DNA Research **8**, 11–22 (2001)

Complete Genome Sequence of Enterohemorrhagic *Eschelichia coli* O157:H7 and Genomic Comparison with a Laboratory Strain K-12

Tetsuya Hayashi,[1,*] Kozo Makino,[2,*] Makoto Ohnishi,[1] Ken Kurokawa,[3] Kazuo Ishii,[4] Katsushi Yokoyama,[2] Chang-Gyun Han,[5] Eiichi Ohtsubo,[5] Keisuke Nakayama,[1] Takahiro Murata,[6] Masashi Tanaka,[3] Toru Tobe,[7] Tetsuya Iida,[8] Hideto Takami,[9] Takeshi Honda,[8] Chihiro Sasakawa,[7] Naotake Ogasawara,[10] Teruo Yasunaga,[3] Satoru Kuhara,[11] Tadayoshi Shiba,[4] Masahira Hattori,[4,12] and Hideo Shinagawa[2]

Department of Microbiology, Miyazaki Medical College, 5200 Kiyotake, Miyazaki 899-1692, Japan,[1] *Department of Molecular Microbiology, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan,*[2] *Genome Information Research Center, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan,*[3] *School of Science, Kitasato University, 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan,*[4] *Institute of Molecular and Cellular Biosciences, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan,*[5] *Department of Bacteriology, Shinshu University School of Medicine, 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan,*[6] *Department of Microbiology and Immunology, Institute of Medical Science,University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan,*[7] *Department of Bacterial Infections, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Osaka 565-0871, Japan,*[8] *Deep-sea Microorganisms Research Group, Japan Marine Science and Technology Center, 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan,*[9] *Graduate School of Biological Science, Nara Institute of Science and Technology, 8916-5 Ikoma, Nara 630-0101, Japan,*[10] *Graduate School of Genetic Resources Technology, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan,*[11] *and Genome Sequencing Team, Human Genome Research Group, RIKEN Genomic Sciences Center, 1-7-22 Shuehiro-cho, Tsurumi-ku, Yokohama-city, Kanagawa, 230-0045 Japan*[12]

(Received 19 January 2001; revised 21 January 2001)

Abstract

Escherichia coli O157:H7 is a major food-borne infectious pathogen that causes diarrhea, hemorrhagic colitis, and hemolytic uremic syndrome. Here we report the complete chromosome sequence of an O157:H7 strain isolated from the Sakai outbreak, and the results of genomic comparison with a benign laboratory strain, K-12 MG1655. The chromosome is 5.5 Mb in size, 859 Kb larger than that of K-12. We identified a 4.1-Mb sequence highly conserved between the two strains, which may represent the fundamental backbone of the *E. coli* chromosome. The remaining 1.4-Mb sequence comprises of O157:H7-specific sequences, most of which are horizontally transferred foreign DNAs. The predominant roles of bacteriophages in the emergence of O157:H7 is evident by the presence of 24 prophages and prophage-like elements that occupy more than half of the O157:H7-specific sequences. The O157:H7 chromosome encodes 1632 proteins and 20 tRNAs that are not present in K-12. Among these, at least 131 proteins are assumed to have virulence-related functions. Genome-wide codon usage analysis suggested that the O157:H7-specific tRNAs are involved in the efficient expression of the strain-specific genes. A complete set of the genes specific to O157:H7 presented here sheds new insight into the pathogenicity and the physiology of O157:H7, and will open a way to fully understand the molecular mechanisms underlying the O157:H7 infection.

Key words: *E. coli* O157:H7; genome sequence; *E. coli* K-12; bacterial pathogenicity; evolution

Communicated by Satoshi Tabata

* Corresponding authors:
T. Hayashi Tel. +81-985-85-0871, Fax. +81-985-6475, E-mail: thayash@fc.miyazaki-med.ac.jp
K. Makino Tel. +81-6-6879-8318, Fax. +81-6-6879-8320, E-mail: makino@biken.osaka-u.ac.jp

1. Introduction

Since enterohemorrhagic *Escherichia coli* (EHEC) O157:H7 was first recognized as a gastrointestinal pathogen in 1982,[1] its occurrence has become a world-

wide public health problem, causing sporadic incidents as well as outbreaks of hemorrhagic colitis. Although other *E. coli* serotypes, such as O26:H11 and O111:NM, share a similar pathogenic potential, most large outbreaks of EHEC infection have been caused by O157:H7.[2] A prominent example is the huge outbreak which occurred in 1996 in primary schools of Sakai City, Osaka prefecture, Japan, where more than 6000 schoolchildren were affected.[3] EHEC causes not only hemorrhagic colitis but also serious complications such as hemolytic uremic syndrome (HUS), sometimes resulting in death. In the Sakai outbreak, approximately 1000 patients were hospitalized with severe gastrointestinal symptoms and about 100 victims had complications of HUS, resulting in 3 deaths. Virulence determinants contributing to the EHEC infection have been partly characterized, but the mechanism by which EHEC causes hemorrhagic colitis and HUS are not fully understood.[4]

We have determined the genome sequence of an O157:H7 strain isolated from the Sakai outbreak. The genome sequence of the benign laboratory strain K-12 MG1655 has already been determined,[5] but, to our knowledge, this strain (referred to as O157 Sakai) is the first pathogenic *E. coli* strain whose genome has been fully sequenced. By comparing the two strains, we identified all chromosomal components specific to each strain as well as those conserved in both strains. These results provided a broad array of whole genome-level information not only for obtaining a complete set of genes potentially related to the pathogenicity of O157:H7 but also for understanding the evolution of *E. coli* strains.

2. Materials and Methods

2.1. Bacterial strain

EHEC 0157:H7 (RIMD 0509952) was isolated from a typical patient during the Sakai outbreak. This strain produces two Shigatoxins, Stx1 and Stx2, and contains two plasmids, pO157 and pOSAK1. The sequences of the plasmids, prophages encoding Stx1 and Stx2, and the seven sets of *rrn* operons were reported previously.[6–9] The physical map of the chromosome was also reported.[10]

2.2. Sequencing and assembly

The initial stage of sequencing was done by the whole genome random shotgun method as described.[6–8] We constructed a pUC18-based library containing 1- to 2-Kb inserts, and sequenced 50,156 clones using a forward-sequencing primer. After assembling the sequence data using phred/phrap/consed,[11,12] we selected two groups of clones: clones having inserts whose sequences started within 1.5 Kb from the ends of contigs and oriented outside, and those having inserts whose opposite ends covered the regions which have ambiguity in the sequence (lower than the evaluation value 20 by phred scoring). A total of 19,969 clones that were selected according to these criteria were sequenced using the reverse primer. This strategy was quite effective in reducing the number of random clones to be sequenced as well as in improving the sequence quality. We also constructed a lambda-based library with ca. 20-Kb inserts. We selected 86 clones that contained the sequences non-homologous to the K-12 sequence at either end of the inserts, and determined the entire sequences of each insert by the random strategy. The obtained sequences were assembled into 111 contigs larger than 1 Kb in size. At this stage, we checked the sequence waves of all the regions that had low quality values by visual inspection, and all regions with any ambiguity (286 regions) were amplified by PCR and reanalyzed by direct sequencing of the PCR products. Subsequently, we performed gap closing by PCR according to the physical map of the chromosome and the results of the systematic gene mapping.[10] The physical map deduced from the whole chromosomal sequence determined in this study agreed with the experimentally determined map, guaranteeing the accuracy of the final assembly.

2.3. ORF prediction, annotation, and sequence comparison with K-12

We first defined all the O157 Sakai-specific sequences larger than 19 bp by comparing the whole chromosomal sequence with that of K-12 MG1655 (Accession no. U00096) using the MUMmer program.[13] Then, the open reading frames (ORFs) in the strain-specific regions and those on the regions conserved in the two strains were identified and annotated separately using Genome Gambler version 1.41,[14] GLIMMER 2.01,[15] and BLAST.[16] In principal, ORFs larger than 150 bp were searched, but there are several exceptions. Conserved ORFs were annotated principally according to the descriptions for K-12 MG1655[5] and to "The *E. coli* Index" (http://web.bham.ac.uk/bcm4ght6/res.html), but eight small conserved ORFs were newly identified in this study. ORFs lying at the junctions of conserved regions and strain-specific regions were manually identified and annotated with a guide by BLAST results. tRNA genes were identified by tRNAscan-SE-1.12,[17] and other small RNAs were identified by BLAST. Paralogous gene families were determined using BLAST under the criteria that at least 60% of query sequences were aligned with at least 30% identity.

3. Results and Discussion

3.1. Overview

The complete sequence of the O157 Sakai chromosome is 5,498,450 bp in length (Fig. 1). Since the strain contains a large virulence plasmid of 92,721 bp (pO157) and a cryptic plasmid of 3306 bp (pOSAK1),[6] the whole

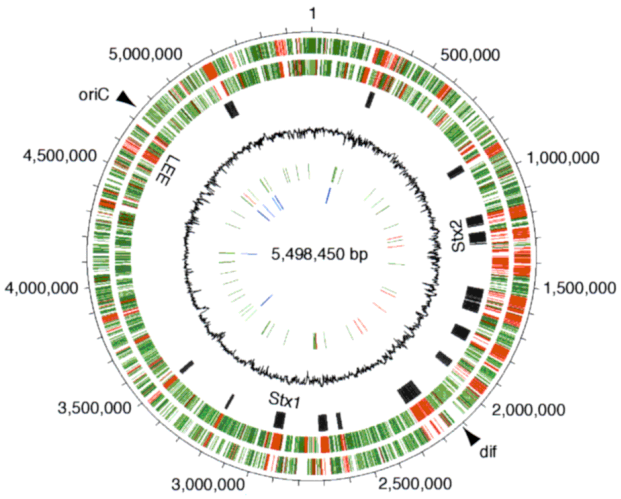

Figure 1. Circular representation of the O157 Sakai chromosome. The outermost circle indicates the chromosomal location in base pairs (each tick is 100 Kb). The second and the third show predicted ORFs transcribed in the clockwise and counterclockwise directions, respectively. ORFs conserved in K-12 are depicted in green and those not present in K-12 in red. The fourth circle shows the locations of ORFs on prophage genomes (Sp1–18). The fifth circle shows the 20-Kb window-average of G+C percent in relation to the mean value of the chromosome. The locations of tRNA and rRNA genes are shown in the sixth and seventh circles, respectively. tRNAs conserved in K-12 are depicted in green, and those absent in K-12 are in red.

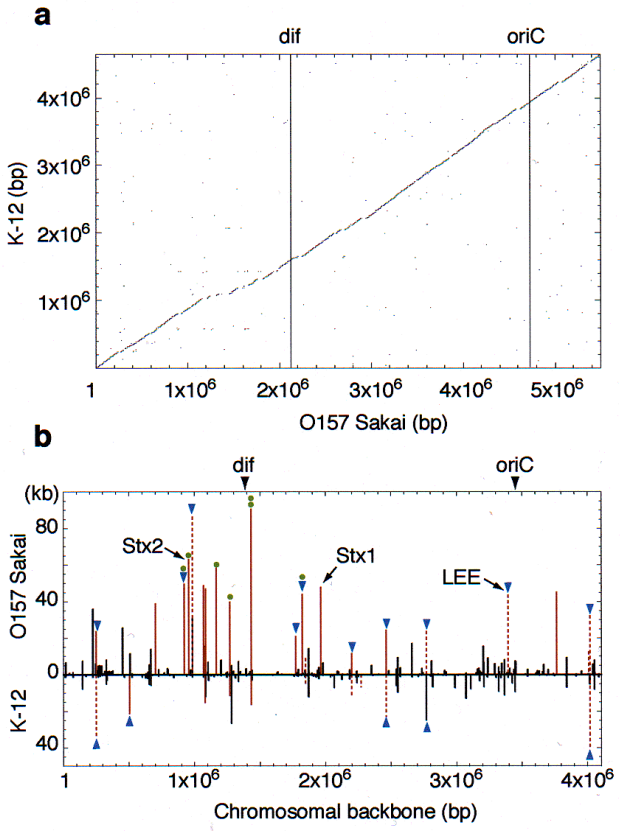

Figure 2. Chromosome comparison of O157 Sakai and K-12 MG1655. a: Dot plot representation of the nucleotide sequence conservation. Dots represent perfectly conserved sequences longer than 19 bp. b: Distribution of the strain-specific loops on the conserved backbone. The horizontal axis represents the backbone location, and the vertical bars the locations and the lengths of loops specific to O157 Sakai (upwards) or K-12 (downwards). Loops composed of prophages and prophage-like elements are depicted by solid and broken red bars, respectively. Phages and phage-like elements integrated into tRNA (or tmRNA) genes and those carrying tRNA genes are indicated by blue triangles and green circles, respectively.

genome size is 5,594,477 bp, being the second largest bacterial genome sequenced so far. The 5.5-Mb chromosome encodes 5361 protein coding regions, 7 sets of rRNAs (16S, 23S, and 5S RNAs), 102 tRNAs, 1 tmRNA, and at least 13 small RNAs including RNase P, 6S RNA, and 4.5S RNA. Protein-coding regions occupy 88.1% of the chromosome and the average length of the ORFs is 904 bp. The G + C content of the entire chromosome is 50.5 mol% (Table 1).

The chromosome length is 859 Kb larger than K-12 MG1655. By comparing the two chromosome sequences, we identified an approximately 4.1-Mb sequence conserved in the two strains. There is no large rearrangement such as translocation or inversion in the conserved regions (Fig. 2a). The level of nucleotide sequence conservation in the 4.1-Mb sequence is remarkable; 98.31% identity with 2027 gaps. Since the two strains are known to belong to distinct *E. coli* lineages,[18,19] the 4.1-Mb sequence probably represents the chromosome backbone conserved in most *E. coli* strains, though it may contain some segments exceptionally common to the two strains but not to others.

The backbone is, however, interrupted by numerous DNA segments of various sizes that are specific to each strain. These segments, that we call "strain-specific loops," are distributed throughout the backbone, but in an uneven manner (Fig. 2b). In O157 Sakai, larger loops were more frequently present in the regions surrounding the replication termination site (*ter*). Although replichores 1 and 2 are almost equal in length in K-12, replichore 1 in O157 Sakai is 290 Kb longer than replichore 2, as has previously been predicted.[10] The importance of the chromosome symmetry has been proposed in enteric bacteria,[20] but this level of asymmetry appears to be permissible in *E. coli*. There are 296 strain-specific loops larger than 19 bp in O157 Sakai (S-loops 1–296) and 325 loops in K-12 (K-loops 1–325). Among these, 203 loops are located at analogous sites on the two chromosomes, but they are different sequences. These sites may represent "hot spots" for integration of foreign DNAs or for recombination. Another striking feature is

Table 1. Genome features of O157 Sakai.

	chromosome					plasmids †		total
	whole	conserved in K-12	strain-specific * whole	strain-specific * prophages	strain-specific * others	pO157	pOSAK1	
length of sequence (bp)	5,498,450	4,105,380	1,393,070	670,944	722,126	92,721	3,306	5,594,477
G+C ratio (%)	50.5	51.1	48.7	50.3	47.2	47.6	43.4	50.4
open reading frame (ORF)	5,361	3,729	1,632	887	745	83	3	5,447
protein coding region (% of genome size)	88.1	-	-	-	-	82.5	52.8	88
average ORF length (bp)	904	979	739	628	888	922	579	904
rRNA (16S-23S-5S)	7	7	0	0	0	0	0	7
tRNA and tmRNA	103	83	20	18	2	0	0	103
non-classical RNA	13	10	3	0	0	0	0	13

* Sum of the strain-specific loops longer than 19 bp.
† The data was taken from Makino et al.[6]

that most of the large loops are prophages or prophage-like elements. In O157 Sakai, 21 of 29 S-loops larger than 10 kb are prophages or prophage-like elements, and the largest loop of 91.8 Kb in size (S-loop108) is composed of two lambda-like phages integrated in tandem.

The total length of S-loops is $1,393,071$ bp, 25.3% of the chromosome. This corresponds to the whole genome size of a Lyme disease pathogen, *Borrelia burgdorferi* (1.44 Mb),[21,22] and is more than twice that of *Mycoplasma genitalium* which has a minimal genome (0.58 Mb).[23] The average G + C content of the S-loops is 48.5 mol%, significantly lower than that for the conserved backbone (Table 1 and Fig. 1), suggesting that many of the loops are of foreign origin. The S-loops are categorized into two groups: loops apparently composed of prophages or phage remnants into one and the rest into another. The G + C content of the latter group is more atypical, while the phage loops have an average G + C content more similar to that of the backbone (Table 1). This is because most regions encoding the phage-essential genes are generally similar to the backbone in base composition. Regions on the phage genomes that are apparently non-essential for phage propagation often exhibit atypical base compositions.

The comparative analysis of codon usage between the genes on the backbone and those on the S-loops also suggests the abundance of foreign genes on S-loops (Fig. 3). Codons that are used frequently in the backbone genes are less frequently used in the S-loop genes. Conversely, codons that are less frequently used in the backbone genes are more frequently used in the S-loop genes. This indicates the atypical codon usage in the S-loop genes.

3.2. Mobile genetic elements

A number of mobile genetic elements were identified on the O157 Sakai chromosome; 20 kinds of insertion sequences (80 copies in total, but 45 are truncated or partially deleted copies) and 18 prophages or phage remnants (Tables 1 and 2 in the Supplement section and Fig. 1). Among the 20 kinds of IS elements, seven species are the ones newly identified in this study. O157 Sakai and K-12 share eight types, but the major IS elements in each strain are completely different. The most abundant IS elements on the O157 Sakai chromosome are IS*629* (19 copies) and the IS*679*-related elements (IS*Ec8*, *682*, and *683*; 16 copies in total).

Of the 18 prophages or phage remnants (Sakai prophages; Sp1–18), 13 are lambda-like phages resembling each other. All these lambda-like phages, including Stx1- and Stx2-transducing phages that corresponded to Sp15 and Sp5, respectively,[7,8] contain various types of deletions and/or insertions of IS elements in the phage-essential regions, and thus are apparently defective. They, however, show surprisingly high similarities to each other even on the nucleotide sequence level (Table 2). It is unknown how these highly homologous sequences were initially brought in and how they are maintained on a single chromosome, but recombination between the prophages may be responsible for the generation of some chimeric phages which share identical or nearly identical segments. Prophages other than the lambda-like phages include a Mu-like phage, a P4-like phage, and the remnants of P2-like and P22-like phages. Taken together, about half of the O157 Sakai-specific sequences (48.2%) are of bacteriophage origin, suggesting the predominant roles of bacteriophages in the evolution of O157:H7. These phages indeed carry not only the Stx genes but also various genes potentially related to the pathogenesis (see below).

In addition to the 18 prophages, six chromosomal regions of O157 Sakai exhibit prophage-like features (Sakai prophage-like elements, SpLE1–6), though they contain no genes with significant homology to known bacteriophage genes, except for those encoding integrase-like proteins (Table 2 in the Supplement section). SpLE1 and 4 share some identities with

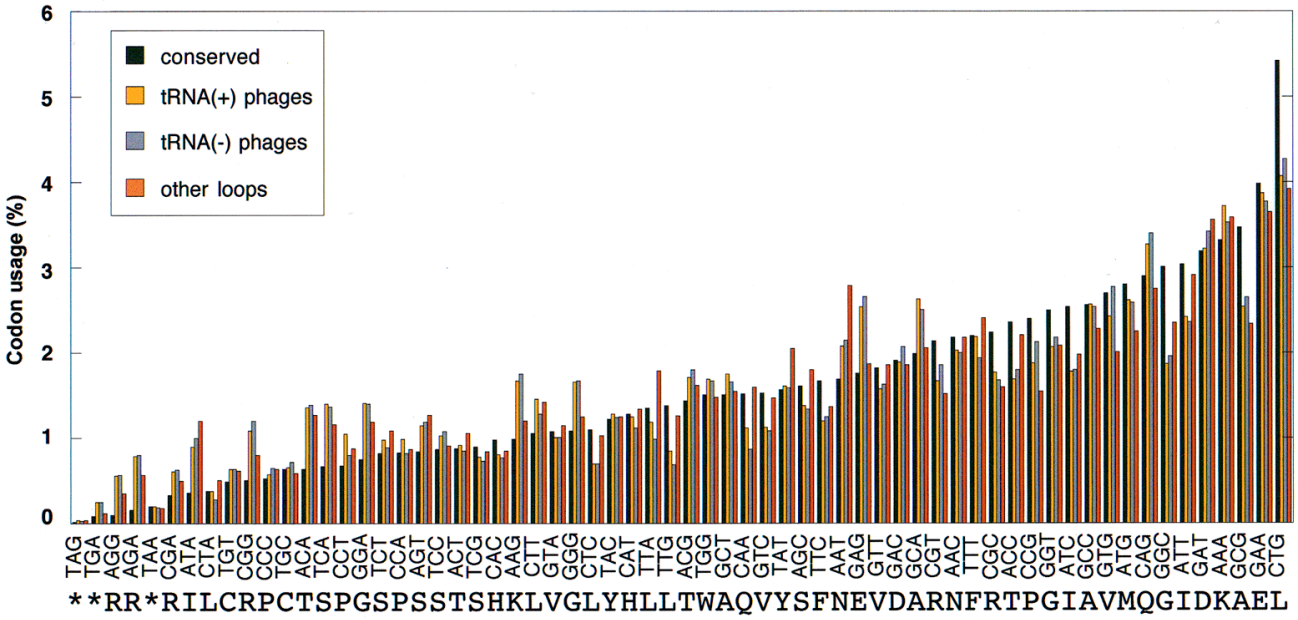

Figure 3. Codon usage analysis of the O157 Sakai chromosomal genes. The genes were divided into four groups: those on the conserved backbone (conserved), on the prophages carrying tRNA genes (tRNA(+) phages), on the prophages not carrying tRNA genes (tRNA(−) phages), on the other S-loops (other loops). Then the codon frequency was calculated for each group. Codons are put in the order of frequency in the genes on the conserved backbone.

"CP4 cryptic prophages," prophage-like elements of K-12.[5,24] The former corresponds to S-loop72 (the second largest loop, 86.2 Kb in size), a part of which has been described as a "tellurite resistance- and adherence-conferring island."[25] The latter includes the LEE locus. Both elements encode P4 integrase-like proteins and are integrated into tRNA genes accompanying phage attachment site-like (*att*-like) sequence duplications. Furthermore, the two SpLEs share several homologous genes with CP4 prophages of K12; such as ECs1405 (on SpLE1)/ECs4538 (on SpLE4)/*yeeV* (on CP4–44), ECs1406/ECs4539/*yeeU*, and ECs1407/ECs4540/*yeeW*, suggesting that both SpLEs are CP4-like elements. SpLE5 encodes a phage integrase-like protein and is integrated into a tRNA gene (*leuX*) with a 26-bp duplication. SpLE3 and SpLE6 also encode P4 integrase-like proteins and are apparently integrated into tRNA genes, but their right end regions containing the *attR* sites have been lost probably by IS insertions and following genetic rearrangements. SpLE2 is a correspondent of CP4–44, but the internal portion is replaced by a different sequence. Although we do not have direct evidence to claim that these elements are really prophages, the above mentioned features suggest that they are at least phage-like mobile genetic elements.

tRNA genes have been repeatedly reported as integration sites for various genetic elements including bacteriophages.[26,27] In O157 Sakai, a total of 10 tRNA or tmRNA genes are used as integration sites for phages or phage-like elements. At two loci, two different phages are integrated in tandem. Each of these loci encodes a single tRNA (or tmRNA) gene except *leuZ* which is located the furthest downstream of the *glyW-cysT-leuZ* tRNA gene cluster. Thus, 36% of the 25 single-tRNA gene loci on the O157 Sakai chromosome are occupied by phages or phage-like elements. In other words, of the 24 phages or phage-like elements, 11 use such single-tRNA gene loci as integration sites, demonstrating that single-tRNA loci are the most favored integration sites for these elements.

K-12 also carries three lambda-like phages (DLP12, Rac, and Qin), a phage-like element (e14) and four CP4 cryptic prophages (CP4–6, -44, -57, and an unnamed one)[5,24] Although the endpoints of the three lambda-like phages have not been exactly defined, we could determine their possible endpoints by sequence comparison with O157 Sakai (Table 3 in the Supplement section). It is noteworthy that Rac is integrated into the same site as that for Sp10 of O157 Sakai, and that Rac and Sp10 share a ca. 21-Kb segment encompassing a region from a putative integrase gene, b1345 in MG1655 (ECs1929 in O157 Sakai), to the 5′ part of *ydaW* (ECs1946), but with several replacements of internal small segments. Qin and Sp12 also share a 3.8-Kb right end segment encoding the *dicF* RNA gene, *dicB* (ECs2284), *ydfD* (ECs2285), and *ydfE* (ECs2286), but the structures of the very ends differ. The *attR*-containing regions of both *qin* and Sp12

Table 2. Nucleotide sequence homologies between the lambda-like phages on the O157 Sakai chromosome.

prophage (length in bp)	Sp9	Sp12	Sp10	Sp6	Sp11	Sp15	Sp4	Sp14	Sp17	Sp5	Sp8	Sp1	Sp3	lambda
Sp9	-	23,302	23,166	17,946	8,433	8,097	7,113	6,483	3,344	1,931	317	0	0	0
(58,175)		(9; 13,038)	(6; 19,508)	(6; 12,315)	(11; 2,074)	(5; 2,157)	(9; 1,952)	(8; 1,953)	(2; 2,842)	(1; 1,931)	(1; 317)	(0; 0)	(0; 0)	(0; 0)
Sp12		-	25,573	21,411	10,339	8,300	9,004	6,329	3,214	2,594	1,143	0	134	335
(46,142)			(9; 13,040)	(7; 9,980)	(7; 5,552)	(6; 3,538)	(7; 2,160)	(5; 1,768)	(3; 1,859)	(2; 1,931)	(1; 1,143)	(0; 0)	(1; 134)	(1; 335)
Sp10			-	18,046	12,079	6,008	9,244	6,134	3,706	2,778	1,063	0	0	0
(51,112)				(6; 6,774)	(9; 2,879)	(3; 2,157)	(6; 4,703)	(4; 2,338)	(3; 2,657)	(2; 1,931)	(1; 1,063)	(0; 0)	(0; 0)	(0; 0)
Sp6				-	3,789	2,552	3,342	5,337	1,257	1,837	0	0	134	0
(48,423)					(3; 2,074)	(2; 1,837)	(6; 2,195)	(5; 2,183)	(3; 863)	(1; 1,837)	(0; 0)	(0; 0)	(1; 134)	(0; 0)
Sp11					-	11,690	10,980	15,500	3,606	3,077	9,095	0	134	7,456
(45,776)						(5; 6,869)	(11; 2,242)	(8; 8,833)	(4; 1,850)	(3; 1,438)	(2; 7,454)	(0; 0)	(1; 134)	(1; 7,456)
Sp15						-	6,099	8,632	1,443	13,004	2,168	992	3,113	6,393
(47,879)							(6; 2,045)	(4; 5,536)	(4; 477)	(10; 3,874)	(2; 1,397)	(2; 568)	(4; 2,316)	(6; 3,702)
Sp4							-	23,055	2,772	179	1,423	0	0	235
(49,650)								(7; 11,153)	(5; 1,688)	(1; 179)	(2; 1,143)	(0; 0)	(0; 0)	(1; 235)
Sp14								-	2,977	0	2,020	0	0	0
(44,029)									(3; 1,682)	(0; 0)	(2; 1,641)	(0; 0)	(0; 0)	(0; 0)
Sp17									-	685	128	0	2,415	1,988
(24,199)										(2; 498)	(1; 128)	(0; 0)	(2; 1,257)	(3; 1,352)
Sp5										-	708	0	2,512	4,637
(62,708)											(1; 708)	(0; 0)	(2; 2,316)	(3; 3,750)
Sp8											-	487	4,743	23,082
(46,897)												(1; 487)	(2; 3,260)	(5; 12,670)
Sp1												-	0	2,453
(10,586)													(0; 0)	(2; 1,885)
Sp3													-	10,469
(38,586)														(7; 3,254)
lambda														-
(48,502)														

Total nucleotide sequence lengths aligned with more than 95% identity by the MUMmer program are presented for each combination. In parentheses, the number of aligned segments and the length of the longest aligned segment are presented. Homologies of each O157 Sakai phage to phage lambda (Accession number; J02459) are also shown.

Table 3. Functions of strain-specific ORFs on the O157 Sakai chromosome.

Class	Number	Percent*	
Function assigned	873	53.5	
Metabolism	60	3.7	(6.9)
Transport	56	3.6	(6.4)
DNA/RNA processing	22	1.3	(2.5)
Regulation	38	2.3	(4.4)
LPS synthesis	17	1.0	(1.9)
Fimbrial synthesis	47	2.9	(5.4)
Virulence-related†	84	5.1	(9.6)
	400	24.5	(45.8)
IS-related	114	7.0	(13.1)
Unclassified	35	2.1	(4.0)
Conserved hypothetical	384	23.5	
Hypothetical‡	375	23.0	
Total	1,632	100	

* The numbers in parenthesis represent the percent of function-assigned ORFs.
† Genes for type III secretion systems are included.
‡ No database hit.

appear to have been deleted.

In addition to the four previously identified CP4 prophages, we have newly identified two phage-like elements on the MG1655 chromosome (K-12 prophage-like element; KpLE1 and 2) (Table 3 in the Supplement section). KpLE1 is integrated into the *argW* tRNA gene with a 16-bp sequence duplication. KpLE2 is apparently integrated into the *leuX* tRNA gene, but the right end region has been deleted. Since K-12 had been originally lysogenized by phage lambda, it contained a total of 11 prophage or phage-related elements, demonstrating again the predominant roles of bacteriophages in generating the genetic diversity among *E. coli* strains.

Another possible mobile genetic element is the *rhs* element.[28] O157 Sakai contains 9 *rhs* elements at 7 loci (Fig. 1 in the Supplement section). Four (*rhsA*, *C*, *D*, and *E*) are conserved in K-12 and two (*rhsF* and *G*) are the same as the elements identified in other *E. coli* strains,[29] but the remaining three are the ones newly identified in O157 Sakai (designated *rhsI*, *J*, and *K*). The significance of diversity and the physiological functions of the *rhs* elements in *E. coli* strains remain to be elucidated.

3.3. RNA genes

O157 Sakai contains seven *rrn* operons (*rrnA-H*) and their locations and directions on the chromosome are the same as those in K-12 (Table 1), though some intraspecific sequence diversities of rRNAs are present between the two strains.[9] In contrast, the compositions of tRNA genes differ remarkably between the two strains (Table 1). Of the 102 tRNA genes of O157 Sakai, 82 are conserved in K-12 but 20 are absent in K-12. Conversely, out of 86 tRNA genes identified in K-12, 4 are absent in O157 Sakai. Two leucine tRNA genes (*leuPV*), and one lysine tRNA gene (*lysQ*) in the *leuQPV* and *lysYZQ* loci are absent in O157 Sakai. In the *argX-hisR-leuT-proM* locus, the *leuT-proM* portion is duplicated in O157 Sakai (or deleted in K-12), yielding two O157 Sakai-specific genes, but the first *proM* of O157 Sakai has become a pseudogene by extensive base changes in the 3′ region. Other 18 tRNA genes that are present only in O157 Sakai reside on the genomes of 7 lambda-like phages, and are essentially the same as the *ileZ-argN-argO* genes identified in phage 933W.[30] However, three genes corresponding to *argN* have undergone extensive base changes and have become pseudogenes. As a consequence, O157 Sakai contains seven copies of *ileZ* (anticodon; CAU), four copies of *argN* (UCG), seven copies of *argO* (UCU). These three tRNA species have been proposed to recognize the Ile codon ATA (*ileZ*), a subset of the CGN family of Arg codons (*argN*), and a subset of the AGN family of Arg and Ser codons (*argO*).[30] They can, thus, recognize the five codons that are used most rarely in the genes on the conserved backbone, but used with a dramatically increased frequency in the genes on the S-loops; ATA, CGA, CGG, AGA, and AGG (Fig. 3). When the codon usage was compared between the genes on the phages carrying the tRNA genes, on the phages not carrying the genes, and on the other loops, there was essentially no difference. This suggests that these phage-encoded tRNAs may be required not only for the efficient expression of the genes on the tRNA genes-bearing phages but also for that on other phages and loops.

3.4. Protein-coding regions

Among the 5361 ORFs identified on the O157 Sakai chromosome, 3,729 are conserved in K-12 (referred to as conserved ORFs) and the remaining 1632 are the ones not present in K-12 (specific ORFs) (Table 1). Table 3 is a simplified presentation of the functions of specific ORFs. The functions of the 873 specific ORFs are predicted by sequence similarity to known proteins and 369 are similar to proteins of unknown functions; the remaining ORFs are unique to O157 Sakai. ORFs for phage-related and IS-related functions occupy large fractions of the function-assigned ORFs (45.7% and 13%, respectively), reflecting the abundance of these genetic elements on the chromosome. Genes with virulence-related functions, including those for fimbrial biosynthesis, also represent a large functional group (15%). Many genes involved in transport and metabolism were identified as well.

Among the genes on the O157 Sakai chromosome, a total of 2292 ORFs constitute 630 paralogous gene families. Out of the 630 families, 345 consisted only of the conserved ORFs (906 ORFs) and 157 of the specific ORFs (637 ORFs), whereas those consisting of both groups of ORFs comprised 128 families (445 conserved

ORFs and 304 specific ORFs). Thus, less than one-fifth of the specific ORFs have paralogs in the conserved ORFs, suggesting again that many of the specific genes were acquired by horizontal gene transfer but not by duplication of preexisting genes. This notion may be further supported by the fact that only 392 specific ORFs, including 46 transposases, have counterparts in COGs (Clusters of Orthologous Groups of proteins; http://www.ncbi.nlm.nih.gov/COG/xindex.html),[31] making a sharp contrast with the finding that 3240 of the 3729 conserved ORFs (86.9%) have counterparts in COGs.

3.5. Virulence-related genes

Adhesion to the tissue surface is the first step for bacteria to establish infection. Fimbria (also called pilus) is an important adhesion apparatus found in many Gram-negative bacteria. Some surface proteins are also used as afimbrial adhesins. On the O157 Sakai chromosome, a total of 14 loci, each encoding a set of genes for fimbrial biosynthesis, were identified. Five loci are conserved in K-12 and four are unique to O157 Sakai. The remaining five are partially conserved in K-12. Similar gene sets are present at the same loci in both strains but have undergone significant sequence changes and/or gene rearrangements (Fig. 2 in the Supplement section). The newly identified loci include a locus encoding a gene set (ECs4426–4431) similar to that for *Salmonella* long polar fimbriae,[32] but no genes for the type IV pili are present in O157 Sakai. Besides the 14 fimbrial gene clusters, at least 14 genes encode adhesin/invasin-like proteins, including two previously identified proteins: gamma-intimin on the LEE locus and the Iha adhesin.[25,33] At present, it is unknown under what circumstances these genes for fimbrial biosynthesis and adhesin-like proteins are expressed and whether they are actually involved in bacterial adherence. The presence of multiple sets of these genes, however, suggests that O157 Sakai may bind/adhere to a broader spectrum of tissues, cells, and molecules encountered in various environments than has been recognized.

It should be also noted that O157 Sakai encodes two proteins belonging to the TrcA chaperon-like protein family, ECs1825 and ECs3485. In enteropathogenic *E. coli* (EPEC), TrcA is encoded in the LIM locus and required for the normal production of the bundle-forming pili and for the full development of the "localized adherence" phenotype.[34] ECs1825 and 3485 are encoded in loci similar or almost identical to LIM, and may participate in the adherence of O157 Sakai as well. Both loci reside on the genomes of lambda-like phages, Sp9 and Sp17, respectively.

A type III secretion system encoded by the LEE locus is responsible for the formation of attaching and effacing (A/E) lesion, an essential step for establishing the EHEC infection.[33,35] We have identified a second locus that encodes a new type III secretion system, designated ETT2 (*E. coli* type three secretion 2). Although the locus resembles the *inv/spa* locus on the *Salmonella* SPI-1 pathogenicity island,[36,37] it encodes only the components of a secretion apparatus but not the secreted effector proteins. It has, however, been shown that the Inv/Spa system translocates not only the proteins encoded on the SPI-1 locus but also those encoded outside the locus.[38,39] Thus, it may be possible that effector proteins, which are translocated via the ETT2 machinery, are also encoded outside the locus. In this regard, it is noteworthy that O157 Sakai encodes several proteins with some similarities to known effector proteins, such as EspF (ECs1126 and 2715). Since the cloned LEE element failed to confer the ability to secrete Esp effector proteins and express the A/E phenotype on K-12,[40] it may be also possible that ETT2 complements the function of the LEE locus. Cloned ETT2 actually shows an ability to secrete EspB in K-12 (Tobe, T. et al., unpublished data).

O157:H7 strains are known to produce Stxs (Stx1 and/or 2) and enterohemolysin. Stxs are deeply involved in the pathogenesis of HUS, one of the life-threatening complications of EHEC infection, though the role of enterohemolysin has not been well elucidated.[41] In O157 Sakai, Stx1 and Stx2 are encoded on the genomes of two lambda-like phages and the enterohemolysin on pO157, as are in other EHEC strains.[6–8] pO157 also encodes a protein resembling LCT toxins,[6] which are the major virulence factors in *Clostridium difficile*, a causative agent of severe antibiotics-associated colitis.[42] In addition to these toxins, we have identified two genes encoding toxin-like proteins, ECs0542 and ECs1283. The former encodes an extremely large protein of 5292 amino acids, which belongs to the RTX toxin family containing repeated glycine-rich motifs (GGXXGXD) at the C-terminal domain.[43] It resides on S-loop42 together with four genes, three of which are probably responsible for its secretion (ECs0540, 0543, 0544). ECs1283 encodes a hemagglutinin/hemolysin-like protein and is followed by a gene encoding a protein belonging to the hemolysin-secretion/activation protein family.[44]

Another unexpected finding is the presence of a large number of genes that may confer an increased capability to survive in phagosome on O157 Sakai. One *lom*-like gene (*lomX*), one copper/zinc-superoxide dismutase (SOD) gene (*sodC*), and two catalase genes (*katE* and *katG*) are present in the conserved region. In addition, the strain contains 11 *lom/rck/pagC*-like genes, 2 copper/zinc-SOD genes, and 2 catalase genes that are absent in K-12. They are all encoded on the genomes of lambda-like phages except for one catalase gene (*katP*) which is carried on pO157. Two copies of *bor* and one copy of *traT* that may be able to confer the serum resistance were also identified. These findings, together with the identification of the Inv/Spa-like type III secretion system, raise a challenging hypothesis that intracellular

and/or invasive phases may exist at some stage in the O157 infection.

Iron transport systems are also important in bacterial pathogenesis, since access to this essential nutrient is severely limited in the host. *E. coli* strains possess multiple iron transport systems.[45] Most genes for iron acquisition that have been identified in K-12 are conserved in O157 Sakai, but the *fec* operon encoding a citrate-dependent iron transport system is missing. The strain, however, contains a set of genes that possibly encodes a different citrate-dependent transport system resembling that of *Synecocystis* spp.[46] Beside, the strain possesses two additional loci for iron acquisition: one is similar to the *afu* operon of *Actinobacillus pleuropneumoniae*,[47] and the other to the *shu* operon of *Shigella dysenteriae* type 1.[48,49] Furthermore, another gene cluster on S-loop83 (ECs1693-1699) may also be involved in iron transport, since the cluster contains the genes for a TonB-dependent outer membrane receptor protein and the proteins with some similarities to the components of iron transport systems. Thus, the iron acquisition ability of O157 Sakai is apparently increased as compared with K-12.

3.6. Transport and metabolism

O157 Sakai contains a number of genes related to transport and metabolic functions that are not present in K-12. Many of the transport systems belong to the ABC transporter family or the phosphoenolpyruvate:carbohydrate phosphotransferase (PTS) system; 25 and 13 ORFs are for the components of ABC transporters and PTS systems, respectively. O157 Sakai possesses operons for transport and utilization of sucrose, urease, and sorbose, but the sorbose operon is disrupted by the insertion of a Mu-like phage. The strain also possesses a glutamate-fermentation system and two aromatic acid degradation systems that are not present in K-12. One of the aromatic acid degradation systems is probably for salicylate degradation and the other is a non-oxidative decarboxylation system similar to the *vdc* system of *Streptomyces* sp.[50] Operons for a multidrug-efflux transport system and for tellurite resistance were also identified. The latter is probably responsible for a high-level resistance to tellurite, one of the phenotypes used for the laboratory isolation of O157 strains.[51] The operon is similar to the *terZ-F* operon on incHI2 and incHII plasmids,[52] and is located on an 86.2-kb CP4-like element (SpLE1) where the urease operon, *traT*, and *iha* also reside. Although substrates were not predicted in other cases, most are apparently for transport and degradation of small molecules. Functional analyses of these strain-specific genes related to transport and metabolic functions are also important because they will provide a line of information essential to develop new selective media for more efficient isolation of O157.

Relatively fewer genes for biosynthetic pathways were identified in specific ORFs than those for degradation. The *rfa* and *rfb* loci encode the genes required for the synthesis of polysaccharide moieties of lipopolysaccharide: the R3 type core and the O157-specific O antigen, respectively.[53,54] On S-loop71 and 225, we have identified two loci that probably encode specialized or modified systems for fatty acid biosynthesis. A locus on S-loop225 encodes a methyltransferase, an acyltransferase and a set of proteins required for elongation cycles in fatty acid synthesis:[55] two acyl carrier proteins (ACPs), a beta-hydroxyacyl-ACP dehydratase, a beta-ketoacy-ACP synthetase (KAS) II, a beta-hydroxydecanoyl-ACP dehydratase, a beta-ketoacyl-ACP reductase, and a protein consisting of two fused KAS II molecules. A locus on S-loop71 also encodes a holo ACP synthetase, a beta-ketoacyl-ACP reductase, a beta-hydroxyacyl-ACP dehydratase, an ACP, an aminomethyl transferase, and a KAS I. Both loci probably participate in the synthesis of fatty acid-containing molecules such as lipids, liposaccharides, or lipoproteins that are specifically produced by O157 Sakai but not by K-12. It is of interest that the second locus are located just downstream of the genes for the hemolysin-like and hemolysin-secretion/activation proteins (ECs1283 and 1284), constituting an operon-like structure. It might be possible that the locus participates in acylation and activation of the hemolysin-like protein as has been demonstrated for the alpha-hemolysin of uropathogenic *E. coli*.[56]

3.7. K-12 genes missing in O157 Sakai

As compared with K-12, O157 Sakai lacks a total of 567 ORFs, including the above mentioned *fec* operon and the genes for utilization of 2-phenylethylamine, xanthosine, D-galactonate, L-idonate, glycolate, and short chain fatty acids. The inability to rapidly ferment D-sorbitol and the lack of beta-glucronidase (GUD) activity are the phenotypes that serve as markers for the laboratory identification of O157:H7.[57,58] In typical *E. coli* strains, utilization of D-sorbitol occurs through a pathway initiated by a specific PTS system encoded by the *gut* operon.[59] In O157 Sakai, both the *gutA* and *gutE* genes encoding the sorbitol-specific PTS enzymes IIC and IIB have authentic frameshifts, and thus sorbitol transport is impaired. The slow sorbitol fermentation may be explained by the finding that the PTS system for D-mannitol can act on D-sorbitol at a low affinity.[59] The *uidA* gene encoding GUD is also disrupted in O157 Sakai by a two-base insertion at nucleotide position 690. This mutation is different from that reported for the *uidA* gene of other O157:H7 strains.[60]

Genes for the general secretion pathway (GSP) identified on the K-12 chromosome do not exist on the O157 Sakai chromosome, but a different set of GSP genes reside on pOI57 instead.[6] All the K-12 genes for DNA restric-

tion and modification (*hsdSMR*, *mrr*, *mcrA*, and *mcrBC*) are not present in O157 Sakai. In the place of *hsdSMR* and *mrr*, a type I system almost identical to the *EcoA* system was identified. Thus, the two strains have completely different sets of DNA restriction/modification systems.

Among the list of two-component regulatory systems identified in K-12,[61] only two systems are missing in O157 Sakai; those encoded by *atoSC* and *ygiYX*. The former is deleted together with other *ato* genes while *ygiY* is split into two parts by the introduction of a premature stop codon. This high level of conservation of the signal transduction systems between the two strains implies that the expression of most strain-specific genes are under the control of global regulatory networks encoded on the conserved backbone, though many of the strain-specific genes are the ones horizontally acquired. This notion gains some support by the finding that most of the K-12 genes involved in global regulatory functions, including all the sigma factors, are encoded on the backbone.[62] Indeed, 38 of the O157 Sakai-specific regulators include no sigma factor, and we could identify only two systems as the O157 Sakai-specific two component regulatory system (ECs0417/0418 and ECs5067/5074).

3.8. Conclusions

Genomic comparison of O157 Sakai with K-12 provided a large amount of information with biological and medical importance. The presence of a well-conserved 4.1-Mb sequence that can be regarded as the chromosome backbone of *E. coli* and numerous strain-specific DNA segments of foreign origins ("strain-specific loops") indicate how the two strains have diversified from a common ancestral lineage. There is no doubt that bacteriophages have played the predominant roles in this process. This mode of diversification probably represents a general pathway of the intraspecific evolution of most *E. coli* strains, though the final verification has to wait until the third or more genomes of the strains belonging to the other lineages are analyzed. Identification of a complete set of genes that are specifically present in O157 Sakai will shed new insights into the pathogenicity and the physiology of O157, and open a way to fully understand the molecular mechanisms underlying the O157 infection and to develop new strategies for prevention, treatment, and surveillance of the infection.

Supplementary information is available in the Supplement section of this issue. It is also available at DNA Research Online [http://www.dna-res.kazusa.or.jp/] and on the authors' World-Wide Web site [http://genome.gen-info.osaka-u.ac.jp/bacteria/o157/]. The chromosome sequence has been deposited DDBJ/EMBL/GenBank under the accession numbers BA000007 and AP002550–AP002569.

Acknowledgements: We thank H. Inokuchi, M. Tsuda, and Y. Nagata for their valuable suggestions, and K. Ohshima and R. Fukawa (Hitachi Instruments Service Co. Ltd.) for their expert technical assistance in sequence determination, A. Ohyama, T. Kozuki, and K. Doga (Mitsui Knowledge Inc. Ltd.) for their assistance in operating the Gambler program. We also thank M. Takahashi, S. Setsu, H. Wakimoto, K. Satoh, K. Hagiwara, Y. Kubota, C. H. Yutsudo, S. Kimura, Y. Sagan, A. Ohmoto, and A. Yoshida for their technical assistance, Sakai City Institute for Public Health for providing the strain used in this work, and Y. Hayashi for her language assistance. We would like to express our special thanks to H. Yoshikawa, Y. Terawaki, T. Nakazawa, and H. Hayashi for their support in initiating and organizing the genome project as well as their encouragement throughout the project. This project was supported by The Japan Society for the Promotion of Science "Research for the Future Program," 97L00101 and JSPS-RETF 00L01411.

References

1. Riley, L. W., Remis, R. S., Helgerson, D. et al. 1983, Hemorrhagic colitis associated with a rare *Escherichia coli* serotype, *N. Engl. J. Med.*, **308**, 681–685.
2. Cohen, M. B. and Giannella, R. A. 1992, Hemorrhagic colitis associated with *Escherichia coli* O157:H7, *Adv. Intern. Med.*, **37**, 173–195.
3. Watanabe, H., Wada, A., Inagaki, Y., Itoh, K., and Tamura, K. 1996, Outbreaks of enterohaemorrhagic *Escherichia coli* O157:H7 infection by two different genotype strains in Japan, *Lancet*, **348**, 831–832.
4. Nataro, J. R. and Kaper, J. B. 1998, Diarrheagenic *Escherichia coli*, *Clin. Microbiol. Rev.*, **11**, 1–60.
5. Blattner, F. R., Plunkett III, G., Bloch, C. A. et al. 1997, The complete genome sequence of *Escherichia coli*, *Science*, **277**, 1453–1462.
6. Makino, K., Ishii, K., Yasunaga, T. et al. 1998, Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak, *DNA Res.*, **5**, 1–9.
7. Makino, K., Yokoyama, K., Kubota, Y. et al. 1999, Complete nucleotide sequence of the defective prophage VT2-Sakai carrying the verotoxin2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak, *Genes Genet. Syst.*, **74**, 227–239.
8. Yokoyama, K., Makino, K., Kubota, Y. et al. 2000, Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak, *Gene*, **258**, 127–139.
9. Ohnishi, M., Murata, T., Nakayama, K. et al. 2000, Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic *Escherichia coli* O157:H7 Sakai strain and an *Escherichia coli* K-12 strain MG1655, *Syst. Appl. Microbiol.*, **23**, 315–324.
10. Ohnishi, M., Tanaka, C., Kuhara, S. et al. 1999, 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 Res.*, **6**, 361–368.
11. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998, Base-calling of automated sequencer traces using phred. I. Accuracy assessment, *Genome Res.*, **8**, 175–185.
12. Gordon, D., Abajian, C., and Green, P. 1998, Consed: A graphical tool for sequence finishing, *Genome Res.*, **8**, 195–202.
13. Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O., and Salzberg, S. L. 1999, Alignment of whole genomes, *Nucleic Acids Res.*, **27**, 2369–2376.
14. Sakiyama, T., Takami, H., Ogasawara, N. et al. 2000, An automated system for genome analysis to support microbial whole genome shotgun sequencing, *Biosci. Biotechnol. Biochem.*, **64**, 670–673.
15. Salzberg, S. L., Delcher, A. L., Kasif, S., and White, O. 1998, Microbial gene identification using interpolated Markov models, *Nucleic Acids Res.*, **26**, 544–548.
16. Altschul, S. F., Madden, T. L., Schaffer, A. A. et al. 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, **25**, 3389–3402.
17. Lowe, T. M. and Eddy, S. R. 1997, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.*, **25**, 955–964.
18. Pupo, G. M., Karaolis, D. K. R., Lan, R., and Reeves, P. R. 1997, Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies, *Infect. Immun.*, **65**, 2685–2692.
19. Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K., and Whittam, T. S. 2000, Parallel evolution of virulence in pathogenic *Escherichia coli*, *Nature*, **406**, 64–67.
20. Bergthorsson, U. and Ochman, H. 1998, Distribution of chromosome length variation in natural isolates of *Escherichia coli*, *Mol. Biol. Evol.*, **15**, 6–16.
21. Fraser, C. M., Casjens, S., Huang, W. M. et al. 1997, Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*, *Nature*, **390**, 580–586.
22. Casjens, S., Palmer, N., van Vugt, R. et al. 2000, A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*, *Mol. Microbiol.*, **35**, 490–516.
23. Fraser, C. M., Gocayne, J. D., White, O. et al. 1995, The minimal gene complement of *Mycoplasma genitalium*, *Science*, **270**, 397–403.
24. Campbell, A. M. 1996, In: Neidhardt, F. C., Curtiss III, R., Ingraham, J. L. et al. (eds) *Escherichia coli* and *Salmonella*, Cellular and Molecular Biology, 2nd Ed., ASM press, Washington DC, pp. 2041–2046.
25. Tarr, P. I., Bilge, S. S., Vary, J. C. Jr. et al. 2000, Iha: a novel O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure, *Infect. Immun.*, **68**, 1400–1407.
26. Cheetham, B. F. and Katz, M. E. 1995, A role for bacteriophages in the evolution and transfer of bacterial virulence determinants, *Mol. Microbiol.*, **18**, 201–208.
27. Hacker, J., Blum-Oehler, G., Möhldorfer, I., and Tschäpe, H. 1997, Pathogenicity islands of virulent bacteria: structure, function, and impact on microbial evolution, *Mol. Microbiol.*, **23**, 1089–1097.
28. Bacheller, S., Gilson, E., Hofnung, M., and Hill, C. W. 1996, In: Neidhardt, F. C., Curtiss III, R., Ingraham, J. L. et al. (eds) *Escherichia coli* and *Salmonella*, Cellular and Molecular Biology, 2nd Ed., ASM press, Washington DC, pp. 2012–2040.
29. Wang, Y-D., Zhao, S., and Hill, C. W. 1998, Rhs elements comprise three subfamilies which diverged prior to acquisition by *Escherichia coli.*, *J. Bacteriol.*, **180**, 4102–4110.
30. Plunkett III, G., Rose, D. J., Durfee, T. J., and Blattner, F. R. 1999, Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product, *J. Bacteriol.*, **181**, 1767–1778.
31. Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. 2000, The COG database: a tool for genome-scale analysis of protein functions and evolution, *Nucleic Acids Res.*, **28**, 33–36.
32. Baumler, A. J. and Heffron, F. 1995, Identification and sequence analysis of *lpfABCDE*, a putative fimbrial operon of *Salmonella typhimurium*, *J. Bacteriol.*, **177**, 2087–2097.
33. Jerse, A. E., Gicuelais, K. G., and Kaper, J. B. 1991, Plasmid and chromosomal elements involved in the pathogenesis of attaching and effacing *Eschelochia coli*, *Infect. Immun.*, **59**, 3869–3875.
34. Tobe, T., Tatsuno, I., Katayama, E., Wu, C-Y., Schoolnik, G. K., and Sasakawa, C. 1999, A novel chromosomal locus of enteropathogenic *Escherichia coli* (EPEC), which encodes a bfpT-regulated chaperone-like protein, TrcA, involved in microcolony formation by EPEC, *Mol. Microbiol.*, **33**, 741–752.
35. Perna, N. T., Mayhew, G. F., Posfai, G., Elliott, S., Donnenberg, M. S., Kaper, J. B., and Blattner, F. R. 1998, Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7, *Infect. Immun.*, **66**, 3810–3817.
36. Collazo, C. M. and Galan, J. E. 1997, The invasion-associated type-III protein secretion system in *Salmonella*, *Gene*, **192**, 51–59.
37. Groisman, E. A. and Ochman, H. 2000, The path to *Salmonella*, ASM News, **66**, 21–27.
38. Hardt, W. D., Urlaub, H., and Galan, J. E. 1998, A substrate of the centisome 63 type III protein secretion system of *Salmonella typhimurium* is encoded by a cryptic bacteriophage, *Proc. Natl. Acad. Sci. USA*, **95**, 2574–2579.
39. Norris, F. A., Wilson, M. P., Wallis, T. S, Galyov, E. E., and Majerus, P. W. 1998, SopB, a protein required for virulence of *Salmonella dublin*, is an inositol phosphate phosphatase, *Proc. Natl. Acad. Sci. USA*, **95**, 14057–14059.
40. Elliott, S. J., Yu, J., and Kaper, J. B. 1999, The cloned locus of enterocyte effacement from enterohemorrhagic *Escherichia coli* O157:H7 is unable to confer the attaching and effacing phenotype upon *E. coli* K-12, *Infect. Immun.*, **67**, 4260–4263.
41. Schmidt, H., Beutin, L., and Karch, H. 1995, Molecular analysis of the plasmid-encoded hemolysin of *Escherichia*

coli O157:H7 strain EDL 933, *Infect. immun.*, **63**, 1055–1061.
42. von Eichel-Streiber, C., Boquet, P., Sauerborn, M., and Thelestam, M. 1996, Large clostridial cytotoxins—a family of glycosyltransferases modifying small GTP-binding proteins, *Trends Microbiol.*, **4**, 375–382.
43. Felmlee, T. and Welch, R. A. 1988, Alterations of amino acid repeats in the *Escherichia coli* hemolysin affect cytolytic activity and secretion, *Proc. Natl. Acad. Sci. USA*, **85**, 5269–5273.
44. Braun, V., Schonherr, R., and Hobbie, S. 1993, Enterobacterial hemolysins: activation, secretion and pore formation, *Trends Microbiol.*, **1**, 211–216.
45. Earhart, C. F. 1996, In: Neidhardt, F. C., Curtiss III, R., Ingraham, J. L. et al. (eds) *Escherichia coli* and *Salmonella*, Cellular and Molecular Biology, 2nd Ed., ASM press, Washington DC, pp. 1075–1090.
46. Kaneko, T., Sato, S., Kotani, H. et al. 1996, Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions, *DNA Res.*, **3**, 109–136.
47. Chin, N., Frey, J., Chang, C. F., and Chang, Y. F. 1996, Identification of a locus involved in the utilization of iron by *Actinobacillus pleuropneumoniae*, *FEMS Microbiol. Lett.*, **143**, 1–6.
48. Wyckoff, E. E., Duncan, D., Torres, A. G., Mills, M., Maase, K., and Payne, S. M. 1998, Structure of the *Shigella dysenteriae* haem transport locus and its phylogenetic distribution in enteric bacteria, *Mol. Microbiol.*, **28**, 1139–1152.
49. Torres, A. G. and Payne, S. M. 1997, Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7, *Mol. Microbiol.*, **23**, 825–833.
50. Chow, K. T., Pope, M. K., and Davies, J. 1999, Characterization of a vanillic acid non-oxidative decarboxylation gene cluster from *Streptomyces* sp. D7, *Microbiology*, **145**, 2393–2403.
51. Zadik, P. M., Chapman, P. A., and Siddons, C. A. 1993, Use of tellurite for the selection of verocytotoxigenic *Escherichia coli* O157, *J. Med. Microbiol.*, **39**, 155–158.
52. Taylor, D. E. 1999, Bacterial tellurite resistance, *Trends Microbiol.*, **7**, 111–115.
53. Amor, K., Heinrichs, D. E., Frirdich, E., Ziebell, K., Johnson, R. P., and Whitfield, C. 2000, Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*, *Infect. Immun.*, **68**, 1116–1124.
54. Wang, L. and Reeves, P. R. 1998, Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes, *Infect. Immun.*, **66**, 3545–3551.
55. Cronan, J. E. and Rock, C. O. 1996, In: Neidhardt, F. C., Curtiss III, R., *Ingraham*, J. L. et al. (eds) *Escherichia coli* and *Salmonella*, Cellular and Molecular Biology, 2nd Ed., ASM press, Washington DC, pp. 612–636.
56. Issartel, J. P., Koronakis, V., and Hughes, C. 1991, Activation of *Escherichia coli* prohaemolysin to the mature toxin by acyl carrier protein-dependent fatty acylation, *Nature*, **351**, 759–761.
57. March, S. B. and Ratnam, S. 1986, Sorbitol-MacConky medium for detection of *Escherichia coli* O157:H7 associated with hemorrhagic colitis, *J. Clin. Microbiol.*, **23**, 869–872.
58. Thompson, J. S., Hodge, D. S., and Borczyk, A. A. 1990, A rapid biochemical test to identify verocytotoxin-positive strains of *Escherichia coli* O157, *J. Clin. Microbiol.*, **28**, 2165–2168.
59. Lin, E. C. C. 1996, In: Neidhardt, F. C., Curtiss III, R., Ingraham, J. L. et al. (eds) *Escherichia coli* and *Salmonella*, Cellular and Molecular Biology, 2nd Ed., ASM press, Washington DC, pp. 307–342.
60. Feng, P. and Lampel, K. A. 1994, Genetic analysis of *uidA* expression in enterohaemorrhagic *Escherichia coli* serotype O157:H7, *Microbiology*, **140**, 2101–2107.
61. Mizuno, T. 1997, Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*, *DNA Res.*, **4**, 161–168.
62. Riley, M. and Labedan, B. 1996, In: Neidhardt, F. C., Curtiss III, R., Ingraham, J. L. et al. (eds) *Escherichia coli* and *Salmonella*, Cellular and Molecular Biology, 2nd Ed., ASM press, Washington DC, pp. 2118–2202.