Complete Genome Sequence of *Alicyclobacillus acidocaldarius* Strain Tc-4-1

Yaping Chen,^{1,2†} Yongzhi He,^{3†} Bing Zhang,¹ Jin Yang,¹ Wenjie Li,¹
Zhiyang Dong,^{3*} and Songnian Hu^{1*}

Key Laboratory of Genome Science and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China¹;
Graduate University of the Chinese Academy of Sciences, Beijing, China²; and State Key Laboratory of Microbial Resources,
Institute of Microbiology, Chinese Academy of Sciences, Beijing, China³

Received 30 June 2011/Accepted 8 July 2011

***Alicyclobacillus acidocaldarius* strain Tc-4-1 was initially isolated from a hot spring in Tengchong, China. This organism is both thermophilic and acidophilic. It can produce heat- and acid-stable enzymes, such as amylase and esterase, which may be important in industry. Here we report the whole genome sequence of the strain.**

The thermoacidophilic Gram-positive bacterium Tc-4-1 was isolated from a hot spring in Tengchong, Yunnan, China. It grows optimally under strictly aerobic conditions at 70°C and pH 3. Phylogenetic analyses showed that strain Tc-4-1 belongs to the genus *Alicyclobacillus*, with the sequence of the 16S rRNA gene exhibiting the highest (99%) identity to that of *Alicyclobacillus acidocaldarius* DSM 452 (GenBank accession number AB059673.1). It has been reported that *A. acidocaldarius* can use several sugars as carbon energy source (2). Many thermoacidophilic glycoside hydrolase genes have been cloned from this organism, such as α -amylase, endoglucanase, β -glucosidase, xylanase, and mannanase. The strain Tc-4-1 can use soluble starch as a carbon energy source. Accordingly, a thermoacidophilic α -amylase with an optimal temperature and pH of 75°C and pH 3 has been cloned from Tc-4-1.

Using the Genome Sequencer FLX system (454), we obtained the genome sequence of Tc-4-1 based on a total of 213,270 reads, which were assembled into 248 large contigs (>500 bp) with the Newbler software. Closure of the gaps between contigs was finished by PCR primer walking using the capillary electrophoresis sequencing method and edited using the Phred/Phrap/Consed package (3–6). Additionally, we also sequenced the strain Tc-4-1 genome with the SOLiD genetic analysis system (Applied Biosystems) to correct potential base errors and improve consensus quality by mapping the SOLiD reads to the genome. The protein-coding genes were predicted using Glimmer 3.0 (1) and annotated with the BLAST tool (NCBI), and the putative genes were annotated using the Clusters of Orthologous Genes (COG) program (15), KEGG (10–12), and InterProScan (16).

The genome contains a circular chromosome of 3,124,048 bp with a G+C content of 61.52%. A total of 3,216 open reading

frames (ORFs), 6 rRNA operons, and 63 tRNA genes were predicted by the Glimmer 3.0 (1), RNAmmer (13), and tRNAscan-SE (14) programs, respectively. The average size of ORFs is 841 bp, comprising ORFs ranging from 37 to 2,731 amino acids. The coding density of the genome is 89.1%. Among all the ORFs, 2,196 (68.3%) are classified according to homology to other documented proteins, and the rest, 1,020 (31.7%), are functionally unknown. The genome encodes 44 ABC transporters, 3 CRISPRs (clustered regularly interspaced short palindromic repeats), and 2 possible CRISPRs (7–9). The strain Tc-4-1 genome has an approximately complete set of genes for bacterial starch and sucrose metabolism pathways. We have also found xylanase, mannanase, β -glucosidase, β -galactosidase, and other glycoside hydrolases in the Tc-4-1 genome which enable its growth on different di- and polysaccharides. Here, β -glucosidase hydrolyzes cellobiose to β -D-glucose. β -Galactosidase catalyzes the hydrolysis of terminal, nonreducing β -D-galactosyl residues from β -D-galactosides. A detailed analysis of the *A. acidocaldarius* Tc-4-1 genome and comparative genome analysis will be performed and will provide insights into the unique metabolism characteristics of this strain.

Nucleotide sequence accession number. The complete genome sequence of *Alicyclobacillus acidocaldarius* strain Tc-4-1 has been deposited in GenBank under accession number CP002902.

The research was supported by the Key Laboratory of Genome Science and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China.

REFERENCES

1. Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**:4636–4641.
2. Di Lauro, B., M. Rossi, and M. Moracci. 2006. Characterization of a beta-glucosidase from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius*. *Extremophiles* **10**:301–310.
3. Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**:186–194.
4. Ewing, B., L. Hillier, M. C. Wendt, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**:175–185.
5. Fouts, D. E. 2006. Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res.* **34**:5839–5851.

* Corresponding author. Mailing address for Songnian Hu: Beijing Institute of Genomics, Chinese Academy of Sciences, No. 7 Beitucheng West Road, Chaoyang District, Beijing 100029, China. Phone: 86-10-82995362. Fax: 86-10-82995401. E-mail: husn@big.ac.cn. Mailing address for Zhiyang Dong: State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, No. 1 West Beichen Road, Chaoyang District, Beijing 100101, China. E-mail: dongzy@sun.im.ac.cn.

† Y. Chen and Y. He contributed equally to this work.

6. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**:195–202.
7. **Grissa, I., G. Vergnaud, and C. Pourcel.** 2008. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **36**:W145–W148.
8. **Grissa, I., G. Vergnaud, and C. Pourcel.** 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**:172.
9. **Grissa, I., G. Vergnaud, and C. Pourcel.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35**:W52–W57.
10. **Kanehisa, M., and S. Goto.** 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**:27–30.
11. **Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa.** 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**:D355–D360.
12. **Kanehisa, M., et al.** 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **34**:D354–D357.
13. **Lagesen, K., et al.** 2007. RNAmmer: consistent and rapid annotation of rRNA genes. *Nucleic Acids Res.* **35**:3100–3108.
14. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of tRNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
15. **Tatusov, R. L., et al.** 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**:41.
16. **Zdobnov, E. M., and R. Apweiler.** 2001. InterProScan: an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**: 847–848.