JOURNAL OF BACTERIOLOGY, Feb. 2010, p. 1165–1166 Vol. 192, No. 4
0021-9193/10/$12.00 doi:10.1128/JB.01543-09


# GENOME ANNOUNCEMENTS

## Complete Genome Sequence of the Wild-Type Commensal *Escherichia coli* Strain SE15, Belonging to Phylogenetic Group B2[▽]

Hidehiro Toh,[1,2] Kenshiro Oshima,[3,4] Atsushi Toyoda,[2,5] Yoshitoshi Ogura,[6,7] Tadasuke Ooka,[7] Hiroyuki Sasamoto,[4] Sang-Hee Park,[8] Sunao Iyoda,[9] Ken Kurokawa,[10] Hidetoshi Morita,[11] Kikuji Itoh,[8] Todd D. Taylor,[1,2] Tetsuya Hayashi,[6,7] and Masahira Hattori[3]*

*Advanced Science Institute, RIKEN, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan[1]; Genomic Sciences Center, RIKEN, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan[2]; Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan[3]; Kitasato Institute for Life Sciences, Kitasato University, 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan[4]; Center for Genetic Resource Information, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan[5]; Frontier Science Research Center, University of Miyazaki, 5200 Kiyotake, Miyazaki 899-1692, Japan[6]; Department of Infectious Diseases, Faculty of Medicine, University of Miyazaki, 5200 Kiyotake, Miyazaki 899-1692, Japan[7]; Graduate School of Agricultural and Life Sciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan[8]; Department of Bacteriology, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku, Tokyo 162-8640, Japan[9]; Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta, Yokohama, Kanagawa 226-0026, Japan[10]; and School of Veterinary Medicine, Azabu University, 1-17-71 Fuchinobe, Sagamihara, Kanagawa 229-8501, Japan[11]*

Received 26 November 2009/Accepted 3 December 2009

***Escherichia coli* SE15 (O150:H5) is a human commensal bacterium recently isolated from feces of a healthy adult and classified into *E. coli* phylogenetic group B2, which includes the majority of extraintestinal pathogenic *E. coli*. Here, we report the finished and annotated genome sequence of this organism.**

The complete genome sequence of *Escherichia coli* SE15 was determined using a combination of 2-kb and 40-kb Sanger libraries and 454 pyrosequencing. We generated 57,600 sequences (ABI 3730xl sequencers) and three sequencing runs (GS20 sequencers). The 454 pyrosequencing reads were first assembled using the Newbler assembler software (4). A hybrid assembly of 454 and Sanger reads was performed using the Phred-Phrap-Consed program (1). Remaining gaps between contigs were closed by direct sequencing of clones. Prediction and annotation of protein-coding genes were performed as described previously (6).

The genome of *E. coli* SE15 consists of a circular 4,717,338-bp chromosome containing 4,338 predicted protein-coding genes and a 122-kb plasmid (pSE15) encoding 150 protein-coding genes. From the multilocus sequence typing analysis based on the nucleotide sequences of seven housekeeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*), SE15 was found to belong to *E. coli* reference collection group B2. In the chromosome, two prophage regions and seven integrative elements are found. Of the predicted protein-coding genes, we could assign 2,883 (64%) to known functions, 1,528 (34%) as conserved hypothetical genes and 77 (2%) as novel hypothetical genes. Of the predicted protein-coding genes on the chromosome, 3,735 (86%) are common to three uropathogenic *E. coli* (UPEC) genomes (CFT073, UTI89, and 536) and 263 (6%) are not identified in any of the three UPEC genomes. The 263 genes include 7 genes for the phosphoenolpyruvate:sugar phosphotransferase system involved in the uptake of carbohydrates, reflecting the adaptation of SE15 to a commensal lifestyle in the intestinal tract. pSE15 shares 121 genes (81%) with a 114-kb plasmid (GenBank accession no. CP000244) of UPEC UTI89, indicating that both plasmids are derived from the same origin.

The chromosome contains six large segments (LSs; >30 kb) designated LSs I to VI, three of which overlap one prophage region and two integrative elements. Each of the six LSs is located at the same locus as at least one of the pathogenicity islands (PAIs) or other insertion regions in the three UPEC genomes. LS II (ECSF_1824 to ECSF_1835) and three PAIs (PAI $IV_{UTI89}$, PAI $IV_{536}$, and $HPI_{CFT073}$) are located at the same loci in each chromosome and share the *ybt* operon encoding the yersiniabactin iron acquisition system, indicating that the ancestral *E. coli* of group B2 strains may have acquired the *ybt* genes. LS III (ECSF_1852 to ECSF_1897), PAI $VI_{UTI89}$, PAI $VI_{536}$, and $PAI_{CFT073\text{-}asnW}$ are located at the same loci in each chromosome. The three PAIs contain the *pks* island encoding multiple nonribosomal peptide synthases and polyketide synthases, whereas LS III in SE15 completely lacks

* Corresponding author. Mailing address: Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan. Phone and fax: 81-47-136-4080. E-mail: hattori@k.u-tokyo.ac.jp.

▽ Published ahead of print on 11 December 2009.

the *pks* island. The commensal *E. coli* strain ED1a also lacks the *pks* island (8), but the commensal *E. coli* strain Nissle 1917 has the *pks* island (5). These data suggest that the presence of the *pks* island may not be common among intestinal commensal strains in group B2. LS V (ECSF_2770 to ECSF_2794) is almost identical to PAI $V_{UTI89}$, which contains the genes cluster for a type II secretion system (*gsp*), group II capsule synthesis (*kps*), and polysialic acid synthesis (*neu*). The *neu* operon between the *kpsFEDUCS* and *kpsMT* genes in PAI $V_{UTI89}$ is responsible for K1 capsule biosynthesis, and this region between the *kpsFEDUCS* and *kpsMT* genes is highly variable in *E. coli* (9). The corresponding region (ECSF_2777 to ECSF_2781) in LS V encodes genes different from those in the *neu* operon in PAI $V_{UTI89}$; differs from the corresponding regions of the CFT073 (K2 serotype), 536 (K15 serotype), and APEC O1 (K1 serotype) strains; and shows no homology with any sequence in public databases.

SE15 lacks many virulence-related genes, whereas UPEC encodes virulence-related factors, including fimbrial adhesins, toxins, capsule, and serum resistance and iron uptake systems. The three UPEC strains have the genes encoding P fimbriae (*pap*), S fimbriae (*sfa*/*foc*), Auf fimbriae (*auf*), and type 1 fimbriae (*fim*), whereas SE15 contains only the *fim* genes and lacked the *pap*, *sfa*/*foc*, and *auf* genes. Amino acid replacements in FimH located at the tip of type 1 fimbriae produce a shift from a commensal-associated trimannose binding phenotype to a urinary tract infection-associated monomannose binding phenotype (7). The other sequenced B2 strains (three UPEC strains, APEC O1, LF82, and ED1a) have Ser-70 and Asn-78 residues in FimH, whereas SE15 has Asn-70 and Ser-78 residues that are conserved in intestinal *E. coli* strains. Of the seven chaperon-usher fimbrial operons in SE15, six (*fim*, *yad*, *yde*, *yeh*, *yfc*, and *yqi*) are conserved in the three UPEC genomes. The one remaining fimbrial operon (ECSF_0163 to ECSF_0166) is specific to SE15. The GC content (42%) of this 5-kb fimbrial region is lower than the average GC content (51%) of the chromosome. UPEC strains contain a greater number of iron acquisition systems than do commensal strains, which may be a consequence of their adaptation to the iron-limiting urinary tract environment (3). SE15 also contains iron uptake system genes encoding siderophore enterobactin, siderophore yersiniabactin, iron transporter (*sit*), and heme (*chu*) systems but lacks genes for siderophore salmochelin, siderophore aerobactin, and novel siderophore (*ireA*), which are encoded by PAIs of UPEC strains. Furthermore, SE15 lacks genes encoding alpha-hemolysin and cytotoxic necrotizing factor, which are known toxins encoded by PAIs of UPEC strains.

It has been pointed out that extraintestinal pathogenic *E. coli* (ExPEC) virulence factors identified in commensal strains of group B2 may facilitate colonization of the human gut and thus act as fitness factors for commensal *E. coli* stains (2). SE15 contains fewer known ExPEC virulence-associated genes than other known commensal strains (ED1a and Nissle 1917) in group B2, suggesting that ExPEC virulence-related genes in the SE15 genome may be necessary for this commensal microorganism to colonize the human gut.

**Nucleotide sequence accession numbers.** The sequence data of the SE15 genome have been deposited in GenBank/DDBJ/EMBL under accession numbers AP009378 (chromosome) and AP009379 (plasmid).

We thank K. Furuya, C. Yoshino, H. Inaba, K. Motomura, and Y. Hattori (University of Tokyo), A. Yamashita, A. Tamura, N. Itoh (Kitasato University), and the technical staff of the Sequence Technology team of RIKEN GSC for technical assistance.

This research was supported by Grants-in-Aid for Scientific Research on Priority Areas, Area 4—Comprehensive Genomics (M.H.), from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## REFERENCES

1. **Gordon, D., C. Desmarais, and P. Green.** 2001. Automated finishing with Autofinish. Genome Res. **11:**614–625.
2. **Le Gall, T., O. Clermont, S. Gouriou, B. Picard, X. Nassif, E. Denamur, and O. Tenaillon.** 2007. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. Mol. Biol. Evol. **24:**2373–2384.
3. **Lloyd, A. L., D. A. Rasko, and H. L. Mobley.** 2007. Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli*. J. Bacteriol. **189:**3532–3546.
4. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380.
5. **Nougayrède, J. P., S. Homburg, F. Taieb, M. Boury, E. Brzuszkiewicz, G. Gottschalk, C. Buchrieser, J. Hacker, U. Dobrindt, and E. Oswald.** 2006. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. Science **313:**848–851.
6. **Oshima, K., H. Toh, Y. Ogura, H. Sasamoto, H. Morita, S. H. Park, T. Ooka, S. Iyoda, T. D. Taylor, T. Hayashi, K. Itoh, and M. Hattori.** 2008. Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. DNA Res. **15:**375–386.
7. **Sokurenko, E. V., V. Chesnokova, D. E. Dykhuizen, I. Ofek, X. R. Wu, K. A. Krogfelt, C. Struve, M. A. Schembri, and D. L. Hasty.** 1998. Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. Proc. Natl. Acad. Sci. U. S. A. **95:**8922–8926.
8. **Touchon, M., C. Hoede, O. Tenaillon, V. Barbe, S. Baeriswyl, P. Bidet, E. Bingen, S. Bonacorsi, C. Bouchier, O. Bouvet, A. Calteau, H. Chiapello, O. Clermont, S. Cruveiller, A. Danchin, M. Diard, C. Dossat, M. E. Karoui, E. Frapy, L. Garry, J. M. Ghigo, A. M. Gilles, J. Johnson, C. Le Bouguénec, M. Lescat, S. Mangenot, V. Martinez-Jéhanne, I. Matic, X. Nassif, S. Oztas, M. A. Petit, C. Pichon, Z. Rouy, C. S. Ruf, D. Schneider, J. Tourret, B. Vacherie, D. Vallenet, C. Médigue, E. P. Rocha, and E. Denamur.** 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet. **5:**e1000344.
9. **Whitfield, C.** 2006. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. Annu. Rev. Biochem. **75:**39–68.