2009

Genomic Instability in Regions Adjacent to a Highly Conserved *pch* Prophage in *Escherichia coli* O157:H7 Generates Diversity in Expression Patterns of the LEE Pathogenicity Island

Zhijie Yang
University of Nebraska-Lincoln

Jaehyoung Kim
University of Nebraska-Lincoln

Chaomei Zhang
University of Nebraska-Lincoln

Min Zhang
University of Nebraska-Lincoln

Joeseph Nietfeldt
University of Nebraska-Lincoln

See next page for additional authors

JOURNAL OF BACTERIOLOGY, June 2009, p. 3553–3568 Vol. 191, No. 11
0021-9193/09/$08.00+0 doi:10.1128/JB.01738-08

Genomic Instability in Regions Adjacent to a Highly Conserved *pch* Prophage in *Escherichia coli* O157:H7 Generates Diversity in Expression Patterns of the LEE Pathogenicity Island▽

Zhijie Yang,[1] Jaehyoung Kim,[1] Chaomei Zhang,[1] Min Zhang,[1] Joeseph Nietfeldt,[1] Carolyn M. Southward,[2] Michael G. Surette,[2,3] Stephen D. Kachman,[4] and Andrew K. Benson[1]*

Department of Food Science and Technology[1] and Department of Statistics,[4] University of Nebraska, Lincoln, Nebraska 68583-0919, and Department of Microbiology and Infectious Diseases[2] and Department of Biochemistry and Molecular Biology,[3] Health Sciences Center, University of Calgary, Alberta, Canada T2N4N1

Received 11 December 2008/Accepted 19 March 2009

The LEE pathogenicity island has been acquired on multiple occasions within the different lineages of enteropathogenic and enterohemorrhagic *Escherichia coli*. In each lineage, LEE expression is regulated by complex networks of pathways, including core pathways shared by all lineages and lineage-specific pathways. Within the O157:H7 lineage of enterohemorrhagic *E. coli*, strain-to-strain variation in LEE expression has been observed, implying that expression patterns can diversify even within highly related subpopulations. Using comparative genomics of *E. coli* O157:H7 subpopulations, we have identified one source of strain-level variation affecting LEE expression. The variation occurs in prophage-dense regions of the genome that lie immediately adjacent to the late regions of the *pch* prophage carrying *pchA*, *pchB*, *pchC*, and a newly identified *pch* gene, *pchX*. Genomic segments extending from the holin S region to the *pchA*, *pchB*, *pchC*, and *pchX* genes of their respective prophage are highly conserved but are nonetheless embedded within adjacent genomic segments that are extraordinarily variable, termed *pch* adjacent genomic regions (*pch* AGR). Despite the remarkable degree of variation, the pattern of variation in *pch* AGR is highly correlated with the distribution of phylogenetic markers on the backbone of the genome. Quantitative analysis of transcription from the *LEE1* promoter further revealed that variation in the *pch* AGR has substantial effects on absolute levels and patterns of LEE1 transcription. Variation in the *pch* AGR therefore serves as a mechanism to diversify LEE expression patterns, and the lineage-specific pattern of *pch* AGR variation could ultimately influence ecological or virulence characteristics of subpopulations within each lineage.

The O157:H7 serotype of *Escherichia coli* is the most common enterohemorrhagic *E. coli* (EHEC) strain in the United States, and like the majority of Shiga toxin-producing *E. coli* strains associated with disease, it possesses a complex attachment phenotype known as attachment and effacement (A/E) (6) The A/E phenotype, shared by EHEC and enteropathogenic *E. coli* (EPEC) strains, is a multistaged process which begins with microcolony formation on the surface of the host cell, followed by the appearance of organized pedestal structures of polymerized actin that originate directly underneath the bacteria. Microcolony formation by EPEC is mediated by the bundle-forming pilus, but this system is absent in EHEC and other A/E populations; consequently, several different potential systems have been proposed to serve similar functions (8, 12, 24, 25, 43, 71, 72, 74).

The complex process of pedestal formation is mediated largely by a type III secretion system encoded within the locus of enterocyte effacement (LEE) pathogenicity island. Comprising nearly 50 different genes, the LEE island is present in the genomes of EPEC, EHEC, and other populations of *E. coli* as well as in the mouse pathogen *Citrobacter rodentium* (7, 35, 36, 48, 83). LEE has been acquired on at least two independent occasions in *E. coli*, where it was inserted into the tRNA genes *selC* (EPEC1 and EHEC1) in one lineage and *pheU* (EPEC2 and EHEC2) in another lineage (35, 39, 54). Within and between the lineages, the LEE island genes themselves have diverged, with unique functions evolving in some of the effectors (35, 48, 65, 78). Functional diversification has also occurred through acquisition of auxiliary effectors in the EHEC and EPEC lineages, largely through their association with highly polymorphic lambdoid prophages external to LEE (30, 73).

In addition to these functional differences between the EPEC and EHEC lineages, transcriptional patterns of LEE between them have also diverged. Though the five primary transcription units, LEE1 to LEE4 and *tir*, are shared by EHEC and EPEC (10, 38), details of their regulation differ. In both EHEC and EPEC, the Ler protein—a DNA-binding protein encoded by the first coding region in the LEE1 operon—functions as a primary positive regulator of the LEE2, LEE3, LEE4, and *tir* transcription units and as a negative regulator of its own transcription (2, 13, 38, 57, 61, 66). A complex network of pathways controls Ler expression, including positive regulation of *ler* transcription from LEE1 by integration host factor,

* Corresponding author. Mailing address: Department of Food Science and Technology, University of Nebraska, 330 Food Industry Complex, Lincoln, NE 68583-0919. Phone: (402) 472-5637. Fax: (402) 472-1693. E-mail: abenson1@unl.edu.

▽ Published ahead of print on 27 March 2009.

GrlA, BipA, and Fis and negative regulation by histone-like nucleoid structuring protein, GrlR, and GadX (2, 5, 7, 13, 14, 16, 62, 63, 67–69, 75).

One primary difference in Ler regulation between EPEC and EHEC lies in the *perABC* and *pchABC* pathways. In EPEC, the pEAF-encoded *perABC* genes also modulate LEE expression through their positive effect on the *LEE1* promoter (15, 38). PerA positively autoregulates transcription of the *perABC* operon (5, 34, 52), whose activity is controlled negatively by the pH-dependent regulator GadX, serving to link LEE expression to pH-dependent signals in EPEC (62). GadX and PerA functions are linked to LEE through PerC, which serves as the actual activator of LEE1 (52).

The *perABC* operon is absent from EHEC, but, instead, *perC*-like genes, termed *perC*-homologous (*pch*) genes, are encoded within cryptic prophages and pathogenicity islands (22, 51). The *pch* genes comprise three different classes based on sequence similarity. The *pchA*, *pchB*, and *pchC* genes form one class. These phage-encoded genes are positive regulators of LEE1, and they are nearly identical in sequence (22, 51). Expression of *pchA*, *pchB*, and *pchC* is differentially regulated by the stringent response and appears to serve as a mechanism for elevating EHEC LEE expression under different physiological conditions, including the growth-regulated expression of LEE in batch culture (42). The other two classes of *pch* genes are encoded by *pchD* (ECs1388) and *pchE* (ECs1588). The role of these genes is less clear, but studies show that they are not able to stimulate LEE1 expression in a K-12 background individually (51).

Our previous octamer-based genome scanning (OBGS) studies of the microevolution of the sorbitol-nonfermenting, β-glucuronidase-negative *E. coli* O157:H7 subpopulations showed that the subpopulations comprise two primary OBGS lineages, termed lineage I and lineage II (26, 27, 79). In the studies reported here, we now show that within these OBGS lineages, specific events have occurred within the *pchD* and *pchE* gene regions as well as within a newly defined *pchABC*-like gene, *pchX*. Genetic studies showed that the *pchA*, *pchB*, *pchC*, and *pchX* genes are carried in conserved regions of lambdoid prophage that are maintained in the populations with high fidelity. In contrast, in the *pch* *a*djacent *g*enomic *r*egions (*pch* AGR) extensive variation was detected. Quantitative analysis of LEE expression showed that variation in the *pch* AGR has substantial effects on the absolute levels and pattern of LEE 1 expression. Based on our results, we propose a model wherein the variation in *pch* AGR generates diversity in LEE expression; we further hypothesize that this may be a general mechanism through which prophage-mediated events adjust virulence gene expression at unlinked loci.

TABLE 1. USA 40 strains[a]

Strain	Lineage	Date	Origin (host)	Location
FRIK1985	II	1991	Cattle isolate	MN
FRIK1990	II	1991	Cattle isolate	NY
FRIK1991	II	1991	Cattle isolate	NY
FRIK1996	II	1991	Cattle isolate	MD
FRIK1999	II	1991	Cattle isolate	NE
FRIK2000	II	1991	Cattle isolate	FL
FRIK2001	II	1991	Cattle isolate	VT
FRIK2004	II	1991	Cattle isolate	MD
NE1487	II	1998	Cattle isolate	NE
FRIK920	II	1996	Cattle isolate	WI
FRIK944	II	1996	Cattle isolate	WI
FRIK957	II	1996	Cattle isolate	WI
FRIK964	II	1996	Cattle isolate	WI
FRIK966	II	1996	Cattle isolate	WI
FRIK1054	II	1996	Cattle isolate	WI
FRIK1123	II	1996	Cattle isolate	WI
FRIK1540	II	1996	Cattle isolate	WI
NE037	II	1997	Human isolate	NE
FDA508	II	1984	Human isolate	NH
FDA517	II	1983	Human isolate	WA
93–001	I	1993	Human isolate	WA
95–003	I	1995	Human isolate	WA
FDA504	I	1982	Human isolate	MI
FDA505	I	1982	Human isolate	MI
FDA506	I	1986	Human isolate	WA
FDA507	I	1982	Human isolate	OR
FDA518	I	1982	Human isolate	CA
FDA520	I	1982	Human isolate	OR
NE018	I	1997	Human isolate	NE
NE047	I	1997	Human isolate	NE
NE050	I	1997	Human isolate	NE
NE098	I	1998	Human isolate	NE
FRIK523	I	1994	Human isolate	WI
FRIK529	I	1994	Human isolate	WI
FRIK533	I	1994	Human isolate	WI
FRIK551	I	1994	Human isolate	WI
FRIK583	I	1994	Human isolate	WI
FRIK1275	I	1996	Cattle isolate	WI
FRIK1986	I	1991	Cattle isolate	MN
FRIK1997	I	1991	Cattle isolate	WA

[a] All strains belonged to serotype O157:H7.

MATERIALS AND METHODS

Strains and plasmids. All strains used in this study are listed in Table 1 and have been described previously (79). They are referred to collectively as the USA 40 and the Australian strain sets. Strains representing the stepwise-descent of the O157:H7 clone complex were obtained from Thomas Whittam (11). The strains were maintained as frozen stocks at −80°C, with working stocks being strictly and routinely maintained on Luria agar for no more than 7 days at 4°C. For expression studies, inocula were grown for 16 h in Luria broth and subsequently inoculated into fresh Luria broth, M9 minimal medium supplemented with glucose, or Dulbecco's modified Eagle's medium (DMEM). Antibiotics were added at concentrations of 50 μg/ml (kanamycin) and 100 μg/ml (ampicillin) where necessary.

High-density OBGS analysis. Comparison of lineage I and lineage II genomes was conducted by high-density OBGS analysis. The USA 40 strain set was used as the basis for comparison. A total of 175 different OBGS primer combinations were used on each strain as described previously (26). The individual OBGS primer sequences are listed in Table 2. OBGS reaction mixtures were electrophoresed on Li-Cor/NEN 4200 global analyzers (Li-Cor, Inc., Lincoln, NE). Images were collected and printed on an Alden 9315 thermal printer and converted to binary files as described previously (26). A composite binary file was created from Microsoft Excel files from each primer combination using the FORMATALL program (27). The composite file was then subjected to two-dimensional hierarchical clustering and self-organizing maps using Gene Cluster and visualized with the Treeview program (9). Band isolations from reactions of interest, cloning, and DNA sequence analysis were performed as previously described (27).

In the instance of polymorphic OBGS amplicons from lineage II strains, the sequence of the cloned polymorphic band was used to identify the nature of the polymorphism by comparison to the EDL933 and Sakai genome sequences (both are lineage I strains). In the instance of polymorphic OBGS amplicons from lineage I strains, PCR primers were designed at 1-kb intervals from the ends of the polymorphic OBGS fragment and used to PCR amplify DNA from strains of both lineages. These cloned fragments were then subjected to DNA sequence analysis. In some instances, insertion events in lineage II are quite large, and only the site of insertion is known.

TABLE 2. Primers used for high-density OBGS analysis

Primer type and name	Sequence
Labeled primers	
OCT2	ATGGCGCTGG
OCT3B	ATGCTGGTGG
OCT4B	ATGCTGGCGG
OCT5	ATTGCTGGCG
OCT6	ATGCGCTGGC
OCT7	ATTGGCGCTG
OCT10	ATCGCTGGTG
OCT14	ATGCTGGCGA
OCT19	ATGCTGGAAG
OCT21	ATGCGCTGGA
OCT22A	ATCTGCGCAA
OCT1	ATCGCTGGCG
OCT8	ATGCTGGCGC
OCT12	ATCTGGCGGC
OCT13	ATCTGGCGCA
OCT15	ATTGGCGGCG
OCT18	ATAACTGGCG
OCT20	ATCTGGCGCG
OCT22B	ATCTGGCGAA
OCT23	ATTGCTGGTG
Unlabeled primers	
OCT1C	GCGCCAGCGT
OCT2C	GCGCCAGCGT
OCT3C	GCCAGCGCCT
OCT4C	GCCGCCAGCT
OCT5C	GCCGCCAGCT
OCT6C	ATGCCAGCGC
OCT7	ATTGGCGCTG
OCT4B	ATGCTGGCGG
OCT5	ATTGCTGGCG
OCT3B	ATGCTGGTGG
OCT7C	ATCAGCGCCA
OCT8	ATGCTGGCAC
OCT8C	ATGCGCCAGC
OCT10	ATCGCTGGTG
OCT10C	ATCACCAGCG
OCT12	ATCTGGCGGC
OCT12C	ATGCCGCCAG
OCT13	ATCTGGCGCA
OCT13C	ATTGCGCCAG
OCT14	ATGCTGGCGA
OCT14C	ATTCGCCAGC
OCT19	ATGCTGGAAG
OCT19C	ATCTTCCAGC
OCT21	ATGCGCTGGA
OCT21C	ATTCCAGCGC
OCT22	ATCTGCGCAA
OCT22C	ATTTGCGCAG

Preparation of cosmid libraries. To delimit precisely the lineage II-specific region of genome difference (RD) in O-island 43 of lineage II (also referred to as the tellurite adhesin island, or TAI [71]), a cosmid library was constructed from FRIK2000, a representative lineage II strain. The cosmid library was prepared in the SuperCos vector (Stratagene) using DNA that was partially digested with Sau3A. Clones containing genome segments spanning the RD were identified by selecting for tellurite resistance because the RD is immediately adjacent to the tellurite resistance genes in the TAI. The cosmid clones were spread on Luria-Bertani agar plates containing 2.5 g/liter potassium tellurite (Sigma, St. Louis, MO). Ten individual tellurite-resistant clones were chosen, and cosmid DNA was prepared for sequence analysis using a Qiagen minprep kit according to the manufacturer's instructions. Restriction analysis and DNA sequencing were further used to identify clones spanning the RD. The junction of the RD was determined by chromosome walking. Primers used in the chromosomal walking were the following: Zp1, 5′-AGGTAATACACCGTAAGAGC-3′; Zp2, 5′-GCTTGGTGCTTACCCGTACACC-3′; and Zp3, 5′-GAGCATCAATGGT CTTTACA-3′. The Zp2 and Zp3 primers were used to confirm the junction.

Nested PCR. Multiple strains were tested for the 10.75-kb deletion in the TAI using a nested PCR assay. The primers were designed internal to the deletion, within the ECs1390 gene (Intfor, 5′-AAACGGGGGGTAGATGAAAG-3′) and in the ECs1377 (Delfor, 5′-TGAATGTGGATTACGTCACCAG3-′) and ECs1394 (Delrev, 5′-GCCAGTTCCACCATACGAAG-3′) genes which flank the deletion. In strains carrying the deletion, PCR from the ECs1377-Delfor and ECs1394-Delrev primers leads to a 653-bp product spanning the junction of the deletion (the >10-kb amplicon in strains with an intact region is not amplified efficiently in the reaction). Strains that carry an intact region produce a 1,021-bp amplicon from the ECs1390-Intfor and ECs1394-Delrev primer pair. PCRs were performed in a total volume of 20 μl containing 20 pmol of each primer, a 100 μM concentration of the deoxynucleoside triphosphates, 1.5 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 1 unit of *Taq* DNA polymerase (Sigma, St. Louis, MO). The PCR cycling protocol consisted of 95°C for 5 min and 29 cycles of 95°C for 30 s, 61°C for 1 min, and 72°C for 1 min, followed by 72°C for 5 min.

Shotgun DNA microarray analysis. Shotgun DNA microarrays were fabricated from libraries of sheared DNA of strains 93-001 (lineage I representative) and FRIK2000 (lineage II representative). DNA was sheared by nebulization, and fragments of ca 1.0 kb were gel purified, ligated to pCR4Blunt-Topo (Invitrogen), and transformed into *E. coli* Top10 cells. A total of 8,000 clones were placed from each library in individual 96-well plates, and cloned inserts from each transformant were PCR amplified using T3 and T7 primers. DNA microarrays were fabricated from the PCR products as described previously (80). Ten strains from each lineage of the USA 40 strain set were then probed on each of the DNA microarrays. Data analysis and sorting of polymorphisms were performed as described previously (80). Clones corresponding to features detecting conserved lineage-specific polymorphisms were sequenced and contigs were determined using Sequencher (Gene Codes Corp.). Contigs of interest were confirmed by Southern blotting using a single gene or cloned segment as a probe from each contig.

Anchored PCR of the *pchABC* cassette. To test for variation in the *pchABC* cassette, anchored PCR was used with primers anchored in highly conserved regions of the S holin gene (Rstxb,5′-ATGGAAAAAATCACAACAGGT-3′, or Rstxb1, 5′-CTGGGGAGTCTGCTGTTT-3′) as described by Unkmeier and Schmidt (76) and the C terminus of the *pchABC* genes (*pch* ProbeR, 5′-TCCT GTCCCTTTATATCGTCG-3′). The PCRs were carried out in 20-μl volumes containing 2 pmol of the *pch* ProbeR primer and the Rstxb or Rstxb1 primer with a 2.5 μM concentration of the deoxynucleoside triphosphates, 12.5 mM $MgCl_2$, and 0.5 units of *Taq* polymerase (Takara). After an initial denaturation for 2.5 min at 95°C, the reaction mixtures were cycled 29 times at 95°C for 30 s, 62.1°C for 45 s, and 72°C for 30 s, followed by a 5-min extension at 72°C. The PCR products were resolved by electrophoresis on a 1% agarose gel.

PFGE and Southern blotting. Polymorphisms associated with *pchABCDE* were detected by Southern blotting of PmeI restriction fragments resolved by pulsed-field gel electrophoresis (PFGE). The gels were run using an initial switch time of 2.2 s and a final switch time of 54.2 s. After being stained, the gels were soaked in 0.2N HCl for 10 min, in the denaturation solution (1.5 M NaCl, 0.5 M NaOH) for 45 min, in neutralization solution (1 M Tris, 1.5 M NaCl, pH 7.4) for 30 min, and in fresh neutralization solution for another 15 min and were then blotted by capillary action to nylon membranes in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). The membranes were probed with internal segments of *pchC* (+24 to + 253), *pchD* (+13 to + 231), and *pchE* (+94 to + 254). Size markers consisted of lambda concatemers (Bio-Rad) which were run on three different lanes of each gel.

Alleles for *pchABCDE*-hybridizing fragments were assigned numerically from smallest to largest relative to the lambda concatemer markers. Fragments of the same length from different strains were given the same allele designation. Using binary representations of each allele occurrence, cluster analysis was performed by unweighted pair-group method using average linkages analysis using PAUP 4.0 β, version 10 (D. Swofford, Sinauer and Associates, Sunderland, MA).

Measurement of LEE expression. Transcription from the *LEE1* promoter was monitored using a reporter gene fusion constructed by cloning a PCR amplicon extending from −690 to +203 of the *LEE1* promoter upstream of the promoterless *luxCDABE* cassette in plasmid pCS26-PAC (4). The PCR primers were tailed with XhoI and BamHI sites to facilitate directional cloning into the pCS26-PAC parental plasmid. The fusion was verified by DNA sequencing.

The LEE1-*luxCDABE* fusion was introduced into O157:H7 strains by electroporation. Transformants were selected on LB medium supplemented with 50 μg/ml kanamycin and were confirmed by plasmid miniprep and agarose gel electrophoresis. Three individual confirmed transformants were chosen from each strain and tested individually. These transformants were tested by growth in

black, clear-bottom 96-well microcultures that were overlaid with mineral oil to prevent drying and were incubated at 37°C and monitored in a Perkin-Elmer Victor 3 multiwell luminometer. Luminescence and optical density measurements at 600 nm (OD_{600}) were obtained at 10-min intervals throughout the experiment. Comparing identical strains grown in the same medium in shake flasks versus microcultures, we observed that the absolute levels of LEE expression differed two- to threefold between the two growth conditions, with higher expression typically observed in the shake flasks. However, the shape of the expression curves and the relative timing of expression with respect to the growth curve were nearly superimposable between the shake flask cultures and the microcultures (data not shown). Overnight cultures used to inoculate the microcultures were grown aerobically in 13-mm culture tubes with shaking for 16 h at 37°C in 3 ml of LB medium supplemented with 50 μg/ml kanamycin. The overnight cultures were then diluted 100-fold into fresh DMEM (supplemented with 50 μg/ml kanamycin) in the 96-well plates with 150 μl of culture covered with 100 μl of mineral oil. Experiments for the three individual transformants from each strain were repeated at least two independent times in a single plate, providing two technical replicates of the three biological replicates, for a total of at least six individual experiments. Luminescence and OD measurements were recorded into Microsoft Excel spreadsheets, and preliminary data analysis was performed in Excel.

Data analysis. Luminescence values, OD values, and background values for luminescence and OD were collected at each time point. Luminescence counts and absorbance were calculated by subtracting background values (obtained from uninoculated wells) from raw values. Relative expression values for the fusions were normalized for cell density by dividing the background-subtracted luminescence counts/absorbance. These values are referred to as the expression values. Prior to statistical analysis, the normalized expression values were log transformed (log base 2).

Statistical analysis was performed by first segmenting the expression curves between 100 to 400 min into six equal periods. The expression values for each occurrence of each allele were then tabulated for each of the expression periods. Multivariate analysis of variance was then performed on the data sets for each time period. Tests for allele-specific effects of the *pch* AGR on expression level were then conducted by two-tailed *t* tests for each of the different time periods. All statistical analyses and data set creations were done in the SAS System for Windows, version 9.00, or Microsoft Excel using the StatistiXL package.

RESULTS

Comparative genome analyses of lineage I and lineage II strains. Our strategy for comparative genome analyses assumes that ancestral events are shared by multiple strains within the lineage while more recent events are shared by only a small number of strains. To this end, a set of 40 strains representing some of the genetic diversity of the two O157:H7 lineages, as well as temporal and spatial diversity in origin (Table 1, USA 40 set), were compared.

Using high-density OBGS analysis, each strain was screened with 175 different OBGS primer combinations, leading to identification of 4,376 polymorphic OBGS products. Of these, 967 were observed specifically within strains from one lineage but not the other. A total of 70 OBGS polymorphisms were unique to and conserved among all strains of a single lineage, whereas the remaining 897 were dispersed among two or more strains in a lineage. The 70 lineage-specific conserved OBGS amplicons and 30 additional lineage-specific OBGS amplicons that were conserved among at least one-half of the strains within a lineage were subsequently cloned and sequenced. Contig analysis yielded 52 unique loci (referred to as OBGS RDs [RD_{OBGS}s]) from the 100 cloned segments. The different RD_{OBGS}s, the genes affected, and the associated polymorphisms are listed in Table 3.

To supplement the high-density OBGS data, comparative genome hybridization experiments were also conducted using DNA microarrays fabricated from amplified shotgun libraries of lineage I and lineage II strains. As shown in Table 3, nine different RDs (comparative genome hybridization RDs, or RD_{CGH}s) were conserved among the lineage I strains (and absent in lineage II strains), whereas only strain-specific alterations were observed when the lineage II array was used. This is in keeping with our contention that the lineage II genome is largely a derived state from a lineage I-like ancestor (27, 79).

Lineage-specific alterations exist in all three classes of *pch* genes. Both the high-density OBGS studies and the microarray-based comparative genome hybridizations identified lineage-specific polymorphisms within loci encoding the three major classes of *pch* genes; RD_{CGH} 4 maps to the region containing *pchD*, RD_{OBGS} 18 maps to *pchE*, and RD_{OBGS} 48 maps to one of the nearly identical *pchA*, *pchB*, or *pchC* genes. The *pchABC*, *pchD*, and *pchE* genes encode three classes of proteins with similarity to PerC, the primary regulator of the LEE island in EPEC. They are dispersed among prophage (*pchA*, *pchB*, *pchC*, and *pchE*) and pathogenicity islands (*pchD*) in the O157:H7 genome. *pchA*, *pchB*, and *pchC* comprise one class, and even though they are encoded within three separate prophages, the 104-amino-acid peptides they specify differ from one another by only one or two residues (residues 6 and 58). They are known to positively regulate LEE through their effect on the LEE1 operon, which encodes the primary regulator of LEE gene expression, Ler (21, 22, 51). Though sharing similarity, the *pchD* and *pchE* genes do not appear to control LEE expression, and their functions are unknown (22).

The RD_{CGH} 4 deletion spans the coding regions of ECs1380 to ECs1391 within the TAI (71), including ECs1388 (*pchD*). ECs1388 is not annotated in the EDL933 genome but is positioned in the duplicated TAIs between the Z1200 and Z1201 genes and between the Z1640 to Z1641 genes. Sequence analysis of a cosmid clone spanning the region of ECs1369 to ECs1395 from the lineage II strain FRIK2000 showed that the deletion spans 10,795 bases, traversing the ECs1378 to ECs1393 coding regions (Fig. 1A). We refer to this specific event as the TAIΔ10795 deletion; nested PCR analysis showed that only lineage II strains from the United States carried the TAIΔ10795 deletion (Fig. 1B and Table 4).

The fragment identified by RD_{OBGS} 18 includes the C-terminal portion of the *pchE* gene (ECs1588). Encoded within a cryptic prophage, *pchE* lies adjacent to the terminase gene in the lysis region of the prophage genome (22). DNA sequence analysis of PCR products resulting from primers that flank the *pchE* gene (gapF and gapR primers) showed that lineage II strains carry a 352-bp deletion starting from +216 of *pchE* (Fig. 1C and Table 4). We have designated this allele *pchEΔ352*.

The third class of *pch* polymorphisms occurred in the *pchABC*-like genes, marked by the RD_{OBGS} 18 (Table 3). This RD, detected by high-density OBGS analysis, yielded an OBGS fragment with sequence similarity to the *pchA*, *pchB*, and *pchC* regulatory genes and was detected in 16 of the 20 lineage II strains. DNA sequence of the cloned OBGS segment deviated from the prophage-borne *pchABC* sequences immediately upstream of the *pch* coding regions. To determine if this apparently unique *pch* gene is similarly positioned in the lysis region of a lambdoid prophage genome, an anchored PCR strategy similar to that of Unkmeier and Schmidt (76) was used with the one primer positioned in the C terminus of the *pch*-

TABLE 3. RDs identified by comparative OBGS analysis and comparative genome hybridization studies of the two *E. coli* O157:H7 lineages[a]

RD	BLAST result and gene function and/or description	Alteration[b]
RD_{OBGS} 1	VT2-Sa Ssb (Ea10); hypothetical protein	SNP in OBGS priming sites (OCT1C)
RD_{OBGS} 2	Hypothetical protein from CP-933C	Large deletion (at least 700 bp) in lineage II
RD_{OBGS} 3	Z1841 to Z1842 from CP-933	ND
RD_{OBGS} 4	Hypothetical protein; holin from *Shigella flexneri* phage V	Large insertion and mismatches in lineage II
RD_{OBGS} 5	NinE; putative methylase; antirepressor; phage VT2-Sa	SNP in OBGS priming sites (OCT21); large deletion and mismatches in lineage II
RD_{OBGS} 6	Hypothetical protein from phage 933W	ND
RD_{OBGS} 7	Nucleotides 301374 to 301721 from CP933H	Large deletion (at least 400 bp) in lineage II
RD_{OBGS} 8	Portal protein from *Shigella flexneri* phage V	Large insertion (at least 400 bp) in lineage II
RD_{OBGS} 9	Hypothetical protein from CP-933C	ND
RD_{OBGS} 10	Antitermination protein; hypothetical protein from phage VT2-Sa	ND
RD_{OBGS} 11	Putative minor tail protein	ND
RD_{OBGS} 12	Putative superinfection exclusion protein; Ssb	ND
RD_{OBGS} 13	Putative integrase from phage TPW22	ND
RD_{OBGS} 14	Putative tail-length tape measure protein from phage HK97	ND
RD_{OBGS} 15	Putative antirepressor; putative endolysin; hypothetical protein	SNP in OBGS priming sites (OCT21)
RD_{OBGS} 16	MokW from 933W	Large deletion (at least 1 kb) in lineage II
RD_{OBGS} 17	Hypothetical protein; Shiga toxin subunit B precursor	Large deletion (at least 1 kb) in lineage II
RD_{OBGS} 18	**ECS1588 PerC-like protein**	**Deletion in lineage II**
RD_{OBGS} 19	Z5935	9-bp insertion in lineage II
RD_{OBGS} 20	Putative fimbrial-like protein	Insertion and mismatches in lineage II
RD_{OBGS} 21	Putative adhesin and penetration protein	SNP in OBGS priming sites (OCT23)
RD_{OBGS} 22	FolD	8-bp duplication in lineage II
RD_{OBGS} 23	Putative phosphotransferase system enzyme I; dihydroxyacetone kinase	Large deletion (at least 1 kb) in lineage II
RD_{OBGS} 24	Hypothetical protein; proline permease II	SNP in OBGS priming sites (OCT19)
RD_{OBGS} 25	Putative ATP-binding protein of ABC transporter	Large deletion (at least 1 kb) in lineage II
RD_{OBGS} 26	D-Ribose periplasmic binding protein; ribokinase	9-bp deletion in lineage II
RD_{OBGS} 27	Regulator of acetyl-coenzyme A synthetase; repressor of *aceBA* operon	18-bp duplication in lineage II
RD_{OBGS} 28	Putative DNA replication protein from *Shigella flexneri*	50 mismatches
RD_{OBGS} 29	IS629/RepE replication protein from pO159	ND
RD_{OBGS} 30	YjgB/LeuX	Deletions and insertions in lineage II
RD_{OBGS} 31	Repeat sequences (158 bp)	ND
RD_{OBGS} 32	Repeat sequences (22 bp)	ND
RD_{OBGS} 33	Repeat sequences (100 bp)	ND
RD_{OBGS} 34	Repeat sequences (13 to 20 bp)	ND
RD_{OBGS} 35	Putative atpase	SNP in OBGS priming sites (OCT1C)
RD_{OBGS} 36	Hypothetical protein; putative kinase	SNP in OBGS priming sites (OCT10)
RD_{OBGS} 37	Putative protein from *E. coli* K-12	SNPs in OBGS priming sites (OCT14)
RD_{OBGS} 38	Hypothetical protein	10-bp insertion in lineage II
RD_{OBGS} 39	Hypothetical protein from *E. coli* K-12	SNP in OBGS priming sites (OCT5C)
RD_{OBGS} 40	Hypothetical protein	SNP in OBGS priming site (OCT6C)
RD_{OBGS} 41	Hypothetical protein from *E. coli* K-12	9-bp insertion in lineage II
RD_{OBGS} 42	Hypothetical protein	Large deletion (at least 1 kb) in lineage II
RD_{OBGS} 43	Hypothetical protein	SNP in OBGS priming sites (OCT12C)
RD_{OBGS} 44	Hypothetical protein from *E. coli* K-12	SNP in OBGS priming sites (OCT6C)
RD_{OBGS} 45	YhcG	77-bp insertion in lineage II
RD_{OBGS} 46	YcbC	SNPs in OBGS priming sites (OCT10, 6C)
RD_{OBGS} 47	No hit	Large insertion (at least 500 bp) in lineage II
RD_{OBGS} 48	**ECs1091, ECs2182, and ECs2737**; ***pch*ABC-like**	**ND**
RD_{OBGS} 49	No hit	Large insertion (at least 500 bp) in lineage II
RD_{OBGS} 50	No hit	Large insertion (at least 350 bp) in lineage II
RD_{OBGS} 51	No hit	Large insertion (at least 500 bp) in lineage II
RD_{OBGS} 52	No hit	Large insertion (at least 300 bp) in lineage II
RD_{OBGS} 53	No hit	Large insertion (at least 150 bp) in lineage II
RD_{CGH} 1	Z0165 (*hemL*)	Deletion in lineage II
RD_{CGH} 2	Z0268 to Z2069 (open reading frames in *Rhs* element)	Deletion in lineage II
RD_{CGH} 3	Z0318 (*pinH* from CP-933H); same as RD_{OBGS} 7	Deletion in lineage II
RD_{CGH} 4	**Z196 to Z1204 (*perC*-like region of TAI)**	**Deletion in lineage II**
RD_{CGH} 5	Z1961 (*prrA*) to Z1962 (*modD*)	Deletion in lineage II
RD_{CGH} 6	Z2185 (hypothetical protein) to Z22186 (hypothetical protein)	Deletion in lineage II
RD_{CGH} 7	Z2207 (hypothetical oxidoreductase)	Deletion in lineage II
RD_{CGH} 8	Z2706 to Z2707 (*ynhG-ynhA*)	Deletion in lineage II
RD_{CGH} 9	ECs4889 to ECs4903 (phage tail region)	Deletion in lineage II

[a] *Pch*-containing RDs are in bold.
[b] ND, not determined.

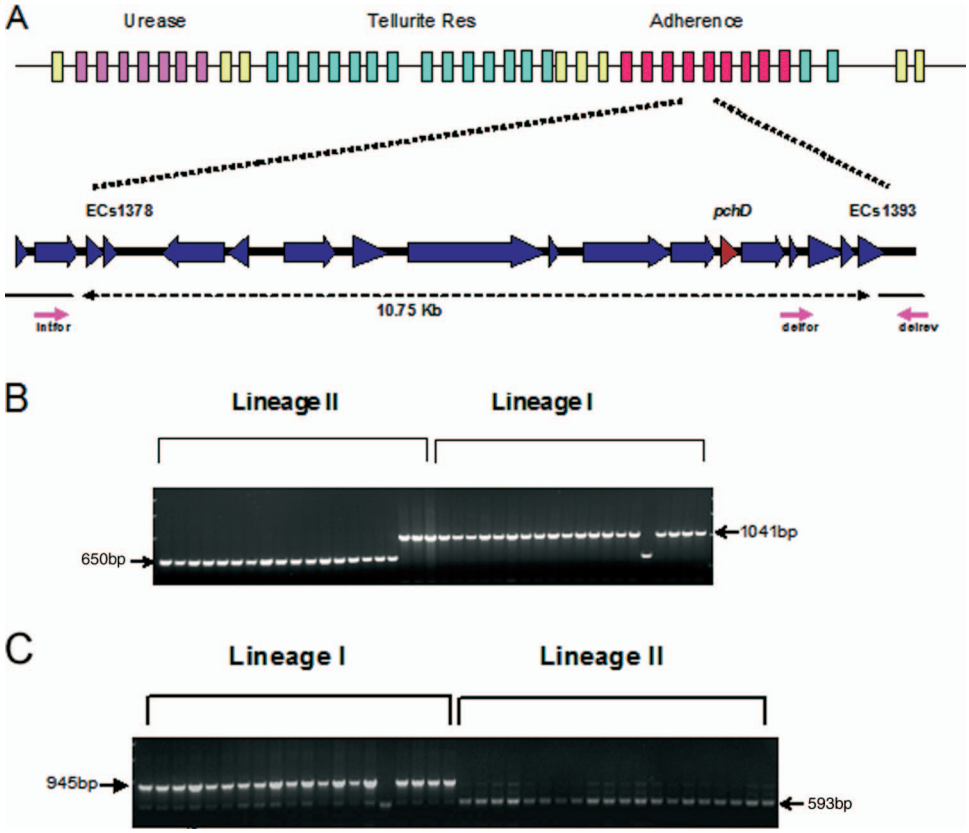

FIG. 1. Illustration of the lineage-specific TAI deletion and nested PCR detection of the deletion. (A) The 10-kb deletion is located in the tellurite resistance and adherence conferring island (TAI), spanning ECs1378 to ECs1394 including the *pchD* encoding gene ECs1388. Nested PCR primers are located in ECs1377, ECs1390, and ECs1394, respectively. (B) Genomic DNA from lineage I and lineage II strains of USA 40 were PCR amplified with the nested PCR primers. The 650-bp amplicons, corresponding to amplification across the junction of the 10,795-base deletion, and the 1,041-bp amplicons, corresponding to the intact region, are indicated by arrows. (C) PCR detection of the lineage-specific deletion in *pchE* in the USA 40 strain set. Genomic DNA from lineage I and lineage II strains of USA 40 were PCR amplified with the gapF and gapR primer set. The 593-bp amplicons, corresponding to the deletion, and the 945-bp amplicons, corresponding to the intact region, are indicated by arrows.

ABC genes and the other primer anchored in a highly conserved region of the S holin gene. Nearly all of the strains tested produced the 3.2-, 2.9-, and 2.7-kb amplicons expected from the *pchB*, *pchA*, and *pchC* prophages, respectively, of the Sakai genome sequence (Fig. 2A and B and 3). Among the lineage II strains, most also produced an additional 1.9-kb amplicon. DNA sequence analysis of the 1.9-kb amplicon showed that it originates from a unique *pchABC*-like prophage that carries an intact *pch* gene having significantly shortened intergenic regions between the *S*, *R*, and *Rz* genes (Fig. 2B), and its sequence matched to the cloned lineage II-specific OBGS band. We refer to this *pch* cassette as *pchX* herein. Only one other *pch*-holin S amplicon was detected, a 3.4-kb amplicon in a single strain (FRIK920) which also lacks the 3.2-kb *pchB* amplicon. We refer to this amplicon as *pchY* and did not characterize it further.

To determine to what extent the *pchABCX*-holin S cassettes are conserved between strains, we next tested for restriction fragment length polymorphisms within the individual cassettes using BstXI restriction digestion of the individual *pchABCX*-holin S PCR products from each strain. This yielded only the collections of expected fragments from the *pchA*-, *pchB*-, *pchC*-, and *pchX*-holin S sequences (data not shown), showing that these cassettes are highly conserved between strains. In contrast, Southern hybridizations of PFGE-separated genomic PmeI restriction fragments revealed a striking degree of diversity when strains were probed with the *pchC* gene (Fig. 2D). Based on the Sakai genome sequence, hybridizing fragments of 208 kb (*pchC* [Sp14]), 205 kb (*pchB* [Sp11]), and 126 kb (*pchA* [Sp4]) were expected, of which the 205-kb and 208-kb fragments comigrate under these PFGE conditions (Fig. 2E). While the pattern from Sakai was consistent with these sizes, other strains yielded an array of fragments ranging from a single band in some strains to four different bands in many strains. The PmeI fragments from the Sakai genome which carry the *pchA* (Sp4), *pchB* (Sp11), and *pchC* (Sp14) genes (Fig. 2E) also carry all or portions of adjacent prophages Sp5 and Sp10, while the *pchC* prophage is flanked by Sp13 and Sp15, implying that the extensive polymorphism is due to insertion, excision, recombination, and/or rearrangement events associated within these adjacent prophages. This finding is consistent with previous microarray studies of 31 different O157:H7 strains showing substantial variation in these same prophage regions (Sp5, Sp10, Sp13, and Sp15) (82). Collectively, this points to a model of highly fluid and rapidly evolving regions that flank highly conserved segments carrying the

TABLE 4. Size of nested PCR amplicons for USA 40 and Australian strain sets

Strain set and name	Lineage	Size of nested PCR product (bp)[a]	Size of gapF/gapR PCR product (bp)[b]	Strain set and name	Lineage	Size of nested PCR product (bp)[a]	Size of gapF/gapR PCR product (bp)[b]
USA 40 strains				FDA517	II	1,041	**593**
93-001	I	1,041	945				
95-003	I	1,041	945	Australian strains			
FDA504	I	1,041	945	AU329	I	1041	945
FDA505	I	1,041	945	AU580a	I	1,041	945
FDA506	I	1,041	945	AU581	I	1,041	945
FDA507	I	1,041	945	AU1812	I	1,041	945
FDA518	I	1,041	945	AU1816	I	1,041	945
FDA520	I	1,041	945	AU1821	I	1,041	945
NE018	I	1,041	945	AU6	II	1,041	**593**
NE047	I	1,041	945	AU14	II	1,041	**593**
NE050	I	1,041	945	AU119	II	1,041	**593**
NE098	I	1,041	945	AU134	II	1,041	**593**
FRIK523	I	1,041	945	AU183	II	1,041	945
FRIK529	I	1,041	945	AU200	II	1,041	**593**
FRIK533	I	1,041	945	AU514	II	1,041	**593**
FRIK551	I	**650**	**593**	AU516	II	1,041	**593**
FRIK583	I	1,041	945	AU525	II	1,041	945
FRIK1275	I	1,041	945	AU539	II	1,041	945
FRIK1986	I	1,041	945	AU571	II	1,041	945
FRIK1997	I	1,041	945	AU623	II	1,041	**593**
FRIK1985	II	**650**	**593**	AU726	II	1,041	**593**
FRIK1990	II	**650**	**593**	AU735	II	1,041	**593**
FRIK1991	II	**650**	**593**	AU739	II	1,041	**593**
FRIK1996	II	**650**	**593**	AU1180	II	1,041	**593**
FRIK1999	II	**650**	**593**	AU1668	II	1,041	945
FRIK2000	II	**650**	**593**	AU1809	II	1,041	**593**
FRIK2001	II	**650**	**593**	AU1810	II	1,041	NP[c]
FRIK2004	II	**650**	**593**	AU1811	II	1,041	**593**
NE1487	II	**650**	**593**	AU1814	II	1,041	**593**
FRIK920	II	**650**	**593**	AU1815	II	1,041	**593**
FRIK944	II	**650**	**593**	AU1817	II	1,041	**593**
FRIK957	II	**650**	**593**	AU1818	II	1,041	**593**
FRIK964	II	**650**	**593**	AU1819	II	1,041	**593**
FRIK966	II	**650**	**593**	AU1820	II	1,041	**593**
FRIK1054	II	**650**	**593**	AU1822	II	1,041	**593**
FRIK1123	II	**650**	**593**	AU1823	II	1,041	945
FRIK1540	II	**650**	**593**	AU151b	II	1,041	945
NE037	II	1,041	**593**	AU195c	II	1,041	945
FDA508	II	1,041	**593**	AU197b	II	1,041	945

[a] Amplicon size of 650 bp (in boldface) corresponds to a deletion of *pchD* in the TAI.
[b] Amplicon size of 593 bp (in boldface) indicates a deletion of *pchE*.
[c] NP, no product.

pchA-, *pchB*-, and *pchC*-, and *pchX*-holin S cassettes. Interestingly, at least two of these fluid regions carry the *stx* prophages (Sp5 and Sp15) that encode major virulence factors in the O157:H7 and other EHEC lineages.

Evolution of the *pch*-holin S cassettes and genomic instability in the O157:H7 genome. Strains representing the stepwise evolutionary model of the EHEC1 O157:H7 lineage (77) were next tested to determine if the *pch*-holin cassettes and the abutting genomic instability were traits that are conserved in the lineage. The *pch*-holin cassettes were determined with the anchored *pch*-holin S PCR strategy, and genomic instability was tested by Southern hybridization with the *pchC* probe. The results showed that the *pchX*-holin S cassette was present in nearly all of the strains tested including all three O55:H7 strains, the descendants of the O157:H-German subclone, and the glucuronidase-positive O157:H7 subclone (Fig. 2C). On the other hand, the *pchA*-, *pchB*-, and *pchC*-holin S cassettes were detected in all of the O157:H7/H$^-$ strains but were absent in the O55:H7 strains. Collectively, these results imply that the *pchX*-holin S cassette was the first to evolve in the lineage, followed by acquisition of the *pchA*-, *pchB*-, and *pchC*-holin S cassettes some time after the serotype switch. These findings are consistent with the microarray-based studies of Wick et al., which detected genes from the Sp4, Sp11, and Sp14 prophages only in strains downstream of the O55:H7-like ancestor (77). Although most markers we have studied are ancestral in OBGS lineage I and derived in lineage II, the *pchX*-holin S cassette is the opposite; it apparently was lost in lineage I strains but persisted in lineage II strains descending within the United States.

The results from Southern hybridizations on these strains using the *pchC* probe showed that the three O55:H7 strains produced two hybridizing bands of similar sizes between strains (Fig. 2D). In contrast, much more diversity was detected in the banding patterns of the O157:H7/H$^-$ strains (Fig. 2D), particularly among strains such as 493/89 and CB2755

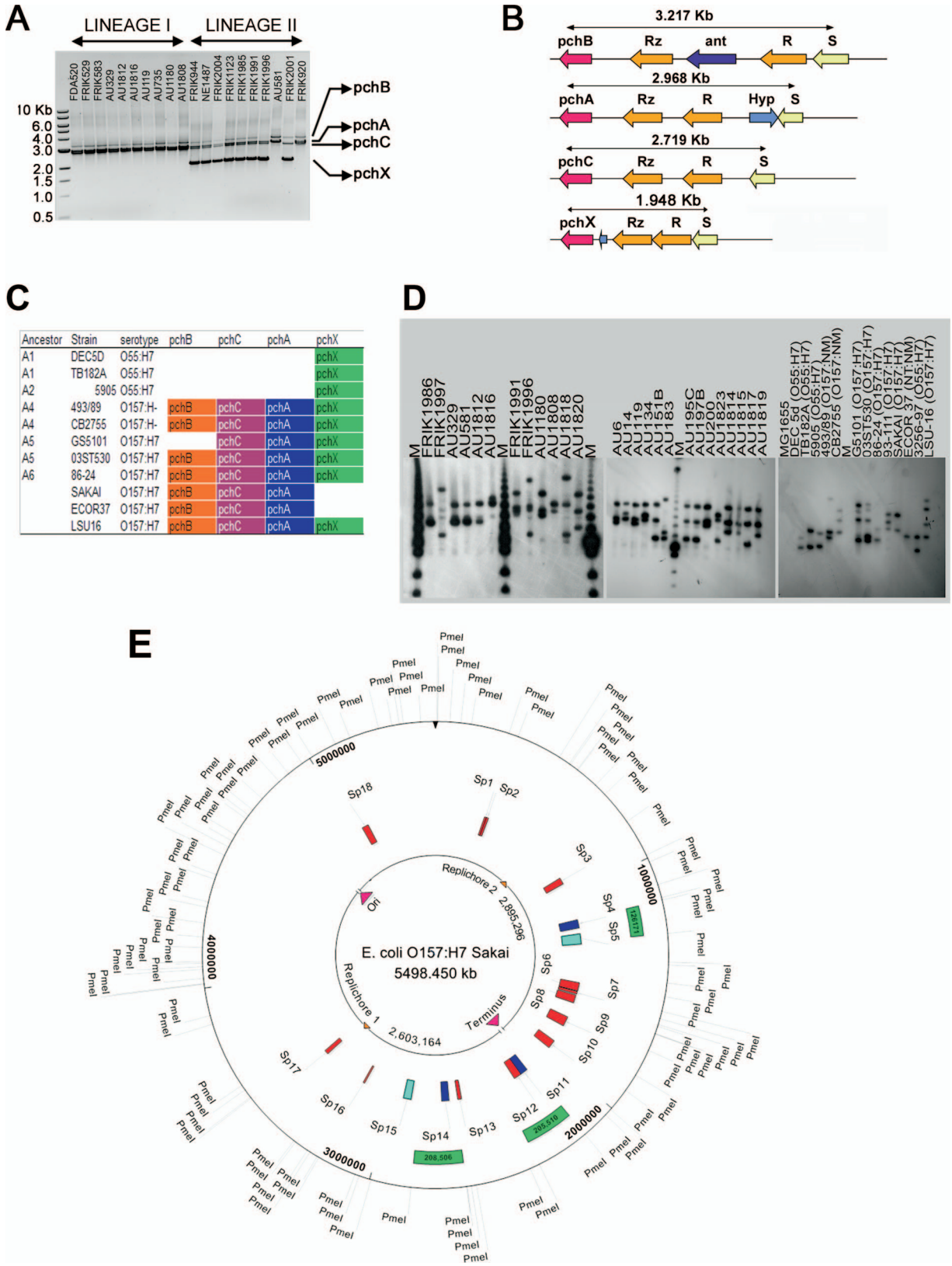

FIG. 2. Mapping *pch* AGR-associated genome alterations in *E. coli* O157:H7 strains. (A) PCR products resulting from the anchored PCR on selected strains using the *stxB* and *pch* ProbeR primers. A 1-kb ladder is shown to the left, and the positions of the bands expected for *pchA*, *pchB*, *pchC*, and the unique *pchX* are indicated at right. (B) Maps illustrate the DNA sequences of the different *pch*-holin S cassettes. The lengths of the PCR products are indicated above the relevant map. (C) Distribution of *pch*-holin cassettes detected in strains representing the stepwise descent

(descending from the A4 ancestor) and 93-111, 86-24, and Sakai (descending from the A6 ancestor) (Fig. 2C). Although more strains need to be examined, these results point toward the conclusion that acquisition of *pchA*, *pchB*, and *pchC* prophages was associated with the emergence of polylysogeny and genomic instability in regions adjacent to these *pch* genes in the subsequent A4 to A6 ancestors.

Genomic instability in the *pch* AGR is congruent with the evolutionary pattern of the genomic backbone. Having observed association of genomic instability with the progressive buildup of *pchX*, *pchA*, *pchB*, and *pchC* in the genome, we next tested whether genomic events in the *pch* AGR show congruence with the microevolutionary patterns. This was accomplished by first mapping the *pch*-holin cassette data and the *pchC* hybridization patterns onto phylogenies of the USA 40 and the Australian strain sets inferred from OBGS (26, 27) and the distribution of lineage-specific polymorphism assay-6 (LSPA-6) alleles (79) (Fig. 3). To simplify this analysis, each of the different *pch* AGR (each unique size of *pchC*-hybridizing genomic segment) was assigned an "allele" number. As expected, the TAIΔ10795 and *pchE*Δ352 deletions were highly biased and were largely limited to lineage II strains. Both the TAIΔ10795 and the *pchX*-holin S cassette were also limited primarily to lineage II strains from the United States. Likewise, the *pch* AGR alleles were remarkably congruent with the inferred phylogeny, including the distribution of individual alleles and allele combinations (haplotypes). *pch* AGR haplotypes were also confined to a given LSPA-6 genotype, with the only exception being the *pch* AGR10.1 *pch* AGR12 *pch* AGR15.2 $pchD^+$ *pchE*Δ352 haplotype that was present in strains AU134 and AU623 belonging to the LSPA-6 genotype 221213 and strains AU1822, AU14, and AU516 belonging to the LSPA-6 genotype 212111. To test the statistical significance of the bias in the distribution of *pch* AGR alleles, a nonparametric Mann-Whitney test was used. This yielded a U statistic of 593.0 with 28 degrees of freedom and was significant at a P value of 0.001, indicating that the distributions are indeed highly biased.

In addition to the lineage bias, the alleles and haplotypes of *pch* AGR also displayed substantial phylogeographic variation (Fig. 3). Certain alleles such as the *pch* AGR2.1, *pch* AGR10.1, and *pch* AGR15.2 were observed only in the Australian strains, as was the *pch* AGR2.1 *pch* AGR10.1 *pch* AGR15.2 haplotype. Likewise, the TAIΔ10795 deletion and the *pch* AGR4 *pch* AGR11 haplotype were observed only in North American strains. Only one haplotype (*pch* AGR4 *pch* AGR12 $pchD^+$ *pchE*+) was found in strains from both continents (Fig. 3) in the lineage I, LSPA-6 genotype 111111 background.

LEE1 expression patterns in the USA 40 in the strain set. Significant strain-strain variation in LEE expression and attachment characteristics has been described by several authors (32, 55, 56, 81); however, such variation has not been systematically associated with any one particular locus. Our identification of the *pchX* cassette and the extensive diversity in *pch* AGR prompted us to test whether variation in the *pchABCX* genotypes or the *pch* AGR has a significant effect on LEE1 expression characteristics. To test for allelic affects, a quantitative genetics approach was used to measure allelic association with the patterns of transcription from the *LEE1* promoter.

To quantify LEE1 expression, a LEE1::*luxCDABE* transcription fusion in the low-copy-number pCS26-PAC vector was introduced into each of the USA 40 strains. The *lux* enzyme system has a half-life of ~15 min in *E. coli* O157:H7 and therefore provides a convenient and noninvasive means for monitoring the expression pattern during growth (64). The LEE1::*luxCDABE* fusion was introduced into each of the *E. coli* O157:H7 strains of the USA 40 set by electroporation, and three individual confirmed transformants were used as biological replicates for measurements of each parental strain. Expression patterns were then determined with a multiwell luminometer using microcultures grown in 96-well plates. Each of the three biological replicates from the 40 different parental strains of the USA 40 set was tested at least twice, and the data were pooled for analysis.

The averaged growth in DMEM and M9 medium and expression data from the entire USA 40 strain set are shown in Fig. 4A and B. Under the microculture growth conditions, the strains grew very uniformly. In M9 medium exponential growth (Fig. 4A) essentially ceases after 110 to 150 min, whereas in DMEM (Fig. 4B) growth continues longer but with a noticeable and highly reproducible change in rate occurring at about 150 min. In M9 medium, LEE1 expression drops during exponential growth, followed by a slight increase at ~130 min as the exponential phase ceases. This is followed by a nearly constant level of expression during stationary phase (out to 400 min). In contrast, expression of LEE1 in DMEM falls early but rises during growth until stationary phase, at about 220 min, at which time growth rate and expression level off (Fig. 4A). An interesting feature of LEE1 expression from the DMEM-grown cells is a highly reproducible trough that occurs during the prestationary period at 70 to 120 min. This expression trough occurred in all but a single strain and corresponds to the period at which the growth rate shifts. The cause of this shift is not clear; glucose analyses showed that glucose concentration was not limiting, as the initial concentration (50 mM)

of the *E. coli* O157:H7 lineage. The ancestor numbering and strain information is from Wick et al. (77), and a colored square indicates presence of the *pch*-holin S PCR product of expected size from the *pchB*-holin S, *pchA*-holin S, *pchC*-holin S, and *pchX*-holin S regions. (D) Autoradiographs of PFGE Southern blots from strains digested with PmeI and probed with ^{32}P-labeled amplicons from internal segments of *pchC*. The strain names are indicated above the relevant lane, with lanes labeled M containing lambda concatemers. The three images in the panel were derived from different PFGE gels and are not necessarily aligned. (E) Circular map of the *E. coli* O157:H7 Sakai genome. The positions of PmeI restriction sites are shown around the circumference. On the innermost ring, the relative positions of the origin and terminus of replication are indicated along with the lengths of the two replichores. Moving outward, the next ring shows the relative positions of the different prophage (Sp1 to Sp18). The *pchA* (Sp4), *pchB* (Sp11), and *pchC* (Sp14) prophage are shown in dark blue, the stx_2 (Sp5) and stx_1 (Sp15) prophage are shown in light blue, and the remainder are shown in red. The next ring shows the relative positions of *pchC*-hybridizing segments (green).

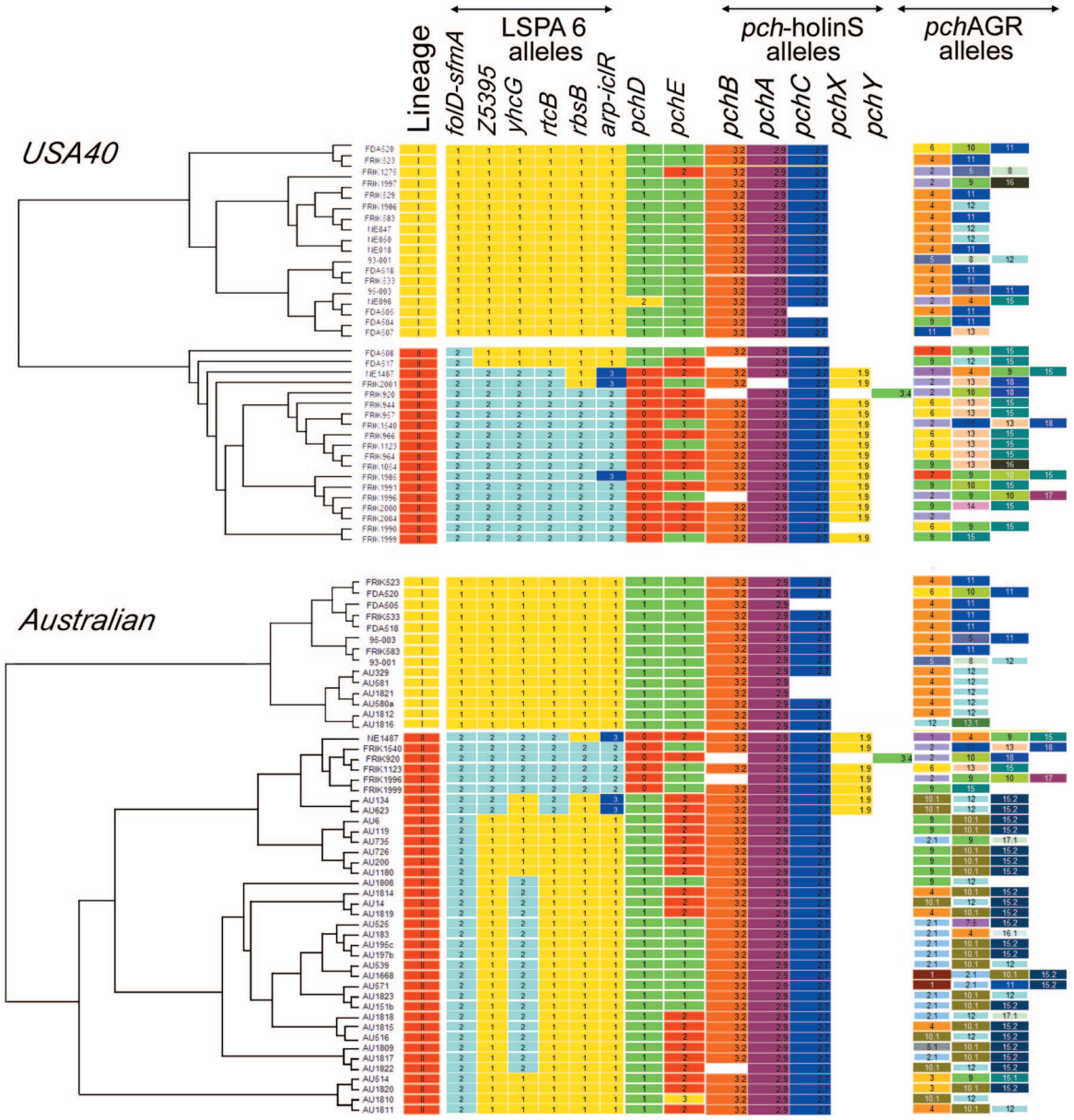

FIG. 3. Distribution of *pch* AGR alleles. The dendrograms were generated by neighbor-joining analysis of OBGS data on the USA 40 strain set and the Australian strain set. The OBGS data and binary files were generated from previously described data sets (26, 27). The USA 40 data set comprised 1,251 binary OBGS characters (453 constant and 161 informative; tree length, 1,081; consistency index, 0.7382; homoplasy index, 0.2618; retention index, 0.8233). The Australian data set comprised 1,159 characters (571 constant and 167 informative; tree length, 1,029; consistency index, 0.6239; homoplasy index, 0.3761; retention index, 0.7364). To the right of the branches, the colored rectangles indicate OBGS lineage (yellow, lineage I; red, lineage II), LSPA loci as described previously in Yang et al. (79), (yellow, lineage I; light blue, lineage II; dark blue, unique), and alleles of *pchD* (*pchD*$^+$, green; TAIΔ10795, red), and *pchE* (*pchE*$^+$, green; *pchEΔ352*, red). Presence of the individual *pchB*-holin S, *pchA*-holin S, *pchC*-holin S, *pchX*-holin S, and *pchY*-holin S is indicated by colored rectangles as identified at the top of the panel. For the *pch* AGR hybridizing segments, each different allele is indicated by a different color and number (positions of the alleles in the four columns are not meaningful).

had fallen to an average of only 38 mM at this point and continued to fall at a similar rate after this time period, implying that the growth rate shift and the expression trough are not due to diauxy. Nonetheless, the diversity in absolute levels of expression and the timing of this expression feature also serve as an additional landmark for quantitative analyses.

Although the apparent LEE1 expression level appears high during the first 100 min of culture (Fig. 4A and B), the data during this portion of the curves are not reliable because the culture density is well below the linear range of the spectrophotometer. To better represent the relationships between expression patterns and phylogenetic lineage or source of the strains, the same data from Fig. 4B was redisplayed, removing the first 100 min of culture and focusing on the 100- to 400-min portion of the expression curves, where both luminescence and spectrophotometric measurements are within the dynamic

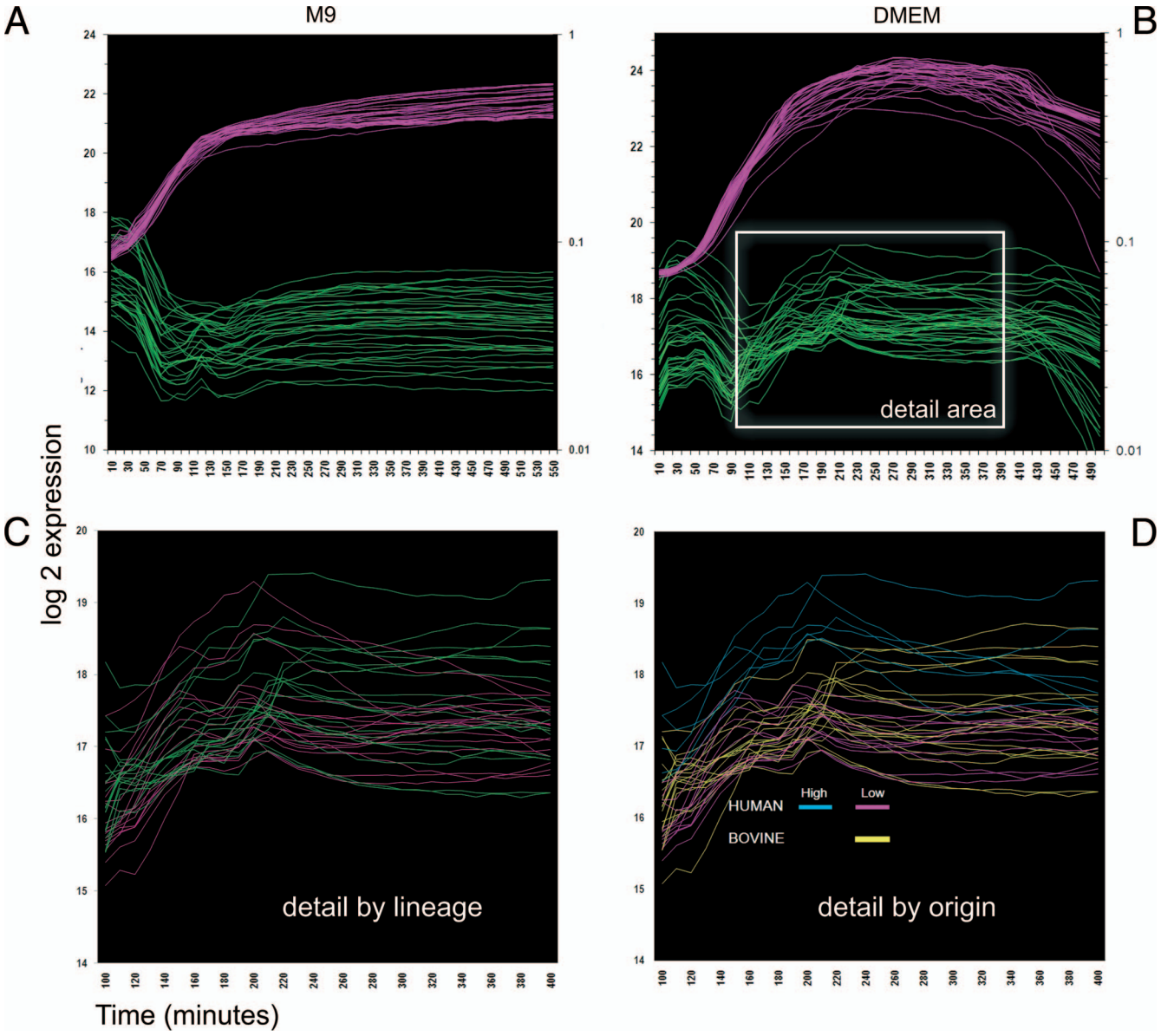

FIG. 4. Expression and growth of LEE1::*luxCDABE* in the USA 40 set of *E. coli* O157:H7 strains. Three different transformants of the LEE1::*luxCDABE* fusion plasmid in each strain were grown individually in M9 medium or DMEM and monitored every 10 min for OD and luminescence. The background-subtracted expression values (luminescence/OD) and $OD_{600\ nm}$ values were log transformed and averaged per parental strain. In panels A and B, time in culture are plotted on the *x* axis, the log-transformed expression values are plotted on the primary *y* axis (left-hand side), and the OD values are plotted on the secondary *y* axis on the right-hand side. The growth (pink) and expression (green) profiles of the averages from each individual strain are plotted for M9 medium and DMEM medium. In panels C and D, the same data from the 100- to 400-min time period of the DMEM-grown cells in panel B were detailed to highlight the relevant portion of the expression profile. In panel C, lineage I strains are shown in green, and lineage II strains are shown in pink. (D) Strains from human clinical samples show both high-level expression and low-level expression; all bovine strains show only low-level expression.

range of the machine. Expression curves were then colored by lineage (Fig. 4C) or by source/host (Fig. 4D). Overall, lineage II strains tend to have higher absolute levels of LEE1 expression, but the expression peaks occur later (Fig. 4C). If the same data set is compared for strain origin (human versus bovine), three main groups of strains emerge (Fig. 4D), with the lineage II human strains ($n = 6$) in general having the highest absolute levels and with the remainder of the human strains grouped with most bovine strains.

Allelic effects of *pch* AGR on LEE1 transcription. If the haplotypes and/or individual alleles of the *pch* AGR contribute significantly to the overall pattern of LEE expression, we would predict that (in the absence of extensive allelic interactions) the stronger alleles would have readily measurable effects on the LEE expression. Because the differences in timing of the expression features (peaks and troughs) were quite variable between strains, statistical analyses were designed to compare different portions and features of the expression

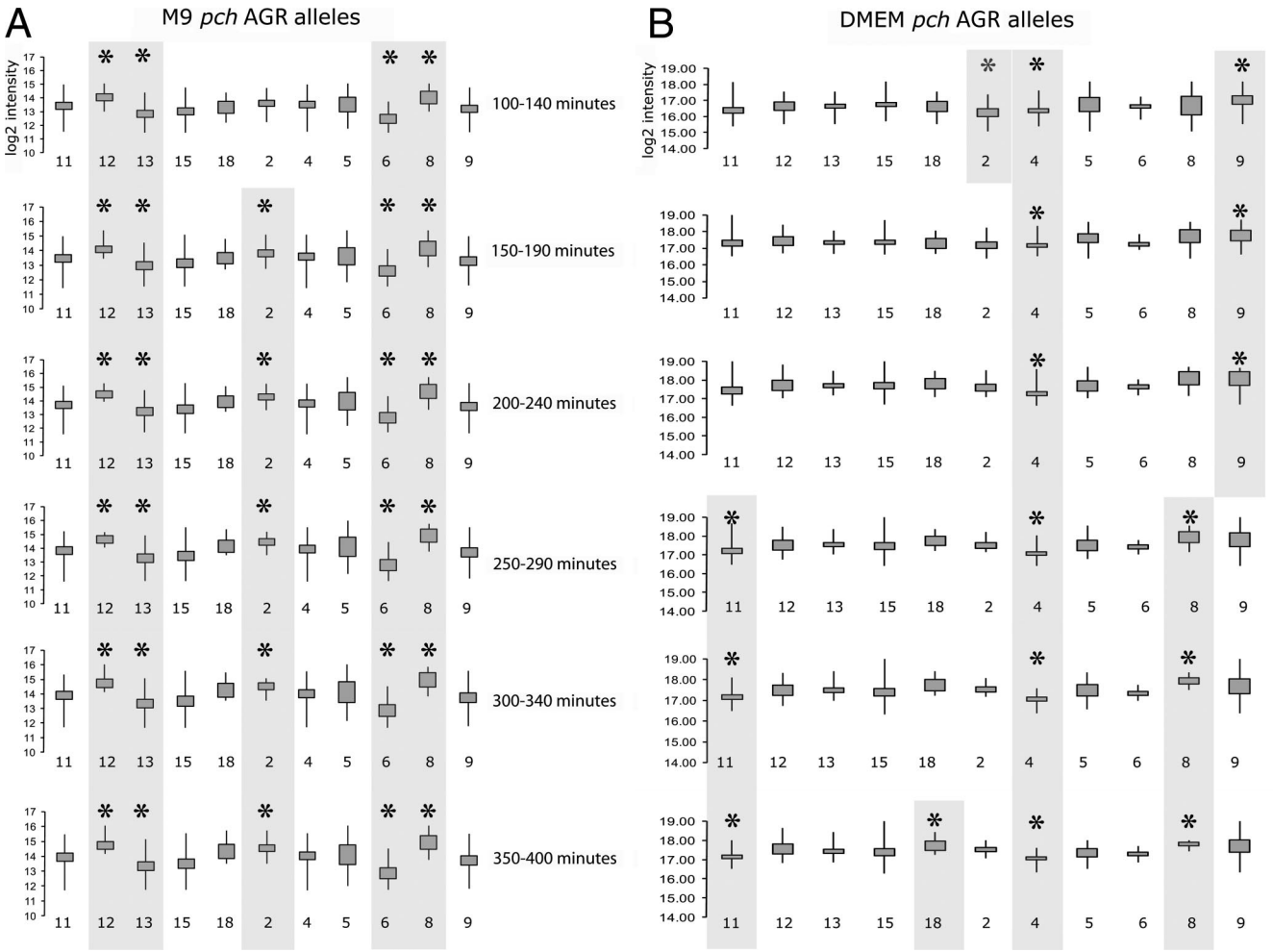

FIG. 5. Box and whisker plots of LEE1 expression values for different portions of the growth curve. The plots in both panels were derived by tabulating individual data points of log-transformed LEE1 expression values (*y* axis) for each *pch* AGR allele (*x* axis) during six 40-min time periods of the growth curves as indicated on the figure. Plots are shown for cells grown in M9 medium (A) and DMEM (B). The vertical bars indicate the range of expression values for each allele while the box denotes the extent of the upper and lower 95% confidence intervals. An asterisk is indicated above data for individual *pch* AGR alleles that show statistical significance (*t* test, $P < 0.01$). These statistically significant alleles are also highlighted.

curves. The first step was to divide the expression curves into six equal time periods between the 100- and 400-min windows. Expression values from each strain were tabulated for each different time period, and these values were then tabulated for each occurrence of the individual *pch* AGR alleles. Box and whisker plots (Fig. 5) showing the range and 95% confidence intervals were used to first visualize the parameters of the data. Collectively, the effects of individual *pch* AGR alleles in cells grown in M9 medium (Fig. 5A) were very consistent with their effects on cells grown in DMEM (Fig. 5B). Alleles were consistently high (e.g., *pch* AGR12 and *pch* AGR8) or consistently low (e.g., *pch* AGR13 and *pch* AGR6) in both media during each of the time periods. Analysis of variance tests on the allele-tabulated data from each time period showed the allele to be highly significant, with *P* values for each of the time periods in both M9 medium and DMEM of <0.0001. The sole exception was the 100- to 140-min period in DMEM, which was still highly significant ($P < 0.012$). Thus, the *pch* AGR alleles contribute significantly to the absolute levels of LEE1 expression under these growth conditions.

To estimate the relative strength of the individual alleles during each time period, *t* tests were used to compare the individual alleles to data from all of the alleles. From the alleles showing statistically significant effects, three broad patterns could be observed. The first pattern is observed with the *pch* AGR8 allele, which was consistently associated with the highest levels of LEE1 expression in either growth medium. The second pattern is exemplified by alleles that were consistently associated with high (*pch* AGR12 and *pch* AGR2) or low (*pch* AGR6 and *pch* AGR13) levels of expression but were statistically significant only in M9 medium. Last were those alleles that were associated with significantly high (*pch* AGR9) or low (*pch* AGR11 and *pch* AGR4) levels of expression only in DMEM. Collectively, these three patterns demonstrate that variation in the *pch* AGR has a substantial affect on LEE1 expression, and they

mark the first type of genomic variation in natural populations of this organism to be associated with effects on LEE expression. The dominance of *pch* AGR8 in both medium types along with the medium- and growth phase-specific effects of other alleles implies that *pch* AGR variation could influence LEE1 expression through multiple pathways.

DISCUSSION

Recent studies of *E. coli* O157:H7 strains have repeatedly shown significant detectable strain-to-strain variation in the levels of proteins that are encoded within the LEE island (37, 53, 55, 56). The genomic basis for this variation is not yet understood, nor is it known how the variation is related to the evolutionary history of the strains, save for the fact that human clinical isolates tend to express higher levels of the proteins. Our studies, which began in a phylogenetic context, now show that the polymorphic *pch* AGR are at least one source of genomic variation that influences LEE expression, and the variation correlates with the phylogeny inferred by OBGS and by markers such as LSPA-6 that lie within the genomic backbone. The comparative genome studies initially revealed variation in four different *pch*-like loci, the TAIΔ10795 (*pchD*), the *pchE*Δ352 *(pchE*), the *pchX*-holin S cassette, and the *pch* AGR. Despite being highly conserved markers within lineage II strains, the TAIΔ10795 (*pchD*), the *pchE*Δ352, and *pchX* alleles showed no statistically significant effects on LEE expression. In contrast, variation in the highly polymorphic *pch* AGR showed substantial effects on LEE expression and had distinct evolutionary patterns.

Several different observations support the hypothesis that the polymorphisms observed in the *pch* AGR are consequences of events adjacent to the *pch*-holin S regions of the *pch* prophage *pchA* (Sp4), *pchB* (Sp11), and *pchC* (Sp14) or in prophage immediately adjacent to them. First, the *pch*-holin S cassettes are highly conserved, which implies that variation detected by our Southern blotting analysis must exist outside of these cassettes. Second, in the Sakai genome sequence, prophage are situated adjacent to the *pch* prophage on the same PmeI fragments, and events in these phage would indeed give rise to polymorphisms in the PmeI fragment sizes. One of these prophage, the Sp5 prophage, is already known to be highly polymorphic. Sp5 is the stx_2 prophage in the Sakai genome, and it is well known that the stx_2 gene can be found on different prophages, some of which occupy different genomic segments in O157:H7 strains (59, 60). Moreover, phage-mediated variation in general is well known as the most common type of genomic variation in O157:H7 strains (3, 28, 29, 44–46, 70, 77). Though we do not know the mechanism for diversity generation in the *pch* AGR, simple excision of existing prophage would not account for the degree of diversity observed in the Southern blots. Therefore, the *pch* AGR likely tolerate a number of different types of events, including recombination. This trait evidently evolved relatively early in the lineage as even the O55:H7 strains displayed variation in the *pch* AGR, presumably in regions adjacent to *pchX* as this is the only *pch*-holin S cassette that was successfully amplified from these strains.

What is the driving force behind variation in the *pch* AGR? Replichore balancing may be one. In our previous study (70) we showed that phage-associated variation in lineage I strains was highly biased to one of the replichores. This replichore (replichore 1) is 290 kb longer than replichore 2 in the Sakai genome sequence, and hence balancing of the genomic length in each replichore by recombination, deletion, or rearrangement may provide a selective advantage for strains undergoing events in *pch* AGR (82). Second, the high degree of similarity in prophage genome segments likely enhances their rates of recombination, serving as hot spots both for intragenic recombination between other similar prophage in the genome and for intergenic recombination between newly infecting phage. The former may also occur in vitro during extensive passage, as was carefully noted recently by Iguchi et al. (19).

Regardless of the underlying mechanisms, these events clearly have the potential to affect LEE expression, thereby coupling the events to virulence and ecological characteristics. The population genetics studies by Manning et al. (33) recently provided evidence that virulence characteristics are evolving in this organism and in particular noted the emergence of specific multilocus single-nucleotide polymorphism (SNP) genotypes. The frequent occurrence of these genotypes among clinical samples along with the severity of associated symptoms implies that they are more virulent. The highly polymorphic *pch* AGR are covariate with respect to the LSPA-6 genomic backbone markers, and we have shown previously that certain LSPA-6 subtypes are also biased among bovine and clinical strains (79, 82). More work, however, is necessary to determine if a relationship exists between the *pch* AGR, the LSPA-6 markers, and the clades identified by SNP typing.

The congruence of the *pch* AGR polymorphisms relative to OBGS and LSPA-6 markers underscores the fact that unique patterns of lysogenization, excision, and/or recombination events in the *pch* AGR occur in each of the OBGS-defined lineages. Assuming that both lineages originated from a common ancestor with a single genomic organization of *pch* AGR, there are two simple explanations for the bias in *pch* AGR alleles and allele combinations. First, the bias could be due to the fact that the two OBGS/LSPA-6 lineages have unique ecologies and are therefore exposed to different pools of phage. Such a phenomenon would explain unique lysogenization or recombination patterns but not necessarily excision. While formally possible, this explanation is not satisfying because even though there is bias in the distribution of some subtypes between bovine and human clinical samples, such bias is far from absolute. In fact, the most abundant subpopulation (the true lineage I, LSPA-6 11111 genotype) can be found at high frequency in both bovine and human clinical samples (79).

If access or exposure to unique phage pools is not an adequate explanation, then the alternative explanation is that the two lineages have different susceptibilities to infection and/or recombination. In this instance, populations of each lineage show a different range of sensitivities to infection and lysogenization/recombination, and the particular "pathways" through which the *pch* AGR can potentially diversify are controlled by or otherwise set into motion by the genomic backbone of the lineage. Thus, simple genomic characteristics such as unique alleles of host factors required for lysogenization or recombination or even a unique prophage segment could affect the efficiency of recombination at the different prophage sites or affect the efficiency of infection by potentially recombining phage. Studies of prophage induction with mitomycin-C have

shown that interactions between phage occur even in lysogenic states (40), and one could envision how this phenomenon could also extend to interaction between prophage and infecting phage.

Though the mechanism of *pch* AGR diversification is not yet understood, several *pch* AGR alleles have clearly demonstrable effects on LEE1 expression. Studies on the other *LEE* promoters also show similar patterns (A. K. Benson, unpublished data), underscoring the fact that the expression pattern of the entire LEE island is affected. Beyond the obvious effects on pathogenesis, unique patterns of LEE expression could also influence the ability of strains to colonize different animal hosts. Differential expression patterns could translate into differential ability to colonize anatomical sites in a host as the subpopulations pass through the gastrointestinal tract. It will therefore be interesting to examine the distribution of *pch* AGR alleles in more diverse strain sets to determine if there are associations with host range or other ecological characteristics.

How could fluidity of the *pch* AGR cause variation in LEE expression? There are two basic models in which this could work. First, the *pch* AGR could encode a regulator(s) that directly modulates LEE expression. Alternatively, the effects could be indirect, mediated by factors within the *pch* AGR that influence LEE regulatory proteins that are encoded elsewhere. In support of the latter, we note that several of the prophage in the *pch* AGR of the Sakai genome carry *ileZ-argN-argO* cassettes specifying tRNAs that can potentially recognize relatively rare arginine and isoleucine codons (18, 49, 50). Although the codons recognized by these tRNAs are rare in the genome, they are highly overrepresented in the prophage-encoded *pchABCX*. In fact, of the 14 arginine codons in Pch, seven are rare codons with three occurrences of the AGA codon (recognized by *argO*) and four occurrences of the CGA codon (recognized by *argN*), including a doublet of the CGA codons near the C terminus. A single occurrence of the ATA codon, recognized by the *ileZ* tRNA, also occurs in the middle of the coding region. Thus, the *ileZ-argN-argO* cassettes in these prophage could stimulate *pchABCX* translation and ultimately enhance LEE1 transcription. Moreover, three of the seven *ileZ-argN-argO* cassettes encoded in prophages of the Sakai genome carry one or more SNPs in *argN* that may inactivate or significantly alter its function (45). Thus, variation in the cassettes themselves as well as the prophage in which they reside may translate into unique patterns in which the cassettes are expressed and into unique efficiency of function, all of which would impact *pch* expression and ultimately LEE expression. Collectively, this could provide a very simple mechanism to explain how variation in *pch* AGR could influence LEE1 expression. In favor of this argument, we note that initial experiments to overexpress *pchC* in a K-12 background have been unsuccessful.

The effects of *pch* AGR variation on LEE expression now represent yet a third phage-associated mechanism of genome variation that influences virulence characteristics, the other two being the variation in stx_2-encoding prophage and the variation in the content of prophage carrying secreted effectors (17, 31). Collectively, these three mechanisms of diversity generation provide the O157:H7 populations with a vast array of combinations of virulence factor alleles and expression patterns. The association of virulence factors and regulatory genes with prophage may represent a very sophisticated means for not only accelerating evolution of new pathogens but also for fine-tuning existing successful combinations of virulence genes and their regulators. The door to this phenomenon likely opened when the O157:H7 lineage became susceptible to polylysogeny. The trait appears to have evolved relatively early in the lineage as even the O55:H7 strains are polylysogenized (77). Though the effects of the polylysogeny on diversification of the genome have been well documented (1, 20, 23, 39–41, 45, 47, 58, 82), it remains unclear why or how the predecessor of the O157:H7 lineages became susceptible to this phenomenon. As we begin to dissect the mechanisms, it will be critical to also assess whether the trait of fine-tuning virulence factor alleles and expression patterns can explain the spread of *E. coli* O157:H7 to new and unexpected vectors of transmission.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the USDA National Research Initiative Competitive Grants Program 2001-35201-10115 to A.K.B.

We are grateful to Nina Murray for editorial assistance and preparation of figures for the manuscript.

REFERENCES

1. **Allison, H. E.** 2007. Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. Future Microbiol. **2:**165–174.
2. **Berdichevsky, T., D. Friedberg, C. Nadler, A. Rokney, A. Oppenheim, and I. Rosenshine.** 2005. Ler is a negative autoregulator of the LEE1 operon in enteropathogenic *Escherichia coli*. J. Bacteriol. **187:**349–357.
3. **Besser, T. E., N. Shaikh, N. J. Holt, P. I. Tarr, M. E. Konkel, P. Malik-Kale, C. W. Walsh, T. S. Whittam, and J. L. Bono.** 2007. Greater diversity of Shiga toxin-encoding bacteriophage insertion sites among *Escherichia coli* O157:H7 isolates from cattle than in those from humans. Appl. Environ. Microbiol. **73:**671–679.
4. **Bjarnason, J., C. M. Southward, and M. G. Surette.** 2003. Genomic profiling of iron-responsive genes in *Salmonella enterica* serovar Typhimurium by high-throughput screening of a random promoter library. J. Bacteriol. **185:** 4973–4982.
5. **Bustamante, V. H., F. J. Santana, E. Calva, and J. L. Puente.** 2001. Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. Mol. Microbiol. **39:**664–678.
6. **Chen, H. D., and G. Frankel.** 2005. Enteropathogenic *Escherichia coli*: unravelling pathogenesis. FEMS Microbiol. Rev. **29:**83–98.
7. **Deng, W., Y. Li, B. A. Vallance, and B. B. Finlay.** 2001. Locus of enterocyte effacement from *Citrobacter rodentium*: sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens. Infect. Immun. **69:**6323–6335.
8. **Donnenberg, M. S., J. B. Kaper, and B. B. Finlay.** 1997. Interactions between enteropathogenic *Escherichia coli* and host epithelial cells. Trends Microbiol. **5:**109–114.
9. **Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA **95:**14863–14868.
10. **Elliott, S. J., L. A. Wainwright, T. K. McDaniel, K. G. Jarvis, Y. K. Deng, L. C. Lai, B. P. McNamara, M. S. Donnenberg, and J. B. Kaper.** 1998. The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. Mol. Microbiol. **28:**1–4.
11. **Feng, P., K. A. Lampel, H. Karch, and T. S. Whittam.** 1998. Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. J. Infect. Dis. **177:**1750–1753.
12. **Frankel, G., A. D. Phillips, I. Rosenshine, G. Dougan, J. B. Kaper, and S. Knutton.** 1998. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. Mol. Microbiol. **30:**911–921.
13. **Friedberg, D., T. Umanski, Y. Fang, and I. Rosenshine.** 1999. Hierarchy in the expression of the locus of enterocyte effacement genes of enteropathogenic *Escherichia coli*. Mol. Microbiol. **34:**941–952.
14. **Goldberg, M. D., M. Johnson, J. C. D. Hinton, and P. H. Williams.** 2001. Role of the nucleoid-associated protein Fis in the regulation of virulence properties of enteropathogenic *Escherichia coli* Mol. Microbiol. **41:**549–559.
15. **Gomez-Duarte, O., and J. Kaper.** 1995. A plasmid-encoded regulatory region activates chromosomal *eaeA* expression in enteropathogenic *Escherichia coli*. Infect. Immun. **63:**1767–1776.

16. **Grant, A. J., M. Farris, P. Alefounder, P. H. Williams, M. J. Woodward, and C. D. O'Connor.** 2003. Co-ordination of pathogenicity island expression by the BipA GTPase in enteropathogenic *Escherichia coli* (EPEC). Mol. Microbiol. **48:**507–521.
17. **Gruenheid, S., I. Sekirov, N. A. Thomas, W. Deng, P. O'Donnell, D. Goode, Y. Li, E. A. Frey, N. F. Brown, P. Metalnikov, T. Pawson, K. Ashman, and B. B. Finlay.** 2004. Identification and characterization of NleA, a non-LEE-encoded type III translocated virulence factor of enterohaemorrhagic *Escherichia coli* O157:H7. Mol. Microbiol. **51:**1233–1249.
18. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.
19. **Iguchi, A., S. Iyoda, J. Terajima, H. Watanabe, and R. Osawa.** 2006. Spontaneous recombination between homologous prophage regions causes large-scale inversions within the *Escherichia coli* O157:H7 chromosome. Gene **372:**199–207.
20. **Iguchi, A., R. Osawa, J. Kawano, A. Shimizu, J. Terajima, and H. Watanabe.** 2003. Effects of lysogeny of Shiga toxin 2-encoding bacteriophages on pulsed-field gel electrophoresis fragment pattern of *Escherichia coli* K-12. Curr. Microbiol. **46:**224–227.
21. **Iyoda, S., and H. Watanabe.** 2005. ClpXP protease controls expression of the type III protein secretion system through regulation of RpoS and GrlR levels in enterohemorrhagic *Escherichia coli*. J. Bacteriol. **187:**4086–4094.
22. **Iyoda, S., and H. Watanabe.** 2004. Positive effects of multiple *pch* genes on expression of the locus of enterocyte effacement genes and adherence of enterohaemorrhagic *Escherichia coli* O157: H7 to HEp-2 cells. Microbiology **150:**2357–2571.
23. **Johansen, B. K., Y. Wasteson, P. E. Granum, and S. Brynestad.** 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.
24. **Jordan, D. M., N. Cornick, A. G. Torres, E. A. Dean-Nystrom, J. B. Kaper, and H. W. Moon.** 2004. Long polar fimbriae contribute to colonization by *Escherichia coli* O157:H7 in vivo. Infect. Immun. **72:**6168–6171.
25. **Karch, H., J. Heesemann, R. Laufs, A. D. O'Brien, C. O. Tacket, and M. M. Levine.** 1987. A plasmid of enterohemorrhagic *Escherichia coli* O157:H7 is required for expression of a new fimbrial antigen and for adhesion to epithelial cells. Infect. Immun. **55:**455–461.
26. **Kim, J., J. Nietfeldt, and A. K. Benson.** 1999. Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. Proc. Natl. Acad. Sci. USA **96:**13288–13293.
27. **Kim, J., J. Nietfeldt, J. Ju, J. Wise, N. Fegan, P. Desmarchelier, and A. K. Benson.** 2001. Ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of sorbitol-negative, beta-glucuronidase-negative enterohemorrhagic *Escherichia coli* O157. J. Bacteriol. **183:**6885–6897.
28. **Kotewicz, M. L., S. A. Jackson, J. E. LeClerc, and T. A. Cebula.** 2007. Optical maps distinguish individual strains of *Escherichia coli* O157:H7. Microbiology **153:**1720–1733.
29. **Kudva, I. T., P. S. Evans, N. T. Perna, T. J. Barrett, F. M. Ausubel, F. R. Blattner, and S. B. Calderwood.** 2002. Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. J. Bacteriol. **184:**1873–1879.
30. **Li, M., I. Rosenshine, H. B. Yu, C. Nadler, E. Mills, C. L. Hew, and K. Y. Leung.** 2006. Identification and characterization of NleI, a new non-LEE-encoded effector of enteropathogenic *Escherichia coli* (EPEC). Microbes Infect. **8:**2890–2898.
31. **Loukiadis, E., R. Nobe, S. Herold, C. Tramuta, Y. Ogura, T. Ooka, S. Morabito, M. Kerouredan, H. Brugere, H. Schmidt, T. Hayashi, and E. Oswald.** 2008. Distribution, functional expression, and genetic organization of Cif, a phage-encoded type III-secreted effector from enteropathogenic and enterohemorrhagic *Escherichia coli*. J. Bacteriol. **190:**275–285.
32. **Low, A. S., N. Holden, T. Rosser, A. J. Roe, C. Constantinidou, J. L. Hobman, D. G. Smith, J. C. Low, and D. L. Gally.** 2006. Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157: H7. Environ. Microbiol. **8:**1033–1047.
33. **Manning, S. D., A. S. Motiwala, A. C. Springman, W. Qi, D. W. Lacher, L. M. Ouellette, J. M. Mladonicky, P. Somsel, J. T. Rudrik, S. E. Dietrich, W. Zhang, B. Swaminathan, D. Alland, and T. S. Whittam.** 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. USA **105:**4868–4873.
34. **Martinez-Laguna, Y., E. Calva, and J. L. Puente.** 1999. Autoactivation and environmental regulation of *bfpT* expression, the gene coding for the transcriptional activator of *bfpA* in enteropathogenic *Escherichia coli*. Mol. Microbiol. **33:**153–166.
35. **McDaniel, T. K., K. G. Jarvis, M. S. Donnenberg, and J. B. Kaper.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc. Natl. Acad. Sci. USA **92:**1664–1668.
36. **McDaniel, T. K., and J. B. Kaper.** 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol. Microbiol. **23:**399–407.
37. **McNally, A., A. J. Roe, S. Simpson, F. M. Thomson-Carter, D. E. Hoey, C. Currie, T. Chakraborty, D. G. Smith, and D. L. Gally.** 2001. Differences in levels of secreted locus of enterocyte effacement proteins between human disease-associated and bovine *Escherichia coli* O157. Infect. Immun. **69:** 5107–5114.
38. **Mellies, J. L., S. J. Elliott, V. Sperandio, M. S. Donnenberg, and J. B. Kaper.** 1999. The Per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). Mol. Microbiol. **33:** 296–306.
39. **Miyamoto, H., W. Nakai, N. Yajima, A. Fujibayashi, T. Higuchi, K. Sato, and A. Matsushiro.** 1999. Sequence analysis of Stx2-converting phage VT2-Sa shows a great divergence in early regulation and replication regions. DNA Res. **6:**235–240.
40. **Muniesa, M., M. de Simon, G. Prats, D. Ferrer, H. Panella, and J. Jofre.** 2003. Shiga toxin 2-converting bacteriophages associated with clonal variability in *Escherichia coli* O157:H7 strains of human origin isolated from a single outbreak. Infect. Immun. **71:**4554–4562.
41. **Muniesa, M., and J. Jofre.** 2000. Occurrence of phages infecting *Escherichia coli* O157:H7 carrying the Stx 2 gene in sewage from different countries. FEMS Microbiol. Lett. **183:**197–200.
42. **Nakanishi, N., H. Abe, Y. Ogura, T. Hayashi, K. Tashiro, S. Kuhara, N. Sugimoto, and T. Tobe.** 2006. ppGpp with DksA controls gene expression in the locus of enterocyte effacement (LEE) pathogenicity island of enterohaemorrhagic *Escherichia coli* through activation of two virulence regulatory genes. Mol. Microbiol. **61:**194–205.
43. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142–201.
44. **Ogura, Y., K. Kurokawa, T. Ooka, K. Tashiro, T. Tobe, M. Ohnishi, K. Nakayama, T. Morimoto, J. Terajima, H. Watanabe, S. Kuhara, and T. Hayashi.** 2006. Complexity of the genomic diversity in enterohemorrhagic *Escherichia coli* O157 revealed by the combinational use of the O157 Sakai OligoDNA microarray and the Whole Genome PCR scanning. DNA Rev. **13:**3–14.
45. **Ohnishi, M., K. Kurokawa, and T. Hayashi.** 2001. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? Trends Microbiol. **9:**481–485.
46. **Ohnishi, M., J. Terajima, K. Kurokawa, K. Nakayama, T. Murata, K. Tamura, Y. Ogura, H. Watanabe, and T. Hayashi.** 2002. Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. Proc. Natl. Acad. Sci. USA **99:**17043–170489.
47. **Osawa, R., S. Iyoda, S. I. Nakayama, A. Wada, S. Yamai, and H. Watanabe.** 2000. Genotypic variations of Shiga toxin-converting phages from enterohaemorrhagic *Escherichia coli* O157: H7 isolates. J. Med. Microbiol. **49:**565–574.
48. **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.
49. **Perna, N. T., G. Plunkett, 3rd, 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. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.
50. **Plunkett, G., III, D. J. Rose, T. J. Durfee, and F. R. Blattner.** 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. J. Bacteriol. **181:**1767–1778.
51. **Porter, M. E., P. Mitchell, A. Free, D. G. Smith, and D. L. Gally.** 2005. The *LEE1* promoters from both enteropathogenic and enterohemorrhagic *Escherichia coli* can be activated by PerC-like proteins from either organism. J. Bacteriol. **187:**458–472.
52. **Porter, M. E., P. Mitchell, A. J. Roe, A. Free, D. G. Smith, and D. L. Gally.** 2004. Direct and indirect transcriptional activation of virulence genes by an AraC-like protein, PerA from enteropathogenic *Escherichia coli*. Mol. Microbiol. **54:**1117–1133.
53. **Rashid, R. A., T. A. Tabata, M. J. Oatley, T. E. Besser, P. I. Tarr, and S. L. Moseley.** 2006. Expression of putative virulence factors of *Escherichia coli* O157:H7 differs in bovine and human infections. Infect. Immun. **74:**4142–4148.
54. **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.
55. **Roe, A. J., D. E. Hoey, and D. L. Gally.** 2003. Regulation, secretion and activity of type III-secreted proteins of enterohaemorrhagic *Escherichia coli* O157. Biochem. Soc. Trans. **31:**98–103.
56. **Roe, A. J., H. Yull, S. W. Naylor, M. J. Woodward, D. G. Smith, and D. L. Gally.** 2003. Heterogeneous surface expression of EspA translocon filaments

by *Escherichia coli* O157:H7 is controlled at the posttranscriptional level. Infect. Immun. **71:**5900–5909.

57. **Sanchez-SanMartin, C., V. H. Bustamante, E. Calva, and J. L. Puente.** 2001. Transcriptional regulation of the *orf19* gene and the *tir-cesT-eae* operon of enteropathogenic *Escherichia coli*. J. Bacteriol. **183:**2823–2833.
58. **Sato, T., T. Shimizu, M. Watarai, M. Kobayashi, S. Kano, T. Hamabata, Y. Takeda, and S. Yamasaki.** 2003. Distinctiveness of the genomic sequence of Shiga toxin 2-converting phage isolated from *Escherichia coli* O157:H7 Okayama strain as compared to other Shiga toxin 2-converting phages. Gene **309:**35–48.
59. **Serra-Moreno, R., J. Jofre, and M. Muniesa.** 2007. Insertion site occupancy by stx_2 bacteriophages depends on the locus availability of the host strain chromosome. J. Bacteriol. **189:**6645–6654.
60. **Shaikh, N., and P. I. Tarr.** 2003. *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. J. Bacteriol. **185:**3596–3605.
61. **Sharma, V. K., and R. L. Zuerner.** 2004. Role of *hha* and *ler* in transcriptional regulation of the *esp* operon of enterohemorrhagic *Escherichia coli* O157:H7. J. Bacteriol. **186:**7290–7301.
62. **Shin, S., M. P. Castanie-Cornet, J. W. Foster, J. A. Crawford, C. Brinkley, and J. B. Kaper.** 2001. An activator of glutamate decarboxylase genes regulates the expression of enteropathogenic *Escherichia coli* virulence genes through control of the plasmid-encoded regulator, Per. Mol. Microbiol. **41:**1133–1150.
63. **Sircili, M. P., M. Walters, L. R. Trabulsi, and V. Sperandio.** 2004. Modulation of enteropathogenic *Escherichia coli* virulence by quorum sensing. Infect. Immun. **72:**2329–2337.
64. **Southward, C. M.** 2004. Gene expression profiling in EHEC. Ph.D. dissertation. University of Calgary, Calgary, Canada.
65. **Sperandio, V., J. B. Kaper, M. R. Bortolini, B. C. Neves, R. Keller, and L. R. Trabulsi.** 1998. Characterization of the locus of enterocyte effacement (LEE) in different enteropathogenic *Escherichia coli* (EPEC) and Shiga-toxin producing *Escherichia coli* (STEC) serotypes. FEMS Microbiol. Lett. **164:**133–139.
66. **Sperandio, V., J. L. Mellies, R. M. Delahay, G. Frankel, J. A. Crawford, W. Nguyen, and J. B. Kaper.** 2000. Activation of enteropathogenic *Escherichia coli* (EPEC) LEE2 and LEE3 operons by Ler. Mol. Microbiol. **38:**781–793.
67. **Sperandio, V., J. L. Mellies, W. Nguyen, S. Shin, and J. B. Kaper.** 1999. Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA **96:**15196–15201.
68. **Sperandio, V., A. G. Torres, J. A. Giron, and J. B. Kaper.** 2001. Quorum sensing is a global regulatory mechanism in enterohemorrhagic *Escherichia coli* O157:H7. J. Bacteriol. **183:**5187–5197.
69. **Sperandio, V., A. G. Torres, B. Jarvis, J. P. Nataro, and J. B. Kaper.** 2003. Bacteria-host communication: the language of hormones. Proc. Natl. Acad. Sci. USA **100:**8951–8956.
70. **Steele, M., K. Ziebell, Y. Zhang, A. Benson, P. Konczy, R. Johnson, and V. Gannon.** 2007. Identification of *Escherichia coli* O157:H7 genomic regions conserved in genotype associated with human infection. Appl. Environ. Microbiol. **73:**22–31.
71. **Tarr, P. I., S. S. Bilge, J. C. Vary, Jr., S. Jelacic, R. L. Habeeb, T. R. Ward, M. R. Baylor, and T. E. Besser.** 2000. Iha: a novel *Escherichia coli* O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure. Infect. Immun. **68:**1400–1407.
72. **Tatsuno, I., M. Horie, H. Abe, T. Miki, K. Makino, H. Shinagawa, H. Taguchi, S. Kamiya, T. Hayashi, and C. Sasakawa.** 2001. *toxB* gene on pO157 of enterohemorrhagic *Escherichia coli* O157:H7 is required for full epithelial cell adherence phenotype. Infect. Immun. **69:**6660–6669.
73. **Tobe, T., S. A. Beatson, H. Taniguchi, H. Abe, C. M. Bailey, A. Fivian, R. Younis, S. Matthews, O. Marches, G. Frankel, T. Hayashi, and M. J. Pallen.** 2006. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. Proc. Natl. Acad. Sci. USA **103:**14941–14946.
74. **Torres, A. G., J. A. Giron, N. T. Perna, V. Burland, F. R. Blattner, F. Avelino-Flores, and J. B. Kaper.** 2002. Identification and characterization of *lpfABCC′DE*, a fimbrial operon of enterohemorrhagic *Escherichia coli* O157:H7. Infect. Immun. **70:**5416–5427.
75. **Umanski, T., I. Rosenshine, and D. Friedberg.** 2002. Thermoregulated expression of virulence genes in enteropathogenic *Escherichia coli*. Microbiology **148:**2735–2744.
76. **Unkmeir, A., and H. Schmidt.** 2000. Structural analysis of phage-borne stx genes and their flanking sequences in Shiga toxin-producing *Escherichia coli* and *Shigella dysenteriae* type 1 strains. Infect. Immun. **68:**4856–4864.
77. **Wick, L. M., W. Qi, D. W. Lacher, and T. S. Whittam.** 2005. Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. J. Bacteriol. **187:**1783–1791.
78. **Wieler, L. H., T. K. McDaniel, T. S. Whittam, and J. B. Kaper.** 1997. Insertion site of the locus of enterocyte effacement in enteropathogenic and enterohemorrhagic *Escherichia coli* differs in relation to the clonal phylogeny of the strains. FEMS Microbiol. Lett. **156:**49–53.
79. **Yang, Z., J. Kovar, J. Kim, J. Nietfeldt, D. R. Smith, R. A. Moxley, M. E. Olson, P. D. Fey, and A. K. Benson.** 2004. Identification of common subpopulations of non-sorbitol-fermenting, β-glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. Appl. Environ. Microbiol. **70:**6846–6854.
80. **Zhang, C., M. Zhang, J. Ju, J. Nietfeldt, J. Wise, P. M. Terry, M. Olson, S. D. Kachman, M. Wiedmann, M. Samadpour, and A. K. Benson.** 2003. Genome diversification in phylogenetic lineages I and II of *Listeria monocytogenes*: identification of segments unique to lineage II populations. J. Bacteriol. **185:**5573–5584.
81. **Zhang, L., R. R. Chaudhuri, C. Constantinidou, J. L. Hobman, M. D. Patel, A. C. Jones, D. Sarti, A. J. Roe, I. Vlisidou, R. K. Shaw, F. Falciani, M. P. Stevens, D. L. Gally, S. Knutton, G. Frankel, C. W. Penn, and M. J. Pallen.** 2004. Regulators encoded in the *Escherichia coli* type III secretion system 2 gene cluster influence expression of genes within the locus for enterocyte effacement in enterohemorrhagic *E. coli* O157:H7. Infect. Immun. **72:**7282–7293.
82. **Zhang, Y., C. Laing, M. Steele, K. Ziebell, R. Johnson, A. K. Benson, E. Taboada, and V. P. Gannon.** 2007. Genome evolution in major *Escherichia coli* O157:H7 lineages. BMC Genomics **8:**121.
83. **Zhu, C., T. S. Agin, S. J. Elliott, L. A. Johnson, T. E. Thate, J. B. Kaper, and E. C. Boedeker.** 2001. Complete nucleotide sequence and analysis of the locus of enterocyte effacement from rabbit diarrheagenic *Escherichia coli* RDEC-1. Infect. Immun. **69:**2107–2115.