S2-3

Genome Analysis of Enterohemorrhagic *Escherichia coli* O157:H7 Derived From Sakai Outbreak in Japan in 1996.

Kozo Makino*, Katsushi Yokoyama, Sigenobu Kimura, Masahiro Hattori[1,2], Tetsuya Hayashi[3], Makoto Ohnishi[3], Ken Kurokawa[4], Teruo Yasunaga[4], Ttakeshi Honda[5], Chihiro Sasakawa[6] and Hideo Shinagawa

Department of Molecular Microbiology, Research Institute for Microbial Diseases, Osaka University, Japan, [1]Human Genome Research Group, RIKEN Genomic Science Center, c/o Kitasato University, Japan, [2]Human Genome Center, Institute of Medical Science, The University of Tokyo, Japan, [3]Department of Bacteriology, School of Medicine, Shinshu University, Japan, [4]Genome Information Research Center, Osaka University, Japan, [5]Department of Bacterial Infections, Research Institute for Microbial Diseases, Osaka university, Japan, [6]Department of Bacteriology, Institute of Medical Science, The University of Tokyo, Japan

We have initiated a project to determine genomic sequences of Entreohemorrhagic *Escherichia coli* (EHEC) O157:H7 strain RIMD 0509952 derived from a large outbreak in 1996 in Sakai city, Japan to identify potentially important genes related to the pathogenicity and to compare the genome organization of this strain with that of *E. coli* K12 previously determined. First, we determined the complete nucleotide sequences of a 93 kb pathogenic plasmid (pO157) commonly identified in EHEC strains world-wide and a 3.3kb plasmid (pOSAKAI) which was uniquely found in some of the Japanese strains. Restriction enzyme maps of *Xba*I and *Bln*I were constructed. The overall sequence

organizations of the two genomes were compared by hybridization using short common sequences as probes. The results indicate that most of the common sequences are arranged in the same order in the two strains although they are interrupted frequently by the sequences unique to the respective genomes. Whole chromosomal sequence is being analyzed by a shot-gun method combined with the analysis of larger chromosome fragments in λ phage library. The total length of the sequenced region is 5.6 Mb, about 1 Mb larger than the K12 genome. The O157 specific sequence and the K12 specific sequence were estimated to be approximately 1.5 Mb and 600 kb, respectively, and they were dispersed along the chromosome map as blocks. For example, the O157 specific 43 kb region of the pathogenicity island LEE which contains many virlence genes is inserted at 82 min of the K12 map. Genes related to bacterial pathogenicity are often transferred horizontally by temperate bacteriophages. The EHEC O157:H7 strain RIMD 0509952 produces two kinds of verotoxins, VT1 and VT2, encoded by the *stx1* and *stx2* genes. The *stx1* and *stx2* genes are integrated in the different but very similar lambdoid prophages. Now, we have practically finished the O157:H7 strain RIMD 0509952 sequencing.

93-kb and 3.3kb Plasmids

EHEC O157:H7 strain RIMD 0509952 possesses two kinds of plasmids: a 93-kb plasmid termed pO157, found in clinical EHEC isolates world-wide and a 3.3kb plasmid termed pOSAK1, prevalent in EHEC strains isolated in Japan. Complete nucleotide sequences of both plasmids have been determined, and the putative functions of the encoded proteins and the *cis*—acting DNA sequences have been analyzed. pO157 shares strikingly similar genes and DNA sequences with F-factor and the transmissible drug-resistant plasmid R100 for DNA replication, copy number control, plasmid segregation,

conjugative functions and stable maintenance in the host, although it is defective in DNA transfer by conjugation due to the truncation and deletion of the required genes and DNA sequences. In addition, it encodes several proteins implicated in EHEC pathogenicity such as an EHEC hemolysine (HlyA), a catalase-peroxidase (KatP), a serine protease (EspP), and type II secretion system. pOSAK1 possesses a ColE1-like replication system, and the DNA sequence is extremely similar to that of a drug-resistant plasmid, NTP16, derived from *Salmonella typhimurium* except that it lacks drug resistance transposons.

Comparative Analysis with K-12

A complete *Xba* I and *Bln* I cleavage map was constructed for the chromosome of the EHEC O157:H7 strain. A comparative chromosome analysis with *E. coli* K-12 strain MG1655 was made. The EHEC chromosome was approximaterly 5600 kb in length, 1 Mb larger than that of MG1655. Despite the marked difference in chromosome length, the location and direction of seven rRNA operons of the EHEC strain were similar to those for MG1655. Overall organization of genes common in both strains is also highly conserved. Chromosome expansion was observed throughout the EHEC chromosome, albeit in an uneven manner. A large portion of the chromosome enlargement was observed in the region surrounding the replication terminus, particularly in a segment containing the *terA* locus. Sample sequencing of 3627 random shotgun clones suggested the presence of approximately 1550 kb strain-specific DNAs on the EHEC chromosome, most of which are likely to be of foreign origin.

Prophage VT2-Sakai

In the EHEC strains, as well as in other VT-producing *E. coli* strains, the toxins are encoded by lysogenic bacteriophages. The EHEC O157:H7 strain RIMD 0509952 did not produce plaque-forming phage particles upon inducing treatments. We have determined the complete nucleotide sequence of a prophage, VT2-Sakai, carrying the *stx2A* and *stx2B* genes

on the chromosome, and presumed the putative functions of the encoded proteins and the *cis*-acting DNA elements based on sequence homology data. To our surprise, the sequences in the regions of VT2-Sakai corresponding to the early gene regulators and replication proteins, and the DNA sequences recognized by the regulators share very limited homology to those of the VT2-encoding 933W phage carried by the EHEC O157:H7 strain EDL933 reported by Plunkett et al. (J. Bacteriol., p1767-1778, **181**, 1999), although the sequences corresponding to the structural components are almost identical. These data suggest that those two phages were derived from a common ancestral phage and that either or both of them underwent multiple genetic rearrangements. An *IS629* insertion was found downstream of the *stx2B* gene and upstream of the lysis gene *S*, and this might be responsible for the absence of plaque-forming activity in the lysate obtained after inducing treatments.

Prophage VT1-Sakai

The phage carrying the *stx1A* and *stx1B* genes of the EHEC O157:H7 strain RIMD 0509952 was induced after treatment with mitomycin C but the plaque formation of the phage was not detected. We have determined the complete nucleotide sequence of the prophage carrying the *stx1* genes, VT1-Sakai, residing in RIMD 0509952. We have identified the integration site of the prophage, which is located within the *yehV* gene at 47.7 min on the chromosome. The location of the *stx1* genes downstream of the *Q* gene in the prophage genome suggests that their expression is regulated by the Q protein, the regulator of the late gene expression of the phage, similarly to that of the *stx1* or *stx2* genes carried by the lambdoid phages reported previously. The sequences of the *N* gene and its recognition sites, *nutL* and *nutR*, are not homologous to those of the phages carrying the *stx* genes thus far reported, but they are very similar to those of bacteriophage φ21. The sequences of the repressor proteins, CI and Cro, that regulate expression of the early genes have low similarities

with those of the known repressors of other phages, and their operator sequences are different from any sequence reported. These data suggest that multiple gene transfers of the *stx* genes have been mediated by bacteriophages with different immunities. Comparison between the sequences of VT1-Sakai and lambda suggests that the ancestor of VT1-Sakai was produced by illegitimate excision, like lambda *gal* and *bio* phages. We found an *IS629* element inserted into a head gene, suggesting that VT1-Sakai in itself is a defective prophage.

Other Futures

The O157 unique 43 kb region of the pathogenicity island LEE contains many virlence genes such as those encoding intimine, intimine receptor, and type III secretion system. One end of LEE is marked by lambdoid phage and it is inserted into the site near the *selC* gene at 82 min of the K12 map. The LEE island contains 54 ORFs, which include 19 putative virulence genes, 13 phage proteins and 18 unknown ORFs. We identified second putative pathogenicity island consisting of 15 ORFs inserted at 64.1 min of the K12 map, which are homologous to pathogenic genes of *Salmonella typhimurium* and *S. enterica*. A cluster of genes, which are likely to make up an operon and encodes putative multi-drug resistance factors, is inserted at 29.1 min of the K12 map.

References

K.Makino, et al. Complete nucleotide sequences of 93-Kb and 3.3Kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. DNA Research 5: 1-9 (1998).

M. Ohnishi, et al. Chromosome of the enterohemorrhagic *Escherichia coli* O157:H7 ; comparative analysis with K-12 MG1655 revealed the acquisition of a large amount of foreign DNAs. DNA Research 6: 361-368 (1999).

K. Makino, et al. Complete nucleotide sequence of the prophage VT2- Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7derived from Sakai outbreak. Genes Gent. Syst. 74: 227-239 (1999).

K. Yokoyama, et al. Complete nucleotide sequence of the defective prophage VT1-Sakai carrying shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from Sakai outbreak. (in preparation).