 12 strains as members of the genus *Agrobacterium*, and report their genome sequences in this work (Table 1).

Unless otherwise noted, methods followed the cited references, kits were used according to the manufacturers' instructions, and bioinformatics tools were used with default settings. For strain isolation, root samples were surface-sterilized, homogenized in distilled water, serially diluted, spread onto Luria Broth (LB) agar, and incubated at 30°C for 48 h to obtain single colonies, as described (3). Colonies were re-streaked at least three times to ensure purity. For DNA extraction, a single colony was cultured in 1 mL LB broth at 30°C for 24 h, then processed sequentially using the Wizard Genomic DNA Purification Kit (Promega) and DNeasy Blood and Tissue Kit (Qiagen).

For shotgun sequencing, libraries were prepared using the SATLite DNA Prep Kit with the SATLite 2.0 Library Prep Automation Device (GeneReach Biotechnology) and sequenced on a NovaSeq X Plus using the 10B Reagent Kit (Illumina) to obtain paired-end 150 bp reads. Raw reads were trimmed using fastp v1.0.1 (4) and assembled using SPAdes v3.15.0 (5) with the "--isolate" option ($k = 21, 33, 55, \text{ and } 77$). Contigs >500 bp were retained. Trimmed reads were mapped to contigs using BWA v0.7.17-r1188 (6), and coverage was assessed using SAMtools v1.9 (7). For species identification, assemblies were compared against NCBI RefSeq genomes (8) with FastANI v1.33 (9). Assemblies were submitted to GenBank and annotated using the NCBI Prokaryotic Genome Annotation Pipeline v6.10 (10). Completeness and contamination were estimated using CheckM v1.2.4 (11). MacSyFinder v2.1.4 (12) was used to identify genes encoding type IV and type VI secretion systems, which are relevant to *Agrobacterium* virulence and ecology, respectively (13). To identify contigs corresponding to oncogenic plasmids, coding sequences from reference plasmids (14, 15) were queried using BLASTP v2.11.0 (16) against annotated coding sequences.

Editor Silvia T. Cardona, University of Manitoba, Winnipeg, Canada

Address correspondence to Hau-Hsuan Hwang, hauhsuan@dragon.nchu.edu.tw, or Chih-Horng Kuo, chk@gate.sinica.edu.tw.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 19 September 2025

Accepted 15 October 2025

Published 5 November 2025

Copyright © 2025 Yen 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/).

TABLE 1 Metadata and genome assembly statistics of the strains included in this study^a

Strain	Species	Similarity to type strain (AF/ANI %)	Collection date	Location	GPS coordinates	Host	SRA accession	Assembly accession	Number of reads (fold)	Coverage (fold)	Contig of contigs N50 (bp)	Assembly size (bp)	GC content (%)	Number of annotated genes	Completeness (%)	Contamination (%)	T6SS genes
NCHU4840	<i>Agrobacterium cavae</i>	85.5/96.2	02-03-2022	Taiwan: Miaoli County, Tongxiao Township	24,495 968 N 120.671 137 E	<i>Bidens pilosa</i>	SRR34682073	GCA_051590745.1	5,511,444 4	159.6	27	422,750	5,048,636 58.2	4,644	97.72	0.68	Complete
NCHU4860	<i>Agrobacterium cavae</i>	84.7/96.2	02-03-2022	Taiwan: Miaoli County, Tongxiao Township	24,495 968 N 120.671 137 E	<i>Urena lobata</i>	SRR34681896	GCA_052261035.1	6,259,710 0	183.5	15	884,046	4,951,969 58.2	4,579	97.76	0.95	Complete
NCHU4975	<i>Agrobacterium cavae</i>	85.7/96.2	02-03-2022	Taiwan: Miaoli County, Tongxiao Township	24,495 968 N 120.671 137 E	<i>Ipomoea pes-caprae</i> subsp. <i>brasilensis</i>	SRR34682131	GCA_052261055.1	5,653,280 2	162.8	29	403,161	5,048,170 58.2	4,647	97.72	0.68	Complete
NCHU4981	<i>Agrobacterium pusense</i>	89.3/98.2	02-03-2022	Taiwan: Miaoli County, Tongxiao Township	24,495 968 N 120.671 137 E	<i>Ipomoea pes-caprae</i> subsp. <i>brasilensis</i>	SRR34692460	GCA_052261075.1	6,124,070 8	166.2	37	654,217	5,355,332 59.1	5,079	99.60	0.77	Absent
NCHU5031	<i>Agrobacterium pusense</i>	88.6/98.2	09-05-2022	Taiwan: Miaoli County, Zhunan Township	24,675 569 N 120.830 987 E	<i>Oenothera lacinata</i>	SRR34685741	GCA_052261095.1	7,600,100 6	206.6	23	1,004,322	5,409,650 59.1	5,115	99.33	0.95	Absent
NCHU5032	<i>Agrobacterium pusense</i>	87.7/97.6	09-05-2022	Taiwan: Miaoli County, Zhunan Township	24,675 569 N 120.830 987 E	<i>Oenothera lacinata</i>	SRR34687319	GCA_052261115.1	6,567,210 0	178.4	30	440,007	5,416,145 59.3	5,172	99.55	0.62	Partial
NCHU5334	<i>Agrobacterium salinitolerans</i>	89.2/99.2	04-06-2022	Taiwan: Tainan City, Qigu District	23,157 924 N 120.077 311 E	<i>Bidens pilosa</i>	SRR34682062	GCA_052261135.1	6,247,710 2	175.8	29	443,794	5,207,595 59.3	4,955	99.71	0.29	Complete
NCHU5640	<i>Agrobacterium cavae</i>	85.7/96.2	22-08-2022	Taiwan: Yilan County, Toucheng Township	24,913 217 N 121.877 840 E	<i>Murdannia keisak</i>	SRR34682063	GCA_052261155.1	5,375,510 8	155.8	43	405,051	5,006,747 58.2	4,677	97.52	1.30	Complete
NCHU5641	<i>Agrobacterium cavae</i>	85.4/96.2	22-08-2022	Taiwan: Yilan County, Toucheng Township	24,913 217 N 121.877 840 E	<i>Murdannia keisak</i>	SRR34685756	GCA_052261175.1	6,621,420 0	193.0	43	405,051	5,006,273 58.2	4,680	97.52	1.30	Complete
NCHU5644	<i>Agrobacterium cavae</i>	85.6/96.2	22-08-2022	Taiwan: Yilan County, Toucheng Township	24,913 217 N 121.877 840 E	<i>Murdannia keisak</i>	SRR34692412	GCA_052261195.1	5,096,430 4	147.9	41	421,673	5,006,542 58.2	4,676	97.79	1.30	Complete
NCHU5645	<i>Agrobacterium cavae</i>	85.2/96.2	22-08-2022	Taiwan: Yilan County, Toucheng Township	24,913 217 N 121.877 840 E	<i>Ipomoea pes-caprae</i> subsp. <i>brasilensis</i>	SRR34685760	GCA_052261215.1	5,920,380 6	170.7	44	405,051	5,006,869 58.2	4,671	97.52	1.30	Complete
NCHU5738	<i>Agrobacterium deltaense</i>	92.0/97.9	14-10-2022	Taiwan: Changhua County, Fangyuan Township	23,928 038 N 120.314 438 E	<i>Pluchea indica</i>	SRR34685761	GCA_052261235.1	5,704,240 6	161.2	41	350,740	5,158,671 59.9	4,914	99.47	1.70	Complete

^a AF, alignment fraction; ANI, average nucleotide identity; T6SS, type VI secretion system.

Results are summarized in Table 1. All strains showed >84% alignment fraction and >96% average nucleotide identity to their respective type strains, allowing confident species assignment (17). Contig counts ranged from 15 to 44, N50 from 351 to 1,004 kb, assembly sizes from 4.95 to 5.42 Mb, GC content from 58.2 to 59.9%, and annotated genes from 4,579 to 5,172. All assemblies showed >97.5% completeness and <2% contamination. No oncogenic plasmid contigs nor type IV secretion system genes were detected. The feasibility of introducing oncogenic plasmids into these strains by conjugation remains to be tested. These genome assemblies provide useful references for future studies of these strains and comparative genomics of *Agrobacterium*.

ACKNOWLEDGMENTS

We thank members of GeneReach Biotechnology Corp., HHH lab, and CHK lab for discussions and technical assistance.

This work was supported by Academia Sinica and the National Science and Technology Council of Taiwan (107-2321-B-005-009, 108-2321-B-005-004, 109-2321-B-005-025, 110-2321-B-005-008, 110-2313-B-005-008, 111-2313-B-005-009-MY3, 114-2313-B-005-004). Additional support was provided by the Innovation and Development Center of Sustainable Agriculture and the Advanced Plant and Food Crop Biotechnology Center through the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education of Taiwan. Illumina library preparation and sequencing were performed by GeneReach Biotechnology Corp (Taichung, Taiwan). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

AUTHOR AFFILIATIONS

¹Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

²GeneReach Biotechnology Corp, Taichung, Taiwan

³Department of Life Sciences, National Chung Hsing University, Taichung, Taiwan

⁴Department of Soil and Water Conservation, National Chung Hsing University, Taichung, Taiwan

⁵Department of Biological Resources, National Chiayi University, Chiayi, Taiwan

⁶Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University, Taichung, Taiwan

⁷Advanced Plant and Food Crop Biotechnology Center, National Chung Hsing University, Taichung, Taiwan

⁸Biotechnology Center, National Chung Hsing University, Taichung, Taiwan

AUTHOR ORCIDs

Hsi-Ching Yen  <https://orcid.org/0009-0006-2346-2197>

Chuan-Wen Ho  <http://orcid.org/0000-0001-6221-124X>

Guo-Zhang M. Song  <http://orcid.org/0000-0002-9266-3690>

Wei-Jen Lin  <http://orcid.org/0000-0002-9217-5267>

Hsing-Juh Lin  <http://orcid.org/0000-0001-8322-7195>

Chieh-Chen Huang  <http://orcid.org/0000-0002-3739-6315>

Hau-Hsuan Hwang  <http://orcid.org/0000-0001-9132-0242>

Chih-Horng Kuo  <http://orcid.org/0000-0002-2857-0529>

FUNDING

Funder	Grant(s)	Author(s)
Academia Sinica		Chih-Horng Kuo
National Science and Technology Council		Hau-Hsuan Hwang
Ministry of Education		Hau-Hsuan Hwang

DATA AVAILABILITY

The raw sequencing reads and draft genome assemblies have been deposited in NCBI GenBank and are publicly available; accession numbers are listed in Table 1.

REFERENCES

- Nester EW. 2014. *Agrobacterium*: nature's genetic engineer. *Front Plant Sci* 5:730. <https://doi.org/10.3389/fpls.2014.00730>
- 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>
- Ho Y-N, Chiang H-M, Chao C-P, Su C-C, Hsu H-F, Guo C, Hsieh J-L, Huang C-C. 2015. In planta biocontrol of soilborne *Fusarium* wilt of banana through a plant endophytic bacterium, *Burkholderia cenocepacia* 869T2. *Plant Soil* 387:295–306. <https://doi.org/10.1007/s11104-014-2297-0>
- Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* 2:e107. <https://doi.org/10.1002/imt2.107>
- Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes *de novo* assembler. *Curr Protoc Bioinform* 70:e102. <https://doi.org/10.1002/cpbi.102>
- 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>
- 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>
- Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Sherry ST, Yankie L, Karsch-Mizrachi I. 2024. GenBank 2024 Update. *Nucleic Acids Res* 52:D134–D137. <https://doi.org/10.1093/nar/gkad903>
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
- 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>
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Néron B, Denise R, Coluzzi C, Touchon M, Rocha EPC, Abby SS. 2023. MacSyFinder v2: improved modelling and search engine to identify molecular systems in genomes. *Peer Community J* 3. <https://doi.org/10.24072/pcjournal.250>
- Weisberg AJ, Wu Y, Chang JH, Lai E-M, Kuo C-H. 2023. Virulence and ecology of agrobacteria in the context of evolutionary genomics. *Annu Rev Phytopathol* 61:1–23. <https://doi.org/10.1146/annurev-phyto-021622-125009>
- Chou L, Lin Y-C, Haryono M, Santos MNM, Cho S-T, Weisberg AJ, Wu C-F, Chang JH, Lai E-M, Kuo C-H. 2022. Modular evolution of secretion systems and virulence plasmids in a bacterial species complex. *BMC Biol* 20:16. <https://doi.org/10.1186/s12915-021-01221-y>
- Weisberg AJ, Davis EW, Tabima J, Belcher MS, Miller M, Kuo C-H, Loper JE, Grünwald NJ, Putnam ML, Chang JH. 2020. Unexpected conservation and global transmission of agrobacterial virulence plasmids. *Science* 368:eaba5256. <https://doi.org/10.1126/science.aba5256>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>
- Konstantinidis KT. 2023. Sequence-discrete species for prokaryotes and other microbes: a historical perspective and pending issues. *mLife* 2:341–349. <https://doi.org/10.1002/mlf2.12088>