

Draft Genome Sequence of *Enterococcus faecium* Strain LCT-EF90

De Chang,^a Yuanfang Zhu,^b Yuanqiang Zou,^b Xiangqun Fang,^a Tianzhi Li,^a Junfeng Wang,^a Yinghua Guo,^a Longxiang Su,^a Jingjing Xia,^a Ruifu Yang,^{b,c} Chengxiang Fang,^b and Changting Liu^a

Nanlou Respiratory Diseases Department, Chinese PLA General Hospital, Beijing, China^a; BGI-Shenzhen, Shenzhen, People's Republic of China^b; and State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China^c

***Enterococcus faecium* is an opportunistic human pathogen, found widely in the human gastrointestinal tract, and can also be isolated from a variety of plants, animals, insects, and other environmental sources. Here, we present the fine draft genome sequence of *E. faecium* LCT-EF90.**

Enterococci are common inhabitants of the human gastrointestinal (GI) tract (4, 9) and can also be cultivated from a variety of plants, animals, insects, and other environmental sources. For a long time, the species *E. faecium* was considered a harmless commensal of the mammalian GI tract and was used as a probiotic (7, 12) added to fermented foods (5); however, some strains have recently been recognized as pathogens (8, 9). *E. faecium* is a Gram-positive bacterium belonging to the family *Enterococcaceae* (10). Strain LCT-EF90 originated from an *E. faecium* strain (CGMCC 1.2136) that was cultured at different temperature (15°C versus 37°C) for more than 4 weeks. Cells occur singly, in pairs, or in chains. This strain has both aerobic and anaerobic cellular respiration pathways.

The genome of *E. faecium* was sequenced with an Illumina HiSeq 2000 instrument according to the manufacturer's instructions. High-molecular-mass genomic DNA from *E. faecium* was used to construct small (500-bp) and large (6-kb) random sequencing libraries. The mean read length is 90 bp for both the 500-bp and the 6,000-bp library. The reads were filtered and assembled into contigs using SOAPdenovo v1.05 (<http://soap.genomics.org.cn/>). Finally, 31 scaffolds consisting of 118 contigs were constructed step by step using all the paired-end information of reads with 120× and 70× genome coverage. The scaffold N_{50} and N_{90} were determined to be 1,498 kb and 108.5 kb, and the longest scaffold was 1,498 kb. The total length of the assembly was 2,773,995 bp, and the average GC content was about 38.24%.

Putative protein-coding sequences were predicted using the Glimmer 3.0 program (3). Overall, there were 2,777 predicted protein-coding sequences (CDSs) with an average gene length of 862 bp. To further verify these gene predictions, all gene functions were determined mainly by BLASTP analysis of sequences in the KEGG (6), COG, Swiss-Prot (12), TrEMBL (1), GO, and NR databases and by manual curation of the outputs of a variety of similarity searches. The results of analysis of COG database sequences showed that there were more genes clustered in the categories "Carbohydrate Transport" and "Metabolism" than in other function clusters. GO annotation analyses of the *E. faecium* genome revealed 20 categories, mainly containing genes for cellular components, binding, transporter activity, and catalytic activity, as well as genes for molecular functions and cellular and physiological processes.

We predicted the transposon sequences using RepeatMasker software (11) and RepeatProteinMasker software and tandem repeat sequences using TRF (Tandem Repeat Finder) (2). We identified different transposable element (TE)-related sequences, with

17 kb in total length, which occupy 0.62% of the assembly. In addition to protein-coding genes, noncoding RNA (ncRNA) sequences were also predicted, including small RNA (sRNA), rRNA, tRNA, snRNA, and micro-RNA. Genome island prediction was performed using IslandPath-DIOMB, SIGI-HMM, IslandPicker, and IslandViewer software. IslandPath-DIOMB and SIGI-HMM are prediction programs based on sequence comparison; IslandPicker is based on genome comparison, and IslandViewer is the combination of the three preceding software programs. In addition, the prophage sequences predicted by Prophinder software and the ACLAME database and CRISPRs predicted from CRISPRFinder software were carried on. Genome island sequences were also obtained, but no prophage sequences or prophage sequences were found.

Nucleotide sequence accession number. This whole-genome sequence has been deposited at DDBJ/EMBL/GenBank under accession number [AJKH00000000](https://www.ncbi.nlm.nih.gov/nuccore/AJKH00000000). The versions described in this paper are the first versions.

ACKNOWLEDGMENTS

This work was supported by the Key Pre-Research Foundation of Military Equipment of China (grant 9140A26040312JB1001), the opening foundation of the State Key Laboratory of Space Medicine Fundamentals and Application, Chinese Astronaut Research and Training Center (grant SMFA11K02), a Special Financial Grant from the China Postdoctoral Science Foundation (grant 201104776), and the National Natural Science Foundation of China (grant 81000018).

REFERENCES

1. Bairoch A, Apweiler R. 1996. The SWISS-PROT protein sequence data bank and its new supplement TrEMBL. *Nucl. Acids Res.* 24:21–25.
2. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580.
3. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
4. Franz CM, Holzapfel WH, Stiles ME. 1999. Enterococci at the crossroads of food safety? *Int. J. Food Microbiol.* 47:1–24.

Received 2 April 2012 Accepted 11 April 2012

Address correspondence to Chengxiang Fang, cxfang@whu.edu.cn, or Changting Liu, liuchangt@gmail.com.

D.C. and Y.Z. contributed equally to this work.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00529-12

5. Franz CM, Stiles ME, Schleifer KH, Holzapfel WH. 2003. Enterococci in foods—a conundrum for food safety. *Int. J. Food Microbiol.* **88**:105–122.
6. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**:D355–360.
7. Lund B, Edlund C. 2001. Probiotic *Enterococcus faecium* strain is a possible recipient of the *vanA* gene cluster. *Clin. Infect. Dis.* **32**:1384–1385.
8. Moellering RC, Jr. 1992. Emergence of *Enterococcus* as a significant pathogen. *Clin. Infect. Dis.* **14**:1173–1176.
9. Murray BE. 1990. The life and times of the enterococcus. *Clin. Microbiol. Rev.* **3**:46–65.
10. Schleifer KR, Kilpper BR. 1984. Transfer of *Streptococcus faecalis* and *Streptococcus faecium* to the genus *Enterococcus* nom. rev. as *Enterococcus faecalis* comb. nov. and *Enterococcus faecium* comb. nov. *Int. J. Syst. Bacteriol.* **34**:31–34.
11. Smit AF. 1996. The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6**:743–748.
12. Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* **278**:631–637.