JB Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequence of *Escherichia coli* LCT-EC106

Tianzhi Li,[a] Fei Pu,[b] Rentao Yang,[b] Xiangqun Fang,[a] Junfeng Wang,[a] Yinghua Guo,[a] De Chang,[a] Longxiang Su,[a] Na Guo,[a] Xuege Jiang,[a] Jiao Zhao,[b] and Changting Liu[a]

Nanlou Respiratory Diseases Department, Chinese PLA General Hospital, Beijing, China,[a] and BGI-Shenzhen, Shenzhen, People's Republic of China[b]

***Escherichia coli* is a Gram-negative, rod-shaped bacterium that is commonly found in the intestine of warm-blooded organisms. Most *E. coli* strains are harmless, but some serotypes can cause serious food poisoning in humans. Here, we present the complete genome sequence of *Escherichia coli* LCT-EC106, which was isolated from CGMCC 1.2385.**

Escherichia coli is a Gram-negative, rod-shaped bacterium that is commonly found in the lower intestine of warm-blooded organisms. Most *E. coli* strains are harmless, but some serotypes can cause serious food poisoning in humans, and they are occasionally responsible for product recalls (15). The harmless strains are part of the normal flora of the gut, and they can benefit their hosts by producing vitamin K2 (2) and by preventing the establishment of pathogenic bacteria within the intestine (12). *E. coli* and related bacteria constitute about 0.1% of gut flora (4), and fecal-oral transmission is the major route through which pathogenic strains of the bacterium cause disease. *E. coli* is the most widely studied prokaryotic model organism and an important species in the fields of biotechnology and microbiology, where it has served as the host organism for the majority of work with recombinant DNA. The whole-genome sequence was described as *Escherichia coli* LCT-EC106, which was obtained from the China General Microbiological Culture Collection Center (CGMCC) as CGMCC 1.2385.

The whole-genome sequence was obtained by Illumina HiSeq 2000 at BGI-Shenzhen, China. Two paired-end libraries were generated with average insert sizes of 500 bp and 6 kb following the manufacturer's instructions. The whole genome was sequenced to average depths of 100× and 50× coverage from the two libraries, respectively, with a read length of 90 bp. The high-quality data from a small-insert-size library were first assembled into 227 contigs using SOAPdenovo v1.06 according to the method described previously (13). Then, these contigs were connected into 38 scaffolds based on paired-end relationships of data from a 6-kb-insert-size library. Putative open reading frames were predicted using Glimmer v3.0 (3). The predicted coding sequences (CDSs) were searched against NR, COG, and KEGG databases. The rRNA and tRNA were identified using RNAmmer (6) and tRNAscan-SE 1.21 (9); the insertion sequence (IS) elements were annotated by ISsaga (14). Virulence genes were determined by aligning to VFDB (1), and antibiotic resistance genes were detected by BLAST to the Antibiotic Resistance Genes Database (7).

The draft genome of *Escherichia coli* LCT-EC106 consists of 5,212,598 bp with a G+C content of 50.38%. A total of 5,075 CDSs were predicted, with an average length of 898 bp. Virulence-associated factors, including hemolysin operon (HlyCABD), Pap operon, IutA, IucDCBA, F1C operon, antigen 43, KpsTM, ClpB, ShlA/HecA/FhaA exofamily protein, yersiniabactin operon (YbtAUTE), and virulence factor MviMN, were discovered in 11 virulence regions based on searching against VFDB. These virulence regions are highly homologous with eight pathogenicity islands (PAIs) of *Escherichia coli* CFT073 (PAI-CFT073-aspV, PAI-CFT073-serX, PAI-CFT073-icdA, PAI-CFT073-asnT, GI-CFT073-asnW, GI-CFT073-cobU, PAI-CFT073-metV, PAI-CFT073-pheV) (5, 8, 10, 11). Multidrug resistance capability was indicated from results based on searching against ARDB (7). A total of 21 antibiotic resistance genes were detected, relating to resistance to β-lactam antibiotics, macrolide antibiotics, aminoglycoside antibiotics, acriflavine antibiotics, and polymyxin.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number AJRI00000000.

ACKNOWLEDGMENTS

This work was supported by the Key Pre-Research Foundation of Military Equipment of China (grant no. 9140A26040312JB1001), the opening foundation of the State Key Laboratory of Space Medicine Fundamentals and Application, Chinese Astronaut Research and Training Center (no. SMFA11K02), the Special Financial Grant from the China Postdoctoral Science Foundation (no. 201104776), and the National Natural Science Foundation of China (no. 81000018).

REFERENCES

1. **Chen L, et al.** 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. **33**:D325–D328.
2. **Conly JM, Stein K.** 1992. The production of menaquinones (vitamin K2) by intestinal bacteria and their role in maintaining coagulation homeostasis. Prog. Food Nutr. Sci. **16**:307–343.
3. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679.
4. **Eckburg PB, et al.** 2005. Diversity of the human intestinal microbial flora. Science **308**:1635–1638.
5. **Guyer DM, Kao JS, Mobley HL.** 1998. Genomic analysis of a pathogenicity island in uropathogenic Escherichia coli CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter-associated bacteriuria and from fecal samples. Infect. Immun. **66**:4411–4417.
6. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
7. **Liu B, Pop M.** 2009. ARDB—Antibiotic Resistance Genes Database. Nucleic Acids Res. **37**:D443–D447.
8. **Lloyd AL, Rasko DA, Mobley HL.** 2007. Defining genomic islands and uropathogen-specific genes in uropathogenic Escherichia coli. J. Bacteriol. **189**:3532–3546.

Received 16 May 2012 Accepted 31 May 2012

Address correspondence to Jiao Zhao, zhaojiao@genomics.cn, or Changting Liu, liuchangt@gmail.com.

T.L. and F.P. contributed equally to this work.

doi:10.1128/JB.00853-12

9. **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.
10. **Oelschlaeger TA, Dobrindt U, Hacker J.** 2002. Pathogenicity islands of uropathogenic E. coli and the evolution of virulence. Int. J. Antimicrob. Agents **19**:517–521.
11. **Rasko DA, Phillips JA, Li X, Mobley HL.** 2001. Identification of DNA sequences from a second pathogenicity island of uropathogenic Escherichia coli CFT073: probes specific for uropathogenic populations. J. Infect. Dis. **184**:1041–1049.
12. **Reid G, Howard J, Gan BS.** 2001. Can bacterial interference prevent infection? Trends Microbiol. **9**:424–428.
13. **Rohde H, et al.** 2011. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. N. Engl. J. Med. **365**:718–724.
14. **Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M.** 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. **34**:D32–D36.
15. **Tarr PI, Gordon CA, Chandler WL.** 2005. Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. Lancet **365**:1073–1086.