Journal of Bacteriology

Genome Sequence of Escherichia coli J53, a Reference Strain for Genetic Studies

Hana Yi, Yong-Joon Cho, Dongeun Yong and Jongsik Chun
J. Bacteriol. **2012, 194(14):3742. DOI: 10.1128/JB.00641-12.**

JB Journals.ASM.org

GENOME ANNOUNCEMENT

Genome Sequence of *Escherichia coli* J53, a Reference Strain for Genetic Studies

Hana Yi,[a] Yong-Joon Cho,[b] Dongeun Yong,[c] and Jongsik Chun[a,b,d]

Institute of Molecular Biology and Genetics, Seoul National University, Seoul, Republic of Korea[a]; Chunlab, Inc., Seoul National University, Seoul, Republic of Korea[b]; Department of Laboratory Medicine and Research Institute of Bacterial Resistance, Yonsei University College of Medicine, Seoul, Republic of Korea[c]; and Advanced Institute of Convergence Technology, Gyeonggi-do, Republic of Korea[d]

***Escherichia coli* J53 (F$^-$ *met pro* Azi^r) is a derivative of *E. coli* K-12 which is resistant to sodium azide. This strain has been widely used as a general recipient strain for various conjugation experiments. Here, we report the genome sequence of *E. coli* J53 (=KACC 16628).**

Since the isolation of the original *Escherichia coli* K-12 strain from a stool sample of a diphtheria patient in 1922 (3), a variety of mutant derivatives of K-12 have been generated for laboratory usage. The auxotrophic mutant strain J5-3 (F$^+$ *met pro*) was derived from K-12 (5), then modified to J53 (F$^-$ *met pro*) (4), and finally developed into J53 (F$^-$ *met pro* Azi^r) by spontaneous mutations (12). This sodium azide-resistant strain, which has been used for over 10 years as a recipient for conjugation experiments in the Department of Laboratory Medicine and Research Institute of Bacterial Resistance, College of Medicine, Yonsei University, was subjected to a genome analysis here.

The genome sequence was determined using a combination of shotgun sequencing from a paired end (read length, 100 bp) in an Illumina Genome Analyzer IIx (4,330,532 reads; 940× coverage) and that from a paired end (insert size, 8 kb) in a Roche Genome Sequencer FLX Titanium system (223,880 reads). The sequencing reads were assembled using the CLC Genomics wb4 (CLCbio), Newbler assembler 2.3 (Roche), and CodonCode Aligner (CodonCode Co.) programs and resulted in 42 contigs. An optical map of the chromosome digested by AflII was produced by OpGen and analyzed using the MapSolver package (OpGen), and the chromosome structure was compared with that of *E. coli* MG1655. The assembled contig sequences were uploaded into the RAST server (2) to predict the open reading frames (ORFs) by using Glimmer 3 (7). The predicted ORFs were annotated by searching against clusters of orthologous group (16) and SEED (8) databases. The genome was 4.5 Mb long, and the G+C content was 50.8 mol%. The genome contained 4,484 predicted protein-coding sequences, one integrated lambda phage across the contigs numbered 25 to 27, 6 copies of rRNA operons, and 67 tRNA genes.

A phylogenetic tree based on average nucleotide index (ANI) values (11) among *E. coli-Shigella* group strains showed that the test strain forms a robust clade with *E. coli* K-12 derivatives, including *E. coli* MG1655 (PRJNA57779) (15), *E. coli* DH1 (PRJNA30031), *E. coli* DH10B (PRJNA58979) (9), *E. coli* W3110 (PRJNA16351) (13), and *E. coli* BW2952 (PRJNA59391) (10). The genomes of the test strain and MG1655 were highly similar with an ANI value of 99.97%, but strain J53 contained a large inversion between nucleotide positions 380898 and 4539582 of the MG1655 genome. In addition, 5 prophage regions and 18 nonhypothetical genes in MG1655 were deleted in strain J53.

The azide resistance of the test strain was explained by a single nucleotide substitution in the *secA* gene (14), namely, SecA A112V. The methionine requirement was explained by a single nucleotide in *metF* (1), namely, MetF E28K. The deletion of the *proBA* operon, known to be responsible for the proline requirement (6), was observed in the test strain. In total, 60 genes of the test strain showed single nucleotide polymorphisms (SNPs) compared with genes in the MG1655 genome. The strain J53 (F$^-$ *met pro* Azi^r), whose genome was sequenced, has been deposited in the Korean Agricultural Culture Collection under the accession number KACC16628.

Nucleotide sequence accession number. The genome sequence was deposited in GenBank under accession number AICK00000000.

ACKNOWLEDGMENTS

This work was supported by the Priority Research Centers Program (2011-0031388) and a National Research Foundation grant (2011-0016498) through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology, Republic of Korea.

REFERENCES

1. **Ahmed A.** 1973. Mechanism of repression of methionine biosynthesis in *Escherichia-coli*. 1. Role of methionine, S-adenosylmethionine, and methionyl-transfer ribonucleic-acid in repression. Mol. Gen. Genet. **123**: 299–324.
2. **Aziz RK, et al.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9**:75. doi:10.1186/1471-2164-9-75.
3. **Bachmann B.** 1996. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p 2460–2488. *In* Neidhardt FC, et al (ed), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, DC.
4. **Clowes RC, Hayes W.** 1968. Experiments in microbial genetics. Blackwell Scientific Publications, Oxford, United Kingdom.
5. **Clowes RC, Rowley D.** 1954. Some observations on linkage effects in genetic recombination in *Escherichia-coli* K-12. J. Gen. Microbiol. **11**: 250–260.
6. **Csonka LN, et al.** 1988. Nucleotide-sequence of a mutation in the proB gene of *Escherichia-coli* that confers proline overproduction and enhanced tolerance to osmotic-stress. Gene **64**:199–205.
7. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying

Received 16 April 2012 **Accepted** 23 April 2012

Address correspondence to Jongsik Chun, jchun@snu.ac.kr.

doi:10.1128/JB.00641-12

bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679.
8. **Disz T, et al.** 2010. Accessing the SEED genome databases via Web services API: tools for programmers. BMC Bioinformatics **11**:319. doi: 10.1186/1471-2105-11-319.
9. **Durfee T, et al.** 2008. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. J. Bacteriol. **190**:2597–2606.
10. **Ferenci T, et al.** 2009. Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12. J. Bacteriol. **191**:4025–4029.
11. **Goris J, et al.** 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. **57**:81–91.
12. **Jacoby GA, Han P.** 1996. Detection of extended-spectrum beta-lactamases in clinical isolates of *Klebsiella pneumoniae* and *Escherichia coli.* J. Clin. Microbiol. **34**:908–911.
13. **Musso R, Di Lauro R, Rosenberg M, de Crombrugghe B.** 1977. Nucleotide sequence of the operator-promoter region of the galactose operon of *Escherichia coli.* Proc. Natl. Acad. Sci. U. S. A. **74**:106–110.
14. **Oliver DB, Cabelli RJ, Dolan KM, Jarosik GP.** 1990. Azide-resistant mutants of *Escherichia coli* alter the SecA protein, an azide-sensitive component of the protein export machinery. Proc. Natl. Acad. Sci. U. S. A. **87**:8227–8231.
15. **Riley M, et al.** 2006. *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. Nucleic Acids Res. **34**:1–9.
16. **Tatusov RL, Koonin EV, Lipman DJ.** 1997. A genomic perspective on protein families. Science **278**:631–637.