Genome Sequence of *Escherichia coli* XH140A, Which Produces L-Threonine

Huilin Yang,¹ Yuling Liao,^{1,2} Bin Wang,¹ Ying Lin,¹ and Li Pan^{1*}

School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China,¹ and
Star Lake Bioscience Co., Inc., Zhaoqing, Guangdong, China²

Received 15 August 2011/Accepted 17 August 2011

Here we report the draft annotated genome sequence of *Escherichia coli* XH140A, which is used to produce L-threonine in industry. The genome sequence will allow the characterization of the molecular mechanisms underlying its beneficial properties.

L-Threonine has been widely used as a supplement in the food, pharmaceutical, and cosmetics industries. Currently L-threonine is produced mainly by microbial fermentation. To date, *E. coli* mutant strains remain the dominant industrial producers of L-threonine (6). Although there are many papers (8, 10, 14) and patents (4, 12) reporting L-threonine-producing strains of *E. coli*, not much is known about the genome and plasmid sequences of these producer strains. Here, we sequenced the genome of *Escherichia coli* XH140A containing a plasmid, a high-yield L-threonine producer in industry.

The genome was sequenced using the Illumina Solexa GA IIx instrument at Beijing Genomics Institute (BGI; Shenzhen, China). A library containing 500-bp inserts was constructed. Sequencing was performed with the paired-end strategy of 75-bp reads to produce 634 Mb of filtered sequences, representing a 136.6-fold coverage of the genome. The sequences were assembled into 184 contigs and 95 scaffolds using the SOAPdenovo package (11). The remaining intrascaffold gaps were closed by Sanger sequencing of PCR products. The complete nucleotide sequences of the plasmid pTHR were determined by Sanger sequencing using the primer walking strategy.

Genome annotation was performed at the NCBI Prokaryotic Genomes Automatic Annotation Pipeline. Open reading frames were identified by Glimmer 3.02 (5) and Genemark (2). The resulting translations were used for a BLASTP (1) search against the GenBank NR database as well as the KEGG (7) and COG (15) databases. tRNA and rRNA genes were identified by tRNAscan-SE (13) and RNAmmer (9), respectively.

The XH140A chromosome is 4,395,442 bp in length, with an average G+C content of 50.80%. A total of 4,159 protein-coding genes, 72 tRNA genes, and 3 rRNA fragments were identified. Eighteen pseudogene were also identified, which suggested that *Escherichia coli* XH140A was selected by genetic engineering treatment. The plasmid pTHR of XH140A is 15,774 bp in length with an average G+C content of 49.07%, which contains 13 protein-coding genes.

In a comparative analysis of *Escherichia coli* XH140A with the reference genome *Escherichia coli* K-12 (3), we found an

important mutation that may explain the molecular mechanisms for the production of L-threonine. A nonsense mutation in the acetolactate synthase 2 catalytic subunit gene *ihvG* blocked the pathway of L-threonine to L-isoleucine, which could make XH140A accumulate L-threonine. Interestingly, the plasmid pTHR did not have the threonine operon *thrABC*, which was different from what has been reported in several public papers and patents about L-threonine-producing strains (4, 6, 12). It is worth noting that the plasmid pTHR contained NAD/NADP transhydrogenase alpha-subunit and beta-subunit genes, which suggested that the balance of NADH and NADPH is very important in L-threonine high-yield strains. Besides these two genes, there were another four genes, encoding aminoglycoside phosphotransferase, phosphoenolpyruvate carboxylase, aspartate ammonia-lyase, and aspartate-semialdehyde dehydrogenase, which may be involved in L-threonine production. Due to this new finding, the genome sequence of *Escherichia coli* XH140A will open new perspectives for breeding other amino acid-producing strains.

Nucleotide sequence accession number. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number AFVX00000000. The version described in this paper is the first version, AFVX01000000.

This research was funded by Guangdong Technological Foundation of China (2010A010500003).

REFERENCES

1. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
2. Besemer, J., and A. Lomsadze, M. Borodovsky. 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.
3. Blattner, F. R., et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–1462.
4. Cho, J. Y., B. C. Lee, D. C. Kim, J. H. Lee, and Y. H. Park. March 2005. Method for producing L-threonine. U.S. patent 20050054076A1.
5. Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**:673–679.
6. Dong, X. Y., P. J. Quinn, and X. Y. Wang. 2011. Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for the production of L-threonine. *Biotechnol. Adv.* **29**:11–23.
7. Kanehisa, M., et al. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**(database issue):D480–D484.
8. Kruse, D., et al. 2002. Influence of threonine exporters on threonine production in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **59**:205–210.

* Corresponding author. Mailing address: School of Bioscience and Bioengineering, South China University of Technology, Guangzhou Higher Education Mega Center, Guangzhou, Guangdong, China, 510006. Phone: 86 20 39380616. Fax: 86 20 39380698. E-mail: btlip@scut.edu.cn.

9. **Lagesen, K., et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**:3100–3108.
10. **Lee, M. H., et al.** 2006. Improved L-threonine production of *Escherichia coli* mutant by optimization of culture conditions. *J. Biosci. Bioeng.* **101**:127–130.
11. **Li, R., et al.** 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**:265–272.
12. **Liaw, H. J., J. S. Bradshaw, Y. Q. Yang, and W. Y. Mao.** March 2010. *Escherichia coli* strains which over-produce L-threonine and processes for the production of L-threonine by fermentation. U.S. patent 7687253B2.
13. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
14. **Ogawa-Miyata, Y., H. Kojima, and K. Sano.** 2001. Mutation analysis of the feedback inhibition site of aspartokinase III of *Escherichia coli* K-12 and its use in L-threonine production. *Biosci. Biotechnol. Biochem.* **65**:1149–1154.
15. **Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**:33–36.