Optical mapping and 454 sequencing of *Escherichia coli* O157:H7 isolates linked to the US 2006 spinach-associated outbreak

Michael L. Kotewicz, Mark K. Mammel, J. Eugene LeClerc
and Thomas A. Cebula

Correspondence

Thomas A. Cebula
thomas.cebula@FDA.HHS.gov

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 for five representative clinical, food-borne and bovine-derived isolates from the 2006 *Escherichia coli* O157:H7 outbreak linked to fresh spinach in the United States showed a common set of 14 distinct chromosomal markers that define the outbreak strain. Partial 454 DNA sequencing was used to characterize the optically mapped chromosomal markers. The markers included insertions, deletions, substitutions and a simple single nucleotide polymorphism creating a *Bam*HI site. The Shiga toxin gene profile of the spinach-associated outbreak isolates (*stx1*[−] *stx2*⁺ *stx2c*⁺) correlated with prophage insertions different from those in the prototypical EDL933 and Sakai reference strains (*stx1*⁺ *stx2*⁺ *stx2c*[−]). The prophage occupying the *yehV* chromosomal position in the spinach-associated outbreak isolates was similar to the *stx1*⁺ EDL933 cryptic prophage V, but it lacked the *stx1* gene. In EDL933, the *stx2* genes are within prophage BP933-W at the *wrbA* chromosomal locus; this locus was unoccupied in the spinach outbreak isolates. Instead, the *stx2* genes were found within a chimeric BP933-W-like prophage with a different integrase, inserted at the *argW* locus in the outbreak isolates. An extra set of Shiga toxin genes, *stx2c*, was found in the outbreak isolates within a prophage integrated at the *sbcB* locus. The optical maps of two additional clinical isolates from the outbreak showed a single, different prophage variation in each, suggesting that changes occurred in the source strain during the course of this widespread, multi-state outbreak.

Received 7 April 2008

Revised 24 June 2008

Accepted 2 July 2008

INTRODUCTION

A number of approaches can be used to identify the causative bacterial strain in a food-borne outbreak, although different techniques vary in their speed, convenience and cost, and most importantly in the depth at which the bacterial genome is interrogated. Pulsed-field gel electrophoresis (PFGE), which relies on the size and banding patterns of large chromosomal restriction fragments, has been a mainstay in public health trace-back investigations (Ribot *et al.*, 2006). A PFGE pattern of a particular genome, resolved into 30–50 fragments with a rare-cutting restriction enzyme, is usually sufficient to implicate a particular bacterial strain in an epidemiological study. However, PFGE has limited resolving power; other microbiological and biochemical characterizations, including single-nucleotide polymorphism (SNP) profiles and DNA microarray analysis, can provide much more detail, albeit at greater expense. For example, when applied to

Escherichia coli O157:H7, DNA microarray studies have generated a wealth of information on genome variability among isolates of this important pathogen (Fukuya *et al.*, 2004; Wick *et al.*, 2005; Ogura *et al.*, 2006; Jackson *et al.*, 2007; Zhang *et al.*, 2006). DNA arrays, however, also have limitations in that they cannot detect the presence of new genes or novel rearrangements of genes, nor can they determine the chromosomal locations(s) of genomic elements. Optical mapping does both.

An optical map is an assemblage of a number of partial restriction fragment maps into a single complete genome restriction map. To generate an optical map, whole chromosomes are spread and immobilized onto treated glass surfaces. The DNA is digested with a restriction enzyme, and across 20–100 long ‘chromosome reads’, automated contiguous restriction fragment size measurements are made. Each contig read is up to one-third of the whole chromosome. These overlapping partial chromosome contigs are assembled by alignment software similar to that used to align nucleotide sequences, using contiguous fragment sizes instead of contiguous bases. The *Bam*HI optical map of a typical *E. coli* O157:H7 isolate

Abbreviations: RFLP, restriction fragment length polymorphism; SNP, single-nucleotide polymorphism.

Four supplementary figures are available with the online version of this paper.

contains 500–600 fragments ranging from 1 to 50 kb spanning the 5 Mbp genome. The contiguous fragments of one optical map can be aligned and compared to the *Bam*HI *in silico* chromosome map of a sequenced reference strain, or to the optical maps of different strains. Changes as small as 2 kb can be mapped (Lim *et al.*, 2001; Chen *et al.*, 2006; Kotewicz *et al.*, 2007).

In the two similar sequenced *E. coli* O157:H7 genomes, many differences were found in unique segments not found in the *E. coli* K-12 genome that have been called O-islands (EDL933, Perna *et al.*, 2001) or S-islands (Sakai strain, Hayashi *et al.*, 2001). Studies have shown that the diversity among individual *E. coli* O157:H7 strains is due, in large part, to variation within prophages (Kudva *et al.*, 2002a, b; Brüssow *et al.*, 2004). Variations among dozens of prophage integrations and excisions, as well as prophage deletions and substitutions, have changed restriction sites and PFGE and PCR profiles for otherwise similar *E. coli* O157:H7 isolates. Notably, large, and often double, chromosomal inversions are prevalent in *E. coli* O157:H7 isolates (Iguchi *et al.*, 2006; Shima *et al.*, 2006; Kotewicz *et al.*, 2007) and the inversions have occurred within resident prophages.

In the United States, *E. coli* O157:H7 causes approximately 75 000 cases of infection annually (Griffin & Tauxe, 1991). In previous outbreaks of *E. coli* O157:H7, an estimated 15–20 % of people infected presented with indications severe enough to require hospitalization, and typically about 2–7 % of cases progress from haemorrhagic colitis to haemolytic-uraemic syndrome (HUS) with kidney failure complications (Su & Brandt, 1995). The 2006 outbreak of

E. coli O157:H7 in the United States associated with the consumption of fresh spinach occurred across a large multi-state area. The outbreak presented a particularly virulent pathology in which 51 % of the patients were hospitalized and 16 % developed HUS (CDC, 2006).

What characteristics defined the *E. coli* O157:H7 strain responsible for the 2006 spinach-associated outbreak in the United States? Here, a set of chromosomal markers including *stx* prophages was used to characterize this highly pathogenic outbreak strain; these markers have medical, epidemiological and forensic implications. The optical maps of five representative food-borne, clinical and bovine isolates from the 2006 spinach-associated outbreak delineated 14 distinct chromosomal markers. Partial genomic 454 DNA sequencing was used to identify each of these chromosomal changes and to define the outbreak strain in molecular detail.

METHODS

Bacterial strains. Sources of the *E. coli* O157:H7 strains used in this study are listed in Table 1. Two completely sequenced strains, 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 a 1996 outbreak in Japan, were used as reference strains (Table 2). Two hundred and eight bovine, food-borne and clinical *E. coli* O157:H7 isolates linked to the 2006 spinach outbreak were examined. Most of the isolates were clinical, obtained from 24 of the 26 states affected by the outbreak; these included isolates provided by State Health Laboratories from 178 of 206 patients most probably sickened from consumption of fresh spinach. In addition, 19 food-borne isolates

Table 1. Spinach outbreak *E. coli* O157:H7 isolates investigated

Spinach outbreak isolate	State*	Source	Optical maps		454 sequencing	
			No. of bands	No. of bases	No. of residues	No. of contigs
EC4042	A	Clinical	524	5 503 378	5 195 295	448
EC4045†	A	Spinach	529	5 523 849	5 195 000	767
EC4076†‡	B	Clinical	530	5 515 363	ND	ND
EC4115§	C	Clinical	524	5 513 658	76 000	98
EC4206	CA	Bovine	525	5 487 566	4 787 000	2700
EC4205	CA	Bovine	ND	ND	5 157 973	1006
EC4009	D	Clinical	ND	ND	5 194 891	277
EC4084	E	Clinical	ND	ND	5 193 526	372
EC4127	F	Clinical	ND	ND	5 181 246	286
EC4148	G	Clinical	ND	ND	656 610	836
EC4191	H	Spinach	ND	ND	5 202 122	268
EC4192	I	Clinical	ND	ND	5 232 373	284

ND, Not determined.

*States other than California (CA; outbreak origin) not identified for reasons of patient confidentiality.

†GenBank through TIGR site.

‡Microarray variant; prophage T deletion.

§Microarray variant; rRNA signal.

||Poor sequence coverage.

Table 2. Reference *E. coli* O157:H7 strains

Reference strain	Origin	Source	Optical/ <i>in silico</i> maps		Completed sequence	
			No. of bands	No. of bases	No. of residues	No. of contigs
EDL933-seq	US	Clinical	645	5 528 446	5 528 446	1
EDL933-OM	—	—	525	5 535 378	—	—
Sakai-seq	Japan	Clinical	639	5 498 451	5 498 451	1
Sakai-OM	—	—	544	5 514 226	—	—

derived from spinach samples implicated in the outbreak were examined. Eleven bovine isolates, obtained from cattle in the Salinas valley, in an area of spinach production implicated in the outbreak, were generously provided by the California Department of Health (USA). The bovine isolates had PFGE patterns similar to food-borne and clinical outbreak isolates associated with the outbreak (Cooley *et al.*, 2007). Isolates were encoded and designated EC4000 to EC4208 to ensure patient confidentiality and compliance with Institutional Review Board (IRB) guidelines. Selection of *E. coli* O157:H7 strains for optical mapping was based on preliminary characterization of the 208 isolates by DNA microarray chip analysis, using the Affymetrix *E. coli* Genome version 2.0 expression array.

Microarray selection of strains for optical mapping. The Affymetrix *E. coli* Genome 2.0 array, which consisted of over 10 000 probe sets for genic and intergenic sequences of four sequenced strains of *E. coli*, was used to determine gene profiles for all the isolates. As the isolate-to-isolate variation in gene content was very low, a clinical and a spinach (food-borne) isolate from the same state were selected as representative isolates for optical mapping. In addition, a single isolate representative of the bovine set and two clinical isolates that displayed array variations were optically mapped. One array variant showed pattern differences in the rRNA gene signals, while another clinical isolate showed the absence of nearly 20 genes, part of the EDL933 cryptic prophage T.

Optical mapping. Optical maps were prepared by OpGen, Madison WI, USA. In brief, following gentle lysis and dilution, high-molecular-mass genomic DNA molecules were spread and immobilized onto derivatized glass slides and digested with *Bam*HI. The DNA digests were stained with YOYO-1 fluorescent dye, and photographed using a fluorescent microscope interfaced with a digital camera. Automated image-analysis software located and sized fragments, and assembled multiple scans into whole-chromosome optical maps. More detail is presented in Zhou *et al.* (2004).

454 DNA sequencing. Partial genomic sequences of the five representative optically mapped isolates were obtained in collaboration with the National Bioforensics Analysis Center (NBFAC). Assembly of sequencing reads into contigs and subsequent ordering of these contigs into scaffolds was performed with GS *De novo* Assembler Software. For partial genome sequencing, five 454 runs were typical, and yielded, on average, five million bases in 277 to 450 contigs.

Other genomic DNA sequences. In addition to the 454 sequencing, in collaboration with TIGR, the nearly complete genomic sequence for spinach isolate EC4045 was used to analyse the structure of the *stx*-containing prophages. This sequence work was supported in part with Federal funds from the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, Department of Health and Human Services, under NIAID Contract N01-AI-30071. Whole-genome shotgun partial genome sequences for

a number of outbreak isolates are available at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome>.

RESULTS

From over 200 initial spinach-associated outbreak isolates from the United States, five representative isolates, including clinical, food-borne and animal (bovine) were optically mapped. Their 520 *Bam*HI fragment optical maps were indistinguishable except for small variable drop-out fragments in the 21–1500 bp range. The average genome size measured from the optical maps for the food-borne isolate EC4045, the clinical isolate EC4042, and the bovine isolate EC4206 was 5 513 658 bp $\pm$ 8 kb, or 5.514 Mbp. The optical map of reference strain EDL933 (525 fragments) was 5.535 Mbp in length relative to its sequence-based (*in silico*) 5.528 Mbp; for Sakai, its optical map was 5.514 Mbp (544 fragments) relative to the *in silico* size of 5.598 Mbp. The accuracy of genome size measurement using an optical map is a function of the sum of the size accuracies of all the fragments measured. The 550–650 *Bam*HI fragments of an *E. coli* O157:H7 chromosome range from 2 to 50 kb, and the sum of accuracies is about $\pm$ 50 kb. There is an additional factor in sizing chromosomes with optical maps. Small restriction fragments, ranging from 21 bp to 1.5 kb, found in the sequence-based maps are not efficiently detected in optical maps. These fragments are called drop-out fragments, and their effect on the overall size determination of a 5 Mbp genome is 20–40 kb (Kotewicz *et al.*, 2007).

An examination of the optical maps of a dozen representative *E. coli* O157:H7 strains shows chromosomal variation at 31 sites (Table 3, A–AE). Fourteen sites in the spinach-associated outbreak strain were different from the sequenced reference strains, EDL933 and Sakai (Table 4), though individual spinach chromosome site markers were found in other strains (Table 3). They were annotated based on their consecutive positions on the chromosome (Fig. 1). The largest chromosomal changes, markers 3–12, are shown in Fig. 2. The figure also illustrates a 429 kb inversion that occurred in the chromosome of EDL933 within cryptic prophages CP933-O and CP933-P relative to the spinach and Sakai chromosomes. This inversion is distinguished by a cross-overlike 'X' of the alignment score lines in the maps (Fig. 2). Changes seen in the chromosomes of outbreak

Table 3. Summary of variations at 31 chromosomal polymorphisms optically mapped in 12 *E. coli* O157:H7 strains

Numbers represent kb of DNA inserted at each site. -, Polymorphism same as Sakai; e, EDL933 variant; s, Sakai variant; sp, spinach variant; cp, cryptic prophage; ins, insertion; r, replacement of several fragments with different sized fragments; IS, insertion sequence; RFLP, restriction fragment length polymorphism; inv, chromosomal inversions – the origins within phages are indicated, L left and R right (the boxed cells indicate the extent of the inversions); ( ), most probable determination; indel, insertion or deletion.

		Site																																		
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE				
		0.264	0.310	0.908	1.093	1.263	1.359	1.488	1.648	1.691	1.728	1.780	1.888	2.159	Map position Mbp from EDL933 sequence										2.826	2.914	2.981	3.276	3.431	3.551	4.633	4.828	5.112	5.173	5.389	5.428
		EDL933 annotated phase or other type of change																																		
		IS	cpH/I	cpK	P4a	cpM	W	P4b	cpN	cpC	cpX	del	cpO	R	cpP	RFLP	ins	ins	cpT	cpU	sbcB	APEC	cpV	argW	RFLP	ins	ins	RFLP	Mu-like	ins	RFLP/ins	ins				
1	Sakai	0	22	38	0	49	s62	87	47	16	45	0	80	50	58	34	0	0	20	s45	18	0	Vs	0	1	6	0	0	49	0	0	0	Sakai			
2	EC503	-	-	-	-	e62	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	0	-	-	-	-	EC503			
3	EC507	-	-	-	-	e62	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	EC507			
4	EDL933	-	-	-	87	45	e62	-	47RFLP	-	54	-	inv1L	e50	inv1R	-	-	-	-	e45	12	-	Ve	-	3	-	-	0	-	-	-	-	EDL933			
5	EC1225	-	-	-	-	93	-	-	-	-	100	-	inv2L	-	inv2R	-	-	-	-	-	-	-	V1225	-	-	-	-	0	-	-	-	-	EC1225			
6	AB1	-	-	-	11	-	-	-	-	-	-	-	inv3L	inv3R	130-150	18+17	-	50	-	-	-	-	-	-	-	-	0	-	-	-	-	AB1				
7	EC536	IS	22+28	-	-	(45)	0	85	-	8	-	del 15	inv4L	inv4R	e58	-	-	-	-	-	74	-	V536	0	3	-	-	RFLP	0	11	-	-	EC536			
8	EC533	-	-	-	-	inv5L	e62	-	-	-	-	-	indel	-	e58	-	-	-	-	inv5R	-	-	V-17	-	-	-	-	0	-	-	-	26	-	EC533		
9	EC502	-	-	-	87	(49RFLP)	inv6L	-	-	-	-	-	indel	-	e58	-	-	-	-	-	12	-	inv6R	-	3	-	-	0	-	-	-	-	-	EC502		
10	EC508	IS	-	36	-	0	-	0	11	-	-	-	120	0	indel	-	-	-	-	-	74	35	(Vs)	56	(.)	-	21	-	0	11	-	-	EC508			
11	EC869	-	22+28	-	77	0	85	-	11	-	del 15	-	68	0	37	-	35	-	-	-	74	30	Ve	-	-	5	14	-	0	-	-	RFLP+2kb	-	EC869		
12	EC1231	IS	11	26	-	49RFLP	0	85RFLP	0	0	r	-	r	50+8	58/ins	-	53	36	45+53	74	-	V1231	-	3	14	RFLP	-	0	47	-	-	10	-	EC1231		

Table 4. Summary of 14 variant chromosomal sites in the 2006 spinach-associated *E. coli* O157:H7 outbreak relative to EDL933 and Sakai reference strains

Spinach outbreak site no.																																	
		1	2	-	3	-	4	-	-	5	6	7	8	9	-	-	-	-	-	10 (stx2c)	11	-	12 (stx2)	-	-	ins	-	13	14	-	-		
		IS seq	cpH/I	cpK	P4a	cpM	W	P4b	cpN	cpC	cpX	del	cpO	cpR	cpP	RFLP	ins	ins	cpT	cpU	sbcB	APEC	cpV	argW	RFLP	ins	ins	RFLP	Mu-like	ins	RFLP/ins	ins	Sakai
1	Sakai	0	22	38	0	49	s62	87	47	16	45	0	80	50	58	34	0	0	20	s45	18	0	Vs	0	1	6	0	0	49	0	0	0	Sakai
13	EC4042	IS	RFLP	-	-	0	-	-	-	11	RFLP	del 15	120	0	sp58	-	-	-	-	-	74	35	-	58	3	-	-	-	0	11	-	-	EC4042
14	EC4045	IS	RFLP	-	-	0	-	-	-	11	RFLP	del 15	120	0	sp58	-	-	-	-	-	74	35	-	58	3	-	-	-	0	11	-	-	EC4045
15	EC4076	IS	RFLP	-	-	0	-	-	-	11	RFLP	del 15	120	0	sp58	-	-	-	0	-	74	35	-	58	3	-	-	-	0	11	-	-	EC4076
16	EC4115	IS	RFLP	-	-	0	-	*	-	11	RFLP	del 15	120	0	sp58	-	-	-	-	-	74	35	V+41	58	3	-	-	-	0	11	-	-	EC4115
17	EC4206	IS	RFLP	-	-	0	-	*	-	11	RFLP	del 15	120	0	sp58	-	-	-	-	-	74	35	-	58	3	-	-	-	0	11	-	-	EC4206

Fig. 1. Positions of spinach chromosomal markers on circular genome map of *E. coli* O157:H7. Optical mapping and 454 sequencing detailed 14 markers on the chromosome of the spinach outbreak strain relative to EDL933 and Sakai: large novel insertions (red); one novel chromosomal deletion (black); prophage variants relative to EDL933 and Sakai and an insertion sequence mediated change (grey); four unoccupied prophage sites (white); and three normal chromosomal landmarks normally found in *E. coli* O157:H7, *oriC*, the LEE island, and the TAI island for tellurite resistance (blue). The 454 sequencing detailed a *stx1*-negative variant at the *yehV* locus (grey), which was not apparent in the optical map.

strains were the same whether their optical maps were aligned either to sequence-based *in silico* or to *de novo* optical maps of EDL933 and Sakai. The small drop-out fragments from 21 bp to 1.5 kb lost in optical maps lower the ability to detect differences only slightly (Kotewicz *et al.*, 2007).

The 14 outbreak-associated chromosomal markers were further characterized by 454 partial genomic sequencing of the five optically mapped isolates and six additional outbreak isolates. Twelve of the chromosomal markers were in prophages. A summary overview of the larger

chromosome changes found in the spinach-associated outbreak strain, markers 3–12, is shown in Fig. 2.

Phage and Shiga toxin gene profiles

Initial PCR and DNA microarray data showed that all the 208 clinical, spinach and bovine isolates were negative for the *stx1* genes and were positive for the *stx2* genes. However, 454 sequencing data indicated that two sets of *stx2* genes were present in all the spinach-associated isolates, *stx2* and a known variant, *stx2c*. The spinach-associated isolates were all *stx1*⁻ *stx2*⁺ *stx2c*⁺.

Fig. 2. Overview of optical maps of spinach outbreak markers 3–12. Comparison of optical maps and sequence-based maps (2 Mbp of the 5.5 Mbp genomes illustrated) of *E. coli* O157:H7 spinach outbreak isolate EC4045 and reference strains Sakai and EDL933. The alignment software highlighted fragments different between strains in white; matching contiguous restriction fragments are highlighted in green. In the aligned maps, the crossed-over alignment lines demonstrate an inversion in the EDL933 chromosome relative to Sakai.

EDL933 and Sakai are *stx1*⁺ *stx2*⁺ *stx2c*⁻. The *stx1* genes are found within the cryptic prophage V at the *yehV* chromosomal locus. The optical maps of the *stx1*⁻ spinach-associated outbreak isolates showed a variant insertion corresponding to the EDL933 *stx1*⁺ prophage V at *yehV*, which more closely resembles the *stx1*⁺ prophage found in Sakai at *yehV*. However, the optical map of the Sakai-like *stx1*⁻ prophage found in the spinach isolates at the *yehV* locus cannot be distinguished from the *stx1*⁺ Sakai prophage. This is because the resolution of optical maps would not detect small (1–4 kb) DNA deletions or substitutions within larger (20–50 kb) fragments. In this case, the optical maps could not detect the difference in the status of the *stx1* genes within this prophage locus (Fig. 3a).

In EDL933 and Sakai, the *stx2* genes are found in a 62 kb non-cryptic prophage W at the *wrbA* locus. For the *stx2*⁺ *stx2c*⁺ spinach isolates, no corresponding prophage insertion at the *wrbA* site was found (Fig. 3b, *wrbA* locus). Note that for EDL933 and Sakai, the optical maps distinguish two different *stx2*⁺ phages. 454 DNA sequences from six outbreak-linked isolates confirmed that the *wrbA* chromosomal attachment site contained no remnant, or

prophage scar sequences in the bacterial chromosome. Instead, the optical maps of the spinach isolates showed three large potential prophage insertions: 56 kb at 2.829 Mbp (at *sbcB*), 33 kb at 2.915 Mbp (at *yegQ*) and 58 kb at 3.276 Mbp near *argW* (Fig. 2, Table 4). The 454 sequence data were sufficient to partially characterize each of these insertions as a prophage.

The 56 kb prophage integration (Fig. 4, marker 10) was bracketed by a 13 bp duplication of the sequence TTTCACGATTACG, and the prophage was integrated at the *sbcB* chromosomal locus. The 454 sequences only allowed the determination of 1 kb of prophage at *attL* and 2 kb of prophage from *attR*. This sequence was sufficient to characterize the integrase proximal to the *attR* site as 98 % homologous to the integrase of an insertion found near *sbcB* in *Shigella flexneri* 2a and *S. flexneri* 5 strains.

The 33 kb insertion near *yegQ* is a prophage flanked by an 18 bp duplication of the sequence TACACGGGCTTA-TTTTTT. The spinach outbreak prophage resembled a prophage found at 2.281 Mbp in the chromosome of an avian pathogenic *E. coli*, APEC O1; the integrase of the outbreak phage was 99 % similar to the integrase of this APEC P2 type phage.

Fig. 3. *wrbA* and *yehV* loci variants in spinach outbreak strains. Aligned fragments are highlighted in green; white fragments indicate non-alignment. (a) Optical maps of the *yehV* region of the spinach outbreak strain EC4045 and Sakai including the prophage V. The sequence-based map of the EDL933 prophage V, which carries the *stx1* gene, is shown below the aligned genomes of the spinach outbreak isolate EC4045 and the reference maps of Sakai and EDL933. In the schematic of the EDL933 prophage V, the absence or presence of the *stx1* genes is indicated by + or -. (b) Optical map of chromosomal marker 4, at the *wrbA* locus; in outbreak strains, there is no prophage and no *stx2* gene at this position. EDL933 and Sakai contain a prophage, W, which carries the *stx2* gene. A schematic of the prophage W map is shown with a polymorphic difference between EDL933 and Sakai.

Fig. 4. Chromosomal markers 10, 11 and 12: new *stx2c* prophage, new APEC-like prophage and a new *stx2* prophage. Aligned fragments are highlighted in green; white fragments indicate non-alignment.

The 58 kb insertion near *argW* (Fig. 4, marker 12), represents an entirely new prophage in the outbreak isolates, next to a truncat *intC* prophage. Twenty-five base-pair *attL* and *attR* sequences, TGTCTCTTAGTT-AAATGGATATAA, bracket the new prophage. The integrase of the new prophage is 99% (1168/1169 bp) homologous to the integrase of the *stx2* converting phage 86 isolated from a pathogenic *E. coli* of the O86 serogroup (GenBank AB255436).

Overall, the map and 454 sequence data demonstrated that the outbreak strains carried variant lambdoid prophages at the *yehV*, *sbcB* and *argW* loci. Because of extensive homologies among the multiple lambdoid phages found in *E. coli* O157:H7, optical mapping and partial 454 genome sequencing were not sufficient to allow the complete resolution of phage sequences, nor were they sufficient for the complete assembly of the bacterial genome sequence around these similar and uncompleted prophage sequences. There were also too many gaps in the contigs of the phages to resolve the Shiga toxin gene arrangements in the 454 sequences. In collaboration with TIGR, however, the partially completed genome sequence of isolate EC4045 demonstrated the absence of *stx1* genes in the V prophage at *yehV*, the presence of *stx2c* genes in the *sbcB* prophage, and the *stx2* genes in the *argW* prophage (see <http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome>). This configuration was very distinct from the sequenced EDL933 and Sakai strains, although components of the *sbcB* prophage described here have been observed in other *E. coli* O157:H7 isolates (Ohnishi *et al.*, 2002).

Other prophage chromosomal markers

Marker 2 was a simple restriction fragment length polymorphism (RFLP) at 0.310 Mbp (Table 4; see also Supplementary Fig. S1, available with the online version of this paper). The 454 sequencing revealed that this new *Bam*HI site was a C→G transversion creating a new *Bam*HI site in the spinach isolates within the cryptic H/I prophage complex (Table 4).

The optical maps indicated three chromosomal markers that appear to be the simple gain or loss of a complete prophage relative to the EDL933 or Sakai strains. The first two, markers 3 and 9, are a P4-like prophage and a prophage designated R in EDL933 (Table 4); both were missing in the spinach isolate (Fig. S2). The third, a Mu-like prophage, marker 13, is not found in the spinach isolate; it is found only in Sakai (Tables 3 and 4, Fig. S3).

Marker 14 was a phage sequence scar (Fig. S3). An 11 kb insertion found in the spinach isolate represented a small piece of a phage and contained sequences homologous to several phage tail genes. There were three additional chromosomal markers found in the optical maps, markers 5, 6, and 7, which were sequence scars within prophages C, X, and O (Fig. 5) in the spinach outbreak strain relative to EDL933. Both the optical maps and preliminary sequence analysis suggested that some of these prophage substitutions are complex.

Non-prophage chromosomal markers

The optical maps revealed two chromosomal markers in the outbreak isolates that were not associated with prophages. The first, marker 1, was a simple RFLP at 0.264 Mbp (Table 4, Fig. S1). 454 sequencing showed that this simple change was a complex rearrangement associated with insertion sequences. The second was a 15.1 kb deletion at 1.782 Mbp (Fig. 5, marker 7). 454 DNA sequencing showed that 15 147 bp found between the two 22 bp sequences ATTCTgCcactctgaAcCGCCC and ATTCTAcGcgcgctgAgCGCCC in EDL933 are deleted in the outbreak isolates, leaving a hybrid of the duplicated 22 bp sequence, ATTCTgCcgcgctgAgCGCCC, at the deletion site. Bases that were the same in both 22 bp sequences on the EDL933 chromosome are capitalized. The 15 kb deletion included 13 EDL933 genes from Z1957, encoding an IS629 transposase, to Z1972, encoding a partial putative adhesion protein. The other genes within the deleted region included ones encoding five hypothetical proteins of unknown function, a putative outer-membrane receptor, a molybdenum transport protein, two putative iron compound ABC transporters, an ATP-binding protein

Fig. 5. Chromosomal markers 4, 5, 6 and 7: altered prophages C and X, a deletion and prophage O. Aligned fragments are highlighted in green; white fragments indicate non-alignment. The chromosomal inversion in EDL933 between the O and downstream P prophages inverts the fragments found (in grey) after the inversion in the O prophage.

and transporter permease, a trehalase, a putative PTS system enzyme I and a putative dihydroxyacetone kinase.

Variant clinical isolates; two single prophage changes

Because their microarray profiles differed from the typical outbreak profile, two spinach-associated clinical isolates were analysed in further detail by optical mapping and 454 sequencing. The microarray profile of isolate EC4076 indicated the probable deletion of a set of genes associated with cryptic prophage CP933-T, and the optical map showed the absence of two *Bam*HI restriction fragments (total deletion of 20 kb) associated with the annotated prophage T in the *in silico* map of EDL933 (Fig. S4). In collaboration with TIGR, the whole-genome shotgun sequence data for this isolate revealed that the 121 bp tandem duplication (AAAAAAACCA CCCGAAGGTG GTTTCACGAC ACTGCTTATT GCTTTGATTT TATTC-TTATC TTTCCTATGG TACCCGGAGC GGGACTTGAA CCCGCACAGC GCGAACGCCG AGGGATTTTA A) flanking prophage T in the reference EDL933 sequence has been reduced to a single 121 bp repeat in EC4076, completely eliminating the 21 118 bp prophage T sequence in this clinical isolate (see <http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome>). From microarray analysis of 208 outbreak-associated isolates only the clinical isolate EC4076 carried the prophage T deletion. EC4076 had all the other outbreak-specific chromosomal markers characterized here and did not have any other differences relative to the archetypal outbreak optical map.

Another clinical isolate, EC4115, showed elevated signals for several rRNA targets in its microarray profile. As was the case with EC4076, the optical map of EC4115 showed the same 14 chromosomal markers characteristic of the spinach outbreak strain. The optical map of EC4115 did not show any other chromosomal abnormalities as had been suggested by the microarray data at rRNA loci. However, the optical map of EC4115 did show one additional chromosomal change. A 41 kb insertion was found in EC4115 at the chromosomal locus *yehV*, at the site of the V-like prophage lacking *stx1* in the outbreak isolates (Fig. S4). Analysis of the whole-genome shotgun sequence data at the NCBI/TIGR website confirmed that identical left and right *att* sites to those of the EDL933

prophage CP933 V were found in the outbreak isolate EC4115. A gap in the sequence contigs in the centre of the prophage prevented the characterization of the 41 kb insertion.

DISCUSSION

Fourteen distinctive chromosome markers were optically mapped in the 2006 spinach-associated US *E. coli* O157:H7 outbreak strain relative to reference EDL933 and Sakai strains. Genomic variability of *E. coli* O157:H7 strains is mainly due to differences within prophages and in chromosomal inversions (Kudva *et al.*, 2002a, b; Brüssow *et al.*, 2004; Iguchi *et al.*, 2006; Kotewicz *et al.*, 2007). Most of the chromosomal markers demonstrated by optical mapping would not have been detected by microarray and simple sequence-based techniques such as SNP analysis. Microarray determinations of the presence/absence of multicopy genes are not simple; the modular nature of these prophages, with similar genes in multiple different chromosomal sites, precludes a simple gene present/absent call. Table 3 shows highly variable sites in *E. coli* O157:H7 strains, which show evidence of repeated integrations and deletions as well as sites of chromosomal inversions. Other prophage sites, more highly deleted and degenerate, represent smaller targets for recombination and appear to be more stable genetic markers.

The *stx* phages represent some of the most important, and yet most variable chromosomal markers (Johansen *et al.*, 2001; Herold *et al.*, 2004; Bielaszewska *et al.*, 2007a). Variation in highly similar toxin genes contributes to strain toxicity (Friedrich *et al.*, 2002; Orth *et al.*, 2007; Manning *et al.*, 2008; Bielaszewska *et al.*, 2007b). The virulence of a particular *E. coli* O157:H7 strain might depend in part on the kind, placement, and/or rearrangement of specific phage sequences (Kohler *et al.*, 2000; Eklund *et al.*, 2002; Ritchie *et al.*, 2003; Karch *et al.*, 2006), all of which affect Shiga toxin gene expression, most notably the anti-terminator Q gene product (Zhang *et al.*, 2005; Koitabashi *et al.*, 2006).

However, as best illustrated by the *stx* prophages, neither optical mapping alone, nor optical mapping in conjunction with partial genomic 454 sequencing, was able to determine the exact location of the *stx* genes. As the limit

of resolution of optical mapping restricts the determination of novel insertions to the size of an average gene, 1–2 kb, further Sanger sequencing and PCR determinations were required to unequivocally locate the *stx* genes within different prophages. The two lambdoid prophages in the outbreak strain EC4045 that contained the Shiga toxin genes, *stx2c* at *sbcB* and *stx2* at *argW*, were both chimeric. Most notably, each was homologous to EDL933 prophages, but contained a non-homologous integrase gene. Recombinant phages carrying assorted integrase genes (integrase cassettes) and different (or null) *stx* cassettes produce different configurations of the toxin genes and the critical phage Q gene regulators, and can be found at different chromosomal insertions sites in different *E. coli* O157:H7 strains (Serra-Moreno *et al.*, 2007).

Instead of recombination assorting genes and modules among phages, the gain or loss of *stx* genes might occur on the chromosome, in prophages, as illustrated by the V-like prophage at *yehV*. In the spinach outbreak strain, phage genes have replaced the *stx1* genes found in EDL933 (or vice versa). A pathway for replacement of phage and *stx1* genes has been substantiated by phylogenetic studies (Shaikh & Tarr, 2003).

As opposed to the highly variable *stx* prophages, small prophages (the probable results of large deletions) appear to be stable markers. The spinach outbreak strain contained five prophage sequence scars, including markers 2, 5, 6 and 8, modified relative to the cryptic EDL933 prophages H/I, C, X and O, respectively, as well as marker 14 (Table 4). Marker 2 is a single base change in the H/I cryptic prophage which created a new *Bam*HI site in the outbreak strain. The optical mapping of nine other *E. coli* O157:H7 strains revealed a possible progenitor to this 23 kb H/I prophage scar, a 28 kb insertion in strains EC869 and EC536 (Table 3 and Kotewicz *et al.*, 2007). Similarly, a 47 kb insertion found in strain EC1231 (5.173 Mbp, Table 3) could be the progenitor of the 11 kb sequence scar found in the spinach outbreak.

Outbreak marker 12, near *yegQ*, was a 33 kb prophage which resembled a P2 prophage found in the sequenced avian pathogenic *E. coli* APEC O1. This prophage has not been previously identified in sequences or optical maps of over a dozen *E. coli* O157:H7 isolates. Sequence analysis of marker 12 strongly suggested that the scar found at *yegQ* in *E. coli* K-12 is the result of the degradation of a similar prophage. In *E. coli* K-12, this left a 629 bp scar with a flanking pair of 18 bp attachment sites (ACACGGGCTT-ATTTTTT). The analysis of sequence data from prophages and flanking regions has begun to clarify how some of these prophages have become potentially long-lived, stable sequence scars and how others are undergoing multiple insertions and deletions.

Between the highly variable prophages and the stable prophage scars was a set of markers with an intermediate level of variation. Some prophage scars may reflect the

emergence of this pathogen from *E. coli* O55:H7 progenitors (Feng *et al.*, 2007). Other markers are useful for tracking changes in *E. coli* O157:H7 between different outbreaks and potentially even within outbreaks. These markers can define specific temporal and geographical changes in the progression of an outbreak; they have important implications for forensic identification and source tracking in outbreak investigations. For example, a deletion has removed prophage T in EC4076, and a new insertion has occurred near prophage V in EC4115. Each of these changes defines a different strain, allowing the subsequent tracing of each variant. We hypothesize that these variants might have arisen during the infection process in the patients. Alternatively, though less likely, they could have been rare isolates co-contaminating the food matrix.

The combination of optical mapping and 454 sequencing accomplished a number of goals: it allowed a broad view of chromosomal variation in *E. coli* O157:H7 (31 optically mapped sites among 20 representative strains); it allowed closure of sequencing contigs across multiple phage repeats (in collaboration with TIGR); and it mapped and detailed the distinctive chromosomal markers that were archetypal for the 2006 US spinach-associated outbreak strain. The increasing amount of information available on chromosome variations in *E. coli* O157:H7 from a number of laboratories allows the design of simple PCR-based assays for complex chromosomal markers. These can be implemented in a number of high-throughput assay formats to determine the strain, gene sets, prophage and *stx* configurations that would fully profile strains and their virulence potential (Manning *et al.*, 2008).

Chromosomal profiles have been used to identify outbreak strains of pathogens such as *E. coli* O157:H7, as evidenced by the use of PFGE for deployment of CDC's PulseNet system since 1996. For more detailed analysis, DNA microarray profiles can be used to characterize outbreak strains by nearly complete genic content. With a resolution of 2 kb across each of the 500–600 contiguous restriction fragments of an enteric bacterial genome, optical mapping provides yet more insights into chromosomal changes and gene acquisitions that neither PFGE nor microarray analysis allows.

ACKNOWLEDGEMENTS

We acknowledge the National Bioforensics Analysis Center (NBFAC) of the Department of Homeland Security for supporting work on optical mapping and 454 partial genome sequencing of *E. coli* O157:H7 strains reported here under the Interagency Agreement Grant no. 224-04-2806. We thank Jonathan Hnath, Kathleen Verratti and Alexander Domingo of NBFAC for 454 sequencing. We wish to particularly thank the California Department of Public Health for the bovine samples, as well as other spinach and clinical isolates, and the 24 State Health Laboratories that generously provided outbreak isolates. We thank FDA colleagues David Lacher, Amit Mukherjee and Scott Jackson for valuable discussions.

REFERENCES

- Bielaszewska, M., Köck, R., Friedrich, A. W., von Eiff, C., Zimmerhackl, L. B., Karch, H. & Mellmann, A. (2007a). Shiga toxin-mediated hemolytic uremic syndrome: time to change the diagnostic paradigm? *PLoS ONE* 2, e1024.
- Bielaszewska, M., Prager, R., Köck, R., Mellmann, A., Zhang, W., Tschäpe, H., Tarr, P. I. & Karch, H. (2007b). Shiga toxin gene loss and transfer in vitro and in vivo during enterohemorrhagic *Escherichia coli* O26 infection in humans. *Appl Environ Microbiol* 73, 3144–3150.
- Brüssow, H., Canchaya, C. & Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 68, 560–602.
- CDC (2006). Ongoing multistate outbreak of *Escherichia coli* serotype O157:H7 infections associated with consumption of fresh spinach – United States, September 2006. *MMWR Morb Mortal Wkly Rep* 55, 1045–1046.
- Chen, Q., Savarino, S. J. & Venkatesan, M. M. (2006). Subtractive hybridization and optical mapping of the enterotoxigenic *Escherichia coli* H10407 chromosome: isolation of unique sequences and demonstration of significant similarity to the chromosome of *E. coli* K-12. *Microbiology* 152, 1041–1054.
- Cooley, M., Carychao, D., Crawford-Miksza, L., Jay, M. T., Myers, C., Rose, C., Keys, C., Farrar, J. & Mandrell, R. E. (2007). Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. *PLoS ONE* 14, e1159.
- Eklund, M., Leino, K. & Siitonen, A. (2002). Clinical *Escherichia coli* strains carrying *stx* genes: *stx* variants and *stx*-positive virulence profiles. *J Clin Microbiol* 40, 4585–4593.
- Feng, P. C., Monday, S. R., Lacher, D. W., Allison, L., Siitonen, A., Keys, C., Eklund, M., Nagano, H., Karch, H. & other authors (2007). Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. *Emerg Infect Dis* 13, 1701–1706.
- Friedrich, A. W., Bielaszewska, M., Zhang, W. L., Pulz, M., Kuczius, T., Ammon, A. & Karch, H. (2002). *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *J Infect Dis* 185, 74–84.
- Fukuya, S., Mizoguchi, H., Tobe, T. & Mori, H. (2004). Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J Bacteriol* 186, 3911–3921.
- Griffin, P. M. & Tauxe, R. V. (1991). The epidemiology of infections caused by *Escherichia coli* O157:H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. *Epidemiol Rev* 13, 60–98.
- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C. G., Ohtsubo, E., Nakayama, K. & other authors (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 8, 11–22.
- Herold, S., Karch, H. & Schmidt, H. (2004). Shiga toxin-encoding bacteriophages – genomes in motion. *Int J Med Microbiol* 294, 115–121.
- Iguchi, A., Iyoda, S., Terajima, J., Watanabe, H. & Osawa, R. (2006). Spontaneous recombination between homologous prophage regions causes large-scale inversions within the *Escherichia coli* O157:H7 chromosome. *Gene* 372, 199–207.
- Jackson, S. A., Mammel, M. K., Patel, I. R., Mays, T., Albert, T. J., LeClerc, J. E. & Cebula, T. A. (2007). Interrogating genomic diversity of *Escherichia coli* O157:H7 using DNA tiling arrays. *Forensic Sci Int* 168, 183–199.
- Johansen, B. K., Wasteson, Y., Granum, P. E. & Brynestad, S. (2001). Mosaic structure of Shiga-toxin-2-encoding phages isolated from *Escherichia coli* O157:H7 indicates frequent gene exchange between lambdoid phage genomes. *Microbiology* 147, 1929–1936.
- Karch, H., Friedrich, A. W., Gerber, A., Zimmerhackl, L. B., Schmidt, M. A. & Bielaszewska, M. (2006). New aspects in the pathogenesis of enteropathic hemolytic uremic syndrome. *Semin Thromb Hemost* 32, 105–112.
- Kohler, B., Karch, H. & Schmidt, H. (2000). Antibacterials that are used as growth promoters in animal husbandry can affect the release of Shiga-toxin-2-converting bacteriophages and Shiga toxin 2 from *Escherichia coli* strains. *Microbiology* 146, 1085–1090.
- Koitaibashi, T., Vuddhakul, V., Radu, S., Morigaki, T., Asai, N., Nakaguchi, Y. & Nishibuchi, M. (2006). Genetic characterization of *Escherichia coli* O157:H7 strains carrying the *stx2* gene but not producing Shiga toxin 2. *Microbiol Immunol* 50, 135–148.
- Kotewicz, M. L., Jackson, S. A., LeClerc, J. E. & Cebula, T. A. (2007). Optical maps distinguish individual strains of *Escherichia coli* O157:H7. *Microbiology* 153, 1720–1733.
- Kudva, I. T., Evans, P. S., Perna, N. T., Barrett, T. J., DeCastro, G. J., Ausubel, F. M., Blattner, F. R. & Calderwood, S. B. (2002a). Polymorphic amplified typing sequences provide a novel approach to *Escherichia coli* O157:H7 strain typing. *J Clin Microbiol* 40, 1152–1159.
- Kudva, I. T., Evans, P. S., Perna, N. T., Barrett, T. J., Ausubel, F. M., Blattner, F. R. & Calderwood, S. B. (2002b). Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. *J Bacteriol* 184, 1873–1879.
- Lim, A., Dimalanta, E. T., Potamousis, K. D., Yen, G., Apodoca, J., Tao, C., Lin, J., Qi, R., Skiadas, J. & other authors (2001). Shotgun optical maps of the whole *Escherichia coli* O157:H7 genome. *Genome Res* 11, 1584–1593.
- Manning, S. D., Motiwala, A. S., Springman, A. C., Qi, W., Lacher, D. W., Ouellette, L. M., Mladonicky, J. M., Somsel, P., Rudrik, J. T. & other authors (2008). Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc Natl Acad Sci U S A* 105, 4868–4873.
- Ogura, Y., Kurokawa, K., Ooka, T., Tashiro, K., Tobe, T., Ohnishi, M., Nakayama, K., Morimoto, T., Terajima, J. & other authors (2006). Complexity of the genomic diversity in enterohemorrhagic *Escherichia coli* O157 revealed by the combinational use of the O157 Sakai oligo DNA microarray and the whole genome PCR scanning. *DNA Res* 13, 3–14.
- Ohnishi, M., Terajima, J., Kurokawa, K., Nakayama, K., Murata, T., Tamura, K., Ogura, Y., Watanabe, H. & Hayashi, T. (2002). Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. *Proc Natl Acad Sci U S A* 99, 17043–17048.
- Orth, D., Grif, K., Khan, A. B., Naim, A., Dierich, M. P. & Würzner, R. (2007). The Shiga toxin genotype rather than the amount of Shiga toxin or the cytotoxicity of Shiga toxin in vitro correlates with the appearance of the hemolytic uremic syndrome. *Diagn Microbiol Infect Dis* 59, 235–242.
- Perna, N. T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J. & other authors (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529–533 (Erratum in *Nature* 410, 240).
- Ribot, E. M., Fair, M. A., Gautom, R., Cameron, D. N., Hunter, S. B., Swaminathan, B. & Barrett, T. J. (2006). Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. *Foodborne Pathog Dis* 3, 59–67.

- Ritchie, J. M., Wagner, P. L., Acheson, D. W. & Waldor, M. K. (2003). Comparison of Shiga toxin production by hemolytic-uremic syndrome-associated and bovine-associated Shiga toxin-producing *Escherichia coli* isolates. *Appl Environ Microbiol* **69**, 1059–1066.
- Serra-Moreno, R., Jofre, J. & Muniesa, M. (2007). Insertion site occupancy by *stx*₂ bacteriophages depends on the locus availability of the host strain chromosome. *J Bacteriol* **189**, 6645–6654.
- Shaikh, N. & Tarr, P. I. (2003). *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. *J Bacteriol* **185**, 3596–3605.
- Shima, K., Wu, Y., Sugimoto, N., Asakura, M., Nishimura, K. & Yamasaki, S. (2006). Comparison of a PCR-restriction fragment length polymorphism (PCR-RFLP) assay to pulsed-field gel electrophoresis to determine the effect of repeated subculture and prolonged storage on RFLP patterns of Shiga toxin-producing *Escherichia coli* O157:H7. *J Clin Microbiol* **44**, 3963–3968.
- Su, C. & Brandt, L. J. (1995). *Escherichia coli* O157:H7 infection in humans. *Ann Intern Med* **123**, 698–714.
- Wick, L. M., Qi, W., Lacher, D. W. & Whittam, T. S. (2005). Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. *J Bacteriol* **187**, 1783–1791.
- Zhang, W., Bielaszewska, M., Friedrich, A. W., Kuczius, T. & Karch, H. (2005). Transcriptional analysis of genes encoding Shiga toxin 2 and its variants in *Escherichia coli*. *Appl Environ Microbiol* **71**, 558–561.
- Zhang, W., Qi, W., Albert, T. J., Motiwala, A. S., Alland, D., Hyytia-Trees, E. K., Ribot, E. M., Fields, P. I., Whittam, T. S. & Swaminathan, B. (2006). Probing genomic diversity and evolution of *Escherichia coli* O157 by single-nucleotide polymorphisms. *Genome Res* **16**, 757–767.
- Zhou, S., Kile, A., Bechner, M., Place, M., Kvikstad, E., Deng, W., Wei, J., Severin, J., Runnheim, R. & other authors (2004). Single-molecule approach to bacterial genomic comparisons via optical mapping. *J Bacteriol* **186**, 7773–7782.

Edited by: S. D. Bentley