JOURNAL OF BACTERIOLOGY, July 2011, p. 3681 Vol. 193, No. 14
0021-9193/11/$12.00 doi:10.1128/JB.05193-11

Genome Sequence of *Escherichia coli* AA86, Isolated from Cow Feces▽

Hana Yi,[1] Yong-Joon Cho,[2] Hor-Gil Hur,[3] and Jongsik Chun[1,2,4]*

Institute of Molecular Biology and Genetics, Seoul National University, Seoul, Republic of Korea[1]; School of Biological Sciences and Institute of Bioinformatics (BIOMAX), Seoul National University, Seoul, Republic of Korea[2]; Department of Environmental Science and Engineering, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea[3]; and Chunlab, Inc., Seoul National University, Seoul, Republic of Korea[4]

Received 28 April 2011/Accepted 10 May 2011

***Escherichia coli* AA86 (=KACC 15541) is an enteric bacterium that was isolated from a sample of healthy cow feces. Its genome sequence revealed that it is most closely related to the human fecal strain *E. coli* SE15 and could be classified under *E. coli* phylogenetic group B2. Here, we report the genome sequence of *E. coli* AA86, consisting of 3 contigs and 2 plasmids.**

The strains of *Escherichia coli*, a commonly found intestinal bacterium of warm-blooded animals (6), can be classified under 4 phylogenetic groups, A, B1, B2, and D (7), which differ in their virulence factors, ecological niches, and life histories (2, 5, 10). In the course of a comparative genomic study of the *E. coli* population structures in domestic animals, a strain named AA86 was isolated from a fecal sample of a healthy cow. This strain belongs to the group B2 of *E. coli*, which contains many strains causing clinical symptoms. No isolates of this group have been reported from cattle yet.

The genome sequence of *E. coli* AA86 was determined using a combination of shotgun sequencings from a single end in an Illumina IIx genome analyzer (27,236,658 reads; 500× fold coverage; mean read length, 93 bp) and from a paired end (insert size = 3 kb) in a Roche Genome Sequencer FLX system (251,179 reads; mean read length, 391 bp). The sequencing reads were assembled using the CLC genomics wb4 (CLCbio), Newbler assembler 2.3 (Roche), and CodonCode Aligner (CodonCode Co.). After multiple rounds of gap-closing steps based on PCR, 5 contigs, including 2 plasmids, were determined. The resultant genome sequence was uploaded into the RAST server (1) to predict the open reading frames (ORFs) by using Glimmer 3 (3). The predicted ORFs were annotated by searching against clusters of orthologous group (8) and SEED databases (4).

The genome excluding the gaps was 5.0 Mb long, and the G+C content was 50.7 mol%. The genome contained 2 plasmids, namely, pAA86S (65 kb) and pAA86L (84 kb); 7 copies of rRNA operons; and 64 tRNA genes. Among 4,447 predicted protein-coding sequences in the chromosome, 1,303 (29.3%) ORFs matched the hypothetical coding sequences of unknown function in the public database. A phylogenetic tree based on 1,749 homologous genes of the *E. coli-Shigella* group showed that strain AA86 is a member of the *E. coli* phylogenetic group B2 and is most closely related to *E. coli* SE15 (9). It is noteworthy that strain SE15 is a human commensal strain, while our isolate was isolated from a healthy cow, implying the possibility of sharing commensal bacteria between human and domestic animals. Although AA86 and SE15 are closely related, there are many genetic elements that can be distinguished between these 2 bacteria. Strain AA86 contains 2 plasmids, while strain SE15 has 1. The smaller plasmid (pAA86S) is similar to the plasmid of strain SE15 (pSE15) because it shares 62 genes out of the 81 genes encoded in pAA86S, but the larger one (pAA86L) showed little relatedness with strain SE15. In the chromosome, 52 nonhypothetical genes, 7 prophage regions, and 2 integrative elements are exclusively present in the strain AA86. The phylogenetic relationship and the different genomic features of the 2 commensal *E. coli* strains may reflect the differential adaptation of these bacteria with a common origin to the different physiopathological conditions of their host.

Nucleotide sequence accession number. The genome sequence was deposited at GenBank under accession number AFET00000000.

This work was supported by the Priority Research Centers Program (2010-0094020) and National Research Foundation grants (2010-0017955 and 2010-0029224) through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology, Republic of Korea.

REFERENCES

1. **Aziz, R. K., et al.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9:**75.
2. **Carlos, C., et al.** 2010. Escherichia coli phylogenetic group determination and its application in the identification of the major animal source of fecal contamination. BMC Microbiol. **10:**161.
3. **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.
4. **Disz, T., et al.** 2010. Accessing the SEED genome databases via Web services API: tools for programmers. BMC Bioinformatics **11:**319.
5. **Gordon, D. M., and A. Cowling.** 2003. The distribution and genetic structure of Escherichia coli in Australian vertebrates: host and geographic effects. Microbiology **149:**3575–3586.
6. **Meays, C. L., K. Broersma, R. Nordin, and A. Mazumder.** 2004. Source tracking fecal bacteria in water: a critical review of current methods. J. Environ. Manage. **73:**71–79.
7. **Selander, R. K., et al.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:**873–884.
8. **Tatusov, R. L., E. V. Koonin, and D. J. Lipman.** 1997. A genomic perspective on protein families. Science **278:**631–637.
9. **Toh, H., et al.** 2010. Complete genome sequence of the wild-type commensal Escherichia coli strain SE15, belonging to phylogenetic group B2. J. Bacteriol. **192:**1165–1166.
10. **Walk, S. T., E. W. Alm, L. M. Calhoun, J. M. Mladonicky, and T. S. Whittam.** 2007. Genetic diversity and population structure of Escherichia coli isolated from freshwater beaches. Environ. Microbiol. **9:**2274–2288.

* Corresponding author. Mailing address: School of Biological Sciences, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. Phone: 82-2-880-8153. Fax: 82-2-874-8153. E-mail: jchun@snu.ac.kr.

▽ Published ahead of print on 20 May 2011.