

## ORIGINAL ARTICLE

**Design and evaluation of PCR primers which differentiate *Escherichia coli* O157:H7 and related serotypes**

W.C. Rice

USDA, ARS, Conservation and Production Research Laboratory, Bushland, TX, USA

**Keywords**

amplicon length polymorphism, cluster analysis, *E. coli* O157, microbial typing, molecular typing, rep-PCR.

**Correspondence**

William C. Rice, USDA, ARS, Conservation and Production Research Laboratory, PO Drawer 10, 2300 Experiment Station Road, Bushland, TX 79012-0010, USA.  
E-mail: William.Rice@ars.usda.gov

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

2008/0047: received 9 January 2008, revised 22 February 2008 and accepted 17 March 2008

doi:10.1111/j.1365-2672.2008.03987.x

**Abstract**

**Aims:** To develop methods to differentiate *Escherichia coli* O157:H7 and related serotypes by the use of amplicon length polymorphism (ALP) analysis based on identifying DNA sequence deletions within highly homologous regions of three sequenced *E. coli* strains.

**Methods and Results:** Potential primer locations along the ancestral genomic backbone were identified and evaluated against three sequenced genomes and then applied to a reference set of pathogenic *E. coli* strains. All 16 primer combinations generated the expected diagnostic fragments as predicted for the *E. coli* K12 MG1655, O157:H7 EDL933, and O157:H7B Sakai genomes.

**Conclusions:** This study defines a collection of primers distributed along the length of the *E. coli* genome that were applied to ALP analysis methods to successfully differentiate between serotypes of *E. coli* O157:H7 and other *E. coli* serotypes.

**Significance and Impact of the Study:** ALP-PCR analysis method was validated as an independent method of classification when compared with that of rep-PCR. The principles underlying ALP analysis can be readily applied for the detection and differentiation of other closely related microbial species because of the abundance of complete DNA sequence data for a large number of microbial genomes.

**Introduction**

*Escherichia coli* O157:H7 is a major food-borne infectious pathogen and, because of its worldwide distribution, is a serious threat to public health systems. *Escherichia coli* O157:H7 was first recognized following two outbreaks in Michigan and Oregon in 1982 (Riley *et al.* 1983) and can cause diarrhoea, hemorrhagic colitis, and haemolytic uremic syndrome. In the United States, approximately 75 000 cases of O157:H7 infections are now estimated to occur annually (Mead *et al.* 1999). Important virulence attributes that have been identified are shiga toxin production (Konowalchuk *et al.* 1997), characteristic attaching and effacing (A/E) lesions and intimin protein production encoded by the *eae* gene (Knutton 1995; Whittam 1998). However, several lines of evidence underscore the fact that other unknown virulence factors may be expressed by *E. coli* O157:H7. Strains of *E. coli* that are

*eae* negative are still capable of colonizing the intestinal epithelium (Su and Brandt 1995; Tzpori *et al.* 1995) and other strains of *E. coli* that show different levels of shiga toxin production have the same clinical manifestation (Paton *et al.* 1999).

In order to better understand the pathogenesis and evolution of *E. coli* O157:H7 and to develop better methods of identification, several groups have completed the DNA sequence determination of two genomes of *E. coli* O157:H7; isolates O157:H7 EDL933 (Perna *et al.* 2001) and O157:H7B Sakai (Hayashi *et al.* 2001). The Sakai strain was implicated in a major enterohaemorrhagic *E. coli* O157:H7 outbreak in Japan (Watanabe *et al.* 1996). A 4.1 Mb ancestral genomic backbone was implicated between the DNA sequence of the two O157:H7 genomes and that of the previously sequenced common laboratory strain of *E. coli* K12 MG1655 (Blattner *et al.* 1997). Evidence for a common *E. coli* genomic backbone

had already been suggested by multilocus sequencing of seven housekeeping genes of *E. coli* O157:H7 (Reid *et al.* 2000). Significant lateral gene flow was observed between the genomes and attributed in part to prophages. Approximately 800 kb of DNA is unique to the O157:H7 serotypes and thus contains the gene function(s) that are responsible for severe infections caused by *E. coli* O157:H7 serotypes. These include putative virulence factors, alternative metabolic capacities, prophages and other new functions.

There are many methods (reviewed, Olive and Bean 1999) each with their respective advantages and disadvantages that have been developed to type and subtype various microbial organisms. The advantages of the sensitivity and speed of polymerase chain reaction (PCR) based approaches have been recognized along with some of the attendant draw backs. DNA sequence determination and analysis is often considered along with pulse field gel electrophoresis as a 'gold standard' for microbial genome identification and typing; however, these approaches require the greatest amount of time and expense.

Herein, the advantages of the wealth of DNA sequence knowledge regarding the *E. coli* O157:H7 and K12 genomes and the power of PCR to rapidly detect and differentiate various *E. coli* O157:H7 genomes and other related 'O' serotypes are combined. A number of papers have been published to detect and/or differentiate between various *E. coli* genomes based on the O antigen coding locus. Some of these *E. coli* serogroups are O15 (Beutin *et al.* 2005), O26 and O111 (Louie *et al.* 1998), O104 (Wang *et al.* 2001), O138 and O139 (Wang *et al.* 2005), O145 (Feng *et al.* 2005) and O111, O113, and O157 (Paton and Paton 1999). They are detected and/or differentiated using gel based methods, whereas real-time PCR assays have been developed for the analysis of *E. coli* serogroups O103 (Perelle *et al.* 2005) and O157 (Fortin *et al.* 2001). All the above-mentioned methods are targeted at specific genes within the O antigen gene locus and are limited in their ability to differentiate among all serogroups. Methods of amplicon length polymorphism (ALP, Ghareyazie *et al.* 1995) in rice genome analysis were developed based on identifying DNA sequence deletions within highly homologous regions. The suitability of the ALP method was evaluated using three sequenced *E. coli* strains. Potential primer locations along the ancestral genomic backbone were identified and evaluated against these three sequenced genomes and then applied to a reference set of pathogenic *E. coli* strains. Results obtained from ALP analysis were compared with those obtained by an established typing method such as repetitive extragenic palindromic-PCR (rep-PCR) based method (Versalovic *et al.* 1991) and these efforts are presented in this study.

## Materials and methods

### Bacterial isolates

The DECA set of *E. coli* strains ( $n = 75$ , when received) contains representative isolates of 15 common diarrheagenic clones (Table 1) based on electrophoretic types (ET). The ET classification provides a clonal framework for assessing the diversity of *E. coli* strains. The DECA set is comprised of 28 serotypes and was obtained along with the O157:H7 EDL933, and the O157:H7B Sakai strains from the National Food Safety and Toxicology Center at Michigan State University (2002) (East Lansing, MI, USA), while the *E. coli* K12 MG1655 strain was obtained from the *E. coli* Genetic Stock Center at Yale University (New Haven, CT, USA). Additional information regarding DECA strains may be obtained from the STEC centre (<http://www.shigatox.net/cgi-bin/stec/index>).

### DNA isolation

DNA was isolated using the PUREGENE DNA Isolation Kit (Gentra Systems, Minneapolis, MN, USA) according to manufacturer's protocol for Gram-negative bacteria. DNA yield was determined using the Hoechst dye method (Cesarone *et al.* 1979) using a VersaFluor™ fluorometer according to manufacturer's protocol (BioRad Laboratories, Hercules, CA, USA).

### DNA homology analysis

Global alignment of *E. coli* strains K12 MG1655 (accession L48811), O157:H7 EDL933 (accession AE005174), and O157:H7B Sakai (accession BA000007) genomic DNA sequences was performed using ALIGN PLUS v4.1 (Sci-ed Software, Durham, NC, USA). Genomic DNA homology analysis was performed as two-way alignments of approximately 2 Mb blocks of DNA using the Fast-Scan program set for maximum score. Initially the *E. coli* K12 MG1655 genome was aligned against O157:H7 EDL933 genome and the O157:H7 EDL933 was aligned against O157:H7B Sakai genome. In this way, major regions ( $\approx 100$ –600 kb) of high homology ( $\geq 97\%$ ) were identified. Three-way global alignments were then conducted using either the K12 MG1655 or the O157:H7 EDL933 genome as the reference genome dependent upon which two genomic regions shared the greatest homology. This method identified deletions within a genome (either the K12 or the two O157 genomes) such that it was possible to amplify a single unique PCR generated diagnostic fragment with respect to either the K12 or the two O157 genomes (two exceptions noted later).

**Table 1** DECA set of strains used in this study

Code	Serotype	Et class	Code	Serotype	Et class
DEC1A	O55:H6	EPEC1r4r6	DEC8D	O111:H11	EHEC2r6
DEC1B	O55:H6	EPECr6	DEC8E	O111:H8	EHEC2r6
DEC1C	O55:H6	EPEC1r6	DEC9A	O26:H11	EHEC2r6
DEC1D	O55:H6	na	DEC9B	O26:HN	EHEC2r6
DEC1E	O55:H6	EPEC1r6	DEC9C	O52:H26	na
DEC2A	O55:H6	EPEC1r4r6	DEC9D	O26:H11	EHEC2r6
DEC2B	O55:HNM	EPEC1r6	DEC9E	O26:H11	na
DEC2C	O55:H6	EPEC1r6	DEC10A	O26:H11	na
DEC2D	O55:H6	EPEC1r6	DEC10B	O26:H11	EHEC2r2r6
DEC2E	O55:H6	EPEC1r6	DEC10C	O26:H11	EHEC2r2r6
DEC3A	O157:H7	EHEC1r6	DEC10D	O26:H11	na
DEC3B	O157:H7	EHEC1r6	DEC10E	O26:H11	EHEC2r6
DEC3C	O157:H7	EHEC1r6	DEC11A	O128a:H2	EPEC2r4r6
DEC3D	O157:H7	EHEC1r6	DEC11B	O128a:H2	DECr6
DEC3E	O157:H7	EHEC1r6	DEC11C	O45:H2	STECr6
DEC4A	O157:H7	EHEC1r6	DEC11D	O128:H2	DECr6
DEC4B	O157:H7	EHEC1r6	DEC11E	O128:H2	DECr6
DEC4C	O157:H7	EHEC1r6	DEC12A	O111:H2	EPEC2r4r6
DEC4D	O157:H7	EHEC1r6	DEC12B	O111:H2	EPEC2r6
DEC4E	O157:H7	EHEC1r6	DEC12C	O111:HNM	EPEC2r6
DEC5A	O55:H7	EHEC1r6	DEC12D	O111:H2	EPEC2r6
DEC5B	O55:H7	EHEC1r4r6	DEC12E	O111:HN	EPEC2r6
DEC5C	O55:H7	EHEC1r6	DEC13A	O128:H7	DECr6
DEC5D	O55:H7	EHEC1r1r4r6	DEC13B	O128:H7	DECr6
DEC5E	O55:H7	EHEC1r6	DEC13C	O128:H7	DECr6
DEC6A	O111:H21	DECr6MLST20r4	DEC13D	O128:H7	DECr6
DEC6B	O111:H12	na	DEC13E	O128:H47	DECr6
DEC6C	O111:H12	DECr6	DEC14A	O128:H21	DECr6
DEC6D	O:H -na	na	DEC14B	O128:H21	DECr6
DEC6E	O111:HNM	DECr6	DEC14C	O128a:H21	DECr6
DEC7A	O157:H43	DECr6	DEC14D	O128:HNM	DECr6
DEC7B	O149:HNM	DECr6	DEC14E	O128:H21	DECr6
DEC7C	O157:H43	DECr6	DEC15A	O111:H21	EPECr6
DEC7D	O157:H43	DECr6	DEC15B	O111:H21	DECr6
DEC7E	O157:HNM	DECr6	DEC15C	O111:H21	DECr6
DEC8A	O111a:HNM	EHEC2r6	DEC15D	O111:H21	DECr6
DEC8B	O111:H8	DECr6	DEC15E	O111:H21	DECr6
DEC8C	O111:HNM	EHEC2r2r6			

na, Not available.

### Polymerase chain reaction

Primers were designed and selected using various programs within PRIMER DESIGN ver. 4.2 (Sci-ed Software, Durham, NC, USA) against K12 MG1655 or O157:H7 EDL933 genomes. Criteria used for primer construction were: length (20 bp), %GC (50–60),  $T_m$  °C (55–80), 3' dimers (<3), dimers-any (<7), stability (>2.0 kcal/s 5' vs. 3'), runs (<3), and repeats (<3). Primers that ranked high in these criteria, that were located within homologous regions flanking small deletions (<1500 bp), and were widely distributed within the *E. coli* genomes were then subjected to Primer Mix analysis (Primer Design v4.2 software) performed against some of the various complete

microbial genomic DNA sequences deposited in GenBank. Primers successfully meeting the above cited criteria were selected for use in PCR analysis and used to generate computer derived amplicons for each genome. Primer sequences (5' to 3') along with their genomic locations and amplicon designations are listed in Table 2.

ALP-PCR reactions (25  $\mu$ l) contained the following: 50 ng of template DNA, 1 $\times$  reaction buffer (Gibco BRL, Grand Island, NY, USA), 1.5 mmol l<sup>-1</sup> MgCl<sub>2</sub>, 200  $\mu$ mol l<sup>-1</sup> of each dNTP, 20 pmol of each primer, and 1 U of Platinum *TaqI* DNA polymerase (Gibco BRL). A BioRad iCycler thermocycler (BioRad Laboratories, Hercules, CA, USA) was used with a temperature profile of 2 min at 95°C, followed by 29 cycles of 30-s

Amplicon	Primer*	Primer sequence 5'-3'	T-An†	Predicted fragment (bp)		
				K12	O157	O157B
A	F50044	ggtgaagtcggtggatgaag	55	1164	1725	1725
	R51207	attggcgacgttcatgggtg				
B	F66504	agcggcaggttgctctaa	58	522	437	437
	R67025	ctccgaagcgtggatctc				
C	F77413	ttaccgacagcgaagtgtg	58	1746	525	525
	R79158	aagtgtgattgcgagcgt				
D	F111117	cagatgcgcacatggcgaat	60	565	449	449
	R111681	agcgtgactggagcgaaga				
E	F127570	cgccgtctggtgatgaagt	55	416	299	299
	R127985	caacggaaggcagcggagta				
F	F184031	cggaggctgtaacgatcaat	53	1121	408	408
	R185151	ttaacgaactgctgcgtacc				
G	F192470	acgctggctcgtgaccataa	55	443	301	301
	R192912	tccatgctacttcagcatc				
L	F638691	caactctggctcgtctctg	55	261	145	145
	R638951	catcatgcaagcggcctctg				
O	F761948	ctgctggaagcgtgtagt	51	374	193	193
	R762321	gcggagtagtacagccataa				
R	F804927	agagcgcgagattatcaagg	52	313	220	220
	R805239	tgcagaggcgaagaagtaag				
OB-A	OB-F2291005	gaccacctgacaggacaata	50	419	257	257
	OB-R2291261	gcctgaattactgggatgtg				
OB-F	OB-F2484561	cacaggttgccgacgcgatt	52	314	558	558
	OB-R2485118	ccaggagagcgtgctgtatt				
OB-I	OB-F5148891	tggcttatgagcgtgaagt	51	748	303	303
	OB-R5149193	ctggttgaaacctgagctgta				
OB-O	OB-F5339630	atcgacggtggtgtcattgt	55	185	398	398
	OB-R5340027	cgtaggcaaggctgttatt				
O-C‡	O157-F596247	gtcgtctcgggtcagcacta	57	0	468	777, 468
	O157-R596714	caggtagcattggcggcaac				
O-E	O157-F810456	cacaccaccgacgcaatct	59	0	624	537
	O157-R812748	cagccgtaccaggtcacttg				

**Table 2** Primer sequences and amplicon designation used in this study

\*F forward primer, R reverse primer, number indicates location on K12 or O157 (O-) genome.

†T-An annealing temperature.

‡Data not used in cluster analysis.

denaturation at 95°C, 30 s at the respective annealing temperature (Table 2), and 1-min extension at 72°C. The 29 cycles were followed by one cycle for 7.5 min at 72°C. Rep-PCR reactions conditions (Woods *et al.* 1992) were similar to ALP-PCR except the annealing reaction was performed at 44°C and 40 pmol of the following primers REP1R-I (5'-IIICGICGICATCIGGC-3') and REP2-I (5'-ICGICTTATCIGGCCTAC-3') were used per reaction. Interspersed repetitive DNA sequence (BOX)-PCR was performed using the BoxA1R primer and BoxA1R PCR reaction conditions were identical to the rep-PCR conditions, except that an annealing reaction was performed at 50°C. Aliquots (5 µl, ALP-PCR; 12 µl, rep-PCR) from each reaction set were run on a 1.4% or 1.5% agarose gel depending on the size of the amplicon generated and stained with ethidium bromide (Sambrook *et al.* 1989).

All PCR assays were repeated at least twice to verify the validity of the primer reaction sets.

#### DNA and data analysis

DNA gel images were captured and analysed by using a computerized video image analysis system (Kodak Digital Science Image Station Model 440CF, Eastman Kodak, Rochester, NY, USA) and exported as TIFF files. DNA molecular weights of observed fragments were determined using image analysis software v3.5.4 (Kodak Digital Science, Rochester, NY, USA). The pGEM and 100-bp DNA Ladder markers (Promega, Madison, WI, USA) and the high-throughput ladders I and II (Biolone, Randolph, MA, USA) were used as size standards. Banding patterns for each isolate were recorded as a categorical code: i.e.

the base pair value present in Table 2 for a K12 MG1655 or for an O157:H7/O157:H7B generated amplicon was used. Missing bands were coded as 0 and novel bands were co-coded as the molecular weight value determined by analysis of agarose gel electrophoresis.

Analysis of the categorical scores from the ALP-PCR reactions was performed using Bionumerics software program version 5.0 (Applied Maths, Austin, TX, USA). A similarity matrix was calculated using the Multi-state coefficient categorical (based on all primer sets in Table 2 except O-C). Analysis of the composite dataset based on rep-PCR assays employing primer sets BoxA1R and REP1R-I\_ REP2-I PCR DNA gel images was performed using various programs within the Bionumerics software program version 4.0 (Applied Maths, Austin, TX, USA). Identical molecular weight standards were included in each gel to allow for normalization of gel images for valid between-gel comparisons of DNA fingerprints. Similarity matrices of densitometric curves from the rep-PCR profiles were calculated using the Dice coefficient. Cluster analysis of the similarity matrices were performed by an unweighted pair group method with arithmetic mean (UPGMA) algorithm (Sneath and Sokol 1973) (data averaged from both experiments). The correlation was expressed as per cent similarity. Two tests (Pearson correlation and Kendall's Tau) were conducted to assess the degree of congruence between the ALP and rep-PCR assay methods.

## Results

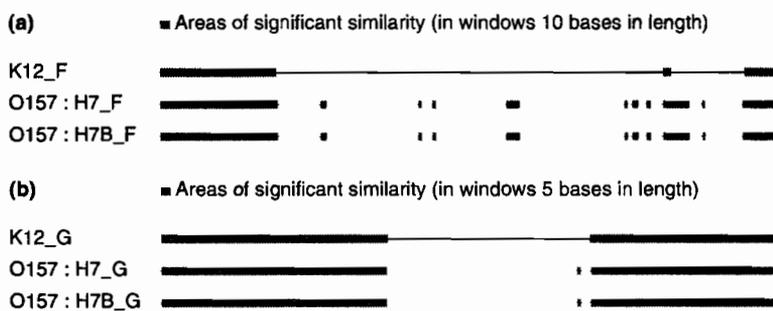
Global alignment of the *E. coli* K12 MG1655, O157:H7 EDL933 and O157:H7B Sakai genomic DNA identified regions of high homology shared by all three genomes that were interspersed with regions of low homology, deletions or both. Computer generated amplicons respective to each primer-genome set were globally aligned to indicate the DNA homology present at each genomic location (e.g. amplicons F and G, Fig. 1). Sixteen primer combinations (Table 2) were selected for their ability to detect and differentiate *E. coli* K12 MG1655, O157:H7

EDL933 and O157:H7B Sakai genomes and were then evaluated against various *E. coli* serotypes within the DECA strain set. These primers span a 2-Mb region surrounding the origin of replication and a 1-Mb region within the middle of the *E. coli* genomes. Results of 'in silico' Primer Mix analysis against selected Gram-negative and Gram-positive microbial genomes indicate that each primer combination is specific only to *E. coli* K12 MG1655, O157:H7 EDL933, and O157:H7B Sakai genomes (Table 3). Either one or the other primer or both would fail to bind at respective annealing temperatures, thus no amplifiable product would be produced.

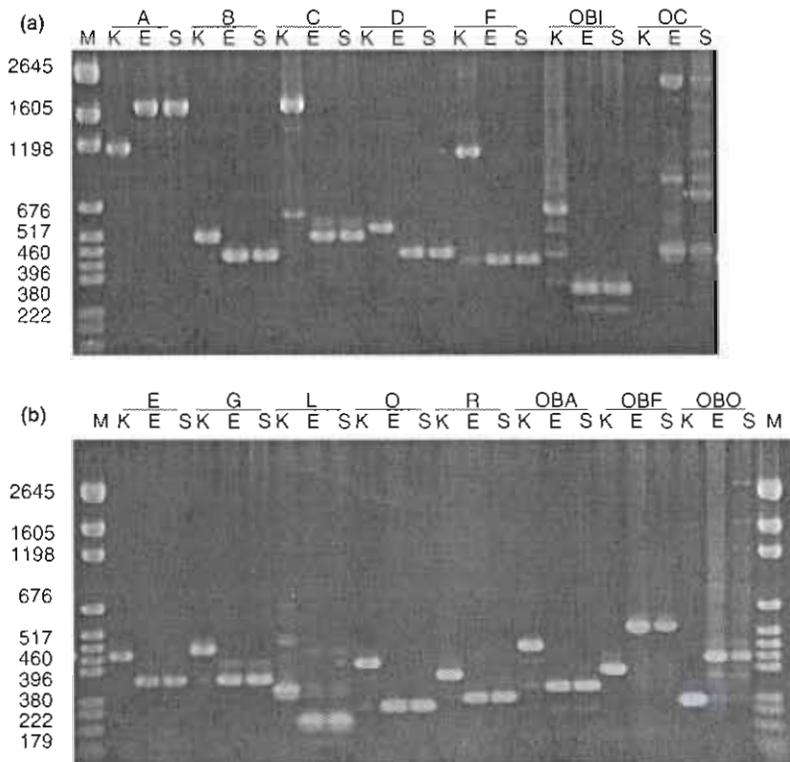
All 16 primer combinations generated the expected diagnostic fragments as predicted for the *E. coli* K12 MG1655, O157:H7 EDL933, and O157:H7B Sakai genomes (Figs 2a,b and 4b). As noted earlier, primer sets O-C and O-E generated a different fragment for each microbial genome. Primer set O-C (Fig. 2a, lanes 24, 25, and 26) identified a region of repeated DNA that contained two copies of the forward primer resulting in generation of two amplicons. Several primer sets (not shown in Table 2) identified genomic regions containing multiple copies of one or more of the primers that could possibly generate six to ten amplicons

**Table 3** Results of primer set design analysis with other microbial genomes

Genome	Accession no.	Predicted bands
<i>Salmonella enterica</i> serovar Typhi	AL513382	None
<i>Salmonella typhimurium</i> LT2	AE006468	None
<i>Listeria monocytogenes</i> EGD	NC-003210	None
<i>Listeria innocua</i> Clp11262	AL592022	None
<i>Pasteurella multocida</i> PM70	AE004439	None
<i>Campylobacter jejuni</i>	AL111168	None
<i>Pseudomonas aeruginosa</i> PA01	AE004091	None
<i>Methanobacterium thermoautotrophicum</i> delta H	NC-000916	None
<i>Bacillus subtilis</i>	NC-000964	None
<i>Bacillus halodurans</i>	NC-002570	None



**Figure 1** (a, b) Homologous regions for *Escherichia coli* K12 MG1655, O157:H7 EDL933 and O157:H7B Sakai genomic amplicons generated by primer sets F and G respectively.



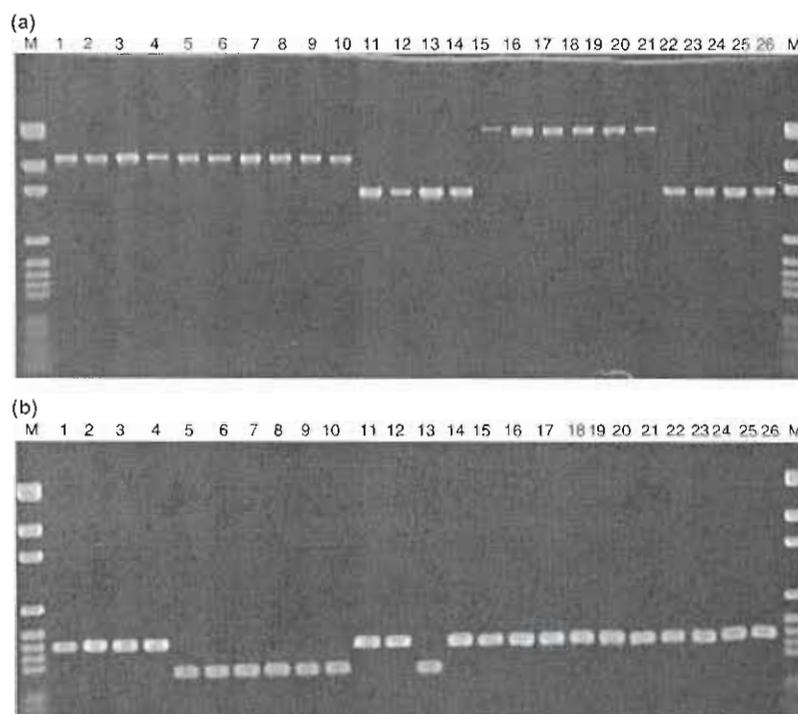
**Figure 2** (a) Amplicons generated by primer sets A, B, C, D, F, OB-I, and O-C for *Escherichia coli* genomes K12 MB 1655 (K), O157-H7 EDL933 (E), and O157 H7B Sakai (S) respectively. (b) Amplicons generated by primer sets E, G, L, O, R, OB-A, OB-F and OB-O for *E. coli* genomes K12 MB 1655 (K), O157.H7 EDL933 (E), and O157 H7B Sakai (S) respectively.

within a single genomic location. Thus, a specific K12 or O157:H7/O157:H7B genomic backbone can be correlated to a genomic region based on the use of these primer sets.

Evidence for the genomic plasticity of the O serotypes in the DECA strain set is indicated by the patterns of bands observed with various primer combinations (e.g. primer sets A and G, Fig. 3a,b). Eight unique bands (i.e., not predicted) were observed in a number of strains (range 1–19) with respect to six of the primer sets utilized (e.g. primer set A, Fig. 3a; Table 4). In addition, five of the primer sets failed to generate amplicons in a number of strains (range 1–36). The most likely explanation for strains that fail to amplify for a given primer set is that the genomic sequence of that strain for that site has undergone a mutation as all strains yield amplicons for the majority of primer sets evaluated. The 468-bp fragment amplified by primer set O-C was present only in serotypes O157:H7 and O55:H7 (Fig. 4a and data not shown). In all other O serotypes, the specific O157:H7B 777-bp fragment was observed. Primer set O-E is also capable of distinguishing among all three *E. coli* genomes, as O157:H7 and O157:H7B DNA generate differently sized amplicons, whereas K12 DNA does not yield a product (Table 2, Fig. 4b, lanes 6, 7, and 8). For the O-E primer set, the 624-bp fragment indicative of O157:H7

backbone was observed in 41 of the DECA strains, while in 9 of the DECA strains the 537 bp fragment or O157:H7B backbone was observed (Fig. 4a and data not shown). The rest of the strains failed to amplify, thus indicating a K12 backbone at this location or alternatively the insertion or transfer of a sequence from another *E. coli* strain. In all O157:H7 and O55:H7 serotypes in the DECA set of strains, the predicted O157:H7 and O157:H7B banding pattern was observed with all primer sets evaluated.

Cluster analysis of the ALP generated banding patterns revealed three main clusters (similarity range of 63–76%) containing a number of distinct subclusters among the DECA set of strains (Fig. 5). One main cluster (similarity of 63.6%) consisted of two subclusters; the first subcluster (mean similarity of 74.5%) contained a total of 27 strains with the EDL933, Sakai, and the O157:H7, O55:H7, O55:H6 and O55:H5 serotypes present. In addition, note that the Sakai strain is differentiated from the EDL933 strain and other O157:H7 strains and that these strains of O157:H7 and O55:H7 serotypes are resolved into both subclusters. Four O157:H43/HNM serotypes formed an outlying nonsignificant cluster diverging with a mean similarity of 60%. The second subcluster of this group (mean similarity of 83%) was comprised mainly of O128, O111, O26 and O45 serotypes with EPEC2r6 as the



**Figure 3** Amplicons generated by primer sets A (a) and G (b) for DECA strains. Lanes M, 1, DEC1A; 2, DEC1B, 3, DEC2A, 4, DEC2B, 5, DEC3A; 6, DEC3B; 7, DEC4A, 8, DEC4B; 9, DEC5A, 10, DEC5B, 11, DEC6A, 12, DEC6B; 13, DEC7A; 14, DEC7B, 15, DEC8A, 16, DEC8B; 17, DEC9A; 18, DEC9B, 19, DEC10A, 20, DEC10B; 21, DEC11A, 22, DEC11B, 23, DEC12A, 24, DEC12B; M.

**Table 4** Primer sets which amplify unique bands or no band is produce

Amplicon	No strains	Unique bands (bp)	Strains no band
A	19	2625	1
B	36	–	+
D	9	1650	–
E	10	–	+
OB-A	1	483	–
OB-E	1	337	–
OB-F	9	1375, 731	10
OB-I	5	–	+
O-C	15	1925, 900	–

dominant ET class. The other two dominant clusters were comprised of serotypes O128, O111, O45, O26, O52, O149 and O11 that formed a mixed distribution of sub-clusters within these main clusters (Fig. 5). The different O serotypes appear to be evenly distributed between these two main clusters and also appear to be influenced by their ET classification.

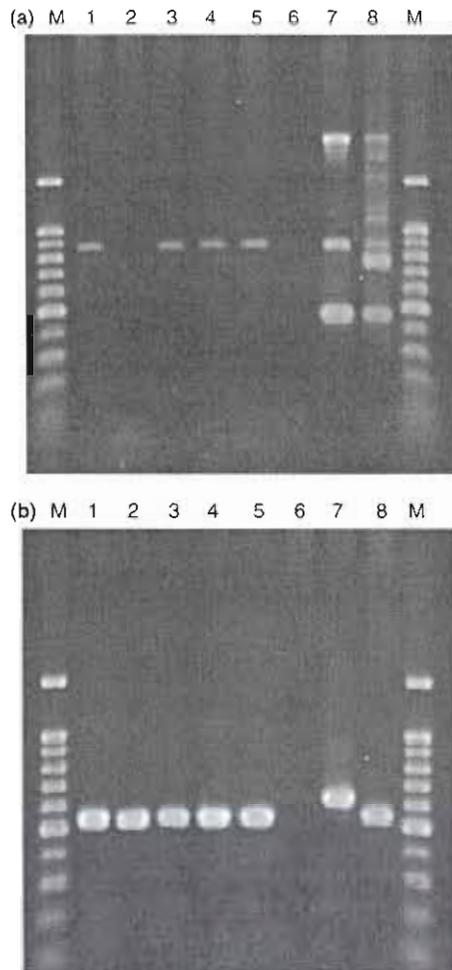
The composite rep-PCR dataset based on primer sets REP1R-I–REP2-I and BoxA1R revealed three main clusters with an observed range of similarities of 44–75% (Fig. 6). One main cluster contained all O157:H7 and O55:H7 serotypes with a mean similarity of 68% while another was comprised of all O55:H5 and O55:H6 serotypes with an overall similarity of 75%. The EDL933 and

Sakai strains clustered together with an overall similarity of 87%. All other serotypes were distributed within the dominant main cluster with an overall similarity of 44%. Strains grouped within this dominant cluster appear to be influenced both by their serotype and ET class as was observed with the ALP dataset. Again O157:H43/HNM serotypes formed an outlying cluster (82% similarity).

An evaluation of the concordance observed between the experimental ALP method described here with a previously established method to measure genetic variation was determined. The Pearson correlation test revealed that 40% congruence was observed between the ALP and composite rep-PCR methods, whereas the Kendall's Tau test indicated that there was a  $28.73 \pm 1.26\%$  agreement between the methods.

## Discussion

This study defined a collection of primers distributed along the *E. coli* genome that can be used to both detect and differentiate between strains of *E. coli* O157:H7 and other *E. coli* serotypes. The 468 bp amplicon generated by primer set O-C was specific only to serotypes O157:H7 and O55:H7 while O-E primer set was capable of differentiating between K12, O157:H7 and O157:H7B. Application of a real-time PCR assay using a molecular beacon that targets the *rfbE* gene enables the detection of only O157:H7 serotype (Fortin *et al.* 2001) and can



**Figure 4** Amplicons generated by primer sets O-C (a) and O-E (b) respectively for DECA strains. Lanes M, 100 bp ladder, 1, DEC5A; 2, DEC5B, 3, DEC5C; 4, DEC5D, 5, DEC5E; 6, K12 MB 1655; 7, O157 H7 EDL933, 8, O157 H7B Sakai; M, 100 bp ladder

successfully differentiate O157:H7 from O55:H7. Additional primer sets employing gel based PCR assays have been developed to detect host specific *vaeA* gene homologs in *E. coli* serotypes O111:H8, O26:H11 and O157:H7 (Louie *et al.* 1998). These primers may prove useful in initial evaluations of large sets of suspect environmental isolates of *E. coli* O157:H7. The utility of primer sets described in this study are that they are capable of defining the extent of a K12 or an O157:H7 ancestral genomic backbone. This is based on the banding patterns that are observed with the various serotypes. Thus, a genetic rela-

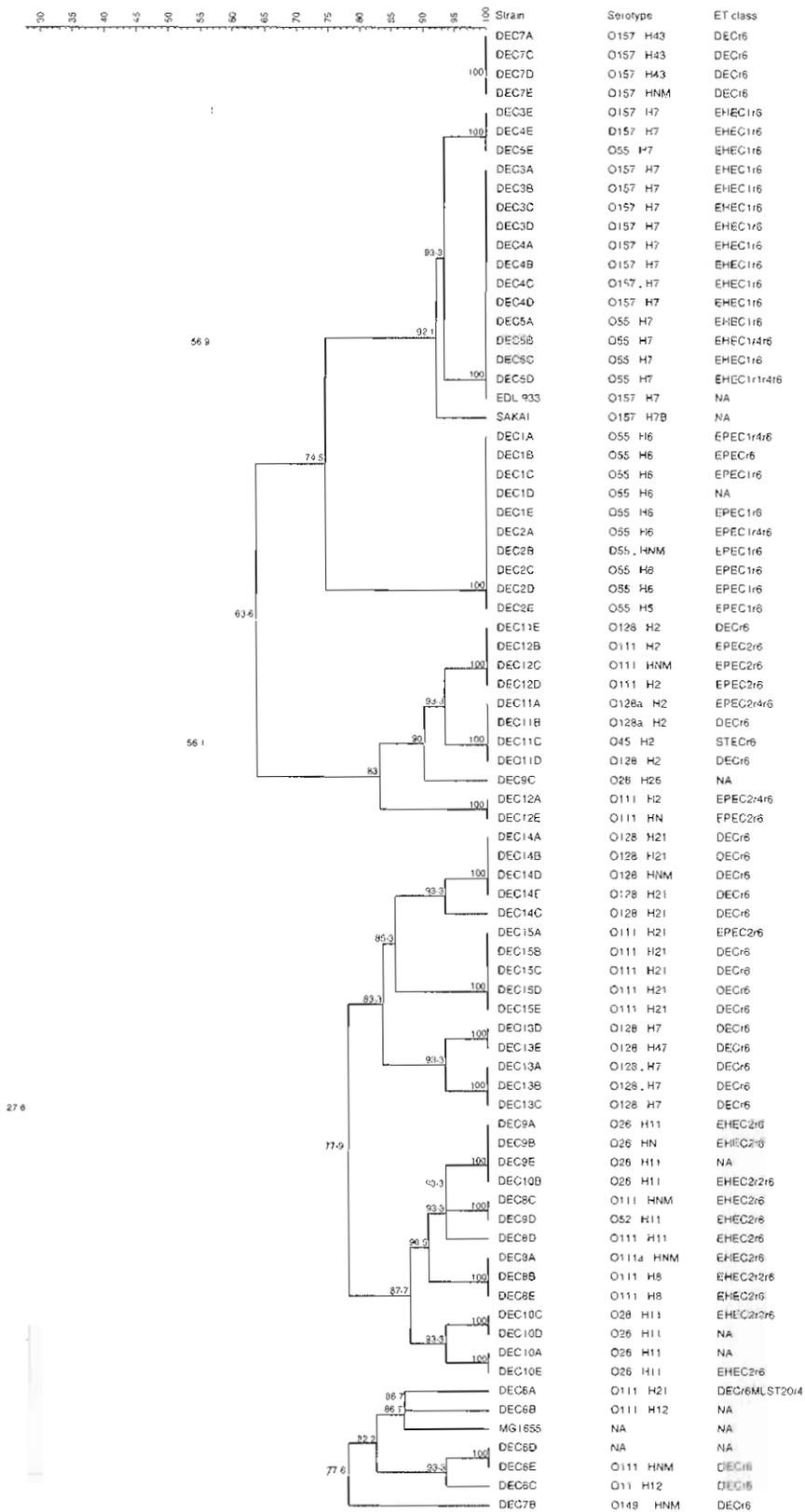
tionship can be established for *E. coli* strains within the individual serotypes. In addition, the unique bands detected in some strains may be subjected to DNA sequence analysis to obtain evolutionary data and may also provide additional information that is useful for subtyping of *E. coli* serotypes.

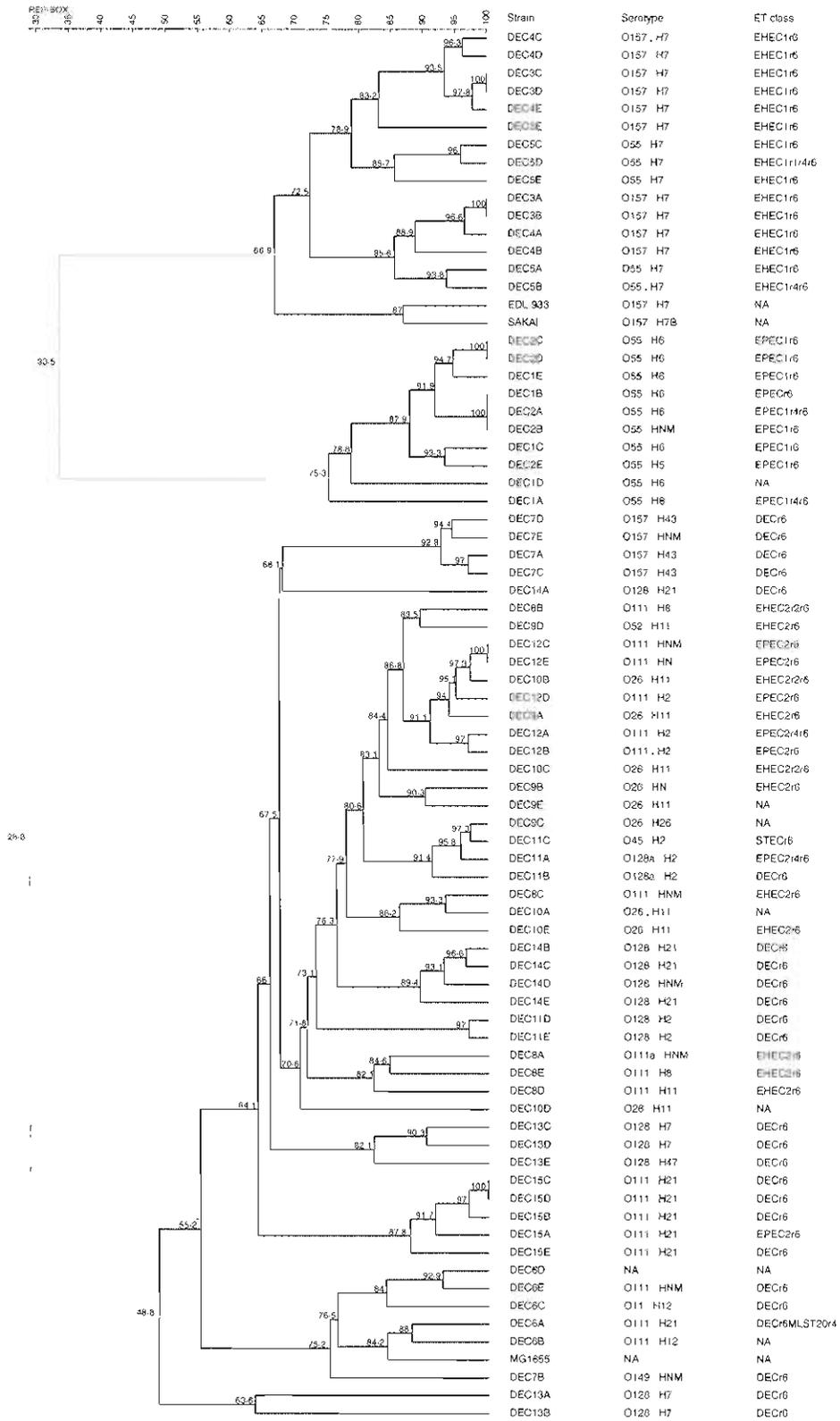
Results obtained from the cluster analysis in this study are consistent with the proposal of *E. coli* O55:H7 as the ancestral parent of *E. coli* O157:H7 (Whittam *et al.* 1993). Additionally, cluster analysis results observed herein are consistent with the evolutionary results obtained by application of MLST of housekeeping genes that employed some of the strains that were used in this study (Reid *et al.* 2000). They also observed high similarity between O157:H7 and O55:H7 with O111 serotypes falling into another group of clustered strains and K12 in another distinct group. The fact that O157:H7, which is highly pathogenic to humans and O55:H7, is less so, suggests that genomic scale sequencing efforts should be directed at O55:H7. Complete DNA sequence analysis of O55:H7 genome should help to further define the evolution of O157:H7 and thus aid in the identification of additional genes involved in virulence. An evaluation of the overall concordance observed between ALP-PCR and rep-PCR methods suggests that there are different evolutionary pressures acting on the genomes of these various *E. coli* serotypes and that ALP method is useful in detecting some of these influences.

The combination of using genomic information (bioinformatics) with ALP-PCR analysis has resulted in a method to detect and differentiate various *E. coli* strains isolated from various sources (human, cow, calf, pig and hamburger) across many locations (over 25 countries and 15 states) into distinct genomic categories (clusters) independent of serotype classification, ET class, housekeeping genes and biochemical traits. Thus, a genetic index based on either K12 or O157 DNA sequences at specific genomic locations can be used to track lineages of various *E. coli* strains. This assessment cannot be made with other rep-PCR methods of analysis as the genomic locations of the repetitive DNA elements within a given strain are unknown. ALP-PCR analysis method was validated as an independent method of classification when compared with that of rep-PCR analysis method and along with other repetitive element PCR based assays may prove useful for high resolution differentiation of genetic relationships between very closely related strains.

Five rep-PCR DNA fingerprinting methods {rep-PCR, enterobacterial repetitive intergenic consensus (ERIC)-PCR, ERIC2-PCR, BOX-PCR and (GTG)<sub>5</sub>-PCR} have

**Figure 5** ALP-PCR dendrogram derived from cluster analysis of a similarity matrix calculated using the Dice coefficient. Cluster analysis was performed using the unweighted-pair group method, arithmetic average. Values are expressed as per cent similarity





**Figure 6** Rep-PCR dendrogram derived from cluster analysis of a similarity matrix calculated using the Dice coefficient. Cluster analysis was performed using the unweighted-pair group method, arithmetic average. Values are expressed as per cent similarity.

been compared against one another for their ability to differentiate faecal *E. coli* from humans, poultry and wild birds (Mohapatra *et al.* 2007). Cluster analysis and discriminant function analysis identified (GTG)<sub>5</sub> as the best method available to differentiate amongst these various *E. coli* isolates. These five rep-PCR methods were subsequently evaluated in a bacterial source tracking (BST) study. The (GTG)<sub>5</sub> method was shown to be the best method to differentiate amongst *E. coli* isolates from natural pond water. The suitability of ERIC-PCR for BST was evaluated against other methods that have traditionally been employed in various BST studies (Casarez *et al.* 2007). The three methods used for comparison against ERIC-PCR were pulsed-field electrophoresis (PFGE), automated ribotyping using HindIII, and Kirby-Bauer antibiotic resistance analysis (KB-ARA). It was concluded that construction of composite datasets, either two- or four-method datasets provided the highest rates of correct classification of isolates. ALP-PCR would appear to be a DNA fingerprinting method that is suitable for BST studies. Application of ALP-PCR (as it represents an independent method) with any of various rep-PCR assays or traditional BST methods mentioned above would contribute to improving rates of correct classification.

The principles underlying ALP-PCR method should be applicable to other micro-organism whose DNA sequence composition has been determined. The development of highly automated large-scale DNA sequencing machines has led to an explosion of microbial genome sequencing projects. Nearly 600 bacterial genomes and almost 50 Archaeal genomes have been sequenced (Liolios *et al.* 2007). *Escherichia coli* strains now represent 11 of these genomes with the addition of eight recently sequence *E. coli* isolates. These newly sequenced *E. coli* strains represent a future opportunity to realize the utility of the ALP-PCR approach.

### Acknowledgements

The author would like to thank Mr Colby Carter for the preparation of media, bacterial DNA and reagents.

### References

- Beutin, L., Tao, J., Feng, L., Krause, G., Zimmermann, S., Glier, K., Xia, Q. and Wang, L. (2005) Sequence analysis of the *Escherichia coli* O15 antigen gene cluster and development of a PCR assay for rapid detection of intestinal and extraintestinal pathogenic *E. coli* O15 strains. *J Clin Microbiol* **43**, 703–710.
- Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474.
- Casarez, E.A., Pillai, S.D., Mott, J.B., Vargas, K.E. and Di Giovanni, G.D. (2007) Direct comparison of four bacterial source tracking methods and use of composite data sets. *J Appl Microbiol* **103**, 350–364.
- Cesarone, C., Bolognesi, C. and Santi, L. (1979) Improved microfluorometric DNA determination in biological material using 33258 Hoechst. *Anal Biochem* **92**, 497–500.
- Feng, L., Senchenkova, S.N., Tao, J., Shashkov, A.S., Liu, B., Shevelev, S.D., Reeves, P.R., Xu, J., Knirel, Y.A. and Wang, L. (2005) Structural and genetic characterization of Enterohemorrhagic *Escherichia coli* O145 O antigen and development of an O145 serogroup-specific PCR assay. *J Bacteriol* **187**, 758–764.
- Fortin, N.Y., Mulchandani, A. and Chen, W. (2001) Use of real-time polymerase chain reaction and molecular beacons for the detection of *Escherichia coli* O157:H7. *Anal Biochem* **289**, 281–288.
- Ghareyazie, B., Huang, N., Second, G., Bennett, J. and Khush, G.S. (1995) Classification of rice germplasm. I. Analysis using ALP and PCR-based RFLP. *Theor Appl Genet* **91**, 218–227.
- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E. *et al.* (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* **8**, 11–22.
- Knutton, S. (1995) Cellular response to enteropathogenic *Escherichia coli* infection. *Biosci Rep* **15**, 469–479.
- Konowalchuk, J., Speirs, J. and Stavric, S. (1997) Vero response to a cytotoxin of *Escherichia coli*. *Infect Immun* **18**, 775–779.
- Liolios, K., Mavromatis, K., Tavernarakis, N. and Kyrpides, N.C. (2007) The genomes on line database (GOLD) in 2008: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* **36**, D475–D479.
- Louie, M., Read, S., Simor, A.E., Holland, J., Ziebell, K., Brunton, J. and Hii, J. (1998) Application of multiplex PCR for detection of non-O157 verocytotoxin-producing *Escherichia coli* in bloody stools: identification of serogroups O26 and O111. *J Clin Microbiol* **36**, 3375–3377.
- Mead, P.S., Slutsker, L., Dietz, V., McCaig, L.F., Bresce, J.S., Shapiro, C., Griffin, P.M. and Tauxe, R.V. (1999) Food-related illness and death in the United States. *Emerg Infect Dis* **5**, 607–625.
- Mohapatra, B.R., Broersma, K. and Mazumder, A. (2007) Comparison of five rep-PCR genomic fingerprinting methods for differentiation of fecal *Escherichia coli* from

- humans, poultry and wild birds. *FEMS Microbiol Lett* **277**, 98–106.
- National Food Safety and Toxicology Center at Michigan State University, East Lansing, MI. (2002) STEC Center, a reference center to facilitate the study of shiga toxin-producing *Escherichia coli*. <http://www.shigatox.net/stec/index.html>.
- Olive, D.M. and Bean, P. (1999) Principles and applications of methods for DNA-based typing of microbial organisms. *J Clin Microbiol* **37**, 1661–1669.
- Paton, A.W. and Paton, J.C. (1999) Direct detection of shiga toxigenic *Escherichia coli* strains belonging to serogroups O111, O157, and O113 by multiplex PCR. *J Clin Microbiol* **37**, 3362–3365.
- Paton, A., Woodrow, M., Doyle, R., Lanser, J. and Paton, J. (1999) Molecular characterization of a Shiga toxigenic *Escherichia coli* O113:H21 strain lacking the *eae* responsible for a cluster of cases of Haemolytic Uremic Syndrome. *J Clin Microbiol* **37**, 3357–3361.
- Perelle, S., Dilasser, F., Grout, J. and Fach, P. (2005) Detection of *Escherichia coli* serogroup O103 by real-time polymerase chain reaction. *J Appl Microbiol* **98**, 1162–1168.
- Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S. *et al.* (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533.
- Reid, S.D., Herbelin, C.J., Bumbaugh, A.C., Selander, R.K. and Whittam, T.S. (2000) Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406**, 64–67.
- Riley, L.W., Remis, R.S., Helgerson, S.D., McGee, H.B., Wells, J.G., Davis, B.R., Hebert, R.J., Olcott, E.S., Johnson, L.M., Hargrett, N.T., Blake, P.A. and Cohen, M.L. (1983) Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *New Engl J Med* **308**, 681–685.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, ed. Nolan, C. Chapter 6, pp. 6.3–6.34. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sneath, P.H.A. and Sokol, R.R. (1973) *Numerical Taxonomy*. San Francisco: Freeman.
- Su, C. and Brandt, L. (1995) *Escherichia coli* O157:H7 infection in humans. *Ann Intern Med* **123**, 698–714.
- Tzpori, S., Gunzer, F., Donnenberg, M., Demontigny, L., Kaper, J. and Donohue-Rolfe, A. (1995) The role of the *eaeA* gene in diarrhea and neurological complications in a gnotobiotic piglet model of enterohaemorrhagic *Escherichia coli* infection. *Infect Immun* **63**, 3621–3627.
- Versalovic, J., Koeuth, T. and Lupski, J.R. (1991) Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acid Res* **19**, 6823–6831.
- Wang, L., Briggs, C.E., Rothmund, D., Fratamico, P., Luchansky, J.B. and Reeves, P.R. (2001) Sequence of the *E. coli* O104 antigen gene cluster and identification of O104 specific genes. *Gene* **270**, 231–236.
- Wang, L., Lie, B., Kong, Q., Steinruck, H., Krause, G., Beutin, L. and Feng, L. (2005) Molecular markers for detection of pathogenic *Escherichia coli* strains belonging to serogroups O138 and O139. *Vet Microbiol* **111**, 181–190.
- Watanabe, H., Wada, A., Inagaki, Y., Itoh, K. and Tamura, K. (1996) Outbreaks of enterohaemorrhagic *E. coli* O157:H7 infection by two different genotype strains in Japan. *Lancet* **348**, 831–832.
- Whittam, T.S. (1998) In *Escherichia coli* O157:H7 and Other Shiga Toxin-Producing *E. coli* Strains ed. Kaper, J.B. and O'Brien, A.D. pp. 195–209. Washington, DC: ASM.
- Whittam, T.S., Wolfe, M., Wachsmith, I., Orsdov, F. and Wilson, R. (1993) Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect Immun* **61**, 1619–1629.
- Woods, C.R., Versalovic, J., Koeuth, T. and Lupski, J.R. (1992) Analysis of relationships among isolates of *Citrobacter diversus* by using DNA fingerprints generated by repetitive sequence-based primers in the polymerase chain reaction. *J Clin Microbiol* **30**, 2921–2929.