INTRODUCTION

Escherichia coli O157 : H7 is a major enteric pathogen, capable of causing major outbreaks of gastrointestinal disease. Approximately 75 000 cases of E. coli O157 : H7 infections occur annually in the United States, making it the principal serotype of enterohaemorrhagic E. coli isolated from patients in the USA (Griffin & Tauxe, 1991). E. coli O157 : H7 infection causes severe abdominal cramping and a watery diarrhoea that progresses to severe bloody diarrhoea (Riley et al., 1983; Griffin et al., 1988). An estimated 15–20 % of people infected with E. coli O157 : H7 present with indications severe enough to require hospitalization and in about 2–7 % of patients, infections progress from haemorrhagic colitis into haemolytic-uraemic syndrome (HUS) with kidney failure complications (Su & Brandt, 1995).

E. coli O157 : H7 infections usually have a food-borne aetiology. Since cattle are a common reservoir of this organism, the vast majority of outbreaks since identification of the pathogen in 1982 have been traced to consumption of meat and dairy products contaminated with this organism (Mead & Griffin, 1998). Recent outbreaks of human disease, in contrast, have been linked to consumption of fresh produce tainted with E. coli O157 : H7. A recent multi-state outbreak in the USA, for instance, was associated with contaminated spinach and caused 199 cases of illness, including three deaths. Among the ill, 51 % were hospitalized and in 16 % of the cases, infection progressed to HUS and kidney failure (www.cdc.gov/foodborne/ecolispinach/100606.htm). The large number of patients hospitalized and high rate of kidney failure suggest that this outbreak was due to a more virulent strain of E. coli O157 : H7. Assessing genetic variation among temporally and geographically diverse strains of E. coli O157 : H7 thus becomes particularly important for understanding its public health consequences and adaptation to new environmental niches, as well as for the practical aspects of surveillance and epidemiological trace-back.

For the epidemiological tracing of clinical isolates to source strains, PFGE has been an effective tool for molecular analysis. Detecting variation among E. coli O157 : H7 strains at the gene and chromosome level, however, requires a more extensive analysis of strains. In the recent past, variation or diversity within bacterial species has been defined largely by genetic polymorphisms – early on, by analysing variants at the protein level as revealed by multilocus enzyme electrophoresis (MLEE, Whittam et al., 1993), and, more recently, by examining variation at the nucleotide level through the use of multilocus sequence typing (MLST, Feil & Enright, 2004; Urwin & Maiden, 2003). Studies have shown differences in the ability to differentiate individual strains by MLST and PFGE. Optical maps distinguish individual strains of Escherichia coli O157 : H7

Michael L. Kotewicz, Scott A. Jackson, J. Eugene LeClerc and Thomas A. Cebula

Division of Molecular Biology, Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, Laurel, MD 20708, USA

Optical maps of 11 Escherichia coli O157 : H7 strains have been generated by the assembly of contiguous sets of restriction fragments across their entire 5.3 to 5.6 Mbp chromosomes. Each strain showed a distinct, highly individual configuration of 500–700 BamHI fragments, yielding a map resembling a DNA ‘bar code’. The accuracy of optical mapping was assessed by comparing directly the in silico restriction maps of two wholly sequenced reference genomes of E. coli O157 : H7, i.e. EDL933 and the Sakai isolate (RIMD 0509952), with the optical maps of the same strains. The optical maps of nine other E. coli O157 : H7 strains were compared similarly, using the sequence-based maps of the Sakai and EDL933 strains as references. A total of 91 changes at 28 loci were positioned and sized; these included complex chromosomal inversions, insertions, deletions, substitutions, as well as a number of simple RFLPs. The optical maps defined unique genome landmarks in each of the strains and demonstrated the ability of optical mapping to distinguish and differentiate, at the individual level, strains of this important pathogen.

A supplementary table listing chromosome polymorphisms found in nine optical maps and in the Sakai strain relative to the EDL933 in silico map is available with the online version of this paper.

Correspondence
Thomas A. Cebula
Thomas.Cebula@fda.hhs.gov

Received 20 November 2006
Revised 8 February 2007
Accepted 12 February 2007
Diversity among \textit{E. coli} et al. oligonucleotide gene sequences (Wick, 2005) provided ample evidence for significant allelic diversity among \textit{E. coli} O157 : H7 strains. For example, when 4000 PCR-amplified genes (Fukiya, 2004) or 6000 50-mer oligonucleotide gene sequences (Wick, 2005) were targeted, and when 1199 specific genes (Zhang, 2006) or 1% of the bacterial genome (Jackson, 2006) were examined, rich allelic diversity was revealed among a relatively homogeneous group of pathogenic \textit{E. coli} strains.

Diversity among \textit{E. coli} serotypes, at the chromosomal level, likewise was uncovered by direct genomic comparisons. For example, relative to the 4.639 Mbp genomic sequence of the prototypic \textit{E. coli} K-12 strain, MG1655, the 5.528 Mbp genome of the pathogenic \textit{E. coli} O157 : H7 strain, EDL933, was ‘peppered’ with insertions and deletions (Perna, 2001). That is, 177 segments of O157 : H7 DNA were defined as ‘O-islands’ because they were found in the O157 : H7 genome, but not in K-12. Correspondingly, 234 ‘K-islands’ were identified by their presence in the MG1655 genome and absence in the EDL933 genome. When the 5.231 Mbp sequence of \textit{E. coli} uropathogenic strain CFT073 was added to this comparison, the complex mosaic structure of \textit{E. coli} pathogens became even more evident (Welch, 2002). While about 39% of the non-redundant proteins were common to all three strains, 46% were unique to a single strain. The two \textit{E. coli} pathogens differed from each other in 900 000 bp of unique sequence, and thus differed from each other by as much as each differed from the commensal strain. Despite the remarkable synteny of enteric bacterial chromosomes, as typified by \textit{E. coli} and \textit{Salmonella}, genomic comparisons have identified the movement and repositioning of large blocks or segments of genes as a major source of diversity. Comparisons of completed sequenced genomes using the NCBI gMap web site underscore the findings that related bacterial genomes contain multi-segmented arrays of genes inserted in different orientations and shuffled in chromosomal positions (http://www.ncbi.nlm.nih.gov/sutils/gmap.cgi; Kotewicz et al., 2003).

Optical mapping, which can scan and assess the architecture of complete bacterial genomes, thus should prove useful for comparative genomics as well as for epidemiology and microbial forensics (Cai, 1998). Optical mapping was pivotal in closing the genome sequence of \textit{E. coli} O157 : H7 EDL933 (Lim, 2001), and its value was further underscored when optical maps were used to compare particular subtypes of \textit{Shigella} with a sequenced strain (Zhou, 2004). In conjunction with subtractive hybridization to identify unique sequences, optical mapping also was used to map the sequences onto the enterotoxigenic \textit{E. coli} H10407 chromosome (Chen et al., 2006). Herein, the chromosomal organization of each of 11 \textit{E. coli} O157 : H7 strains was examined by \textit{BamHI} optical mapping. A \textit{BamHI} optical map for a typical 5–5.5 Mbp \textit{E. coli} O157 : H7 chromosome presents as 500–700 ordered restriction fragments, with the number and arrangement of fragments varying with each strain examined. As restriction digestion is carried out on genomic DNA affixed to a glass substratum, the molecular size of contiguous \textit{BamHI} fragments is determined and precisely mapped along the chromosome. In contrast, PFGE, used commonly in molecular epidemiology studies (Ribot, 2006), typically produces 50 genomic restriction fragments ordered only by fragment size within the gel, not by map position on the genome. Because of its greater resolution and specificity for strain identification, optical mapping holds promise for both the public health and forensic communities (Cebula, 2005). To this end, we present data to show that complex events, including inversions, deletions and insertions occurring within an individual chromosome, can be detected using optical mapping. Optical maps can serve as a unique DNA ‘bar code’ identifier for a particular microbe, not only at the species or subspecies level, but also at the individual strain or isolate level.

\section*{METHODS}

\textbf{Bacterial strains.} Sources of the \textit{E. coli} O157 : H7 strains used in this study are listed in Table 1. Reference strains that have been completely sequenced were EDL933 (GenBank AE005174), isolated from ground hamburger associated with a 1982 outbreak in the United States; and Sakai, RIMD 0509952 (GenBank BA000007), a clinical sample isolated during an outbreak in Japan in 1996. Selection of \textit{E. coli} O157 : H7 strains for optical mapping was based on cladistic analysis of single nucleotide polymorphisms, which allowed grouping of the strains into different clades (Cebula, 2005). The FDA isolates of EDL933 and Sakai were assigned strain numbers EC1275 and EC1276, respectively. To clarify the comparisons of the optical maps of these isolates with the \textit{in silico} maps derived from sequenced EDL933 and Sakai isolates, the EC1275 and EC1276 designations have been retained in this report.

\textbf{Optical mapping.} Optical maps were prepared by OpGen. Briefly, for \textit{E. coli} O157 : H7 isolates, \textit{in silico} analysis of sequenced strains allowed the selection of an appropriate restriction enzyme (\textit{BamHI}) to give an optimized number of fragments and size distribution for optical mapping. Strains were grown and embedded in low-melting-point agar. Upon lysis and dilution, high molecular mass, genomic DNA molecules were spread and immobilized onto derivatized glass slides and digested with \textit{BamHI}. After restriction digestion, a small gap in the DNA at the precise location of the restriction endonuclease cleavage site is left. The DNA digests were stained with YOYO-1 fluorescent dye, and photographed with a fluorescence microscope interfaced with a digital camera. Automated image-analysis software located and sized fragments, based on YOYO-1 binding and assembled multiple scans, into whole-chromosome optical maps. The order of the DNA fragments was retained on the optical mapping surface, and the mass of each fragment was determined by comparing the fluorescence intensity measurements of each DNA fragment to known standards that were added to the DNA sample prior to loading onto the glass surface. The single DNA molecule restriction map reads were assembled into partial maps, or contigs, in a process similar to shotgun sequence assembly. A number of overlapping partial contigs

\url{http://mic.sgmjournals.org}
were further assembled into an optical map that spans the entire bacterial genome. The depth of coverage minimizes mapping error and the overlapping cascades of partial maps create continuity across the entire genome (Zhou et al., 2004).

**RESULTS**

The BamHI optical map of a typical *E. coli* O157 : H7 strain is a contiguous set of 500–700 restriction fragments displayed graphically as an array of sized and ordered fragments that resemble a bar code. Quantitative analysis was performed by tabulating changes in restriction fragments relative to the reference genomes, EDL933 and Sakai. The optical mapping software allows the construction of sequence-based reference maps with any DNA sequence, and allows different restriction enzymes to be selected for the sequence-based maps. For PFGE analysis, on the other hand, a rare-hitter enzyme like *Xba*I is used to partition the bacterial chromosome into 40–60 resolvable fragments on a pulsed field gel. The difference in resolution between *Xba*I and BamHI bacterial genome maps is striking (Fig. 1a). The accuracy of optical mapping was demonstrated by comparing the *in silico* BamHI map and the optical BamHI map of EDL933 (EC1275, Fig. 1b). With the exception of the loss of a number of small fragments (dropout fragments, discussed in the next section) the optical map recapitulates the sequence-based map for both EDL933 and Sakai strains.

**Optical map accuracy**

There were 96 fragments missing from the EC1276 (Sakai) optical map compared to the 639 expected fragments found in the *in silico* Sakai map (Table 2). Of the 96 ‘dropout’ fragments, all were less than 2000 bp in size; a majority (60) was smaller than 600 bp. That is, of the fragments expected from the sequence of Sakai, 60/639 fragments in the 21–600 bp range, 22/30 fragments in the 601–1000 bp range, 10/39 fragments in the 1–1.5 kb range and 4/36 in the 1.5–2.0 kb range were not detected in the Sakai optical map (EC1276, Table 2). A similar set of results was found for EDL933. There were 112 dropout fragments in the optical map of EDL933 (EC1275, Table 2) compared to the *in silico* map. None of the 61 expected fragments in the 21–600 bp range were detected. Just as was the case for the optical map of Sakai, there was an increase in the mapping of expected fragments as they increased in size (Table 2). Optical mapping detected 100 % of expected fragments above 2 kb.

Relative to the *in silico* maps of the sequenced EDL933 (EC1275) and Sakai (EC1276) strains, the loss of smaller fragments in the nine optical maps of unsequenced strains was quantified over a representative 1 Mbp region. In a generally quiet portion of the genome from bp 1–1 000 000, the optical maps of the nine unsequenced strains were generally similar and largely devoid of the insertions, deletions and inversions found beyond 1.0 Mbp.

---

Table 1. Strains and summary of their genomic properties from *in silico* and optical maps of sequenced reference strains EDL933 and Sakai

<table>
<thead>
<tr>
<th>Strain no.</th>
<th>FDA designation</th>
<th>Alternative strain designation/source</th>
<th>Total no. of bands</th>
<th>Genome size, kb</th>
<th>Inversion type†</th>
<th>O-island sites of inversion</th>
<th>Prophages flanking inversion</th>
<th>Inversion size, kb (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EC1275</td>
<td>EDL933 ATCC 43895</td>
<td>525</td>
<td>5535.378</td>
<td>EDL933</td>
<td>57, 71</td>
<td>O, P</td>
<td>429 (7.80)</td>
</tr>
<tr>
<td>2</td>
<td>EC1276</td>
<td>Sakai ATCC BAA460</td>
<td>544</td>
<td>5514.226</td>
<td>Sakai</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>EC503</td>
<td>EC177/P. Feng</td>
<td>507</td>
<td>5340.242</td>
<td>Sakai</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>EC507</td>
<td>ATCC 35150</td>
<td>522</td>
<td>5440.014</td>
<td>Sakai</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>EC869</td>
<td>FSIS45750/R.Buchanan</td>
<td>546</td>
<td>5465.250</td>
<td>Sakai</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>EC1231</td>
<td>GS-101/T.Whittam</td>
<td>549</td>
<td>5568.840</td>
<td>Sakai</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>EC1225</td>
<td>93-1111/P.Tarr</td>
<td>524</td>
<td>5510.735</td>
<td>EDL933</td>
<td>57, 71</td>
<td>O, P</td>
<td>409 (7.70)</td>
</tr>
<tr>
<td>8</td>
<td>AB1</td>
<td>A. Benson</td>
<td>565</td>
<td>5579.147</td>
<td>EDL933‡</td>
<td>57, R</td>
<td>O, R</td>
<td>258 (4.70)</td>
</tr>
<tr>
<td>9</td>
<td>EC536</td>
<td>86-17/P. Tarr</td>
<td>515</td>
<td>5326.110</td>
<td>EDL933‡</td>
<td>57, R</td>
<td>O, R</td>
<td>233 (4.20)</td>
</tr>
<tr>
<td>10</td>
<td>EC533</td>
<td>86-24/P. Tarr</td>
<td>549</td>
<td>5451.142</td>
<td>EDL933‡</td>
<td>44, 79</td>
<td>M, U</td>
<td>1482 (26.80)</td>
</tr>
<tr>
<td>11</td>
<td>EC502</td>
<td>EC1211/P. Feng</td>
<td>548</td>
<td>5547.921</td>
<td>EDL933‡</td>
<td>45, 93</td>
<td>W, V</td>
<td>1656 (29.90)</td>
</tr>
</tbody>
</table>

*Reference, sequenced strains.

†Comparisons of a number of genomes indicated that *Escherichia coli* O157 : H7 EDL933 contains a chromosomal inversion relative to *E. coli* K-12 and Sakai. Thus, the Sakai chromosome is considered the ground state.

The inversion in this strain appears to have occurred in a strain like EDL933 containing an inversion between the O-islands 57 and 71, prophages O and P, upon which either an external larger flanking inversion (EC1233 and EC502) or smaller internal inversion (EC1225 or EC536) has occurred between the indicated prophages; i.e. a double inversion.
in their genomes (Fig. 2). This area of the chromosome allowed a reasonable comparison of the dropouts from strain to strain. Of the 110 restriction fragments in this portion of the genome, 25 showed dropouts among the nine strains; eight dropouts were below 600 bp in size, nine were 600–1000 bp and eight were from 1001 to 2000 bp.

Thus, as was found for the EDL933 and Sakai optical maps relative to their \textit{in silico} maps, no fragments below 600 bp were detected in the nine optical maps of unsequenced strains. In the optical maps of the nine unsequenced strains, five demonstrated all eight expected fragments in the 1–2 kb size range. Each of the remaining four maps was

<table>
<thead>
<tr>
<th>Fragment size range</th>
<th>EDL933</th>
<th>Sakai</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of fragments expected (sequence)</td>
<td>No. of fragments in optical map</td>
<td>% of fragments mapped</td>
</tr>
<tr>
<td>21–600 bp</td>
<td>61</td>
<td>0</td>
</tr>
<tr>
<td>601–1000 bp</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>1001–1500 bp</td>
<td>35</td>
<td>20</td>
</tr>
<tr>
<td>1501–2000 bp</td>
<td>39</td>
<td>30</td>
</tr>
<tr>
<td>2001–3000 bp</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>\textit{BamH}I fragments missing</td>
<td>112</td>
<td>96</td>
</tr>
<tr>
<td>Total expected bp missing</td>
<td>35 014</td>
<td>20 309</td>
</tr>
<tr>
<td>Total bp, sequence</td>
<td>5 528 446</td>
<td>5 498 451</td>
</tr>
<tr>
<td>Total bp, optical map</td>
<td>5 535 378</td>
<td>5 514 226</td>
</tr>
</tbody>
</table>
missing either one, two, three, or four of the expected eight fragments. Overall, for 25 expected fragments, the mean loss was 17.5; expected fragment loss ranged from 16 to 21 fragments among the nine strains. For the 1 Mbp portion of the genomes examined, this represented a mean loss of 10.9 kb, ranging from 7.5 to 15.8 kb among the nine strains. Based on fragments expected from the sequenced genomes, the optical maps showed an estimated loss of 44 kb for each genome. For the two sequenced reference strains, the optical maps showed 112 dropout fragments totalling 35.014 kb for EDL933 (EC1275) and 96 dropout fragments totalling 20.309 kb for Sakai (EC1276) (Table 2).

Whereas factors like the optical limit for physical detection dominate mapping the smallest fragments, 21–600 bp, other factors contribute to dropout frequency for larger fragments, including biophysical restraints. That is, the loss of small fragments from the substratum during digestion, staining and washing steps; sites that remain refractory to restriction digestion; and clustering of three or more restriction sites within 4 kb of each other contribute to small fragment loss or dropouts. Expected BamHI restriction fragments larger than 2.0 kb, however, were routinely and uniformly identified and mapped relative to the reference sequences. For these reasons, we discounted fragments of less than 2.0 kb and turned our attention to the reproducible portions of the optical maps.

There were a number of differences between the optical and the in silico maps of EDL933. In EDL933, if the 112 small drop-out fragments (21 bp to 2 kb) are subtracted from the sequence-based total of 645 fragments, there were 533 in silico fragments relative to the 525 optical fragments. The difference of 12 was a combination of BamHI polymorphisms (or failure to digest at six positions), which created a net loss of six fragments by fusion of fragments, and two and four fragment differences each in cryptic prophages O and V in runs of small fragments. These same types of similarities and differences held true for the Sakai sequence-based and optical maps. Determining whether sequencing errors or recombination between and among duplicated prophages would explain the seven differences awaits further investigation. For Sakai, the identity between the optical and in silico maps was stronger. Disregarding the 96 dropout fragments, there was only a single difference between the optical (EC1276) and in silico chromosome maps of Sakai. The difference represented a probable single restriction site change; the remaining 540 out of 541 restriction fragments were comparable in size and position in the optical and in silico maps of Sakai.

Chromosome size

Relative to the sequence-based fragment sizes, the standard deviation of measuring a 10 kb fragment in optical maps of 11 strains was about 300 bp or 3%. The standard deviation for 20–30 kb fragments was typically 600 bp or 2%. The variability of the total chromosome length derived from the sum of 530 measured fragments becomes a mean of means with a variability (v) that can be approximated by the formula \( v = \frac{1}{\sqrt{n}} \times \frac{1}{\sqrt{2}} \) = 4.6 kb or about 0.1% for a 5 Mbp genome. The estimated effect of fragment measurement accuracy (5 kb) was smaller than the calculated effect of dropout fragments (30–40 kb) on the overall length determination for a 5 Mbp genome. Whereas the reference sequence of the Sakai strain was 5 498 451 bp, the optical map size was 5 514 226 bp. For the EDL933 strain, the reference sequence was 5 528 446 bp, and the optical map was 5 535 378 bp. The optical map chromosome lengths for the two reference strains were larger than the sequence-based chromosome sizes by 16 and 7 kb. The chromosomal sizes measured for the nine additional strains of E. coli O157 : H7 ranged from 5.326 Mbp (EC536) to 5.579 Mbp (AB1), a size difference of 250 kb (Table 1).

These data, coupled with the comparison of each of the individual BamHI maps, demonstrated the rich diversity that exists among E. coli O157 : H7 strains. The optical maps visually demonstrated chromosomal similarities and differences. Optical map alignment software semiquantitatively defines the similarities between genomes. In this study of 11 E. coli O157 : H7 strains, restriction fragment alignments were used to define similarities. Because of the extensive annotation of the sequenced EDL933 and Sakai sequences, similarities and

---

**Fig. 2.** Aligned optical maps for E. coli O157 : H7 strains and in silico reference maps. In the pair-wise alignments, lines connecting two chromosomal maps indicate a discontinuity in the alignment of fragments. Chromosomal inversions are indicated by crossed alignment lines between paired maps and are highlighted in yellow. Unaligned restriction fragments, representing differences between two aligned chromosomes, are shown in white; green indicates aligned restriction fragments. Blue restriction fragments indicate marker prophage fragments; in some alignments red prophage fragments indicate that more than two copies are found in two aligned genomes. The extent of green fragments in the optical maps reflects similarities based on alignments of restriction fragments, spanning up to hundreds of contiguous fragments. At the bottom of the panels, for reference, the aligned in silico maps of EDL933 and Sakai are presented. The 8% chromosomal inversion, detected in the sequenced EDL933 strain relative to the sequenced Sakai strain, is indicated in yellow. A Mbp scale is shown at the bottom of each panel as well as a reference alignment of the in silico maps of EDL933 and Sakai. (a) The optical map of each of the four strains indicated in Table 1, whose chromosomes resemble the Sakai genome and do not contain an inversion, was aligned with the in silico map of Sakai. The optical map of the FDA isolate of Sakai (EC1276) was aligned with the in silico Sakai map. (b) The optical map of each of the five strains indicated in Table 1 containing a chromosome inversion was aligned with the in silico map of EDL933. The optical map of the FDA EDL933 isolate (EC1275) was aligned with the EDL933 sequence-based in silico map.
Optical mapping *E. coli* O157 : H7 strains

(a)

(b)
differences identified in the sequences were verified in the optical maps of the two reference strains and, by inference, in the optical maps of the other isolates. All differences in the sequence-based maps were seen in the optical maps of the reference strains. Verification of chromosome similarities by Southern analyses was not performed; however, the presence of the normal genetic complement of *E. coli* O157 : H7 (some 5000 genes) was verified for each isolate by DNA microarray analysis (data not shown). Several deletions demonstrated in optical maps were verified by DNA microarray analysis (data not shown).

Fig. 2 shows sets of pairwise alignments of the *Bam*HI optical maps of the strains examined in this study to reference *in silico* maps. The alignments demonstrate that although all of the strains showed large regions of similarity by contig analysis (green fragments), extensive regions of dissimilarity (white fragments) as well as five notable chromosomal inversions were found (yellow fragments, Fig. 2). Among the 11 *E. coli* O157 : H7 optical maps, there were 91 differences found at 28 chromosome sites or loci. The polymorphic sites were numbered consecutively from the chromosome origin; they were defined both by the closest numbered *Bam*HI reference fragment and by the position of the fragment on the EDL933 chromosome (see Supplementary Table S1, available with the online version of this paper).

**Simple fragment size polymorphisms**

Comparison of the sequence-based EDL933 (645 fragments) and Sakai (639 fragments) *in silico* maps showed five restriction site polymorphisms and 13 simple fragment differences greater than 100 bp between equivalently positioned fragments. Most of these polymorphisms represented changes in fragment sizes below the limit of resolution and below the 2 kb limit of reliability. Only two of the differences between EDL933 and Sakai were useful for differentiating individual strains, polymorphic sites 21 and 25 (Supplementary Table S1). Polymorphic site 21 showed a difference between EDL933 restriction fragment e374 (2589 bp) and the comparable Sakai fragment s367 (3902 bp). The optical maps of EDL933 and Sakai showed fragments of 2111 bp and 4263 bp, respectively, in the two strains. Analysis of optical maps of the nine other strains revealed two classes of fragments at this chromosomal position; five strains harboured EDL933-like fragments of 2544 ± 178 bp and four contained Sakai-like fragments of 4077 ± 181 bp. The sequence-based polymorphic site 25 showed an EDL933 *in silico* fragment e600 of 20.654 kb and a comparable Sakai *in silico* fragment s593 of 59.412 kb. Optical maps of the nine remaining O157 : H7 strains showed a 20.652 ± 0.377 kb EDL933-like fragment at this position. The number of useful simple fragment polymorphisms was too small to allow significant strain differentiation.

Two additional simple polymorphisms at sites 24 and 28 were found when the optical maps of the remaining nine strains were compared (Supplementary Table S1). These two simple discriminatory sites contained RFLPs. As detailed in the next section, although some RFLPs initially appeared simple, examination showed that some RFLPs were indicative of polymorphic sites where complex chromosomal changes and rearrangements had occurred. They were evident only when a greater number of strains was examined. Notably, many of the complex polymorphic sites were within prophages undergoing substitutions, inversions and possible truncations. The number and variability of these complex polymorphisms allowed strain differentiation.

**Complex chromosomal polymorphisms; inversions**

Complex alignment differences were frequent in the optical maps of the *E. coli* O157 : H7 strains and they were used successfully to discriminate individual strains. Inversions earmarked the first class of complex differences. The original genomic sequences and the *in silico* maps of EDL933 and Sakai strains demonstrated a chromosomal inversion spanning 430 kb, or 7.8 % of EDL933 relative to the Sakai and *E. coli* K-12 chromosomes. This inversion was prominent in the aligned *in silico* maps of EDL933 and Sakai (Fig. 2a). This inversion is flanked by prophages O and P at EDL933 O-islands 57 and 71.

Notably, when nine optical maps were compared with Sakai and EDL933, five showed chromosomal inversions (Fig. 2b). In each case, fragments corresponding to annotated prophages were found flanking the chromosomal inversions (Table 1). A 7.7 % inversion, distinctly different from, but similar in size and location to the 7.8 % inversion found in EDL933, was mapped in EC1225. Although differences between similar inversions might be due to assorted phages lysogenizing at the same insertion sites prior to or after inversion, they could also be explained by independent recombination events occurring at different sites within homologous regions of the flanking prophages, thus creating discrete configurations of restriction fragments at the junctions.

There were four other inversions. The chromosomes of the four strains appear complex when their optical maps are aligned with Sakai, but appear simple when aligned with EDL933. The chromosome structures of these strains are the result of double inversions. This is shown schematically for EC502 in Fig. 3(a). The order of inversion is not critical, but the structures observed in the optical maps can be most simply explained by an inversion occurring within another inversion (Fig. 3a). Fig. 3(b) illustrates this inversion within an inversion with the optical map alignments of EC502 with the Sakai and EDL933 chromosome maps. For EC502, the large inversion extended from the EDL933-equivalent position of O-island 45, which contains prophage W in EDL933, to the equivalent position of EDL933 O-island 93, the site for prophage V. For EC502, in the alignment with Sakai, within the large inverted segment (yellow), the internal inversion (green) has regained its orientation and alignment to Sakai.
shown aligned with EDL933, the entire inversion in EC502 is yellow, since EDL933 contains the internal inversion. For the other large inversion in strain EC533, the overall structure is very similar, except that the inversion extended from the equivalent positions of EDL933 prophage M (O-island 44) to prophage U (O-island 79). The two large inversions represented 1.656 kb (30%) of the genome in EC502 and 1.482 kb (27%) in EC533. These large inversions encompass the smaller 429 kb EDL933 inversion.

For the other two inversions, in the chromosomes of AB1 and EC536, a smaller inversion is found within the EDL933 inversion, but the schematic model applies equally; only its scale is changed. In AB1, a 258 kb (4.7%) chromosomal inversion and in EC536, a 233 kb (4.2%) inversion has occurred within the 429 kb (7.8%) EDL933 inversion. Both inversions were flanked by prophage O located within O-island 57 and prophage R. Whereas the inversions found in AB1 and EC536 could have been considered the same at first glance, differences in the restriction fragments of flanking prophages suggest that they are distinct. Parenthetically, prophage R is found in E. coli K-12, and is not considered or annotated as an O-island.

**Complex chromosomal differences; substitutions, insertions and deletions**

Optical mapping alignments of the various strains also showed that portions of the chromosome were not aligned over stretches up to 150 kb (strain EC1231, Fig. 2a, strain

---

**Fig. 3.** Double chromosomal inversions. (a). Schematic representation of double inversion. A smaller chromosomal inversion (as found in EDL933 relative to Sakai and K-12) is shown first, illustrated by the smaller crossed lines. A second, larger inversion encompassing the smaller inversion is then shown. The double inversion shows (as found in EC502) the smaller initial inversion restored to its original orientation flanked by two separate inverted segments. Aligned chromosomes are shown in green, inverted segments are shown in yellow; however, the state of inversion depends on the pair of chromosomes aligned. These inversions are not drawn to scale. (b) Optical maps demonstrating double chromosomal inversions. The optical maps of EC502 aligned with the in silico map of Sakai and EC502 aligned with the in silico map of EDL933. The inversion within an inversion is apparent in the alignment with Sakai. Fragments within mini-optical maps of O-island prophages W and V flanking the larger inversion are highlighted in blue. A Mbp scale is shown at the bottom of the panel; the optical map inversions are to scale.
AB1, Fig. 2b). Many of the non-aligned fragments occurred at positions known in EDL933 to contain cryptic prophages. Supplementary Table S1 summarizes the prophage deletions and substitutions at novel chromosomal positions, including a number of novel insertions whose size and fragment distributions are suggestive of known prophages. Comparing each of the strains with the in silico EDL933 map, the extent of each change was measured and positioned relative to the EDL933 sequence. A total of 91 chromosomal mapping incongruities were identified at 28 unique loci within the 11 strains examined (Supplementary Table S1). Relative to strain EDL933, each strain exhibited chromosomal variation at seven to 18 sites, yielding a unique map with a different assortment of polymorphisms, including the five distinct chromosomal inversions. Two of the most variable sites are those occupied by the stx1 and stx2 prophages in EDL933, prophage W (stx2) at O-island 45 and prophage V (stx1) at O-island 93. Examples of each of the major types of variation are illustrated and detailed in the next section for strain EC536.

**Complex chromosomal differences in strain EC536**

The major chromosomal polymorphisms identified by the optical map of EC536 are shown in Fig. 4. An insertion containing five BamHI restriction fragments, totalling 26.9 kb, occurs in EC536 within a set of fragments corresponding to O-island 8 (Fig. 4a). In EDL933, O-island 8 contains two cryptic prophages H and I, spanning bp 300 060 to 323 540. Two chromosomal polymorphisms in Fig. 4. Seven chromosomal polymorphisms found in EC536 relative to EDL933. BamHI mini-optical maps were prepared for relevant O-island sequences from the complete annotated genome sequence of EDL933 and aligned to whole chromosomes. A Mbp reference scale is shown at the bottom of each panel. (a) O-islands 8, 43 and 45. EC536 contains five inserted fragments (27 kb) within O-island 8, encompassing the cryptic prophage complex H/I in EDL933. At polymorphic site 3, there are no fragments corresponding to a P4-like prophage (O-island 43 in EDL933), nor are there any restriction fragments at polymorphic site 5, representing prophage W (O-island 45 in EDL933, containing the Shiga toxin II genes) in EC536. (b) Three complex changes and a simple polymorphism in EC536. At polymorphic sites 7–9, equivalent to EDL933 sequences at 2.6 to 2.7 Mbp, the complex cryptic prophage islands 50-51-52 show a different set of fragments in EC536. The substitutions are indicated by the two sets of non-aligned white fragments in the two genome maps shown above the O-island maps. A inversion in EC536, spanning 4.2% of the EDL933 chromosome, is indicated by an X formed by crossover lines in the alignment and yellow highlighting. The inversion is flanked by fragments comparable to EDL933 prophage O (polymorphic site 10) and prophage R (polymorphic site 11). In a pair of aligned genomes, red fragments, usually prophages, indicate that more than two copies are found. At polymorphic site 18, at 2.8 Mbp, an insertion of six restriction fragments totalling 56 kb has occurred. Finally, at polymorphic site 20, an additional BamHI site is found within a probable prophage in the chromosomal position occupied in EDL933 O-island 93 by stx1 prophage V.
EC536 involved loss of prophages relative to their chromosomal positions in EDL933. Whereas EDL933 harbours a P4-like prophage sequence at O-island 43 (bp 1 060 000) and prophage W at O-island 45 (bp 1 310 000), optical mapping demonstrated that EC536 was devoid of prophages at these sites (Fig. 4a). The absence of a prophage at one chromosomal location does not mean, however, that the prophage or a related prophage is not located elsewhere on the chromosome. An example is EDL933 prophage W, which carries the Shiga-like toxin II gene, stx, at the chromosomal position annotated O-island 45 in EDL933. Although strain EC536 contains the stx2 gene as determined by PCR using stx2-specific primers, as well as by DNA microarray analysis, microarray results suggest that a number of prophage W genes in strain EC536 are absent (S. Jackson, I. Patel & M. L. Kotewicz, unpublished results). Since the O-island 45 site of prophage W is unoccupied in EC536 (Fig. 4a), the stx2 gene is probably located at a different chromosomal location. Five candidate sites were identified from the optical map: the insertion in the H/I prophage described above; the substitution within O-island complex 50, 51 and 52; both prophages, O and R, flanking the chromosomal inversion found in EC536; the phage-like insertion near O-island 79; and the altered prophage at O-island 93 (Fig. 4b). The latter two sites, chromosomal polymorphisms 18 and 20, represent prime candidates for the new location of stx2 in EC536. At site 18, a 55.7 kb insertion of six additional fragments, located 39.3 kb upstream of the O-island 79 site in EDL933, resembles the stxl prophage V found in EDL933. At site 20, the BamHI polymorphism is found within fragments representing O-island 93, prophage CP-933 V containing the stxl gene in EDL933. The stx2 gene found in EC536 may have replaced the stxl gene.

**DISCUSSION**

Optical mapping offers sufficient resolution and detail to differentiate closely related strains of *E. coli* O157 : H7. The differences observed in the optical maps suggest that prophage artefacts resident in an individual genome, the result of specific temporal and geographical phage encounters (insertions, deletions, rearrangements), can be useful biological markers. The variety and frequency of chromosomal changes revealed by optical mapping further support the growing list of prophage effects on chromosome structure across a wide range of enteric bacteria (Brussow et al., 2004).

Many of the chromosomal differences found in the optical maps of nine *E. coli* O157 : H7 strains relative to two reference strains were associated with known prophages. There are 22 entries for integrases in the annotated EDL933 genome (Perna et al., 2001; see GenBank AE005174). Only one fully competent prophage, BP-933W, is found in EDL933 (Plunkett et al., 1999), indicating, by the measure of integrase genes, that 21 defective prophages are present in the chromosome. The other integrase genes are present within truncated phage genomes, presumed to be defective because of missing blocks of genes and their inability to form complete phage particles when induced. The number of defective prophages annotated within the Sakai and EDL933 genomes suggests that since the time their intact progenitors took up residence in each pathogen, extensive editing by deletion has occurred to ‘lock in’ prophage remnants in different *E. coli* O157 : H7 strains. This is consistent with hypotheses about mechanisms to prevent the continual expansion of bacterial genomes (Ochman, 2005).

Marked heterogeneity among prophages and their integration sites in different *E. coli* O157 : H7 strains has been observed by genomic PCR scanning (Ohnishi et al., 2002), and further refined by combining PCR scanning with microarray analysis for adjudging presence or absence of specific gene sequences (Ogura et al., 2006). Although both optical mapping and DNA microarray interrogation of novel strains rely heavily on knowledge of previously sequenced genomes, the latter technique is incapable of detecting novel insertions. In contrast, as evidenced from the present study, this is a major strength of optical mapping. Optical mapping complements PCR- and cloning-based strategies for genome sequencing, facilitating closure of bacterial genomes like *E. coli* O157 : H7 when multiple copies of homologous phages might hinder sequence assembly (Perna et al., 2001).

The optical maps of nine *E. coli* O157 : H7 strains presented in the current study contained 91 differences at 28 polymorphic chromosomal loci relative to the EDL933 and Sakai sequence-based maps, and relative to each other. Three of the 28 polymorphic loci showed simple changes, such as the gain or loss of a BamHI site, but 25 sites contained more complex differences. Fifteen of the loci demonstrating complex changes in the optical maps were O-islands previously annotated as prophage or cryptic prophage sequences, and seven others are candidates for novel prophage insertion sites. Moreover, at several of the latter sites, the optical maps showed positions where a large insertion had occurred in one strain, whereas smaller insertions were found in other strains. For example, at polymorphic site 18, EC1231 contained a 97 kb insertion relative to EDL933; other strains, however, contained smaller insertions at this position, such as EC869 (61 kb), EC536 (56 kb) and a number of strains with 6 kb insertions. This suggests that phage acquisition and truncation occurred at these sites (Supplementary Table S1, polymorphic loci 18, 22 and 26). No other means has yet been used to demonstrate that the shorter insertions are truncated phages, and it is possible that smaller insertions have occurred at these sites rather than the truncation of large prophages. Whether these sites result from strain-specific scars of phage insertion and truncation, or other insertions, determining the likely origins of these changes through sequence analysis will be of interest.

In addition to altering the primary structure, multiple homologous prophages have another effect on chromosome structure. Five chromosomal inversions were found
among the nine strains examined. The optical maps revealed the sizes and locations of each inversion, and in each case the inversion was flanked by prophages. The positions of the ends of the inversions suggested that prophages served as targets for recombination, as has been found for inversions between artificially constructed chromosomal repeats (Miesel et al., 1994) or inversions found between copies of rRNA operons (Liu & Sanderson, 1995) and insertion elements in Salmonella (Alokam et al., 2002). The large number of cryptic prophages found in EDL933 and Sakai supply extensive 10–90 kb regions for homologous recombination. The inversion mapped for EC1225 was similar to the EDL933 inversion, and was flanked by prophages O and P (EDL933 O-islands 57 and 71). Careful examination of the prophage restriction fragments flanking the inversion in EC1225 suggested that either this inversion occurred within different restriction fragments than occurred in EDL933, i.e. it was a different inversion, or that different but related prophages were resident prior or subsequent to the inversion.

In the other four strains carrying inversions, AB1, EC536, EC533 and EC502, alignment of each optical map with the in silico chromosomal map of Sakai suggested a complex set of events over hundreds of kb. On the other hand, the alignments of these strains with the in silico EDL933 map suggested that, for all four strains, two simple chromosomal inversions in each strain might explain the chromosomal structures. For EC533 and EC502, two large inversions (26.8 and 29.9 % of the chromosome) flanking an EDL933-like inversion (7.8 % of the genome) explained their chromosomal maps. This does not exclude the reverse order of occurrence of a double inversion, that is, a novel large inversion followed by an EDL933-like inversion. The double inversions were flanked by prophage pairs, M/U or W/V (Table 1). For AB1 and EC536, a small 4.7 % inversion and a 4.2 % inversion within the larger 7.8 % EDL933-like inversion most simply explained their chromosomes. The prophage pair O and R was involved in these smaller inversions. As the flanking restriction fragments were different in EC536 and AB1, it is possible that they were independent inversions. Each of the four inversions is described most simply as an inversion within a chromosome containing an EDL933-like inversion, i.e. a double inversion.

There was concern about the preponderance of inversions found in the optical maps of these strains. Was it possible that the optical mapping assembly software created an artefact, inversions based on the assembly of similar flanking prophages? Two points refute this explanation. The prophages range in size from 10 to 90 kb. In contrast, the single molecule ‘reads’ performed in optical mapping are much longer. The minimum size of each molecule read is 200 kb, the mean size of the assembled molecules is 350 kb, and it is common to detect single molecules that extend over 600 kb. Additionally, optical maps are assembled to very high coverage; the minimal coverage was 30× and mean coverage was 50–100×. As a consequence, the overlap between molecules in the assembly is also very high, so that when two molecules are aligned, there is at least 150 kb of shared regions. In this context, even a relatively large prophage of 90 kb would be contained entirely within a single molecule. In order to overlap two molecules, the prophage would have to be flanked with at least 50 kb of additional information. Because of the length of the single mapped molecules, the accuracy of optical mapping is extremely robust in the presence of prophages, even when a similar prophage is found in multiple locations within a single genome.

The second point arguing against the chromosomal inversions being an artefact was independent PFGE and Southern experiments published during the preparation of this manuscript. Carefully detailed PFGE analysis of inversions occurring during passage of E. coli O157 : H7 EDL933 in the laboratory (Iguchi et al., 2006) found most of the inversions identified by optical mapping. An additional inversion between an EDL933 P4-like prophage M located at O-island 43 and another P4-like prophage at O-island 48 was also found after laboratory passage of EDL933. The large inversion in the optical map of strain EC533, representing 27 % of the EDL933 chromosome, was not found in the laboratory passage studies. Both optical mapping and PFGE results suggest that as additional strains are examined, additional inversions between other pairs of homologous phages will probably be found. There are, of course, chromosomal constraints that will modulate inversion formation, e.g. the positions of the origin and termination of chromosome replication (Iguchi et al., 2006).

In addition to showing that a single isolate of EDL933, during subculture in the laboratory, produced a number of different PFGE profiles derived from chromosomal inversions (Iguchi et al., 2006), three other E. coli O157 : H7 strains, including Sakai, showed changes in PFGE patterns after repeated subculture and prolonged storage (Shima et al., 2006). Chromosomal inversions in these strains were implicated in some of the changed PFGE profiles. Notably, a strain missing the stx1 prophage exhibited fewer changes in its PFGE profile than did strains carrying both stx1 and stx2 phages.

Overall, the optical map data presented here and other evidence (Iguchi et al., 2006; Shima et al., 2006) suggest that chromosomal inversions are ongoing and frequent in O157 : H7 strains. Although the frequency of inversions cannot be estimated from optical maps of temporally and geographically dispersed strains, the five sets of inversions in EDL933 identified in the laboratory studies arose during 20, 30 and 50 subcultures (Iguchi et al., 2006). Regarding the frequency of inversions and prophage changes in chromosomes, dissimilarities in the stx1 and stx2 gene profiles of particular O157 : H7 strains have been demonstrated to be not only the result of the heterogeneity in prophages containing the toxin genes but also due to the
unexpected variability in their integration sites (Shaikh & Tarr, 2003). Moreover, polymorphic PCR amplification patterns were an early indicator that insertions and deletions dominated the differences between O157 : H7 strains (Kudva et al., 2002a, b); many of these polymorphisms were found within prophages. The optical map data confirm earlier suggestions that polymorphisms in O157 are not sequence-based changes in XbaI or BamHI sites, but rather that PFGE polymorphisms reside within changing prophage and chromosome profiles.

Thus, chromosomal inversions and phage trafficking are major contributors to outbreak-specific PFGE profiles. The real-world exposure of an O157 strain to environmental stress, to host selection and to teeming populations of competing phages alters the chromosomes of this pathogen in a feedlot-by-feedlot, patient-by-patient and temporally and geographically specific manner. The scars of prophage deletions represent locked-in variants that accumulate and which will be documented over the next few years by sequencing of E. coli O157 : H7 strains. The first of these chromosomal scars was found by comparing EDL933 and Sakai genomic sequences.

It is thus not surprising that different sets of phages in different insertion sites create a backdrop where, in the course of an outbreak, recombinational scrambling of phages, exacerbated by antibiotic treatments causing induction of prophages (Zhang et al., 2000; Kohler et al., 2000) or stress and SOS repair (Kimmitt et al., 2000), would further increase the variety of phages found (Schmidt, 2001; Herold et al., 2004; Bielaszweska et al., 2006). The occurrence of multiple inversion events demonstrated by the optical mapping in four strains indicates that one inversion might shuffle phage genes between a defective and infective prophage, thus altering the configuration of virulence genes typified by stx1 and stx2 genes. The initial recombination that translocated phage and virulence genes could be masked by the occurrence of a second inversion within the same pair of phages, or exposed by a second inversion involving a different phage pair. While changes in the toxin structural genes themselves can moderate toxin production, so too can changes in transcriptional regulation of genes (Zhang et al., 2005; Koitabashi et al., 2006), further emphasizing how recombinational events might potentiate or attenuate virulence.

Genomic comparisons of an assortment of microbes have not only revealed the extensive diversity that exists within a bacterial species, but they also are helping to challenge the very concept of a bacterial ‘species’ itself. For example, although there are only 75 000 single nucleotide polymorphisms (SNPs) within the homologues of E. coli O157 : H7 strain EDL933 and E. coli K-12 strain MG1655, nearly a million base pairs of unique DNA distinguish these two strains. Moreover, as individual strains of O157 : H7 show insertions and deletions of hundreds of thousands of base pairs, while exhibiting only a few hundred SNPs, methods that assess both rates of mutation and recombination must be used to assess the genesis, evolution and lineage of a particular strain. Optical mapping offers a relatively economical and rapid method for analysing insertions, deletions, inversions and rearrangements – those anomalies that ultimately shape the organization and circumscribe the gene order of an individual genome or strain.

The resolution and detail that optical maps provide underscore their usefulness for molecular epidemiological studies. The method should also prove useful for microbial forensics investigations, which require technologies that discriminate between closely related strains or that detect insertions in engineered strains (Budowle et al., 2005). In this respect, optical mapping offers a tractable solution for triaging microbial strains for complete genome sequencing.

ACKNOWLEDGEMENTS

We acknowledge the National Bioforensics Analysis Center (NBFAC) of the Department of Homeland Security for supporting work on optical mapping of E. coli O157 : H7 strains reported here. We thank Adam Briska from OpGen for discussions on optical mapping accuracy and assembly software. We acknowledge the following colleagues who have supplied E. coli strains used in this work: Dr Andrew Benson, University of Nebraska, Lincoln, NE; Dr Robert Buchanan, US FDA, College Park, MD; Dr Peter Feng, US FDA, College Park, MD; Dr Phillip Tarr, Washington University, St Louis, MO; and Dr Thomas Whittam, Michigan State University, East Lansing, MI. We would like to dedicate this publication to the memory of Dr Harrison (Hatch) Echols. One of us (M. L. K.) is greatly indebted to Hatch and the Berkeley bacteriophage lambda group he led.

REFERENCES


Edited by: J. Parkhill