

## Strains of *Escherichia coli* O157:H7 Differ Primarily by Insertions or Deletions, Not Single-Nucleotide Polymorphisms

Indira T. Kudva,<sup>1</sup> Peter S. Evans,<sup>2†</sup> Nicole T. Perna,<sup>3</sup> Timothy J. Barrett,<sup>4</sup> Frederick M. Ausubel,<sup>5,6</sup> Frederick R. Blattner,<sup>2</sup> and Stephen B. Calderwood<sup>1,7\*</sup>

Division of Infectious Diseases<sup>1</sup> and Department of Molecular Biology,<sup>5</sup> Massachusetts General Hospital, Boston, Massachusetts 02114; Department of Microbiology and Molecular Genetics<sup>7</sup> and Department of Genetics,<sup>6</sup> Harvard Medical School, Boston, Massachusetts 02115; Laboratory of Genetics<sup>2</sup> and Department of Animal Health and Biomedical Sciences,<sup>3</sup> University of Wisconsin—Madison, Madison, Wisconsin 53706; and Foodborne and Diarrheal Diseases Branch, Centers for Disease Control, Atlanta, Georgia 30333<sup>4</sup>

Received 17 September 2001/Accepted 31 December 2001

***Escherichia coli* O157:H7 (O157) strains demonstrate varied pulsed-field gel electrophoresis patterns following *Xba*I digestion, which enable epidemiological surveillance of this important human pathogen. The genetic events underlying PFGE differences between strains, however, are not defined. We investigated the mechanisms for strain variation in O157 by recovering and examining nucleotide sequences flanking each of the *Xba*I restriction enzyme sites in the genome. Our analysis demonstrated that differences between O157 strains were due to discrete insertions or deletions that contained the *Xba*I sites polymorphic between strains rather than single-nucleotide polymorphisms in the *Xba*I sites themselves. These insertions and deletions were found to be uniquely localized within the regions of the genome that are specific to O157 compared to *E. coli* K-12 (O islands), suggesting that strain-to-strain variation occurs in these O islands. These results may be utilized to devise novel strain-typing tools for this pathogen.**

*Escherichia coli* O157:H7 (O157) is an important human pathogen, implicated in several outbreaks in the United States, Europe, and Japan, as well as in sporadic infections (7, 9). In humans, the primary clinical manifestation of infection is bloody diarrhea, which may progress to the hemolytic-uremic syndrome (9). O157 apparently evolved from an enteropathogenic *E. coli* O55:H7 ancestor, bearing the locus for enterocyte effacement, by the acquisition of at least the bacteriophage-encoded Shiga toxin type 1 (*stx*<sub>1</sub>) and/or 2 (*stx*<sub>2</sub>) gene, a virulence plasmid, and the transition of somatic antigen O55 to O157 (6). The complete genomic sequence of one strain of O157 has recently been determined (15). However, different strains of this pathogen have different patterns by pulsed-field gel electrophoresis (PFGE) following *Xba*I digestion (8), and the molecular events underlying this strain variation in O157 are unknown. In this study, we recovered the nucleotide sequences flanking each *Xba*I restriction enzyme site in the O157 genome and used these sequences to examine genomic differences across a well-defined strain set of O157 isolates.

### MATERIALS AND METHODS

**Bacteria.** O157 strain 86–24, streptomycin resistant and originally isolated from a human in an outbreak in the state of Washington, was kindly provided by A. D. O'Brien. O157 strain EDL933, a human isolate from an outbreak in Michigan, was obtained from the American Type Culture Collection (Manassas, Va.) (ATCC 43895). Strain EDL933 is the O157 isolate that has been sequenced at the University of Wisconsin—Madison (15). In addition, 44 isolates of O157,

2 each from 22 different outbreaks collected by the Centers for Disease Control and Prevention (CDC), were also included in this study. The isolates from different outbreaks had different PFGE patterns, suggesting genetic heterogeneity among them. Forty-two of the 44 isolates were recovered from cases of human infection.

**Design of primer pairs amplifying O157 *Xba*I sites.** Genomic DNAs from O157 strains 86–24 and EDL933 were initially fragmented using *Sau*3AI (strain 86–24) or *Nla*III (strain EDL933), followed by self-ligation (Fig. 1). The circularized DNA was digested with *Xba*I to linearize only fragments containing an internal *Xba*I site. These fragments were cloned into pBluescribe (Stratagene, La Jolla, Calif.) and sequenced, and the 22 distinct insert sequences were used to design primer pairs (prefixed IK) flanking different *Xba*I restriction sites in these two reference strains. Direct examination of the O157 strain EDL933 genome sequence revealed an additional 18 *Xba*I sites that had not been recovered as described above, and chromosomal sequences around these sites were used to design an additional 18 primer pairs with the prefix IKB, yielding a total of 40 primer pairs for the study.

**PCR conditions.** Colony lysates were prepared by boiling colonies suspended in sterile distilled water, followed by centrifugation at 4°C. Each O157 strain template was tested with each individual primer pair. PCR was carried out on the GeneAmp PCR system 2400 thermal cycler (PE Biosystems, Foster City, Calif.), using 10 µl of colony lysate, 200 pmol of each primer, 800 µM deoxynucleoside triphosphates, 1× diluted Ex *Taq* enzyme buffer, and 2.5 U of TaKaRa Ex *Taq* DNA polymerase. The hot-start PCR technique (4) was employed in combination with a touchdown PCR profile (5). To create this profile, an amplification segment of 20 cycles was set in which the annealing temperature started at 73°C with touchdown at 53°C at the end of those cycles. Then, another amplification segment of 10 cycles was set, using the last annealing temperature of 53°C.

**Evaluation of amplicons.** Amplicons obtained by PCR were purified using the Qiaquick PCR purification kit and digested with *Xba*I to confirm the presence of an *Xba*I site within the amplicon. Undigested and digested DNA fragments were resolved on a 4% agarose gel prepared with a combination of 3% Nusieve GTG agarose (FMC BioProducts, Rockland, Maine) and 1% agarose (Shelton Scientific Inc., Shelton, Conn.) and stained with ethidium bromide. Sequencing of purified amplicons was done at the DNA Sequencing Core Facility, Department of Molecular Biology, Massachusetts General Hospital.

**Southern blotting.** DNA was fractionated by agarose gel electrophoresis, transferred to Hybond-N+ membranes (Amersham Pharmacia Biotech, Inc., Piscataway, N.J.), UV cross-linked to the membrane using a Stratalinker (Stratagene), and hybridized with the appropriate probe, labeled using the ECL direct

\* Corresponding author. Mailing address: Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA 02114. Phone: (617) 726-3811. Fax: (617) 726-7416. E-mail: scaldwood@partners.org.

† Present address: Institute of Human Virology, Baltimore, MD 21201.

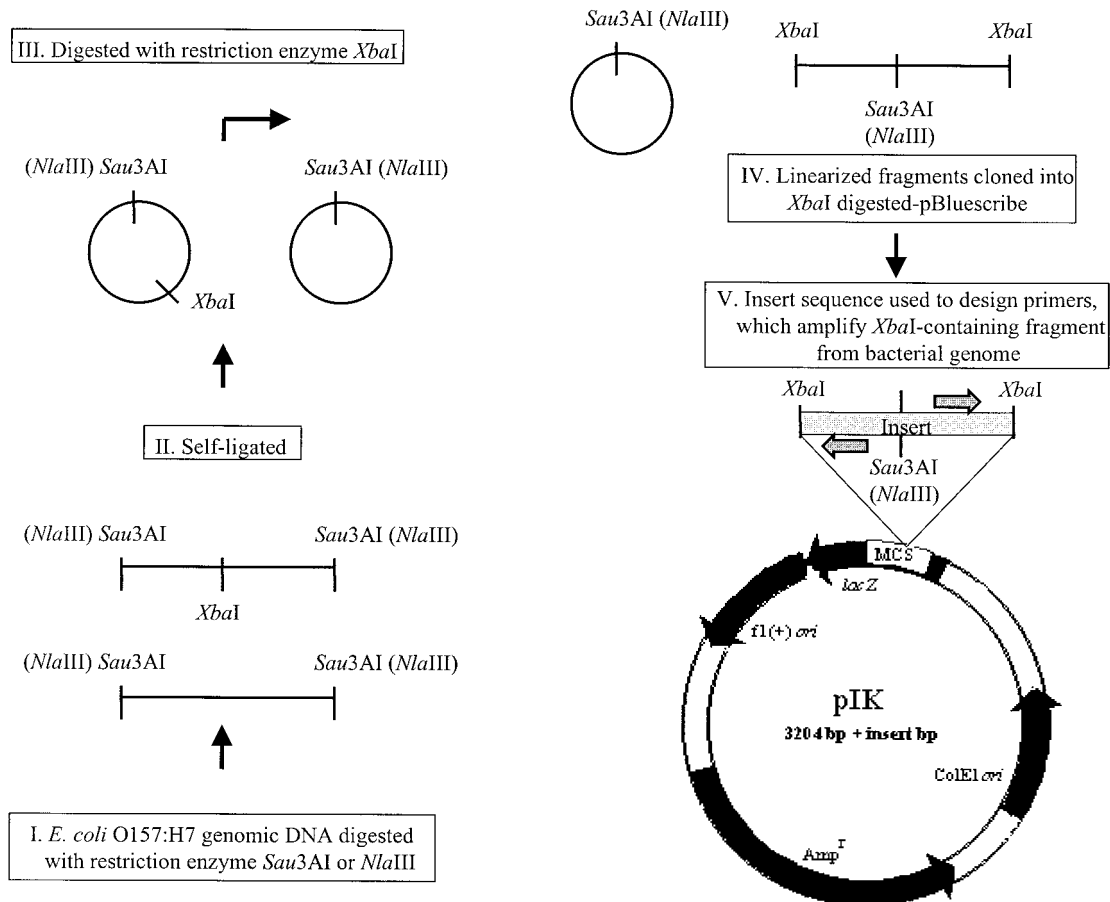


FIG. 1. Protocol for the design of primer pairs. Genomic DNA fragments derived from O157 strains 86–24 and EDL933, containing an *XbaI* restriction site, were selectively cloned into pBluescribe. DNA was initially fragmented using *Sau3AI* (strain 86–24) or *NlaIII* (strain EDL933) and self-ligated. The circularized DNA was then digested with *XbaI* to linearize only fragments containing an internal *XbaI* site. Cloning of these fragments resulted in plasmids of various sizes that were prefixed pIK. The insert sequences were determined and used to design primer pairs, shown as divergent shaded arrows, which flank *XbaI* restriction sites in the bacterial genome. MCS, multiple cloning site.

nucleic acid labeling and detection system (Amersham Pharmacia). Hybridization at 42°C and posthybridization washing of the blots was done in accordance with the ECL kit manual. Autoradiographs were prepared by exposure of processed blots to Scientific Imaging X-OMAT AR film (Eastman Kodak Company, Rochester, N.Y.).

**Data analysis.** Statistical analysis was performed using the EpiInfo6 (CDC) software. The significance of differences in proportions was calculated with Fisher's exact test. DNA G+C content was determined using the Wisconsin Package version 10.0 (Genetics Computer Group, Madison, Wis).

## RESULTS

We initially focused on two well-characterized O157 reference strains, EDL933 and 86–24; a total of 40 *XbaI* sites were identified between the genomes of these two strains. Primer pairs were designed that flanked each of these 40 unique *XbaI* sites and amplified approximately 200- to 400-bp genomic fragments containing the sites. The presence or absence of an amplicon with each primer pair in each strain, as well as the presence or absence of an *XbaI* site within each amplicon, was assessed by PCR, *XbaI* digestion, and agarose gel electrophoresis. The majority of the primer pairs (36 of 40) amplified *XbaI*-containing DNA fragments of identical sizes from both strains. However, two primer pairs derived from strain EDL933

failed to yield an amplicon with strain 86–24, and two primer pairs derived from strain 86–24 did not yield amplicons when strain EDL933 DNA was used as the template, suggesting that this primer set could distinguish between the strains.

**Analysis of 44 isolates of O157 for genomic differences.** We then used these 40 primer pairs to analyze 44 O157 isolates, 2 isolates each from 22 different outbreaks, collected by the CDC. Thirty-two of the 40 primer pairs produced identical results in all 44 isolates, generating amplified products of identical sizes and containing an internal *XbaI* site. None of the 40 primer pairs generated an amplified product that lacked an *XbaI* site from any strain, which might have suggested a single-nucleotide polymorphism in the restriction site itself. Eight primer pairs produced polymorphic results across the isolate set, amplifying identically sized products with an *XbaI* site in some isolates but failing to amplify any product in others. In these cases, the presence or absence of an amplicon by PCR correlated with the presence or absence of a hybridizing fragment by Southern blot analysis, using the control PCR amplicon to probe digested genomic DNA (not shown). A single exception was observed with one amplicon (IK8) as a probe. This fragment hybridized to genomic DNA from all strains by

Southern blotting, irrespective of the PCR results. Further evaluation revealed partial overlap of this amplicon with the IS629 *tnp* gene, which is widely distributed over the O157 genome (see below).

The DNA sequences amplified by the 40 primer pairs were analyzed using the GenBank database (BLAST search program; National Center for Biotechnology Information) and the O157 strain EDL933 genome sequence database (University of Wisconsin [http://www.genome.wisc.edu]). Of the 40 O157 *Xba*I-containing genome sequences amplified by the primer pairs, 18 were homologous to *E. coli* strain K-12 genome sequences (referred to as backbone sequences) (15) and 22 were in regions of the O157 chromosome not shared with K-12 (referred to as O islands) (15). All 8 of the polymorphic regions (present in some but not in other O157 isolates) were localized to O islands, while only 14 of the 32 regions conserved across all isolates tested were in O islands ( $P < 0.01$ ). This difference suggests that the major genomic differences between O157 strains occur in these O island sequences.

**Molecular analysis of genomic differences between strains.** We analyzed three of the eight polymorphic regions in more detail to gain insight into the mechanisms underlying strain differences. Additional primers were designed from either EDL933 or 86–24 genomic sequences to amplify and sequence regions upstream or downstream from or across the polymorphic regions being evaluated. Our analysis confirmed that all three regions examined, defined by primer pairs IK8A and -B, IKB3A and -B, and IK118A and -B, were polymorphic in different O157 isolates because of discrete insertions or deletions in the genome that contained *Xba*I sites rather than because of single-nucleotide polymorphisms in the *Xba*I sites themselves.

**(i) Polymorphism attributable to an insertion in the virulence plasmid.** Polymorphism between isolates for the *Xba*I-containing fragment amplified by IK8A and -B, for example, was a consequence of a small insertion in the virulence plasmid. Using the primer pair IK8A and -B, an amplicon referred to as IK8 was obtained from O157 strain 86–24 but not from strain EDL933. This amplicon was mapped to the virulence plasmid pO157 in strain 86–24. The IK8 sequence extended from a region of unknown function into a transposase gene (IS629 *tnp*) on the plasmid (Fig. 2a). The region of unknown function occurred as a 635-bp insertion in the DNA between the resolvase (*redF*) and IS629 *tnp* genes in strain 86–24 compared to the sequence of the same region in pO157 from O157 strain EDL933 (Fig. 2a; GenBank accession no. AF074613) (3). The insertion in strain 86–24 contains an *Xba*I site and is identical to a similarly situated insertion in the virulence plasmid of the O157 strain isolated from the outbreak in Sakai, Japan (GenBank accession no. AB011549) (11).

For further analysis of the IK8 amplicon, primer pairs IK8C and -D, IK8E and -F, and IK8G and -H were designed to amplify sections of *redF*, IS629 *tnp*, and the insertion in strain 86–24 (Fig. 2b). Identical amplicons were obtained from strains 86–24 and EDL933 using the first two sets of primers, indicating conservation of the respective genes on both plasmids (Fig. 2c); these amplicons did not digest with *Xba*I. On the other hand, an amplicon was obtained with IK8G and -H only from strain 86–24 (Fig. 2c), and it contained an *Xba*I site (data not shown). The primer combination of IK8C and -F was

used to amplify the entire length of this region in both strains. The size difference in the resulting amplicons (1.2 kbp from strain 86–24 and 613 bp from strain EDL933) confirmed the earlier observation that only pO157 from strain 86–24 contained a 635-bp insertion between bp 850 and 851 of pO157 in strain EDL933 (Fig. 2a). BLAST search analysis revealed no homologies for the inserted sequence in strain 86–24.

The same primer pairs were used to analyze additional isolates of O157 (G5320, G5327, G5303, and G5323), chosen randomly from the CDC isolates that did not yield an amplicon with primer pair IK8A and -B. Amplicons derived from isolates G5320 and G5327, using primer pair IK8C and -F, were the same size as that from strain EDL933 (613 bp), indicating the absence of an insertion (Fig. 2a). Amplicons from isolates G5303 and G5323 using these primers were 1.3 kbp in size, but these amplicons did not contain an internal *Xba*I site (Fig. 2a). Failure to obtain amplicons from isolates G5303 and G5323 with primer pairs IK8A and -B and IK8G and -H showed that isolates G5303 and G5323 contained an insertion different from that in 86–24. The sequences flanking the point of insertion were, however, identical for all isolates tested, including 86–24, G5303, and G5323 (Fig. 2a). BLAST search analysis revealed that the insert in isolates G5303 and G5323 had 99% homology to three open reading frames (ORFs), L0013, L0014, and L0015, in the locus for enterocyte effacement pathogenicity island of O157 strain EDL933 (14). These three ORFs comprise IS *Ec8* in strain EDL933, an insertion element similar to IS *Rm14* present in *Rhizobium* and *Agrobacterium* plasmids (19); however, the homologous insert in isolates G5303 and G5323 contained only part of the L0015 ORF and not the complete insertion sequence element. The G+C content for the sequences shared among all isolates (Fig. 2a) was 51%, while that for the inserted sequence in strain 86–24 was 33% and that for the inserted sequence in strains G5303 and G5323 was 55%. The lower G+C content of the insert in strain 86–24 suggests it is possibly of heterologous origin (1, 2). These observations suggested that polymorphisms between different strains of O157 can reflect the acquisition or loss of discrete segments of DNA in the genome, at least some of which may be of heterologous origin.

**(ii) Polymorphism attributable to a substitution-insertion in a lysogenic bacteriophage.** Similar analysis of the *Xba*I-containing fragment amplified by IKB3A and -B linked the polymorphism in this region to a substitution-insertion in a lysogenic bacteriophage. Using the primer pair IKB3A and -B, an amplicon was obtained from O157 strain EDL933 but not from strain 86–24. This amplicon, referred to as IKB3, was mapped to the lysogenic bacteriophage 933W in strain EDL933 (GenBank accession no. AF125520) (16). The IKB3 sequence overlapped a 2,091-bp insertion, containing an internal *Xba*I site, which was present between the antiterminator protein (*N*) and repressor protein (*cI*) genes in phage 933W (Fig. 3a). This insertion replaced a 1,439-bp sequence, located at exactly the same site on a similar bacteriophage in O157 strain 86–24 but which lacked an *Xba*I site (Fig. 3a). Four additional isolates, G5290, G5325, G5296, and G5301, chosen randomly from the CDC isolates that did not yield an amplicon with primer pair IKB3A and -B, were analyzed using a primer pair, IKB3E and -J, that would amplify the entire length of this substitution-insertion (Fig. 3a). No amplicons were obtained from isolates

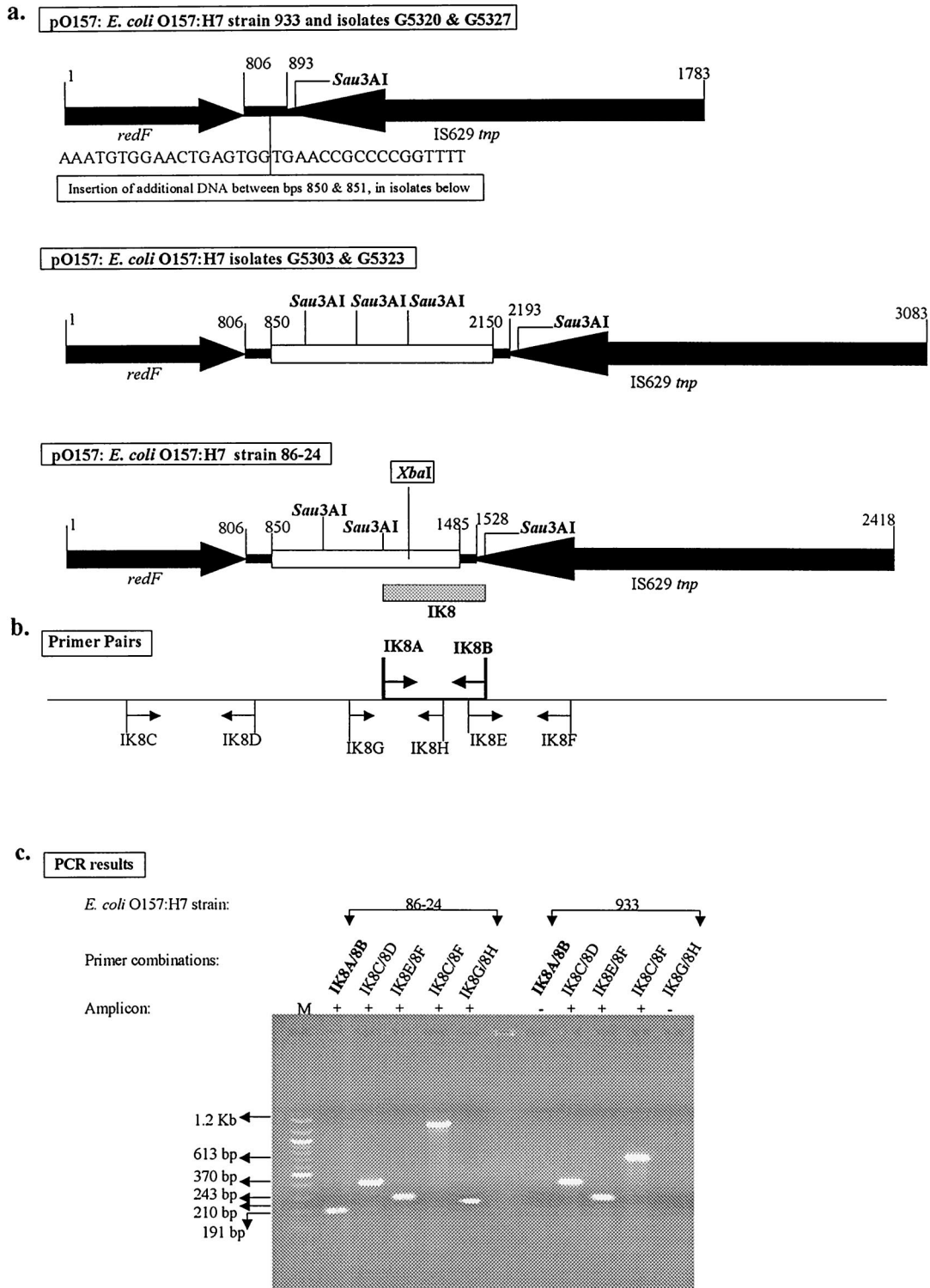


FIG. 2. *Xba*I restriction site polymorphism in O157 strains attributable to an insertion in the virulence plasmid. (a) Comparison of pO157 DNAs from strain EDL933 (933); isolates G5320, G5327, G5303, and G5323; and strain 86-24. The arrows indicate the directions of transcription of the designated genes. Identical regions are shown as solid, and the inserts that differ among the strains are shown as open. The insertions in isolates G5303 and G5323 are identical but differ from that in strain 86-24. The insertion in strain 86-24 contains an *Xba*I site. Fragment IK8 (shaded), amplified by primer pair IK8A and -B, maps to a 635-bp insertion in an intergenic region in pO157 DNA from strain 86-24. The insertions in isolates G5303 and G5323 occur at positions in the intergenic region identical to that in strain 86-24. (b) Original primers (shown in boldface) and additional primers used for further analysis of the polymorphisms between strains. The primers are in direct alignment with the regions in pO157 DNA from strain 86-24 used to design them. (c) Agarose gel electrophoresis pattern of amplicons derived using these primer pairs. The results confirm the polymorphisms between strains 86-24 and EDL933 diagrammed in panel a. M, molecular size marker (100-bp DNA ladder; New England Biolabs, Inc., Beverly, Mass.); +, presence of amplicon; -, absence of amplicon.



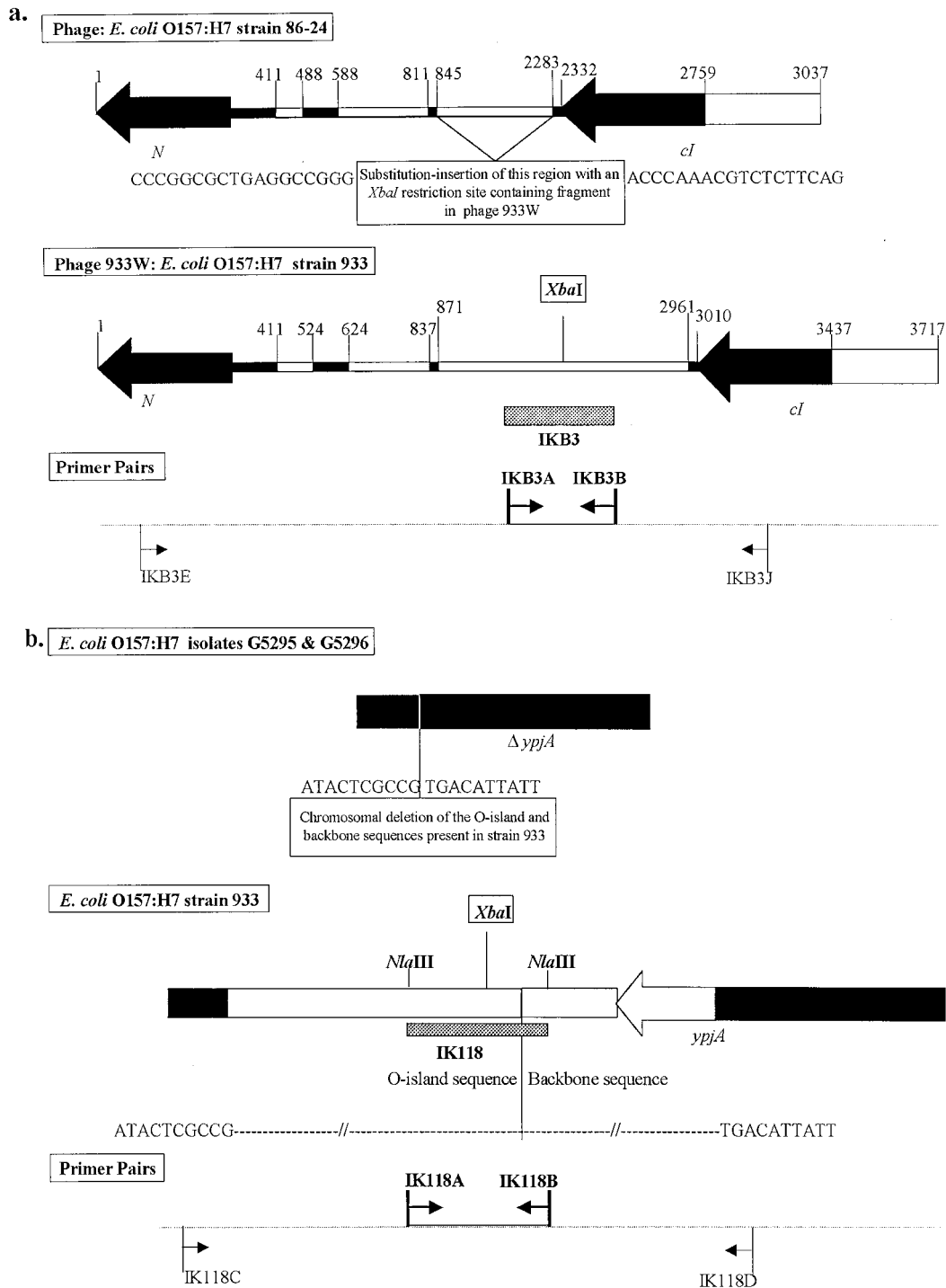


FIG. 3. Diagrammatic representation of other *Xba*I restriction site polymorphisms identified in O157 strains. (a) Polymorphism attributable to a substitution-insertion in a lysogenic bacteriophage. Lysogenic phage DNAs from O157 strain 86-24 and strain EDL933 are compared. Identical regions are shown as solid, and regions that differ between the two strains are shown as open. Strain EDL933 contains a 2,091-bp substitution-insertion, containing an *Xba*I restriction site, between the *N* and *cI* genes in place of a 1,439-bp fragment without an *Xba*I site at an identical position in strain 86-24. Fragment IKB3 (shaded), amplified by the primer pair IKB3A and -B, maps to the substituted region within phage 933W from strain EDL933. The sequences flanking the substitution-insertion are identical between the two strains. Original primers (shown in boldface) and additional primers used for further analysis of this polymorphism between the strains are depicted. The primers are in direct alignment with the regions in phage 933W used to design them. (b) Polymorphism attributable to a chromosomal deletion. Chromosomal DNA segments from O157 isolates G5295 and G5296 and strain EDL933 are compared. Identical regions are shown as solid, and regions that differ between the strains are shown as open. Fragment IK118 (shaded), amplified by primer pair IK118A and -B, maps to a chromosomal region at an O island-backbone junction in strain EDL933 and contains an *Xba*I restriction site in the O island sequence. Isolates G5295 and G5296 have a deletion in this region, resulting in the loss of the sequence containing the *Xba*I restriction site. Original primers (shown in boldface) and additional primers used for further analysis of this polymorphism between the strains are depicted. The primers are in direct alignment with the regions in the DNA from strain EDL933 used to design them.

TABLE 1. Description of regions polymorphic among *E. coli* O157:H7 isolates and associated with O islands in the strain EDL933 genome

| Amplicon derived from <i>E. coli</i> O157:H7 isolates | Length of associated O island in strain EDL933 (bp) | Position of <i>Xba</i> I site from one end of O island (bp) | Description of O island   | Relation of O island to <i>E. coli</i> K-12 genome |
|---|---|---|---|--|
| IKB3  | 61,664  | 11,088  | Stx2-encoding prophage BP-933w  | Insertion  |
| IK118   | 21,681  | 21,637  | Cryptic prophage CP-933Y  | Replaces unrelated sequences in K-12               |
| IKB5  | 49,798  | 36,431  | Cryptic prophage CP-933R  | Partial homology to cryptic prophage Rac of K-12   |
| IK114   | 44,434  | 8,367   | Large island adjacent to <i>leuX</i> ; includes a putative site-specific integrase or recombinase, several IS elements, putative helicases, and numerous unknowns | Replaces unrelated sequences in K-12               |
| IK123   | 80,502  | 35,859  | Cryptic prophage CP-933O  | Replaces unrelated sequences in K-12               |
| IK127   | 21,120  | 19,318  | Cryptic prophage CP-933T  | Insertion  |

G5325, G5296, and G5301, indicating that this region in these isolates is even more divergent than the region in 86–24 is from the region in EDL933. This was confirmed by additional PCRs using primer pairs designed to amplify various segments of the region between IKB3E and IKB3J, which also failed to yield amplicons from the three isolates. In contrast, the primer pair IKB3E and -J yielded an amplicon from isolate G5290 that was identical in size to that from strain 86–24 and lacked an *Xba*I site. Thus, this region has at least three variants in the O157 population studied.

**(iii) Polymorphism attributable to a chromosomal deletion.**

Analysis of a third *Xba*I-containing fragment, amplified by IK118A and -B, which also differed among isolates demonstrated a polymorphism linked to a deletion in the chromosome. Using the primer pair IK118A and -B, an identical amplicon containing an *Xba*I site was obtained from most O157 isolates tested. This amplicon, referred to as IK118, was mapped to a chromosomal DNA segment in O157 strain EDL933 that extended across a junction between O island and backbone sequences (Fig. 3b). The backbone sequence contained the putative transport gene *ypjA* (GenBank accession no. AE000350) (14, 18). While this entire region, along with its *Xba*I site, was conserved in most of the O157 isolates tested, no amplicons were obtained from isolates G5295 and G5296 using IK118A and -B. O157 strain EDL933 and isolates G5295 and G5296 were analyzed using the primer pair IK118C and -D that amplified across part of the O island and backbone sequence into the 3' end of *ypjA* (Fig. 3b). A 2.5-kbp amplicon containing an *Xba*I site was obtained from strain EDL933. In contrast, isolates G5295 and G5296 yielded a 700-bp amplicon that lacked an *Xba*I site, resulting from a deletion of the O island-backbone sequence junction, including part of the 3' end of *ypjA* (Fig. 3b). This deletion in G5295 and G5296 may have been caused by the excision of a prophage in these isolates; cryptic prophage genes have been identified in the O island region adjacent to this O island-backbone junction in O157 strain EDL933 (Table 1) (15).

In addition to IK8, IKB3, and IK118, the remaining five polymorphic regions were also found in O islands absent from the K-12 genome. Six of the eight polymorphic regions were present in strain EDL933, and the availability of the genome sequence of this strain allowed us to determine the properties of the O islands containing these six regions (Table 1). The remaining two polymorphic regions (IK8 and IK25) were present in strain 86–24 but not in the sequenced strain EDL933, so we were not able to define their larger genomic context.

## DISCUSSION

Acquisition of foreign DNA through horizontal DNA transfer is a hallmark of enterobacterial genome evolution (10, 12, 17). The divergence of *E. coli* from *Salmonella enterica* was shaped by the acquisition of approximately 755 genes through at least 234 horizontal-transfer events (10, 12). Comparison of the O157 strain EDL933 genome with that of *E. coli* strain K-12 indicates similar evolution of the O157 genome through horizontal-transfer events (13, 15, 17). The sequenced O157 genome contains 1.34 Mbp of DNA that is not present in strain K-12 and lacks 530 kbp of DNA present in K-12 (15). The O157-specific DNA in this strain is found in 177 O islands, which have a lower G+C content than the DNA sequences shared between O157 and K-12 (15). The mechanisms underlying differences in genomic structure among the sequenced O157 strain, EDL933, and other strains of this pathogen, however, have not previously been examined in detail, and it was not known whether strains of O157 would differ from each other by point mutations (as for many other bacteria that are closely related) or would differ by insertions and deletions (as O157 differs from *E. coli* K-12).

In this study, we observed that the presence or absence of polymorphic *Xba*I sites in the genomes of O157 strains that differed by PFGE reflected the presence or absence of discrete *Xba*I-containing segments of DNA in the individual genomes rather than point mutations in the *Xba*I sites themselves. The inserted sequences containing the polymorphic *Xba*I sites were often small and usually neither encoded a functional ORF nor disrupted a preexisting ORF. One exception was the chromosomal deletion observed in isolates G5295 and G5296, which resulted in the loss of 732 bp in the 3' end of *ypjA*. However, this deletion apparently did not affect either the viability or pathogenicity of these isolates, as they were recovered from human infections. The inserted sequences analyzed did not resemble known, intact insertion sequences, transposons, or bacteriophages. However, several of the inserted sequences were found within O islands that contained nearby cryptic prophage genes, suggesting that phage-mediated events may underlie their acquisition or loss. Sequences that characterize mutational hot spots or other composition variations (20) were not observed in the sequences flanking the insertion points, although each of a set of insertions occurred at exactly the same nucleotide position between strains. The 44 O157 isolates could be differentiated solely by the presence or absence of the eight regions of DNA containing the polymorphic *Xba*I sites.

This method of classifying O157 strains should, in principle, be similar to strain-typing methods that employ PFGE analysis of *Xba*I-digested genomic DNA.

#### ACKNOWLEDGMENTS

We acknowledge technical assistance provided by G. Joel DeCastro and Francis O'Neill.

I.T.K. is the recipient of a training grant from the National Institutes of Allergy and Infectious Diseases (T32 AI07061). P.S.E. was supported by the Centers for Disease Control/APHL Emerging Infectious Diseases Fellowship Program.

#### REFERENCES

- Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–1474.
- Boerlin, P. Evolution of virulence factors in Shiga-toxin-producing *Escherichia coli*. *Cell. Mol. Life Sci.* **56**:735–741.
- Burland, V., Y. Shao, N. T. Perna, G. Plunkett, H. J. Sofia, and F. R. Blattner. 1998. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res.* **26**:4196–4204.
- Dieffenbach, C. W., and G. S. Dveksler. 1995. PCR primer—a laboratory manual. Cold Spring Harbor Laboratory Press, New York, N.Y.
- Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker, and J. S. Mattick. 1991. “Touchdown” PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**:40008.
- Feng, P., K. A. Lampel, H. Karch, and T. S. Whittam. 1998. Genetic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. *J. Infect. Dis.* **177**:1750–1753.
- Griffin, P. M., S. M. Ostroff, R. V. Tauxe, K. D. Greene, J. G. Wells, J. H. Lewis, and P. A. Blake. 1988. Illnesses associated with *Escherichia coli* O157:H7 infections. A broad clinical spectrum. *Ann. Intern. Med.* **109**:705–712.
- Harsono, K. D., C. W. Kaspar, and J. B. Luchansky. 1993. Comparison and genomic sizing of *Escherichia coli* O157:H7 isolates by pulsed-field gel electrophoresis. *Appl. Environ. Microbiol.* **59**:3141–3144.
- Kaper, J. B., and A. D. O'Brien. 1998. *Escherichia coli* O157:H7 and other Shiga toxin-producing *E. coli* strains. ASM Press, Washington, D.C.
- Lawrence, J. G., and H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* **95**:9413–9417.
- Makino, K., K. Ishii, T. Yasunaga, M. Hattori, K. Yokoyama, C. H. Yutsudo, Y. Kubota, Y. Yamaichi, T. Iida, K. Yamamoto, T. Honda, C. G. Han, E. Ohtsubo, M. Kasamatsu, T. Hayashi, S. Kuhara, and H. Shinagawa. 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.
- Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299–304.
- Ohnishi, M., C. Tanaka, S. Kuhara, K. Ishii, M. Hattori, K. Kurokawa, T. Yasunaga, K. Makino, H. Shinagawa, T. Murata, K. Nakayama, Y. Terawaki, and T. Hayashi. 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.
- Perna, N. T., G. F. Mayhew, G. Posfai, S. Elliott, M. S. Donnenberg, J. B. Kaper, and F. R. Blattner. 1998. Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* **66**:3810–3817.
- Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamouisis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7. *Nature* **409**:529–533.
- Plunkett, G., III, D. J. Rose, T. J. Durfee, and F. R. Blattner. 1999. Sequence of Shiga toxin 2 phage EDLW from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J. Bacteriol.* **181**:1767–1778.
- Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam. 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406**:64–67.
- Rudd, K. E. 1998. Linkage map of *Escherichia coli* K-12, edition 10: the physical map. *Microbiol. Mol. Biol. Rev.* **62**:985–1019.
- Schneiker, S., B. Kosier, A. Pühler, and W. Selbitschka. 1999. The *Sinorhizobium meliloti* insertion sequence (IS) element IS *Rm14* is related to a previously unrecognized IS element located adjacent to the *Escherichia coli* locus of enterocyte effacement (LEE) pathogenicity island. *Curr. Microbiol.* **39**:274–281.
- van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh. 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* **62**:275–293.