Complexity of the Genomic Diversity in Enterohemorrhagic
Escherichia coli O157 Revealed by the Combinational Use of
the O157 Sakai OligoDNA Microarray and the Whole Genome
PCR scanning

Yoshitoshi Ogura,1,2 Ken Kurokawa,3 Tadasuke Ooka,2 KouSuke Tashiro,4 Toru Tobe,5
Makoto Ohnishi,6 Keisuke Nakayama,2 Takuya Morimoto,1 Jun Terajima,6 Hafuo Watanabe,6
Satoru Kühara,4 and Tetsuya Hayashi1,2,*

Division of Bioenvironmental Science, Frontier Science Research Center, University of Miyazaki,
Miyazaki, Japan1, Division of Microbiology, Department of Infectious Diseases, Faculty of Medicine,
University of Miyazaki, Miyazaki, Japan2, Laboratory of Comparative Genomics, Graduate School of
Information Science, Nara Institute of Science and Technology, Nara, Japan3, Laboratory of Molecular
Gene Technology, Department of Genetic Resources Technology, Faculty of Agriculture, Kyushu University,
Fukuoka, Japan4, Division of Applied Bacteriology, Graduate School of Medicine, Osaka University,
Osaka, Japan5, and Department of Bacteriology, National Institute for Infectious Diseases, Tokyo, Japan6

(Received 29 November 2005; revised 21 December 2005; published online 22 February 2006)

Abstract

Escherichia coli O157, an etiological agent of hemorrhagic colitis and hemolytic uremic syndrome, is one of
the leading worldwide public health threats. Genome sequencing of two O157 strains have revealed that the
chromosome is comprised of a 4.1 Mb backbone shared by K-12 and a total of 1.4 Mb O157-specific sequences.
Most of the large O157-specific sequences are prophages and prophage-like elements, which have carried
many virulence genes into the O157 genome. This suggests that bacteriophages have played the key roles in
the emergence of O157. The Whole Genome PCR Scanning (WGPScanning) analysis of O157 strains, on the
other hand, revealed a high level of genomic diversity in O157. Variation of prophages has also been suggested
as a major factor generating such diversity. In this study, we analyzed the gene content of O157 strains, by an
oligoDNA microarray, using the same set of strains as examined by the WGPScanning method. Although
most of the strains were typical O157 : H7, they differed remarkably in gene composition, particularly in those
on prophages, and we identified more than 400 ‘variably absent or present’ genes which included virulence-
related genes. This confirms the role of prophages in generating the genomic diversity, and raises a possibility
that some level of variation in potential virulence is present among O157 strains. Fine comparison of the two
data sets obtained by microarray and WGPScanning provided much further details on the O157 genome
diversity than illustrated by each method alone, indicating the usefulness of this combinational approach in
the genomic comparison of closely related strains.

Key words: E. coli O157; genomic diversity; microarray; whole genome PCR scanning

1. Introduction

Escherichia coli is a constituent of normal microflora
in intestinal tracts, but certain types of E. coli strains are
associated with diseases in human and animals.1,2 Among
these pathogenic E. coli strains, enterohemorrhagic E. coli
O157 not only causes large outbreaks of hemorrhagic colitis
but also numerous small outbreaks and sporadic cases,
as is regarded as one of the major worldwide public health
concerns.3 The genome sequence determination


The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access
version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University
Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its
entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org
of an O157 strain, RIMD 0509952 (referred to as O157 Sakai) and the genomic comparison with a nonpathogenic strain, K-12 MG1655, have revealed that the O157 chromosome is comprised of a 4.1 Mb backbone common with K-12 and a total of 1.4 Mb strain-specific sequences. O157 Sakai-specific sequences (referred to as S-loops) are inserted at various sites on the chromosome backbone, and encode more than 1600 of O157 Sakai-specific genes. Of importance is that most of the large S-loops are prophages and prophage-like elements. We have identified 18 prophages (Sp1-18) and 6 prophage-like elements (SpLE1-6). They comprise about two thirds of the O157 Sakai-specific sequences and have carried many virulence-related genes including Shiga toxin (Stx) genes (stx1 and stx2) into O157. Similar findings were also obtained by the genome sequencing analysis of another O157 strain, EDL933.

More recently, we analyzed, based on the genome sequence of O157 Sakai, the whole genome structures of eight O157 strains that displayed diverse virulence-related genes including Shiga toxin (Stx) genes (stx1 and stx2) into O157. This suggested that acquisition of these bacteriophage genes have played the key roles in the emergence of O157. Similar findings were also obtained by the genome sequencing analysis of another O157 strain, EDL933.

2.2. Design of the O157 Sakai microarray

Oligonucleotide probes were prepared for all the protein-coding genes on the O157 Sakai genome (5447 genes in total). Principles for the probe design were 60 mer in length and two probes for each gene. However, O157 Sakai contains 542 repeated genes, most of which are derived from IS elements or 13 lambda-like prophages sharing various length of DNA segments with almost identical sequences. As for these repeated genes, we prepared a single probe for each repeated gene family. Such probes totaled to 151 probes, each having two or more targets of ≥90% sequence identity on the O157 Sakai genome (referred to as ‘multiple hit probes’). As for the singleton genes (4905 genes), we were able to design two different 60 mer probes for 4173 genes, but only one 60 mer probe for 725 genes and one 45 mer probes for 7 genes. In total, 9620 probes were synthesized (Sigma-Aldrich Japan, Tokyo, Japan), and each probe was spotted onto poly-L-Lysine-coated glass slides (SD10011; Matsunami Glass Ind., Osaka, Japan) using SPBIO (Hitachi Software Engineering, Tokyo, Japan). As the negative controls, we included 20 oligonucleotides prepared from 10 yeast genes. Probe sequences are shown in Supplementary Table 1 available at www.dnareseach.oxfordjournals.org.

2.3. DNA labeling and hybridization

Genomic DNAs from test strains and the reference (O157 Sakai) were fluorescently labeled with Cy3 and Cy5, respectively, according to the following protocol. First, 3 μg of genomic DNA was labeled with aminoallyl-modified dUTP (Sigma) using the Bioprime DNA Labeling System (Invitrogen). DNAs were not sheared or digested by restriction enzymes prior to the

2. Materials and Methods

2.1. Bacterial strains, growth condition and DNA extraction

Eight O157 strains examined in this study were 980938 (referred to as #2), 980706 (#3), 990281 (#4), 980551 (#5), 990570 (#6), 981456 (#7), 982243 (#8) and 981795 (#9). They were human isolates obtained in Japan in 1998 as described previously. All were negative for Sorbitol fermentation (SOR−) and β-glucuronidase activity (GUD−) whose activities were examined after 24 h incubation at 37°C on SIB II agar plates (Kyokuto Pharmaceutical Industrial, Tokyo, Japan) and on ES Colimark agar plate (Eiken Kizai, Tokyo, Japan), respectively. As for strain #2, however, a weak positive reaction for β-glucuronidase was detected after 40 h incubation. O157 Sakai (RIMD 0509952) was used as the reference in all hybridization experiments. This strain was isolated in a large outbreak in Sakai city, Japan, which occurred in 1996, and has been sequenced by our group. The strain is available at the American Type Culture Collection (ATCC BAA-460). Escherichia coli K-12 MG1655 was kindly provided by Dr H. Mori (Nara Institute of Science and Technology). Cells were grown to the stationary phase at 37°C in Luria–Bertani medium. Genomic DNA was purified using the Genomic-tip 100/G and Genomic DNA buffer set (Qiagen) according to the manufacturer’s instruction.

In this study, we constructed an oligoDNA microarray according to the genome sequence of O157 Sakai. We first validated the accuracy of the comparative genomic hybridization (CGH) analysis using this oligoarray by a test experiment with K-12 and O157 Sakai, and then performed a gene composition analysis of O157 using the same set of strains examined by the WGPScanning. We further made a fine comparison of the microarray data with WGPScanning data to understand the genomic diversity of O157 more in detail.
labeled. The aminooallyl-labeled DNA was purified by Microcon YM-30 (Millipore), dried in a speed-vac and resuspended in 10 μl of 50 mM NaHCO₃. After adding 10 μl of dimethyl sulfoxide solution containing Cy3 or Cy5 monofunctional reactive dye (Amersham), the sample was incubated at room temperature in the dark for 1 h to allow the dye to couple with DNA. The fluorescently labeled DNA was finally purified by the Qiagix PCR purification kit (Qiagen) into 30 μl elution buffer provided by the manufacturer.

The Cy5-labeled and the Cy3-labeled DNA preparations (15 μl for each) were mixed with a 90 μl hybridization solution containing 6.4× SSC, 0.64% SDS and 1.3 mg/ml yeast tRNA. After incubation at 96°C for 2 min, the denatured sample was applied to a microarray slide and incubated on an ArrayBooster hybridization apparatus (Advalytix AG, Brunthal, Germany) at 50°C for 16 h. The slide was then washed twice in 2× SSC/0.2% SDS at room temperature for 10 min, twice in 0.2× SSC/0.2% SDS at 50°C for 5 min and twice in 0.2× SSC at room temperature for 10 min. Finally, the slide was briefly rinsed with ethanol, dried by centrifugation and scanned with a FLA-8000 scanner (Fuji Photofilm, Tokyo, Japan). The obtained data were analyzed by the ArrayVision 8.0 software (Imaging Research). Each strain was examined twice with the labeled DNAs prepared independently for each hybridization.

2.4. Data analysis

In our experimental condition, signal intensities from the negative controls were almost the same as that of the local background (LBG) in both channels. Thus, spots with reference signal intensities lower than the LBG plus 5 SD or with some spotting abnormalities were categorized as ‘low quality’ (LQ). Signal intensities of other spots were corrected by subtracting the LBG, were categorized as ‘low quality’ (LQ). Signal intensities from the two probes. In such cases, the gene was categorized as ‘uncertain’ (U), P or A, respectively.

Finally, the presence or absence of each gene was determined according to the data from each spot. Because most genes were represented by two different probes on our microarray, inconsistent results could also be obtained from the two probes. In such cases, the gene was categorized as P when the judgments of two probes were P/U or P/LQ, as A when they were A/U or A/LQ, and as U when they were P/A or U/LQ. Processed data sets were displayed in the genomic order using the TREEVIEW program.¹⁰ Raw signal intensities from each spot and processed data are presented in Supplementary Tables 1 and 2 available at www.dnares.oxfordjournals.org, respectively.

The presence or absence of 37 genes, which were always categorized into LQ on the current version of O157 Sakai microarray, were determined by PCR. The PCR amplification was performed using the Ex Taq PCR kit (TAKARA Bio, Japan) and 10 ng of template DNA with 30 cycles of 20 s at 98°C/30 s at 60°C/30 (or 120) s at 72°C. The list of genes analyzed by PCR and the primer sequences are shown in Supplementary Table 3 available at www.dnares.oxfordjournals.org.

3. Results

3.1. Evaluation of the sensitivity and specificity of CGH analysis using the O157 Sakai oligoDNA microarray

The array we constructed covered all of the 5447 protein-coding genes on the O157 Sakai genome; 3729 on the chromosome backbone, 1632 on S-loops, and 86 on pO157 and pOSAK1 plasmids.¹¹ Before applying the microarray to the CGH analysis of O157 strains, we evaluated its sensitivity and specificity using K-12 MG1655 as a test strain.

We first examined the sequence identities of each probe to the K-12 genome sequence by the FASTA program, and classified the probes into two groups; those with ≥90% identity into ‘Conserved in K-12’ (CK) probes and others into ‘Specific to Sakai’ (SS) probes. Each gene was then classified into CK, ‘Partially Conserved in K-12’ (PCK; genes represented by a CK probe and an SS probe), or SS genes according to the probe category (Supplementary Table 1). Next, the data of the microarray analysis were compared with those from the in silico analysis. In this comparison, repeated gene families, represented by ‘multiple hit probes’ having two or more targets of ≥90% sequence identity on the O157 Sakai genome, were analyzed separately from the singleton genes.

As for the singleton genes, 96.9% of CK genes were correctly judged as ‘present’, and 97.8% of SS genes as ‘absent’ (Table 1). Most of the K-12 genes that gave incorrect results (false negative, false positive and uncertain) contained slightly divergent target sequences with 3–6 base mismatches with CK probes or weak homologies to the SS probes (70–90% identity). On the other hand, the presence/absence determination of the repeated gene families conserved in K-12 was somewhat problematic. While all families with the copy number ratio of >0.5 were judged as ‘present’, many families with the copy number ratios of ≤0.5 were judged as ‘absent’ or ‘uncertain’. This indicates that the repeated
genes judged as ‘absent’ by the microarray analysis include those actually absent and those with reduced copy numbers. Thus, in the following CGH analyses of O157 strains, we decided to categorize such repeated genes into ‘uncertain due to the multiple target sequences in O157 Sakai’ [U(M)] and to consider only the repeated genes judged as ‘present’ as having the same or similar (or higher) copy numbers in the test strain.

3.2. Overview of the gene content analysis of O157 strains

Using the O157 Sakai microarray, we analyzed the gene content of eight O157 strains that were previously
analyzed by the WGPScanning method. The data are summarized in Fig. 1 and Tables 2 and 3. As shown in Table 2, all the singleton genes identified in O157 Sakai were shared by at least one strain, but as many as 431 genes displayed variable distributions (or highly divergent sequences). Most of these ‘variably absent or present genes’ (referred to as VAP genes) belonged to the SS genes; 389 of the 1153 SS singleton genes were variably present. As shown in Fig. 1, these genes appeared very frequently on prophages and prophage-like elements; among the 640 singleton genes on these genetic elements, 350 (54.7%) displayed variable distributions. In sharp contrast, CK singleton genes exhibited a high level of conservation; only 1% (37 genes) exhibited variable distributions (Table 2). Among these, 20 genes were again located on prophage regions shared by O157 and K-12. In particular, a region of Sp10 (corresponding to a part of the Rac prophage in K-12) contained 11 VAP genes. These included recE and recT genes (ECs1933 and ECs1934) involved in the RecA-independent recombination and double-strand break repair.

Among the eight strains, strain #2 lacked the largest number of genes (at least 307 genes), and the number of missing singleton genes reduced in the following order; #7, #6, #4, #8, #3, #9 and #5 (Table 3). This pattern well correlated with the level of structural diversification observed in the WGPScanning analysis (Fig. 1, and also refer to Ref. 23). Most of the repeated gene families identified in O157 Sakai were also conserved (present in the same or similar copy number) in strains #3, #5, #8 and #9. These data suggest that the four strains are more closely related to O157 Sakai in terms of gene content and genome structure.

3.3. O157 Sakai-specific genes with variable distributions in O157

Classification of the SS genes according to their functions revealed that genes belonging to specific categories contained more VAP genes compared to other categories (Table 4). The abundance of VAP genes in ‘transcription’, ‘replication, recombination, and repair’, ‘genes with unknown or uncharacterized functions’ and ‘unclassified genes’ was remarkable. This appears to be a reflection of the fact that most VAP genes are on prophages or prophage-like elements, which contain many genes for transcriptional regulation, replication, repair and phage-specific functions as well as a number of uncharacterized genes.

A notable finding was that 16 virulence-related genes exhibited variable distributions (Table 5). They included genes for Stx1, an HmwA-like protein, a TraT-homologue, a HeCBlike protein. HmwA is known to be involved in adhesion in Haemophilus influenzae, TraT in serum resistance in enteric bacteria and HeCB in hemolysin activation in Neisseria meningitides. One strain lacked the efa-I gene which is involved in the adherence of non-O157 EHEC to cultured epithelial cells and in the inhibition of lymphocyte proliferation and pro-inflammatory cytokine synthesis in EPEC (called lifA in EPEC). In O157, the efa-I gene is split into two genes (ECs3860 and 3861), but it has been reported that disruption of ECs3860 (efa-I') results in reduced expression and secretion of LEE (the locus of enterocyte effacement)-encoded proteins and in reduced adherence to cultured cells. The pchD and pchE genes encoding PerC-like regulators were also missing in four strains and one strain, respectively. Since the other pch genes (pchA, pchB and pchC) have been shown to regulate the LEE gene expression, pchD and pchE may be involved in the regulation of virulence-related genes as well.

More importantly, several genes encoding type III secretion system (TTSS) effectors exhibited variable distributions. Recently, several proteins encoded on the non-LEE loci have been identified as effectors secreted by the LEE-encoded TTSS (non-LEE encoded effectors) in EPEC, EHEC and Citrobacter rodentium. O157 Sakai contains at least 19 non-LEE encoded effectors or their homologues. Among these, nleA (also named espI, ECs1812), an nleF-homologue (ECs1815), an nleG-homologue (ECs1828) and an nleH-homologue (ECs1814) were variably present among the eight O157 strains. Functions of these effectors have not been well elucidated, but it has been shown that NleA/EspI injected into the host cell localizes to the Golgi apparatus.

Many virulence-related genes were also encoded on the pO157 plasmid, but all were conserved except for katP (pO157,76). Of interest was that 18 out of the 83 genes on pO157 were missing in Strain #2 (Fig. 1). Since the pO157 plasmid of #2 is almost the same in size as those from other strains (data not shown), several parts of the plasmid have probably been replaced by different or highly divergent sequences. pOSAK1, a small cryptic plasmid, was present only in two strains.

3.4. Comparison of two datasets obtained by microarray and WGPScanning analyses

The CGH analysis using microarray provides the information on the gene composition but not their genomic positions, thus translocation events specific to the test strain cannot be detected. It also provides no information on strain-specific insertions. Conversely, WGPScanning detects possible loci where such genetic events have taken place while the presence or absence of each gene cannot be determined. Therefore, in order to know more details on the genomic diversity of O157, we compared the two datasets, one from the present microarray analysis and the other from the previous WGPScanning (Fig. 1).

In the segments that exhibited no size variation in the WGPScanning analysis (indicated in gray in Fig. 1), no
8 Complexity of the Genomic Diversity in *E. coli* O157 [Vol. 13,
deletion of genes was detected by microarray except for only five segments. This confirmed that these regions possess the same genomic structures as those in Sakai. Among the five segments where deletions of a gene(s) were observed, four were derived from lambda-like prophage regions, implying that some parts of these prophages have been replaced by DNA fragments with same sizes but with different or highly divergent sequences.

By the WGPScanning analysis, a total of 35 segments with larger size reduction (≥2 kb smaller than that of O157 Sakai) were identified from the 8 O157 strains (indicated in yellow in Fig. 1). In most cases (29 segments), deleted genes were identified. The regions with gene deletion, however, were often associated with repeated genes derived from IS elements or lambda-like prophages, and thus precise boundaries of the deletion events were not defined in many cases. In six segments,
Table 4. Categorization and conservation of SS genes in the eight O157 strains

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conserved</td>
</tr>
<tr>
<td>On the chromosome</td>
<td></td>
</tr>
<tr>
<td>Energy production and conversion</td>
<td>6</td>
</tr>
<tr>
<td>Amino acid transport and metabolism</td>
<td>4</td>
</tr>
<tr>
<td>Carbohydrate transport and metabolism</td>
<td>37</td>
</tr>
<tr>
<td>Coenzyme transport and metabolism</td>
<td>5</td>
</tr>
<tr>
<td>Lipid transport and metabolism</td>
<td>15</td>
</tr>
<tr>
<td>Translation</td>
<td>1</td>
</tr>
<tr>
<td>Transcription</td>
<td>30</td>
</tr>
<tr>
<td>Replication, recombination and repair</td>
<td>38</td>
</tr>
<tr>
<td>Cell wall/membrane biogenesis</td>
<td>32</td>
</tr>
<tr>
<td>Cell motility</td>
<td>45</td>
</tr>
<tr>
<td>Post-translational modification, protein turnover, chaperones</td>
<td>3</td>
</tr>
<tr>
<td>Inorganic ion transport and metabolism</td>
<td>25</td>
</tr>
<tr>
<td>Secondary metabolites biosynthesis, transport and catabolism</td>
<td>7</td>
</tr>
<tr>
<td>Signal transduction mechanisms</td>
<td>10</td>
</tr>
<tr>
<td>Intracellular trafficking and secretion</td>
<td>16</td>
</tr>
<tr>
<td>Defense mechanisms</td>
<td>8</td>
</tr>
<tr>
<td>General function prediction only</td>
<td>40</td>
</tr>
<tr>
<td>Function unknown</td>
<td>56</td>
</tr>
<tr>
<td>Not in COGs</td>
<td>328</td>
</tr>
<tr>
<td>On plasmids</td>
<td>58</td>
</tr>
<tr>
<td>Total</td>
<td>764</td>
</tr>
</tbody>
</table>

aGenes that were absent at least in one strain.
bOnly the singleton genes were analyzed.
cGenes on the chromosome were classified according to the COG database.

no apparent gene deletion was detected, indicating that several genes on these segments have probably translocated to other genomic loci.

Among the 60 segments with smaller size reduction (<2 kb), which were indicated in light yellow in Fig. 1, we detected deletion of a gene(s) only in seven segments by the microarray analysis. Most of the segments with undetectable deletions (50 out of 53) contained IS elements and/or parts of lambda-like prophages, indicating that small deletions involving these elements frequently occur on O157 genomes. In the remaining three segments, small deletions in the regions that were not represented by probes or some translocations have probably taken place.

Besides the above-mentioned genomic segments that apparently contain deletions, 20 loci have been identified by the WGPScanning analysis, where prophages or prophage-like elements are integrated in Sakai but deleted (or not integrated) in some other strains (indicated by blank in Fig. 1). These loci were classified into three groups according to the patterns of gene conservation observed in the microarray analysis. The first group included the Sp7 region of strain #2, Sp13 (P2-like phage) of strains #3 and #9, and Sp18 (Mu-like phage) of strains #2, #3, #5, #6, #7, #8 and #9. In these regions, all or nearly all genes were absent. We concluded that these genetic elements are completely missing in these strains.

The second group included the Sp18 region of strain #4 and SpLE1 of strain #6. In both cases, most genes on the elements were conserved, indicating that a prophage or prophage-like element almost identical to Sp18 or SpLE1 is present in other locus. We have already identified alternative integration sites for each element (Fig. 1).

The third group included the Sp5 region of strains #2, #3, #4, #6, #7 and #8, Sp1/Sp2 of strain #8 and Sp7 of strain #3. In these regions, genes were partially conserved, exhibiting mosaic conservation patterns. This indicates that these strains contain, in some other loci, prophages which are significantly diverged but with some structural similarity to Sp5, Sp1/Sp2 or Sp7. Of particular interest was Sp5, the Stx2 phage of O157 Sakai. The Sp5 region was one of the regions with the highest level of variation, and we have already identified the alternative integration sites for the Stx2 phages in these six strains (Fig. 1). These Stx2 phages are thus different from Sp5 not only in the integration site but also in the genomic structure.

A total of 83 segments yielded PCR products larger than those of Sakai in the WGPScanning analysis as shown in Fig. 1, where the segments with larger increments (≥2 kb) are indicated in blue, and those with smaller increments (<2 kb) in light blue. These segments must contain some insertions. In addition, as many as 120 segments were not amplified by PCR (indicated in red in Fig. 1), in which large insertions or some other types of large genomic rearrangements likely have taken place. In these cases, if all genes on a segment were conserved, the segment most likely contains a simple insertion. By comparing the two datasets, 66 such segments were identified. Among these, we have identified seven new phage integration sites including those for Stx1 and Stx2 phages (Fig. 1). In other cases, more complicated genetic events must have occurred.

3.5. Unusual signal reductions around the IS insertion sites

During the data comparison of microarray and WGPScanning analysis, we noticed that somewhat unusual signal reductions occurred in the ETT2 region. In strains #2, #4 and #7, the region yielded larger PCR products, but a block of genes were judged as absent in the microarray analysis. Sequencing analyses of these regions, however, indicated that all genes were completely conserved but one copy of IS Ec8 was newly inserted in
A closer inspection of the microarray data revealed that, although the IS\textsubscript{Ec8} insertion sites were different between the strains, the genes judged as absent were all located in the 2–4 kb regions surrounding each IS\textsubscript{Ec8}, and that their signal ratios were just slightly lower than the cutoff value (Fig. 2). This phenomenon was specific to the ETT2 region, and observed even after the labeled DNA was fragmented by sonication. Although we do not know the mechanism underlying this phenomenon, some caution is required to interpret the microarray data when a block of genes exhibit a moderate level of signal reduction.

### 4. Discussion

Presence of an unexpectedly high degree of structural diversity in O157 genomes has been revealed by our previous analysis using the WGPScanning method.\(^8\) The results also suggested that the major factor generating such diversity is the variation of prophages and prophage-like elements. However, it remained to be elucidated how the genomic diversity detected by WGPScanning affects the gene content of each strain. In order to know the variation of their gene composition, we constructed an oligoDNA microarray according to the genome sequence of O157 Sakai,\(^10\) and performed the CGH analysis of the same set of O157 strains as examined by the WGPScanning.

By this analysis, we identified 431 VAP genes that were variably present in the 8 O157 strains, and found that most of them were on S-loops, particularly on the prophage regions. This indicates that a remarkable level of variation in the gene repertoire exist among the O157 strains, and confirms the role of bacteriophages as the key players in the genome diversification of O157. Very recently, Wick \textit{et al.}\(^26\) have reported the results of CGH analyses of several types of O157 strains and its close relatives, including typical O157:H7 strains (GUD\(^+\), SOR\(^+\)), atypical O157:H7 strains (GUD\(^-\), SOR\(^-\)), O157:H\(^-\) German clones (GUD\(^+\), SOR\(^-\)) and O55:H7 strains (the closest relatives of O157:H7\(^27\)). They used a multigenome array containing 50 mer oligonucleotide probes that target 6176 ORFs on the chromosomes of K-12, O157 EDL933 and O157 Sakai. By this analysis, they identified the Sakai and K-12 genes that were gained or lost during the emergence of O157:H7 from its O55:H7-like ancestor. Many of these genes were parts of prophages and prophage-like elements, suggesting complex histories of these elements during the evolution of O157.\(^26\) What we should emphasize here is, however, that all O157 strains but strain #2 examined in our present study belonged to the typical O157 group. Even if the 89 genes that were absent only in strain #2 were excluded, 350 genes still exhibited variable distribution. This indicates that there is a remarkable variation in gene content even among typical O157 strains.

Identification of these VAP genes is of medical importance in that they could be used as valuable genetic markers to discriminate O157 strains. For example, a mini-DNA chip featuring these genes would become a useful molecular tool for epidemiological studies of

### Table 5. Virulence-related genes that were variably present among the eight O157 strains

<table>
<thead>
<tr>
<th>ORF number</th>
<th>Location</th>
<th>Gene product</th>
<th>Description/function</th>
<th>References</th>
<th>Conserved in</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECs0350</td>
<td>S-loop 23</td>
<td>HmwA (\textit{Haemophilus influenzae})—like protein</td>
<td>Adhesion</td>
<td>13</td>
<td>7/8</td>
</tr>
<tr>
<td>ECs0743</td>
<td>S-loop 51</td>
<td>Outer membrane usher protein</td>
<td>Adhesion (fimbrial synthesis)</td>
<td>5</td>
<td>7/8</td>
</tr>
<tr>
<td>ECs1236</td>
<td>Sp5</td>
<td>Lom-like protein</td>
<td>Adhesion/resistance to host immune response</td>
<td>5</td>
<td>5/8</td>
</tr>
<tr>
<td>ECs1312</td>
<td>SpLE1</td>
<td>TraT</td>
<td>Resistance to host immune response</td>
<td>5</td>
<td>7/8</td>
</tr>
<tr>
<td>ECs1382</td>
<td>SpLE1</td>
<td>HecB (\textit{Neisseria meningitidis})—like protein</td>
<td>Toxin activation</td>
<td>15</td>
<td>5/8</td>
</tr>
<tr>
<td>ECs1388</td>
<td>SpLE1</td>
<td>PchD</td>
<td>PerC-like regulator</td>
<td>19</td>
<td>4/8</td>
</tr>
<tr>
<td>ECs1388</td>
<td>Sp7</td>
<td>PchE</td>
<td>PerC-like regulator</td>
<td>19</td>
<td>7/8</td>
</tr>
<tr>
<td>ECs1812</td>
<td>Sp9</td>
<td>NleA/EspI</td>
<td>Non-LEE TTSS effector</td>
<td>24,25</td>
<td>7/8</td>
</tr>
<tr>
<td>ECs1814</td>
<td>Sp9</td>
<td>NleH (\textit{Citrobacter rodentium})—like protein</td>
<td>Non-LEE TTSS effector</td>
<td>22</td>
<td>4/8</td>
</tr>
<tr>
<td>ECs1815</td>
<td>Sp9</td>
<td>NleF (\textit{C. rodentium})—like protein</td>
<td>Non-LEE TTSS effector</td>
<td>22</td>
<td>4/8</td>
</tr>
<tr>
<td>ECs1828</td>
<td>Sp9</td>
<td>NleG (\textit{C. rodentium})—like protein</td>
<td>Non-LEE TTSS effector</td>
<td>22</td>
<td>7/8</td>
</tr>
<tr>
<td>ECs2973</td>
<td>Sp15</td>
<td>Stx1B</td>
<td>Toxin</td>
<td>2</td>
<td>6/8</td>
</tr>
<tr>
<td>ECs2974</td>
<td>Sp15</td>
<td>Stx1A</td>
<td>Toxin</td>
<td>2</td>
<td>6/8</td>
</tr>
<tr>
<td>ECs3860</td>
<td>SpLE3</td>
<td>Efa1 (N-terminal half)</td>
<td>Adhesin and other</td>
<td>16,18</td>
<td>7/8</td>
</tr>
<tr>
<td>ECs3861</td>
<td>SpLE3</td>
<td>Efa1 (C-terminal half)</td>
<td>Adhesin and other</td>
<td>16,18</td>
<td>7/8</td>
</tr>
<tr>
<td>pO157\textsubscript{76}</td>
<td>pO157</td>
<td>KatP</td>
<td>Catalase/peroxidase</td>
<td>11</td>
<td>6/8</td>
</tr>
</tbody>
</table>
Figure 2. Unusual signal reduction associated with insertions of ISEc8. Distribution of signal ratios (relative to O157 Sakai) from the probes representing the ETT2 regions of strains #2, #4, #7 and #8 is shown. Genes in the ETT2 region (ECs3703–3737) are depicted by black arrows, and genes on ISEc8 newly inserted in strains #2, #4 and #7 by gray arrows. Positions of each gene and probe are drawn to the O157 Sakai genome sequence. Horizontal broken lines indicate cutoff lines for the presence or absence determination in each strain. We have confirmed that all the probe sequences exhibiting unusual signal reduction are identical to those of Sakai. As seen in strain #8, no signal reduction was observed in other strains without ISEc8-insertion.
O157. Of more importance may be that the VAP genes include a considerable number of virulence-related genes. Although the variation of \( stx1 \) and \( stx2 \) genes among O157 strains is known, additional 14 potentially virulence-related genes were found to display variable distributions. They include four non-LEE TTSS effectors, indicating that, although the LEE locus encoding the TTSS machinery is highly conserved, the repertoire of its effectors differs significantly among O157 strains. Some metabolic genes, such as those for urease production (ECs1321–1326) and iron acquisition (ECs1693–1699), which may be indirectly involved in the pathogenesis of O157, were also found to be variably present. These data raise a possibility that some variation in the potential virulence and infectivity exists among O157 strains, and we need to analyze more O157 strains from various sources, including environmental isolates.

The phylogenetic position of strain #2 in the O157 lineage is somewhat obscure. But, since weak \( β \)-glucuronidase activity was detected after long cultivation, the strain may belong to the GUD\(^+\) O157:H7 group. The largest number of VAP genes was detected in this strain, and it shared many, if not all, features that Wick et al.\(^{26}\) described as being specific to GUD\(^+\) strains. For example, several gene blocks, such as ECs4134–4139, ECs1611–1619 and ECs3860–3861, were deleted (or highly divergent) only in this strain. The gene content on the pO157 plasmid of strain #2 also differed significantly from other strains. Unfortunately Wick et al.\(^{26}\) did not examine the pO157 genes, but the difference in pO157 may be another genomic feature that discriminates the GUD\(^+\) group from typical O157 : H7 strains.

In the present study, we examined the same set of O157 strains as those were analyzed previously by the WGPSScanning method. This provided a good opportunity to compare the data obtained by the two methods (Fig. 1). Although the microarray is now widely used as a major tool in comparative genomics, a fine comparison of the two datasets rather highlighted the weakness of microarray and emphasized the usefulness of a combinational approach with WGPSScanning. Microarray actually provided valuable genome-wide information on the genome composition in each test strain, but almost no information on strain-specific insertions and translocations. In the case of O157, particular cautions are required in interpreting the microarray data because of the presence of same or very similar prophages as well as many IS elements. Even with such difficulties, all simple insertions as well as many translocations and replacements have been identified by comparing the two datasets. This approach also identified many small deletions that could not be detected by microarray alone. IS elements were frequently associated with such small deletions, suggesting that IS elements may be another key player to diversify O157 genomes. The combinational use of microarray and WGPSScanning overcame the weakness of each method alone, and provided us more details and thus a much complex view on the genomic diversity of O157 strains.

The variation of prophages among O157 strains was not fully illustrated even with this combinational approach. But, as most strikingly shown for the Stx2-transducing phages, the data clearly indicated that prophages of O157 exhibit an extremely high level of strain-to-strain variations. These variations appeared to have been generated not only by simple insertion, deletion or translocation of prophages but also by more complex genetic events involving local recombination and deletion of prophage segments. One such example has been documented by Shaikh and Tarr\(^{28}\) in the prophages integrated into the \( yehV \) locus (corresponding to Sp15, the Stx1-transducing phage of O157 Sakai). Of importance again is that most of the strains examined here were typical O157 : H7 strains. This indicates that such dynamic turnover of prophages is still actively taking place, and as we have proposed previously, O157 strains are functioning as a kind of ‘phage factory’, releasing a variety of new recombinant bacteriophages into the environment.\(^6\) This process should be significantly involved in the generation of variable gene contents of O157 strains detected by the CGH analysis.

Acknowledgements: This work was supported by Research for the Future Program funded by the Japan Society for the Promotion of Science (JSPS-RFTF00L01411), Grant-in-Aids for Scientific Research on Priority Areas and the 21st Century COE (Centers of Excellence) Program (Life Science) from the Ministry of Education, Culture, Sports, Science and Technology, Japan and a grant from the Yakult Foundation. We thank Dr Taku Ohshima for his valuable advice, Yoshiyuki Maki for his assistance in primer design, Akemi Yoshida and Yumiko Takeshita for their technical assistance, and Yumiko Hayashi for her language assistance.

Supplementary Data: Supplementary Data can be found online at http://dnaresearch.oxfordjournals.org.

References


