



Complete Genome Sequence of *Curtobacterium* sp. Strain YC1, Isolated from the Surface of *Nostoc flagelliforme* Colonies in Yinchuan, Ningxia, China

Yongjun Wei (魏 勇军)^a, Xiang Gao (高 翔)^b

^aKey Laboratory of Advanced Drug Preparation Technologies, Ministry of Education, School of Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, Henan Province, China

^bSchool of Food and Biological Engineering, Shaanxi University of Science and Technology, Xi'an, Shaanxi Province, China

ABSTRACT The genus *Curtobacterium* belongs to the family *Microbacteriaceae*, within the phylum *Actinobacteria*. This genus includes a wide range of Gram-positive species associated with plants and soils. Here, we report the genome sequence of a new strain, *Curtobacterium* sp. strain YC1, which was isolated from the surface of *Nostoc flagelliforme* colonies. The genome of this strain contains one chromosome and one plasmid, and its size is 3.4 Mb.

Nostoc flagelliforme is a type of terrestrial filamentous cyanobacterium distributed in the arid and semiarid steppes of the western and northwestern parts of China (1). It appears as a hair-like colony on the soil surface. It has been used as a healthy food for more than 2,000 years, and disordered exploitation during the past century has caused a sharp decline in its availability. The State Council of China banned the collection and trade of *N. flagelliforme* in 2000.

A previous study suggested that *Curtobacterium* species could generate oxidative stress to inhibit plant growth (2). We thus attempted to isolate and identify those epiphytic microorganisms on the surface of *N. flagelliforme* colonies collected on the eastern side of Helan Mountain in Yinchuan, Ningxia, China (3). The epiphytic microorganisms were washed from the *N. flagelliforme* colonies with sterilized water. The collected water was plated onto Luria-Bertani (LB) solid medium supplemented with 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal) (used to detect β -galactosidase activity), and one strain that showed β -galactosidase activity was isolated. The strain was identified as a new *Curtobacterium* strain by 16S rRNA gene sequencing and was named *Curtobacterium* sp. strain YC1.

One done of the isolated *Curtobacterium* sp. strain YC1 was cultivated in LB medium and collected for DNA extraction using the DNeasy PowerMax soil kit (Qiagen). The Illumina HiSeq 4000 platform was used to obtain the draft genome assembly of *Curtobacterium* sp. YC1, and the PacBio RS II platform was used to help obtain the full genome assembly. The same DNA was used for Illumina and PacBio sequencing at the Beijing Genomics Institute (BGI) (Shenzhen, China). Paired-end sequencing (2 × 150 bp) was used for Illumina sequencing, and the total read number was 8,353,300. The total Illumina data represented 1,148 Mbp, and the coverage was 339×. For PacBio sequencing, a total of 91,229 reads were obtained, and the read N_{50} was 19,715 bp. The total PacBio data represented 1,280 Mbp, and the coverage was 378×. The program pbdagcon (<https://github.com/PacificBiosciences/pbdagcon>) was used for self-correction of the PacBio sequencing results (4), and draft genome unitigs were assembled with Celera Assembler v8.3 against a high-quality corrected circular consensus sequence subread set (5). GATK v1.6-13 and the SOAP tool package (SOAP2, SOAPsnp, and SOAPindel) were used to improve the accuracy of the

Citation Wei Y, Gao X. 2021. Complete genome sequence of *Curtobacterium* sp. strain YC1, isolated from the surface of *Nostoc flagelliforme* colonies in Yinchuan, Ningxia, China. *Microbiol Resour Announc* 10:e01467-20. <https://doi.org/10.1128/MRA.01467-20>.

Editor David A. Baltrus, University of Arizona

Copyright © 2021 Wei and Gao. 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/).

Address correspondence to Yongjun Wei, yongjunwei@zzu.edu.cn, or Xiang Gao, xianggao@sust.edu.cn.

Received 23 December 2020

Accepted 11 February 2021

Published 11 March 2021

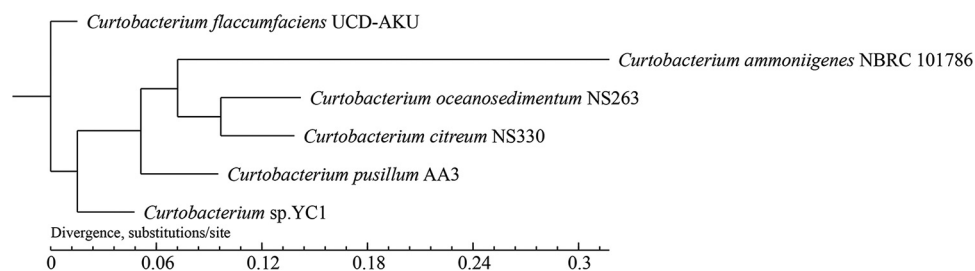


FIG 1 Core/pan-gene-based phylogenetic tree of *Curtobacterium* sp. strain YC1 and its closest relatives. MUMmer v3.22 and BLASTp were used to calculate the synteny of five selected *Curtobacterium* species and *Curtobacterium* sp. strain YC1. The core/pan genes of the six *Curtobacterium* species were clustered by the CD-HIT program for rapid clustering of similar proteins with a threshold of 50% pairwise identity and 0.7 length difference cutoff in amino acids. The gene family was built from the genes of the six *Curtobacterium* species with several software tools. Finally, the phylogenetic tree was constructed by TreeBeST v1.9.2 using the neighbor-joining method based on the gene family clustering with Muscle software v3.8.31. The results suggested that *Curtobacterium* sp. YC1 is a new *Curtobacterium* strain.

genome sequences by making single-base corrections (6, 7). The bacterial plasmid database was used to identify plasmid sequences in the genome using SOAP2 software. Genes in the genome were predicted by Glimmer3 (<http://ccb.jhu.edu/software/glimmer/index.shtml>) with hidden Markov models. The KEGG, Clusters of Orthologous Groups of proteins (COG), nonredundant, Swiss-Prot, and Gene Ontology (GO) databases were used for functional gene annotation (8). Noncoding RNA sequences were recognized with the tools tRNAscan-Se v1.3.1, RNAmmer v1.2, and the Rfam database v9.1 (9–11). Default parameters were used for all software unless otherwise specified.

One circular chromosome assembly and one circular plasmid assembly composed the final assembly of the *Curtobacterium* sp. YC1 genome. The length of the chromosome is 3,301,309 bp, with a GC content of 71.68%. The length of the plasmid is 77,217 bp, with a GC content of 66.99%. There are 3,270 genes in the chromosome and the plasmid, and the average gene length is 938 bp. The length of all the genes represented 90.79% of the whole genome. Among the 3,270 genes, the Swiss-Prot database identified 1,119 genes, the COG database identified 2,315 genes, the nonredundant database identified 3,083 genes, the KEGG database identified 1,683 genes, and the GO database identified 1,918 genes. A total of 84 genes were identified with the carbohydrate-active enzyme (CAZy) database, and 52 glycoside hydrolase genes were annotated (12). There are 46 copies of tRNAs, 3 copies of 5S rRNAs, 3 copies of 16S rRNAs, 3 copies of 23S rRNAs, and 2 copies of small RNAs. Moreover, 6 clustered regularly interspaced short palindromic repeat (CRISPR) regions were predicted with CRISPRFinder (<https://crispr.i2bc.paris-saclay.fr/>). The phylogenetic analyses suggested that *Curtobacterium* sp. YC1 is a new *Curtobacterium* strain (Fig. 1). The public version of this genome was annotated using PGAP through NCBI (13).

Data availability. The complete genome sequences (one chromosome and one plasmid) of *Curtobacterium* sp. YC1 were deposited in GenBank under the accession numbers CP066341 and CP066342, with BioProject accession number PRJNA683987, BioSample accession number SAMN17050104, and SRA accession numbers SRR13298322 and SRR13638075.

ACKNOWLEDGMENT

This work has received partial support from the Key Project of Natural Science of Shaanxi Province, China (grant number 2020JZ-51).

REFERENCES

- Gao K. 1998. Chinese studies on the edible blue-green alga, *Nostoc flagelliforme*: a review. *J Appl Phycol* 10:37–49. <https://doi.org/10.1023/A:1008014424247>.
- Pugazhendhi A, Radhakrishnan R, Duc PA. 2019. *Curtobacterium* sp. MA01 generates oxidative stress to inhibit the plant growth. *Biocatal Agric Biotechnol* 20:101274. <https://doi.org/10.1016/j.bcab.2019.101274>.

3. Gao X, Xu H, Ye S, Liang W. 2016. A proposal on the restoration of *Nostoc flagelliforme* for sustainable improvement in the ecology of arid steppes in China. *Environments* 3:14. <https://doi.org/10.3390/environments3020014>.
4. Mosher JJ, Bowman B, Bernberg EL, Shevchenko O, Kan J, Korlach J, Kaplan LA. 2014. Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. *J Microbiol Methods* 104:59–60. <https://doi.org/10.1016/j.mimet.2014.06.012>.
5. Denisov G, Walenz B, Halpern AL, Miller J, Axelrod N, Levy S, Sutton G. 2008. Consensus generation and variant detection by Celera Assembler. *Bioinformatics* 24:1035–1040. <https://doi.org/10.1093/bioinformatics/btn074>.
6. O'Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, Bodily P, Tian L, Hakonarson H, Johnson WE, Wei Z, Wang K, Lyon GJ. 2013. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Med* 5:28. <https://doi.org/10.1186/gm432>.
7. Hurgobin B. 2016. Short read alignment using SOAP2. *Methods Mol Biol* 1374:241–252. https://doi.org/10.1007/978-1-4939-3167-5_13.
8. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
9. Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol* 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.
10. 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. <https://doi.org/10.1093/nar/gkm160>.
11. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res* 31:439–441. <https://doi.org/10.1093/nar/gkg006>.
12. Liang J, Mai W, Wang J, Li X, Su M, Du J, Wu Y, Dai J, Tang Q, Gao J, Liu Y, Tang J, Wei Y. 2021. Performance and microbial communities of a novel integrated industrial-scale pulp and paper wastewater treatment plant. *J Clean Prod* 278:123896. <https://doi.org/10.1016/j.jclepro.2020.123896>.
13. 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>.