

# Genome Sequence of *Thermus thermophilus* ATCC 33923, a Thermostable Trehalose-Producing Strain

Ling Jiang,<sup>a,b</sup> Ming Lin,<sup>b</sup> Xiaotong Li,<sup>b</sup> Huaiyan Cui,<sup>b</sup> Xian Xu,<sup>b</sup> Shuang Li,<sup>b</sup> He Huang<sup>b</sup>

College of Food Science and Light Industry, Nanjing University of Technology, Nanjing, China<sup>a</sup>; College of Biotechnology and Pharmaceutical Engineering, State Key Laboratory of Materials-Oriented Chemical Engineering, Nanjing University of Technology, Nanjing, China<sup>b</sup>

***Thermus thermophilus* ATCC 33923 contains a thermostable enzyme that can efficiently catalyze the conversion of maltose into trehalose. Here we report a 2.15-Mb assembly of its genome sequence and other useful information, including the coding sequences (CDS) responsible for biological processes such as DNA replication, DNA repair, and RNA maturation.**

Received 17 June 2013 Accepted 24 June 2013 Published 25 July 2013

**Citation** Jiang L, Lin M, Li X, Cui H, Xu X, Li S, Huang H. 2013. Genome sequence of *Thermus thermophilus* ATCC 33923, a thermostable trehalose-producing strain. *Genome Announc.* 1(4):e00493-13. doi:10.1128/genomeA.00493-13.

**Copyright** © 2013 Jiang et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to He Huang, [biotech@njut.edu.cn](mailto:biotech@njut.edu.cn).

Trehalose ( $\alpha$ -D-glucopyranosyl-1,1- $\alpha$ -D-glucopyranoside) is a naturally occurring nonreducing disaccharide that is widely spread throughout mycobacteria, bacteria, archaea, yeast, fungi, and higher orders of the plant and insect kingdom (1). This simple disaccharide can serve as an energy reservoir and a protectant from various physical and chemical stresses, as well as an additive to food, cosmetics, and medicinal and biological reagents (2). The wide range of applications of trehalose has increased the interest of many researchers into the development of economically feasible production systems. Trehalose synthase (TSase) (EC 5.4.99.16) employs an inexpensive substrate, maltose, and allows one-step formation of trehalose by an intramolecular rearrangement of the  $\alpha$ -1,4-linkage of maltose to the  $\alpha$ -1,1-linkage of trehalose. This pathway is thought to be the most promising process for industrial manufacture of trehalose (3).

Up to now, TSases were found mainly in bacteria (4), and only TSase from the *Thermus* strain was characterized as a thermostable enzyme (5). The intrinsic stability of this thermostable enzyme and its resistance to denaturing physical and chemical factors are considerable advantages in industrial processes. The trehalose produced by *Thermus thermophilus* strain ATCC 33923 showed its optimal activity levels at 70°C, and the maximum yield was up to 72% for 24 h of the reaction, which is higher than most of the previously published results (6, 7). Therefore, investigation of the genetic information and characteristics of *T. thermophilus* ATCC 33923 is desired to further elucidate this mechanism. Knowledge of the genome sequence and bioinformatics will be of great help in this regard. In addition, genome-scale analysis has proven useful for metabolic engineering applications (8).

Here we present the draft genome sequence of strain *T. thermophilus* ATCC 33923, obtained using an Illumina HiSeq 2000 system, which was performed by Shanghai Majorbio Bio Pharm Technology Co., Ltd. The reads were assembled with Velvet (9), and the sequence was annotated using the RAST annotation server (10). A library containing 300-bp inserts was constructed. Sequencing was performed based on the paired-end strategy of 101-bp reads to produce 541 Mb of filtered sequences, represent-

ing a 252.03-fold coverage of the genome. The sequence of *T. thermophilus* ATCC 33923 is 2,147,217 bases with a G+C content of 69.41%, which was assembled into 117 contigs and 94 scaffolds. It contains 2,270 open reading frames (ORFs), 46 tRNA genes, and 2 rRNA genes identified by Glimmer 3.02 (11), Genemark (12), tRNAscan-SE (13), and RNAmmer (14).

According to the genomic analysis, strain *T. thermophilus* ATCC 33923 may have a powerful resistance to physical and chemical agents, as there were 4 ORFs relate to DNA repair and 13 ORFs related to the thermostable protein. Additionally, the vitamin B<sub>12</sub> and carotenoid biosynthesis enzymes were also annotated in the *T. thermophilus* ATCC 33923 genome sequence. Further studies will be performed to confirm their functions, and a complete genome sequence will be included in the future to reveal the unique molecular characteristics of strain *T. thermophilus* ATCC 33923.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession number [AQOS000000000](https://www.ncbi.nlm.nih.gov/nuccore/AQOS000000000). The version described in this paper is the first version, with accession number [AQOS010000000](https://www.ncbi.nlm.nih.gov/nuccore/AQOS010000000).

## ACKNOWLEDGMENTS

This work was supported by the National Basic Research Program of China (number 2012CB725204), the National Natural Science Foundation of China for Young Scholars (numbers 21225626, 21106064, and 20936002), and the National High Technology Research and Development Program of China (numbers 2012AA021700 and 2012AA022101).

## REFERENCES

1. Elbein AD, Pan YT, Pastuszak I, Carroll D. 2003. New insights on trehalose: a multifunctional molecule. *Glycobiology* 13:17–27.
2. Schiraldi C, Di Lernia I, De Rosa M. 2002. Trehalose production: exploiting novel approaches. *Trends Biotechnol.* 20:420–425.
3. Chen YS, Lee GC, Shaw JF. 2006. Gene cloning, expression, and biochemical characterization of a recombinant trehalose synthase from *Picrophilus torridus* in *Escherichia coli*. *J. Agric. Food Chem.* 54:7098–8104.
4. Avonce N, Mendoza-Vargas A, Morett E, Iturriaga G. 2006. Insights on the evolution of trehalose biosynthesis. *BMC Evol. Biol.* 6:109–123.

5. Wang JH, Tsai MY, Lee GC, Shaw JF. 2007. Construction of a recombinant thermostable  $\beta$ -amylase-trehalose synthase bifunctional enzyme for facilitating the conversion of starch to trehalose. *J. Agric. Food Chem.* 55:1256–1263.
6. Chen YS, Lee GC, Shaw JF. 2006. Gene cloning, expression, and biochemical characterization of a recombinant trehalose synthase from *Picrophilus torridus* in *Escherichia coli*. *J. Agric. Food Chem.* 54:7098–7104.
7. Zhu YM, Wei DS, Zhang J, Wang YF, Xu HY, Xing LJ, Li MC. 2010. Overexpression and characterization of a thermostable trehalose synthase from *Meiothermus ruber*. *Extremophiles* 14:1–8.
8. Nölling J, Breton G, Omelchenko MV, Makarova KS, Zeng QD, Gibson R, Lee HM, Dubois J, Qiu DY, Hitti J, Wolf YI, Tatusov RL, Sabathe F, Doucette-Stamm L, Soucaille P, Daly MJ, Bennett GN, Koonin EV, Smith DR. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* 183:4823–4838.
9. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
10. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formisano 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://www.biomedcentral.com/1471-2164/9/75>.
11. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* 23:673–679.
12. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29:2607–2618.
13. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
14. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108.