Complete Genome Sequence of *Xanthomonas campestris* pv. campestris Strain 17 from Taiwan

Yi-Chung Liu,a Shiuh-Chaun Wang,b Yu-Jen Yu,g Kit-Man Fung,c Ming-Te Yang,c Yi-Hsiung Tseng,c Shih-Feng Tsai,d H. Sunny Sun,e Ping-Chiang Lyu,f Shan-Ho Choug

Institute of Population Sciences, National Health Research Institutes, Miaoli, Taiwana; Center for Biomedical Resources, National Health Research Institutes, Miaoli, Taiwanb; Institute of Molecular Biology, National Chung Hsing University, Taichung, Taiwanc; Institute of Molecular and Genomic Medicine, National Health Research Institutes, Miaoli, Taiwand; Institute of Molecular Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwane; Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwanf; Institute of Biochemistry and NCHU Biotechnological Center, National Chung Hsing University, Taichung, Taiwang

Xanthomonas campestris pv. campestris 17 is a Gram-negative bacterium that is phytopathogenic to cruciferous plants in Taiwan. The 4,994,426-bp-long genome consists of 24 contigs with 4,050 protein-coding genes, 1 noncoding RNA (ncRNA) gene, 6 rRNA genes, and 55 tRNA genes.

Address correspondence to Ping-Chiang Lyu, pclyu@mx.nthu.edu.tw, or Shan-Ho Chou, shchou@nchu.edu.tw.

*X*anthomonas campestris pv. campestris is phytopathogenic to cruciferous plants and results in worldwide agricultural losses. It causes black rot, which is one of the most serious plant diseases in Taiwan (1). The warm and wet local climate is suitable for plant infection by *X. campestris* pv. campestris and the development of disease. The cultivation of black rot-resistant cultivars has long been recognized as an important control strategy. Meanwhile, an investigation of the genome sequence from the local strain can be used to discover its genetic and gene expression profiles to prevent black rot disease.

X. campestris pv. campestris 17 was isolated from the leaves of the infected plants in Taiwan. Its genome sequencing was initially carried out using the Sanger sequencing method in 2002. After data assembly, 575 contigs consisting of 5,228,421 bp with approximately 10-fold coverage were obtained. The draft sequences were found to contain either 6,327 or 3,811 genes, as predicted by the Glimmer 2.0 (2) or ORPHEUS (3) program, respectively, using *X. campestris* pv. campestris ATCC 33913 as a reference. To improve the previous sequencing result, the genome of *X. campestris* pv. campestris 17 was resequenced on a GAIIx Illumina platform in 2012. Genome assembly was conducted using the SSAKE software (4) and Velvet assembler (5). However, some gaps were still present, which were further completed by PCR and cloning, followed by Sanger sequencing. A total length of 4,994,426 bp, comprising 24 contigs, with a contig N_{50} of 722,084 bp and 506.7-fold coverage, were obtained. The genome was found to be enriched with G+C content (65.1%). The sequence was annotated using the RAST server (6) and the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genome/annotation_prok), which predicted 4,215 genes, 4,050 coding sequences (CDSs), 1 noncoding RNA (ncRNA) gene, 6 rRNA genes, and 55 tRNA genes. The genome information presented here will help discover the mechanism of *X. campestris* pv. campestris-plant interactions and allow for further investigation of the *X. campestris* pv. campestris functional and structural genomics.

Nucleotide sequence accession numbers. The whole-genome sequence has been deposited in GenBank under the accession number CP011946. The version described in this paper is version CP011946.1.

ACKNOWLEDGMENTS

REFERENCES

1. **Leu YS, Deng WL, Yang WS, Wu YF, Cheng AS, Hsu ST, Tzeng KC.** 2010. Multiplex polymerase chain reaction for simultaneous detection of *Xanthomonas campestris* pv. campestris and *X. campestris* pv. raphani. Plant Pathol Bull **19:**137–147.
2. **Delcher A, Harmon D, Kasif S, White O, Salzberg S.** 1999. Improved microbial gene identification with Glimmer. Nucleic Acids Res **27:** 4636–4641. http://dx.doi.org/10.1093/nar/27.23.4636.
3. **Frishman D, Mironov A, Mewes H-W, Gelfand M.** 1998. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. Nucleic Acids Res **26:**2941–2947. http://dx.doi.org/10.1093/nar/26.12.2941.
4. **Warren RL, Sutton GG, Jones SJM, Holt RA.** 2007. Assembling millions of short DNA sequences using SSAKE. Bioinformatics **23:**500–501. http://dx.doi.org/10.1093/bioinformatics/btl629.
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.