IAI Journals.ASM.org

Comparative Genomics Provides Insight into the Diversity of the Attaching and Effacing *Escherichia coli* Virulence Plasmids

Tracy H. Hazen,[a,b] James B. Kaper,[b] James P. Nataro,[c] David A. Rasko[a,b]

University of Maryland School of Medicine, Institute for Genome Sciences, Baltimore, Maryland, USA[a]; University of Maryland School of Medicine, Department of Microbiology and Immunology, Baltimore, Maryland, USA[b]; University of Virginia School of Medicine, Department of Pediatrics, Charlottesville, Virginia, USA[c]

Attaching and effacing *Escherichia coli* (AEEC) strains are a genomically diverse group of diarrheagenic *E. coli* strains that are characterized by the presence of the locus of enterocyte effacement (LEE) genomic island, which encodes a type III secretion system that is essential to virulence. AEEC strains can be further classified as either enterohemorrhagic *E. coli* (EHEC), typical enteropathogenic *E. coli* (EPEC), or atypical EPEC, depending on the presence or absence of the Shiga toxin genes or bundle-forming pilus (BFP) genes. Recent AEEC genomic studies have focused on the diversity of the core genome, and less is known regarding the genetic diversity and relatedness of AEEC plasmids. Comparative genomic analyses in this study demonstrated genetic similarity among AEEC plasmid genes involved in plasmid replication conjugative transfer and maintenance, while the remainder of the plasmids had sequence variability. Investigation of the EPEC adherence factor (EAF) plasmids, which carry the BFP genes, demonstrated significant plasmid diversity even among isolates within the same phylogenomic lineage, suggesting that these EAF-like plasmids have undergone genetic modifications or have been lost and acquired multiple times. Global transcriptional analyses of the EPEC prototype isolate E2348/69 and two EAF plasmid mutants of this isolate demonstrated that the plasmid genes influence the expression of a number of chromosomal genes in addition to the LEE. This suggests that the genetic diversity of the EAF plasmids could contribute to differences in the global virulence regulons of EPEC isolates.

Enterohemorrhagic *Escherichia coli* (EHEC) and enteropathogenic *E. coli* (EPEC) are diarrheagenic types of *E. coli* that can be collectively termed attaching and effacing *E. coli* (AEEC) due to similarities in their virulence mechanisms (1). EHEC can cause more severe illness, including mild to severe bloody diarrhea and hemolytic-uremic syndrome (HUS), resulting in kidney failure and death, with the majority of cases occurring in the developed world (1, 2). Meanwhile, EPEC is a leading cause of lethality associated with diarrhea in young children in developing countries (3, 4).

A major component of the virulence of AEEC is the locus of enterocyte effacement (LEE) genomic island, which encodes a type III secretion system (T3SS) (1, 5, 6). Similar to other diarrheagenic types of *E. coli*, such as enteroaggregative *E. coli* (EAEC), uropathogenic *E. coli* (UPEC), and enterotoxigenic *E. coli* (ETEC), the virulence mechanisms of AEEC involve genes carried by phage and plasmids (1, 7–9). The ability of EHEC to cause severe illness has been attributed to the presence of Shiga toxin genes (1, 2). In addition, nearly all of the O157:H7 EHEC isolates associated with food-borne illness in the United States carry the virulence-associated plasmid pO157 (92 kb), which encodes an α-hemolysin (*hlyA*) (10-12) and additional putative virulence factors (13, 14). Similarly, the severity of illness attributed to typical EPEC has been associated with a virulence-associated plasmid designated the EPEC adherence factor (EAF) plasmid, encoding the bundle-forming pilus (BFP) (15–17). EPEC strains carrying BFP are classified as typical EPEC (tEPEC), while those that do not contain BFP are considered atypical EPEC (aEPEC) (18, 19). tEPEC strains carrying BFP were demonstrated to be more likely to cause diarrhea (20) and were associated with more severe illness (3).

Two EAF plasmids have been sequenced to date, pMAR2 (95 kb), from the EPEC1 isolate E2348/69 (21, 22), and pB171 (69 kb), from the EPEC2 isolate B171 (23). Both of these plasmids contain the BFP gene, which encodes a type IV pilus conferring localized adherence to the surface of intestinal epithelial cells (8, 17, 18, 24). A difference in overall size of the two sequenced EAF plasmids can be attributed to the presence of genes for conjugal transfer contained on pMAR2, whereas pB171 lacks these genes (21, 23). Other than the sequences of the two EAF plasmids, investigations of the genetic diversity of the EAF-like plasmids have been limited to molecular-type analyses in comparison with pMAR2 (25) and the sequence characterization of a limited number of pEAF genes (26–30).

The unexpected genomic diversity recently demonstrated by phylogenomic analysis of 114 AEEC isolates (31) raised the question of how much genetic diversity exists among the virulence-associated plasmids in these isolates. Therefore, in the current study, we used comparative genomics to investigate the genetic diversity of AEEC virulence-associated plasmids in a collection of 210 AEEC isolates (see Table S1 in the supplemental material). The EHEC and EPEC plasmid genes identified in these genomes exhibited additional genetic diversity compared to the previously characterized EHEC and EPEC virulence plasmids. These findings

Received 10 June 2015 **Returned for modification** 10 July 2015
Accepted 26 July 2015
Accepted manuscript posted online 3 August 2015
Citation Hazen TH, Kaper JB, Nataro JP, Rasko DA. 2015. Comparative genomics provides insight into the diversity of the attaching and effacing *Escherichia coli* virulence plasmids. Infect Immun 83:4103–4117. doi:10.1128/IAI.00769-15.
Editor: V. B. Young
Address correspondence to David A. Rasko, drasko@som.umaryland.edu.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/IAI.00769-15.

doi:10.1128/IAI.00769-15

TABLE 1 AEEC plasmids analyzed in this study

ID	Plasmid size (kb)	No. of plasmids		BFP[a]	Phylogenomic lineage	Phylogroup	FIB group	Accession no.
		≥20 kb	<20 kb					
Isolates								
303289	95.3[c]	1	3	+	EPEC1	B2	II	JHRJ00000000
403116	95.8	1	0	+	EPEC1	B2	I	JHTJ00000000
C581-05	98[c]	2	3	+	EPEC4	B2	II	AIBE00000000
100414[d]	ND[b]	3	2	±	EPEC4	B2	II	JHQX00000000
401140	98[c]	2	2	+	EPEC5	A	II	JHRM00000000
401091	101.6[c]	1	0	+	EPEC7	B1	I	JHSI00000000
401588	99.4[c]	1	2	+	EPEC8	B2	II	JHSK00000000
302053	98[c]	2	2	+	EPEC9	B2	II	JHRG00000000
100329	90[c]	1	1	±	EPEC10	A	I	JHRT00000000
Previously characterized AEEC plasmids								
pMAR2 (E2348/69)	97.9	1	1 or 2	+	EPEC1	B2	I	NC_011603.1
pB171 (B171)	68.8	2	1	+	EPEC2	B1	I	NC_002142.1
pO157 (EDL933)	92	1	1	−	EHEC1	E	II	NC_007414.1
pO103 (12009)	75.5	1	0	−	EPEC2	B1	II	NC_013354.1
pO26_1 (11368)	85.1	2	2	−	EHEC2	B1	II	NC_013369.1
p101-1 (101-1)	153.5	ND	ND	±	EAEC	A	None	AAMK02000010

[a] +, presence of all or nearly all of the BFP genes; ±, presence of a few of the BFP genes, including *bfpA*; −, absence of all of the BFP genes.
[b] ND, not determined.
[c] Approximate EAF plasmid size of unclosed plasmids based on sequence analysis of the draft genome and by gel visualization of plasmid profiles.
[d] EPEC isolate containing an EAF-like plasmid with BFP genes that have ~72% nucleotide identity to genes of the prototype EAF plasmid pMAR2.

highlight the genetic diversity of the AEEC virulence plasmids, especially the EAF plasmids of EPEC, and suggest that there may be variation in EPEC virulence that results from the genetic diversity of the EAF plasmids.

MATERIALS AND METHODS

Bacterial strains and plasmids. The genome sequences analyzed in this study are listed in Table S1 in the supplemental material. The representative AEEC plasmids analyzed in this study are listed in Table 1. Bacteria were grown overnight in Luria-Bertani (LB) broth at 37°C with shaking or on LB agar plates at 37°C. Transcriptional analyses included the wild-type typical EPEC isolate E2348/69, a E2348/69 Δ*perABC* mutant (32), and the E2348/69 strain JPN15, which is missing the entire EAF plasmid (5, 20). The E2348/69 Δ*perABC* mutant was generated as previously described by replacement of the *perABC* region with a chloramphenicol resistance cassette (32).

Phylogenomic analysis. Phylogenomic analysis of 210 AEEC genomes and a diverse collection of 25 *E. coli* and *Shigella* genomes was performed as previously described (33, 34) (see Table S1 in the supplemental material). These isolates are identified as attaching and effacing *E. coli* (AEEC) based on the molecular presence of the locus of enterocyte effacement (LEE) region and have not been functionally examined for the attaching and effacing phenotype. The isolates were further classified as enteropathogenic *E. coli* (EPEC) by the presence of the bundle-forming pilus genes (*bfp*) or enterohemorrhagic *E. coli* (EHEC) by the presence of the Shiga toxin genes. Additional classification that was not based on these canonical virulence genes was accomplished by comparing the core genomes using phylogenomic analysis. Single nucleotide polymorphisms (SNPs) were detected relative to the completed genome sequence of the phylogroup F laboratory isolate *E. coli* IAI39 using the In Silico Genotyper (ISG) (34), which uses MUMMER v.3.22 (35) for SNP detection. The SNP sites that were identified in all genomes analyzed were concatenated and used for phylogenetic analysis as previously described (33). A maximum-likelihood phylogeny with 100 bootstrap replicates was generated using RAxML v.7.2.8 (36) and visualized using FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Alignment and phylogenetic analyses of genes. The gene phylogenies of *bfpA*, *perA*, FIB *repA*, and FIIA *repA* phylogenies were constructed with nucleotide sequences that were identified in each of the 210 AEEC genomes analyzed (see Table S1 in the supplemental material). The nucleotide sequences were aligned in MEGA5 (37) using ClustalW (38). A maximum-likelihood phylogeny was constructed using the Kimura 2-parameter model of distance estimation (39) with 1,000 bootstrap replications. The accession numbers for gene sequences of previously described alleles that are publicly available in GenBank are included here as references.

BSR analysis. BLAST score ratio (BSR) analysis was performed as previously described (40). Genes with similarity to those carried by the EHEC and EPEC virulence plasmids pO157 (NC_007414.1), pO26_1 (NC_013369.1), pO103 (NC_013354.1), and pMAR2 (NC_011603.1) were identified in each of the AEEC genomes using TBLASTN (41). The predicted amino acid sequence encoded by each gene on the plasmids was compared to each of the AEEC genomes by TBLASTN (41). The BSR of each gene compared to each genome was determined by dividing the TBLASTN bit score by the score obtained when the reference sequence is compared to itself. The plasmid genes that were identified with significant similarity had BSRs of ≥0.8, genes that were present but divergent had BSRs of ≥0.4 but <0.8, and genes that were absent had BSRs of <0.4.

The genetic similarity of the pMAR2 protein-encoding genes and pseudogenes to the genomes of representative EPEC isolates of diverse phylogenomic lineages was investigated using BLASTN (42) BSR (40) analysis. The circular display of the BLASTN BSR analysis of the similarity of pMAR2 genes to regions present in the genomes of EPEC isolates representing each of the EPEC phylogenomic lineage was generated using Circos 0.65 (43).

In silico detection of the incompatibility region used for typing (Inc typing) (44, 45) in each of the AEEC genomes analyzed in this study was performed by using BLASTN and determining the BSRs as described above. Briefly, the nucleotide sequences of each of plasmid genes were compared to each of the genomes using BLASTN (42), and the BSRs were calculated as described above.

Plasmid analyses. The contigs that exhibited similarity to the EAF plasmid were identified in the draft genome sequences of selected EPEC isolates chosen to represent the diverse EPEC phylogenomic lineages listed in Table 1. The plasmid profiles of these EPEC isolates were determined as previously described using an acid-phenol extraction method (46, 47). Briefly, each isolate was grown overnight in 5 ml of LB broth at 37°C with shaking (225 rpm). The supercoiled plasmid DNA was then isolated as previously described (46, 47). The supercoiled plasmid DNA was separated from chromosomal DNA by electrophoresis on a 1% agarose gel run at 100 V/cm for 4 h and visualized by staining with ethidium bromide.

Plasmid-free colonies were obtained for three isolates (401195, 401588, and 302053) after each was streaked to single colonies on an LB agar plate and grown overnight at 37°C. Colonies were inoculated into 5 ml of LB broth and grown overnight at 37°C with shaking (225 rpm), and the plasmid profiles were obtained as described above. The EAF plasmid-carrying and plasmid-free colonies were determined to have the same genomic background using enterobacterial repetitive intergenic consensus (ERIC) PCR analysis as previously described using primers ERIC1R (5′-ATGTAAGCTCCTGGGGATTCAC-3′) and ERIC2 (5′-AAGTAAGT GACTGGGGTGAGCG-3′) (48, 49). The ERIC PCR profiles were visualized by running at 100 V/cm for 1 h and 15 min on a 1% agarose gel that contained ethidium bromide.

RNA-Seq. RNA sequencing (RNA-Seq) was used to examine the global transcription of the EPEC prototype isolate E2348/69, the E2348/69 Δ*perABC* mutant (32), and the EAF plasmid-free variant of E2348/69, JPN15 (5, 20). The RNA was isolated and sequencing and analysis was performed as previously described (50). The isolates were grown overnight at 37°C in LB and inoculated by 1:100 dilution into 50 ml of Dulbecco's modified Eagle's medium (DMEM) supplemented with 4.5 g/liter of glucose or into Luria-Bertani (LB) broth. The cultures were then incubated at 37°C with shaking until they reached an optical density at 600 nm (OD_{600}) of 0.5, corresponding to exponential growth. Two biological replicates were generated for each EPEC isolate and medium type, and the total RNA was purified from the cells using the Ribopure bacteria kit (Ambion). Residual contaminating DNA was removed from the RNA by DNase treatment using the Turbo DNA-free kit (Ambion). The purified RNA was verified to be DNA free using qPCR with primers to detect the conserved chromosomal gene *rpoA*, which encodes a subunit of the RNA polymerase. The RNA was submitted for library construction using the Ovation prokaryotic RNA-Seq system (NuGen), and the samples were multiplexed to generate between 20 and 40 million reads for each sample using the Illumina HiSeq2000. The differential expression of genes in two RNA-Seq samples was determined using DESeq (51). The genes that were considered to have significant differential expression met the following criteria: a minimum read count of 10 in at least one of the samples, $\log_2$ fold change (LFC) of ≥ 2 or ≤ -2, and a false discovery rate (FDR) of 0.05 or lower determined using DESeq v. 1.5.24 and R-2.15.2. Circular displays of the significant LFC values calculated for each RNA-Seq sample comparison for genes of the chromosome and the plasmid pMAR2 were generated using Circos 0.65 (43). The relative expression trends of selected genes identified by RNA-Seq analysis were verified by quantitative reverse transcriptase PCR (qRT-PCR) analysis as previously described (50, 52) using primers listed in Table S2 in the supplemental material.

Accession numbers. The accession numbers of genomes included in this study are listed in Table S1 in the supplemental material. The RNA-Seq data were submitted to GEO under the accession number GSE69375.

RESULTS AND DISCUSSION

Phylogenomic analysis and identification of plasmids in AEEC isolates. Phylogenomic analysis of a collection of 210 AEEC genomes and 25 diverse *E. coli* and *Shigella* genomes demonstrated that the AEEC isolates' genomes occur in four of the previously defined *E. coli* phylogroups (53, 54) (B1, B2, A, and E) (Fig. 1). The AEEC genomes represent a global collection that includes previously characterized EHEC and EPEC isolates from the Statens Serum Institute (SSI) culture collection and the diarrheagenic *E. coli* (DEC) collection (31, 55) and also EPEC isolates from young children in Africa that were enrolled in the Global Enteric Multicenter Study (GEMS) (3, 56) (T. H. Hazen, M. S. Donnenberg, S. Panchalingam, M. Antonio, A. Hossain, I. Mandomando, J. B. Ochieng, T. Ramamurthy, B. Tamboura, S. Qureshi, F. Quadri, A. Zaidi, K. L. Kotloff, M. M. Levine, E. M. Barry, J. B. Kaper, D. A. Rasko, and J. P. Nataro, submitted for publication) (see Table S1 in the supplemental material). There were a total of 75,504 SNP sites present in all of the genomes analyzed, which were concatenated and used to construct the phylogeny (Fig. 1). Of the 210 AEEC isolates analyzed, 170 (81%) were identified in 10 different lineages (O157 EHEC1/O55 AEEC, EHEC2, EPEC1, EPEC2, EPEC4, EPEC5, and EPEC7 to EPEC10) that each contained five or more genomes per lineage (Fig. 1). The remaining 40 (19%) AEEC isolates were outside any of the previously characterized phylogenomic lineages and may occupy as-yet-unnamed EPEC phylogenomic lineages that could be identified following the sequencing and analysis of additional AEEC genomes (Fig. 1).

The *in silico* detection of the plasmid- or phage-carried EPEC and EHEC virulence genes (*stx*, *bfpA*, and *ehxA*) demonstrated that of the 210 AEEC genomes included in this study, 145 contained at least one or more of these genes. There were a total of 30 genomes that contained Shiga toxin genes (Fig. 1). Of the stx^+ isolates, 15 were in the EHEC1 phylogenomic lineage, while 12 were in the EHEC2 phylogenomic lineage, and three were in the EPEC2 phylogenomic lineage (Fig. 1). There were 40 AEEC genomes that had a region with similarity to the EHEC hemolysin gene (*ehxA*) that is often carried by the EHEC virulence plasmid, pO157, or other plasmids in non-O157 EHEC (Fig. 1; also, see Table S1 in the supplemental material), including 24 genomes that also contained *stx*. Meanwhile, the *bfpA* gene of the EPEC virulence plasmid was detected in 98 of the AEEC isolates (Fig. 1; also, see Table S1). Interestingly, there was one isolate (H.I.8.) that contained both *stx* and *bfpA*; however, there were no isolates in this study that contained *bfpA* and *ehxA*, which are both typically contained on plasmids (Fig. 1; also, see Table S1). Since the Shiga toxin genes and BFP genes are typically carried by different types of mobile genetic elements (a phage and a plasmid, respectively), it may be more likely to observe co-occurrence of BFP and Shiga toxin genes, rather than virulence factors carried on two different plasmids. These findings highlight the incongruent nature of the pathovar assignment, which is based on limited virulence-associated gene identification, and the current identification of the phylogenomic lineages, which is based on complete genomic content.

The identification of the virulence-associated genes *ehxA* and *bfpA*, which are usually harbored by plasmids, in 138 of the 210 AEEC genomes raised questions regarding the diversity of the plasmids involved in the mobility of these virulence genes. Therefore, we investigated the replicon diversity of the plasmids present in each of the AEEC genomes by *in silico* identification of the incompatibility types HI1, HI2, I1, X, L/M, N, FIA, W, Y, P, A/C, T, K/B, and B/O, which are part of a PCR multiplex-based Inc typing scheme (44, 45) (see Table S1 in the supplemental material). The IncP marker was one of the most prevalent incompatibility types, identified in 34 (16%) of the AEEC genomes (see Table S1). Other incompatibility types identified in the AEEC genomes were HI1, HI2, I1, FIA, Y, K/B, and B/O (see Table S1). These other incompatibility types may not be associated with vir-

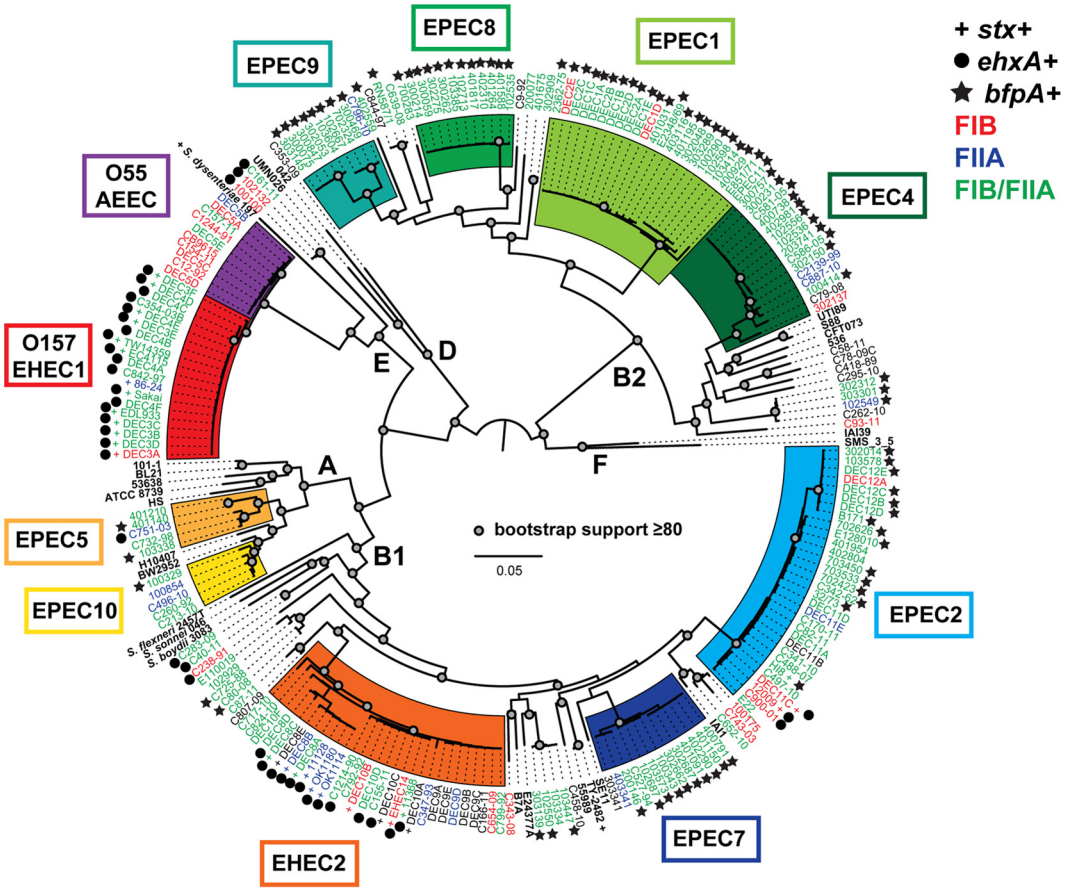

FIG 1 Single nucleotide polymorphism (SNP)-based phylogenomic analysis of the 210 AEEC genomes analyzed in this study compared to a reference collection of 25 diverse *E. coli* and *Shigella* genomes. SNPs were detected relative to the completed genome sequence of the laboratory isolate *E. coli* IAI39 with the In Silico Genotyper (ISG) (34). There were 75,504 SNP sites that were identified in all of the genomes analyzed, and these were concatenated into a representative sequence for each genome and used to generate a phylogeny. A maximum-likelihood phylogeny with 100 bootstrap replicates was constructed using RAxML v.7.2.8 (36). The genome names are colored by the plasmid *rep* type that was detected in the genome sequence. Red indicates that they contained only the FIB and not the FIIA *repA* gene, blue indicates that they contained only the FIIA and not the FIB *repA* gene, and green indicates that the genome contained both the FIB and FIIA *repA* genes. The symbols next to each genome indicate the presence of the genes for Shiga toxin, *stx*, the EHEC hemolysin, *ehxA*, and the major subunit of the bundle-forming pilus, *bfpA*. The clades in the phylogeny are colored by phylogenomic lineage, and the previously designated *E. coli* phylogroups are indicated by the designations A, B1, B2, D, E, and F (53, 54).

ulence factors in the AEEC genomes; however, they may carry genes conferring other phenotypes related to disease outcomes, such as antimicrobial resistance. Several sequenced EHEC and ETEC plasmids have been described with the I1 incompatibility type (57), and a previous study identified IncI1 in typical EPEC isolates that exhibited antimicrobial resistance (58).

Of particular interest are the FIB and FIIA incompatibility types, which are the incompatibility types of many of the previously characterized *E. coli* virulence plasmids (57). The presence of plasmids belonging to the FIB and FIIA plasmid families was investigated by *in silico* identification of the FIB and FIIA *repA* genes of the EAF plasmid pMAR2 (21, 22) (Fig. 1 and 2). The pMAR2 *repA* gene of the FIB incompatibility type was identified (TBLASTN BSR ≥ 0.7) in 171 (81%) of the AEEC genomes analyzed in this study (see Table S1 in the supplemental material). The FIIA *repA* gene of pMAR2 was identified in 164 (78%) of the AEEC genomes (see Table S1). There were 147 (70%) AEEC isolates that contained both the FIB and FIIA *rep* gene (see Table S1), and of these AEEC isolates, 100 were also *bfpA*$^+$. Previous plasmid characterization identified both the FIB and FIIA replicons on the BFP-encoding EPEC virulence plasmid (22, 23, 57). However, in these draft genomes it is not known whether these *rep* genes are on the same plasmid or whether either or both are associated with *bfpA*. Meanwhile, 24 (11%) of the 210 AEEC isolates contained only the FIB *repA*, and 17 (8%) contained only the FIIA *repA*. Only two AEEC isolates were *bfpA*$^+$ and contained only FIB, and only one isolate was *bfpA*$^+$ and contained only FIIA. It is possible that these isolates may possess the missing FIB or FIIA and that it is not represented in the draft genome assembly. This significant correlation of the presence of *bfpA* in genomes that contain both the FIB and FIIA *rep* genes suggests that BFP is largely associated with the FIB/FIIA plasmid family.

The replication protein-encoding gene, *repA*, is one of the few genes common to most AEEC virulence plasmids. The phylogenetic diversity of this gene has previously been used as a measure of the phylogenetic diversity and ancestral history of *E. coli* virulence plasmids (57). Phylogenetic analysis of the nucleotide sequences of the FIB *repA* (Fig. 2A) and the FIIA *repA* genes (Fig. 2B) demonstrated for lineages such as O157 EHEC1 that the *repA* genes exhibit lineage specificity. Analysis of the FIB *repA* gene

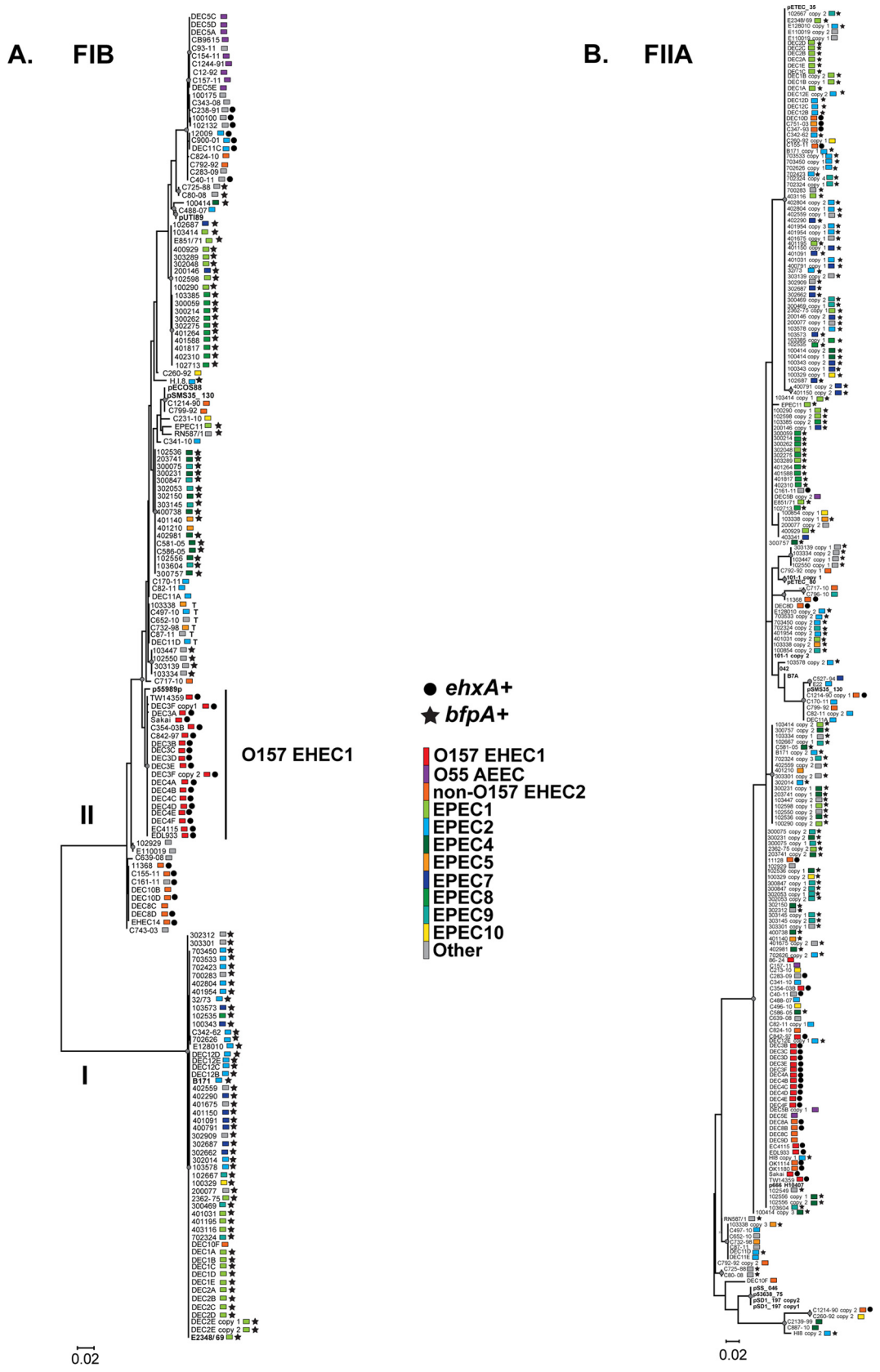

FIG 2 Phylogenetic analysis of the FIB replication protein-encoding gene, *repA* (A), and the FIIA replication protein-encoding gene, *repA* (B), identified in all AEEC genomes analyzed in this study. The nucleotide sequences were aligned in MEGA5 (37) using ClustalW (38). A maximum-likelihood phylogeny was constructed in MEGA5 (37) using the Kimura 2-parameter model and 1,000 bootstrap replications. All major nodes of the tree are supported by bootstrap values of ≥50 and are indicated on the phylogeny by a circle. The scale bar represents the distance of the number of nucleotide substitutions per site. The symbols next to each genome indicate the presence of the genes for EHEC hemolysin, *ehxA*, and the major subunit of the bundle-forming pilus, *bfpA*. The clades in the phylogeny are colored by phylogenomic lineage.

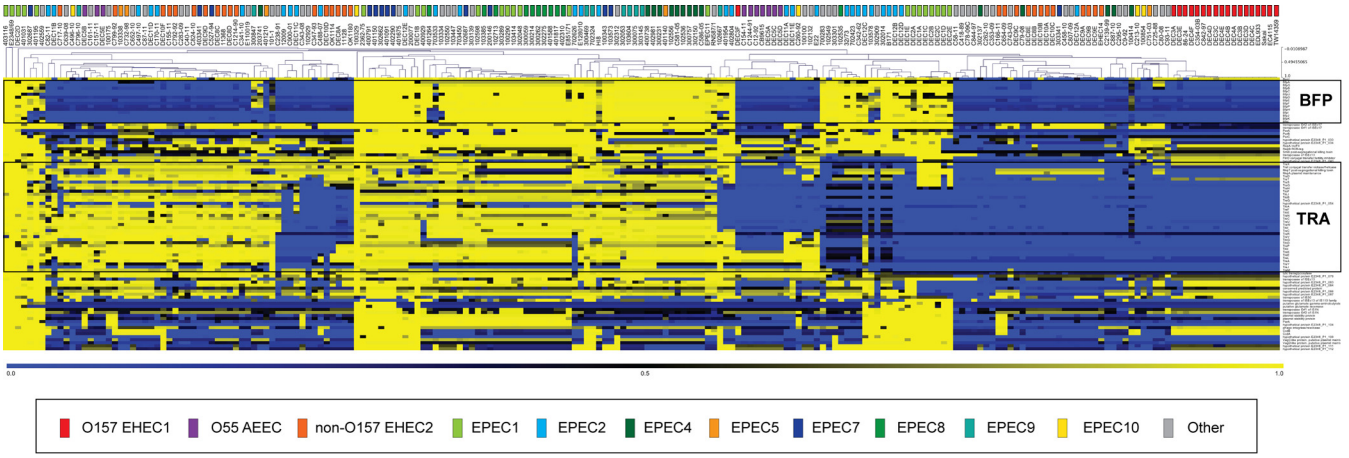

FIG 3 *In silico* detection of genes that have similarity to those of the EAF plasmid, pMAR2 (21, 22). TBLASTN (41) BSR (40) analysis of proteins encoded by the EAF plasmid pMAR2 (21, 22) of the EPEC1 lineage prototype isolate E2348/69, compared to the 210 AEEC genomes analyzed in this study, was carried out. Each square represents the BSR of a single gene compared to each genome. Each row is a gene, while each column is a genome. Hierarchical cluster analysis was performed using the Pearson correlation coefficient of MeV (76–78). The colored rectangle above each genome indicates the phylogenomic lineage that each genome belongs to in Fig. 1. The colors of the heatmap indicate the BSRs corresponding to the presence (yellow), divergence (black), or absence (blue) of each of the plasmid genes in each AEEC genome.

from AEEC isolates that occurred in EPEC phylogenomic lineages demonstrated that the EAF plasmids have greater sequence diversity and a more complex evolutionary history than the previously sequenced prototype EAF plasmids (21, 23, 57). The FIB *repA* sequences of the characterized EAF plasmids pB171 (23) and pMAR2 (21, 22) were previously described to occupy a phylogenetic group different from that of the other *E. coli* virulence plasmids, such as pO157 from the O157:H7 EHEC (57). In the current study, we observed that the FIB *repA* from the sequenced prototype plasmids, together with the FIB *repA* sequences from additional genomes in the EPEC1 and EPEC2 phylogenomic lineages, formed a separate phylogenetic group (group I) (Fig. 2A) from the other *E. coli* virulence plasmids analyzed in this study (group II) (Fig. 2A). Surprisingly, there were many FIB *repA* sequences from the EPEC genomes that were identified in group II with the other *E. coli* virulence plasmids (Fig. 2A). Overall, the virulence genes (BFP and *hlyA*) were identified in subgroups that also exhibited phylogenomic lineage specificity, except for those of group I, which was a mixture of genomes belonging to diverse phylogenomic lineages (EPEC1, EPEC2, EPEC7, EPEC8, and EPEC10) and additional, as-yet-uncharacterized lineages. Nearly all of the FIB *repA* sequences of group I were from genomes that were bfp^+ (Fig. 2A). An exception is the FIB *repA* of DEC10F, which was present in group I; however, the genome was identified in the non-O157 EHEC2 phylogenomic lineage but lacked *stx* and *bfp* (Fig. 2A).

In contrast to the single copy of the FIB *repA* gene identified in nearly all of the AEEC genomes analyzed, the FIIA *repA* sequences were identified as multiple distinct copies in some of the genomes (Fig. 2B). This suggests that more than one plasmid carrying the FIIA *repA* co-occur in selected AEEC isolates, or that some plasmids may contain multiple copies of the FIIA *repA* gene. Unlike the lineage and virulence gene specific subgroups identified in the FIB *repA* phylogeny, the FIIA *repA* phylogeny was composed of several large groups containing *repA* sequences from isolates of diverse phylogenomic lineages. This suggests that the IncFIIA plasmids are more widely distributed among AEEC isolates than are the previously characterized EHEC and EPEC plasmids. Further studies would be required to determine whether the multiple FIIA *repA* genes identified are involved in the replication and maintenance of each plasmid.

Overall, these analyses highlight the similarity of the AEEC virulence plasmids to other plasmids carried by AEEC and other *E. coli* isolates that do not contain any of the known virulence factors that are typically carried by plasmids. These findings also demonstrate a greater diversity of plasmid replicons associated with the EPEC virulence plasmid than previously recognized.

Comparative genomics of the EAF plasmids and their similarity to other AEEC virulence plasmids. To investigate how much genetic diversity there is among AEEC virulence plasmids, we compared protein-encoding genes of several previously sequenced plasmids to the AEEC genomes. *In silico* identification of genes carried by the EAF plasmid pMAR2 in the 210 AEEC genomes using TBLASTN BSR (40) demonstrated that there were genes with significant similarity to pMAR2 genes in genomes of diverse EPEC phylogenomic lineages but also in selected non-O157 EHEC2 lineage genomes (Fig. 3). Unlike some of the non-O157 EHEC2 genomes, the O157 EHEC1 genomes had few genes with any significant similarity to pMAR2 (Fig. 3). Hierarchical cluster analysis of the BSRs demonstrated there was similar plasmid gene content among EPEC genomes that contained the BFP genes (Fig. 3). A separate large group consisted of EPEC and non-O157 EHEC2 genomes that did not contain the BFP genes; however, they did contain genes with significant similarity to other regions of the pMAR2 plasmid (Fig. 3). This suggests that these genomes may contain plasmids that previously contained BFP and have since lost these genes, or alternatively, they may contain a plasmid that is an ancestor of the EAF plasmids. The *in silico* identification of the pMAR2 genes also illustrated that differences exist in the genetic content of EAF plasmids from genomes of the same phylogenomic lineage (Fig. 3). The sequenced EPEC1 EAF plasmid, pMAR2, contained genes for conjugative transfer (Tra) (21, 22), while the sequenced EAF plasmid, pB171, from an EPEC2 isolate (23) did not contain the Tra genes. However, sev-

eral EPEC1 genomes analyzed in this study also did not contain Tra genes (Fig. 3), highlighting the variation in the genetic composition of EAF plasmids within a single phylogenomic lineage.

The *in silico* identification of the hemolysin encoding EHEC virulence plasmids demonstrated that while highly conserved among the O157 EHEC isolates, which are clonal in nature, the plasmid-associated genes of the non-O157 EHEC isolates exhibited greater genetic diversity (see Fig. S1 in the supplemental material). The majority of the genes carried by the virulence plasmid pO157 of the O157:H7 EHEC isolate EDL933 were identified with significant similarity in the O157 EHEC1 (see Fig. S1A in the supplemental material). Genes of the pO157 plasmid encoding a type II secretion system were also identified in the O55 AEEC isolates, although other genes from this plasmid were not detected in these isolates (Fig. 1A). Detection of the protein-encoding genes of the hemolysin-encoding plasmid pO26_1 (NC_013369.1) from the non-O157 EHEC isolate 11368 (59) demonstrated that genes with significant similarity to those carried by this plasmid were present in other non-O157 EHEC isolates and several other AEEC isolates not included in the named phylogenomic lineages (see Fig. S1B in the supplemental material). *In silico* detection of the protein-encoding genes of plasmid pO103 (NC_013354.1) from the *E. coli* isolate 12009 (59) further demonstrated the similarity among plasmid genes of EPEC and EHEC (see Fig. S1C in the supplemental material). *E. coli* isolate 12009 is considered a non-O157 EHEC isolate based on virulence gene content (LEE$^+$, *stx*$^+$, and lacking *bfp*); however, phylogenomic analysis placed this isolate within the EPEC2 phylogenomic lineage (Fig. 1) (31). Interestingly, the genome of 12009 formed a small subgroup with two additional genomes (DEC11C and C900-01) (Fig. 1). The genomes of 12009 and DEC11C were the only genomes identified in the EPEC2 phylogenomic lineage that were *stx*$^+$. Also, 12009, DEC11C, and C900-01 had the only genomes of the EPEC2 lineage that carried the α-hemolysin gene, *hlyA*, as well as other pO103 plasmid genes. Among the pO103 genes with similarity in O55 AEEC and O157 EHEC1 genomes were genes encoding a type II secretion system (see Fig. S1C in the supplemental material). Meanwhile, the only genes of the EHEC plasmids (pO157, pO26_1, and pO103) that were identified with significant similarity in EPEC genomes were involved in plasmid replication, partitioning, conjugal transfer, or conserved hypothetical proteins. *In silico* detection of the EHEC plasmid genes demonstrated the plasmids present in the O157:H7 isolates are highly conserved, while there is greater variability among the plasmids from the O55 AEEC isolates and the non-O157 isolates.

Overall, comparison of previously characterized EPEC and EHEC plasmids to the AEEC isolates analyzed in this study demonstrated that virulence-associated genes of the plasmids, such as the bundle-forming pilus, hemolysin, and type II secretion system genes, exhibited significant similarity among isolates belonging to the same pathovar. Meanwhile, conserved plasmid genes encoding putative proteins involved in replication, conjugal transfer, or plasmid addiction systems were detected with similarity in both EPEC and EHEC.

Comparison of EAF plasmids from diverse EPEC phylogenomic lineages. The two EAF plasmids that have been sequenced were from prototype EPEC isolates of two phylogenomic lineages (EPEC1 and EPEC2) (21–23); however, the genome sequencing of additional LEE$^+$ *bfpA*$^+$ isolates demonstrated that there are many additional EPEC phylogenomic lineages (31) (Hazen et al., submitted). To further investigate the sequence variation of EAF plasmids, we compared the EAF plasmids identified in the genome sequences of representative EPEC isolates from multiple diverse phylogenomic lineages (Table 1). Identification of only some of the genes carried by pMAR2 in each of the genomes of these representative isolates demonstrated the considerable genetic diversity of the EAF plasmids (Fig. 4). Among the genes that were not identified in some of the genomes were those for conjugative transfer, which was previously demonstrated to be absent from the EAF plasmid pB171 (23). Initially, the BFP operon also appeared to be absent from the genome of isolate 100414; however, upon closer inspection, this isolate was identified to contain BFP genes that have significant nucleotide sequence divergence (approximately 72% nucleotide identity) compared to the prototype EAF plasmid, pMAR2 (22) (Fig. 4; also, see Fig. S3 in the supplemental material).

The diversity of the EAF plasmid from EPEC isolates within the same lineage was also determined for two EPEC1 isolates and one EPEC7 isolate relative to the EPEC1 prototype EAF plasmid pMAR2 (21, 22) (see Fig. S2 in the supplemental material). EPEC isolate 403116 contained an FIB *repA* gene that was in the same phylogenetic group (group I) as the prototype plasmid pMAR2 (Fig. 2A; also, see Fig. S2 in the supplemental material). Also, there was significant similarity over nearly all of the genes of the completed EAF plasmid identified in the genome sequence of 403116 compared to pMAR2 (see Fig. S2). Meanwhile, EPEC isolate 303289 was also from the EPEC1 phylogenomic lineage; however, this isolate contained an FIB *repA* gene that was more related to *repA* sequences of other *E. coli* plasmids in phylogenetic group II than to the previously sequenced EAF plasmids (Fig. 2A). Similarly, although the *bfp* genes were highly conserved, the other genes carried by the EAF plasmid exhibited considerable sequence divergence or were absent compared to pMAR2 (see Fig. S2). In contrast, the EAF plasmid of EPEC isolate 401091 from the EPEC7 phylogenomic lineage had greater sequence similarity over many of the non-BFP genes carried by its EAF plasmid compared to pMAR2 than was observed for the EAF plasmid from 303289 (see Fig. S2).

To investigate the variability of the BFP operon present in the *bfpA*$^+$ AEEC isolates analyzed, we identified regions with genetic similarity to protein-encoding genes of the BFP operon using TBLASTN BSR (see Fig. S3 in the supplemental material). Hierarchical cluster analysis of the identifiable genetic similarity demonstrated that there was some phylogenomic lineage specificity for many of the genomes (see Fig. S3 in the supplemental material). However, there were also genomes that had BFP genetic similarity that was inconsistent with other isolates belonging to the same phylogenomic lineage (see Fig. S3). For example, some genomes contained additional protein-encoding genes (ORF18 to -20 and the TrcP gene) that were previously identified on the B171 EAF plasmid, while other genomes from the same phylogenomic lineage did not carry these genes (see Fig. S3B). In addition, there were several genomes that contained a truncated *bfpA* and thus were missing the majority of the BFP operon. However, these genomes did contain genes with similarity to the plasmid regulator *perA* (see Fig. S3). This was similar to a previous study that also described EPEC isolates which had a truncated BFP operon with only *bfpA* but which also contained *per* genes (60). The *bfp* and *per* genes of previously characterized EAF plasmids are separated by an insertion sequence element (21–23), and it is possible that the

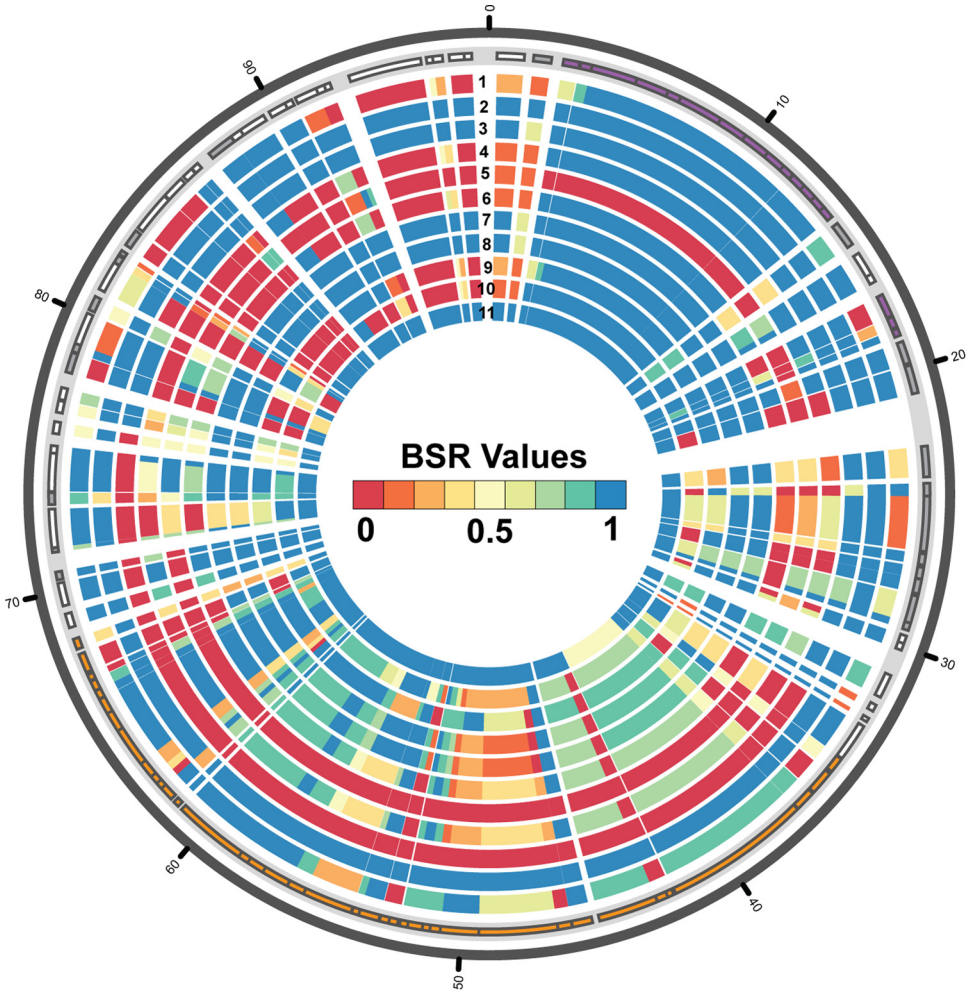

FIG 4 Circular display of genes with similarity to those of pMAR2 in the genomes of EPEC isolates from diverse phylogenomic lineages. The nucleotide sequences of predicted protein-encoding genes and pseudogenes of pMAR2 (21, 22) were compared to selected EPEC genomes and the previously sequenced EPEC2 plasmid pB171 (23), using BLASTN (42). A BSR was generated for each predicted protein-encoding sequence and pseudogene compared to each genome as previously described (40), and the plot was generated using Circos (43). Genes are indicated in the light gray outer track of the plot as rectangles of different sizes according to their coordinates and are colored purple to indicate BFP-associated genes, orange to indicate the conjugal transfer region, white to indicate protein-encoding genes, and gray to indicate pseudogenes. The genomes in the plot, from the outer BSR track to the inner track, are as follows: EPEC1 isolate 303289 (track 1), EPEC1 isolate 403116 (track 2), EPEC2 plasmid pB171 (track 3), EPEC4 isolate C581-05 (track 4), EPEC4 isolate 100414 (track 5), EPEC5 isolate 401140 (track 6), EPEC7 isolate 402290 (track 7), EPEC7 isolate 401091 (track 8), EPEC8 isolate 401588 (track 9), EPEC9 isolate 302053 (track 10), and EPEC10 isolate 100329 (track 11). The BSR scale is in the middle of the figure, with blue indicating genes that were present with significant similarity, yellow indicating divergent sequences, and red indicating genes that were absent.

bfp and *per* genes may undergo recombination events that are independent of each other.

Analysis of the genetic diversity of BFP genes demonstrated that the *bfpA* gene exhibited the greatest sequence variation of this gene cluster (see Fig. S3 in the supplemental material). Previous studies used sequence analysis of *bfpA* (27–29) and the plasmid regulator *perA* (26, 29, 30) to investigate the diversity of typical EPEC. Phylogenetic analysis of the nucleotide sequence diversity of *bfpA* demonstrated there was less phylogenomic lineage specificity among *bfpA* genes identified in the current AEEC genomes than we have previously described for other virulence-associated genes of EPEC, such as the LEE (31) (see Fig. S4A in the supplemental material). There was some observable lineage specificity for the *perA* sequences analyzed, such as that observed for the EPEC8 isolates (see Fig. S4B in the supplemental material). However, there were other phylogenetic groups that contained *perA* sequences from genomes of multiple phylogenomic lineages (see Fig. S4B in the supplemental material).

Overall, these findings indicate that the EAF plasmid has genetic diversity that is inconsistent with a single acquisition and stable maintenance by EPEC isolates over time. In some instances EPEC isolates have likely lost their ancestral EAF plasmid and have acquired another EAF plasmid originating from a different lineage of *E. coli*. Also, the EAF plasmid has likely undergone many genetic modifications, including recombination of the BFP genes and other genes on the plasmid.

RNA-Seq analysis of the contribution of the EAF plasmid to the virulence regulon of E2348/69. To investigate the transcriptional contribution of the EAF plasmid, pMAR2, to the virulence regulon of E2348/69, we used RNA-Seq to characterize the differ-

TABLE 2 RNA-Seq analysis of the total number of genes that were differentially expressed in E2348/69 and its plasmid mutants

Sample 1	Sample 2	No. of genes[a]		
		Increased (LFC ≥ 2)	Decreased (LFC ≤ −2)	Total altered
E2348/69 in LB	E2348/69 in DMEM	201	267	468
E2348/69 in LB	Δ*perABC* mutant in LB	30	58	88
E2348/69 in DMEM	Δ*perABC* mutant in DMEM	56	34	90
Δ*perABC* mutant in LB	Δ*perABC* mutant in DMEM	345	335	680
JPN15 in LB	Δ*perABC* mutant in LB	365	367	732
JPN15 in DMEM	Δ*perABC* mutant in DMEM	59	134	193
E2348/69 in LB	JPN15 in LB	313	322	635
E2348/69 in DMEM	JPN15 in DMEM	79	91	170
JPN15 in LB	JPN15 in DMEM	265	410	675

[a] LFC is the $\log_2$ fold change, determined as the normalized read counts of sample 2 divided by those of sample 1.

ential gene expression of wild-type E2348/69 and two plasmid mutants (the Δ*perABC* mutant [32] and JPN15, which lacks the EAF plasmid [20]). These isolates were examined under conditions that were previously described to increase virulence gene expression (61, 62). There were 468 genes of E2348/69 that had altered expression when grown in DMEM compared to LB, 201 genes that were induced and 267 genes that were repressed (Table 2 and Fig. 5A; also, see Data Set S1 in the supplemental material). Of these 468 differentially expressed genes, 456 (97%) were located on the chromosome (Fig. 5B). The LEE region exhibited increased expression during growth in DMEM compared to LB, and the flagellar motility gene *fliC* had decreased expression, which is consistent with previous findings (62) (see Data Set S1 in the supplemental material). Of the other regions that had increased expression, there were many O-antigen biosynthesis genes, whose products form components of the lipopolysaccharide (LPS) (*galE*, *galF*, *gne*, *wcaM*, *wcaL*, *wcaK*, *wzxC*, and *wzy*) (63) (see Data Set S1 in the supplemental material). Another expression pattern that was consistent with previous studies was the increased expression of genes in the BFP operon and the plasmid regulator (*perABC*) of the EAF plasmid pMAR2 (Fig. 6) (64). However, a surprising finding was that no additional genes of pMAR2 exhibited statistically significant differential expression under these growth conditions (Fig. 6).

The plasmid regulator *perABC* was previously demonstrated to be involved in the regulation of the BFP genes under growth conditions that promote virulence gene expression (64) (Fig. 6). The *perABC* operon has also been demonstrated to regulate the expression of the chromosomal LEE genes (65); however, it was not known how many additional chromosomal genes are influenced by the presence of this regulator. The role of *perABC* in expression of additional pEAF and also of chromosomal genes was investigated using RNA-Seq of a Δ*perABC* deletion mutant compared to wild-type E2348/69. There were 88 genes that were differentially expressed in the Δ*perABC* strain compared to wild-type E2348/69 when grown in LB (see Data Set S2 in the supplemental material) and 90 genes that were differentially expressed when the isolates were grown in DMEM (Table 2; also, see Data Set S3 in the supplemental material). There were 34 genes that were differentially expressed in both LB and DMEM for the Δ*perABC* strain compared to wild-type E2348/69. This similar number of differentially expressed genes for the Δ*perABC* mutant compared to the wild type whether they were grown in LB (88 genes) or DMEM (90 genes) suggests that the majority of the observed changes in gene expression during growth in DMEM compared to LB are not dependent on *perABC* (Table 2). As expected, expression of genes in the BFP operon is decreased in the Δ*perABC* mutant compared to the wild type whether the isolate is grown in LB or DMEM, which is consistent with previous reports that *perA* is required for expression of BFP (64). The lipopolysaccharide (LPS) region, which had increased expression for wild-type E2348/69 grown in DMEM compared to LB, had increased expression in the Δ*perABC* mutant compared to wild-type E2348/69 in both LB and DMEM (Fig. 5A; also, see Data Sets S1 to S3 in the supplemental material). The increased expression of both *bfpA* (BFP) and *wzy* (LPS) was verified by qRT-PCR analysis (see Fig. S4 in the supplemental material). This finding suggests that that plasmid regulator decreases the expression of these LPS biosynthesis genes. However, further studies are required to determine what role this potential change in phenotype plays during establishment of a host infection or interactions with other bacteria.

Interestingly, there were two plasmid genes outside the BFP operon that exhibited increased expression in the Δ*perABC* mutant compared to the wild-type E2348/69 isolate when grown in DMEM but not LB (Fig. 6; also, see Fig. S5 in the supplemental material). One of these genes was previously described as a regulator of conjugative transfer, *traM* (Fig. 6). TraM serves as an anti-activator of the quorum-sensing dependent activator *traR* and prevents TraR from increasing conjugation of the plasmid (66–69). The *traM* gene of pMAR2 had increased expression for the Δ*perABC* mutant compared to wild-type E2348/69 when grown in DMEM (Fig. 6; also, see Fig. S5 in the supplemental material), demonstrating that *perABC* likely represses expression of *traM* in wild-type E2348/69 under virulence conditions. This suggests that in addition to increasing expression of the BFP operon (64) and LEE region (65, 70) during virulence, *perABC* increases the expression of genes involved in conjugation by reducing the anti-activator function of *traM*. In addition, the quorum-sensing-dependent transcriptional activator *traR* had decreased expression in the Δ*perABC* mutant grown in DMEM compared to LB (Fig. 6A), suggesting that under the virulence-promoting growth conditions of DMEM and in the absence of *perABC*, TraM increases the anti-activation function of *traR*. Although we observed expression of other pMAR2 Tra genes, there was no statistically significant difference in expression of these genes during growth in DMEM compared to LB for wild-type E2348/69. Further research is necessary to determine whether the presence of *perABC* facilitates expression of Tra and increased

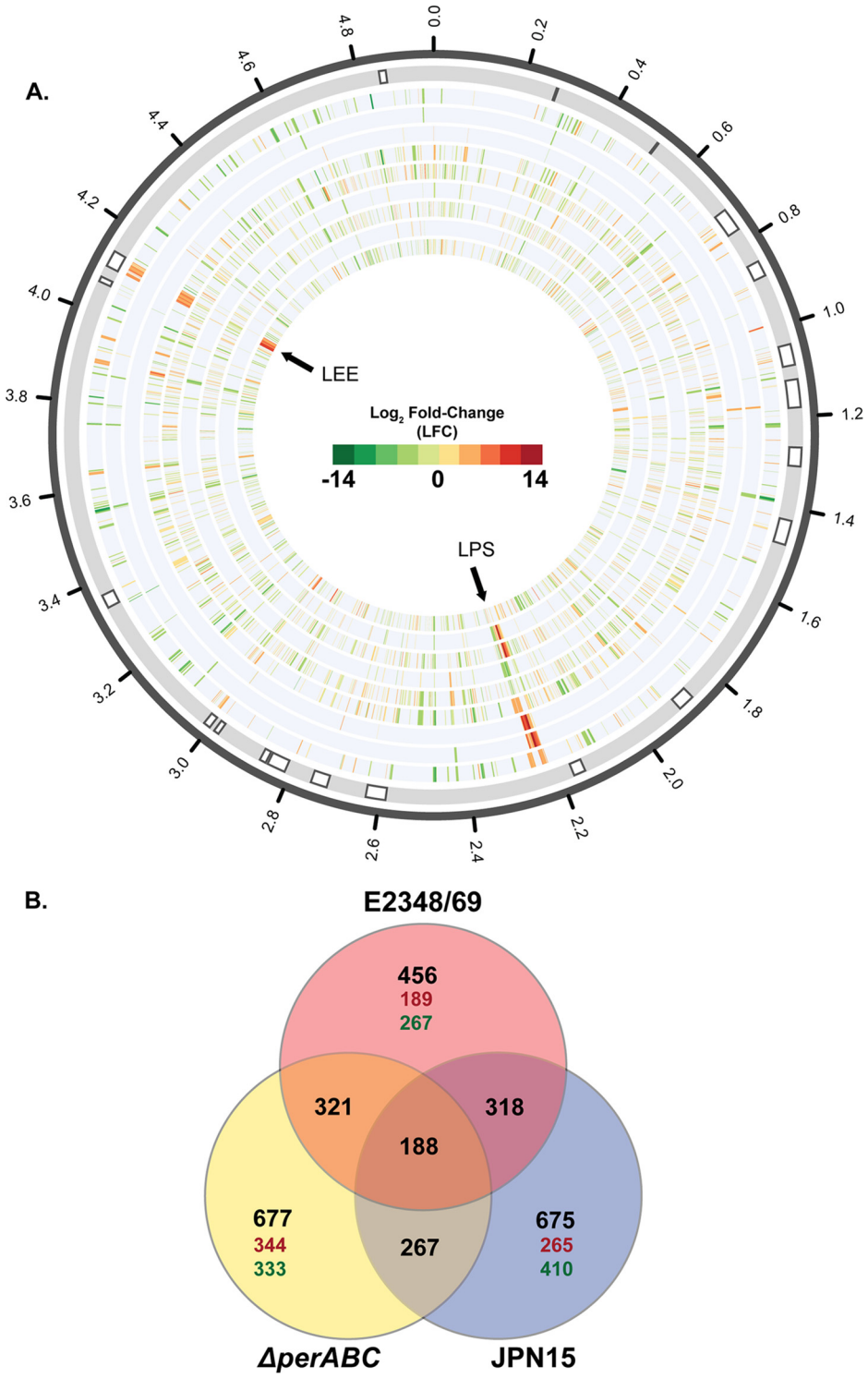

FIG 5 RNA-Seq analysis of the global transcription of the E2348/69 chromosome. (A) Circular display of the differential expression analysis of genes on the E2348/69 chromosome (22). RNA-Seq was used to determine changes in the global transcription of wild-type E2348/69, an E2348/69 Δ*perABC* mutant, and the isogenic plasmid-free E2348/69 isolate, JPN15. The color scale indicates $\log_2$ fold change (LFC) from −14 (green) to 14 (red). The outermost track denotes the location of previously identified prophage and insertion element regions of E2348/69 (indicated by white boxes) (22). The tracks of differential expression data are as follows, from the outer track to the innermost track: wild-type E2348/69 in DMEM versus wild-type E2348/69 in LB (track 2), the Δ*perABC* mutant in LB versus wild-type E2348/69 LB (track 3), the Δ*perABC* mutant in DMEM versus wild-type E2348/69 in DMEM (track 4), the Δ*perABC* mutant in DMEM versus the Δ*perABC* mutant in LB (track 5), the Δ*perABC* mutant in LB versus JPN15 in LB (track 6), the Δ*perABC* mutant in DMEM versus JPN15 in DMEM (track 7), JPN15 in LB versus wild-type E2348/69 in LB (track 8), JPN15 in DMEM versus wild-type E2348/69 in DMEM (track 9), and JPN15 in DMEM versus JPN15 in LB (track 10). (B) Venn diagram with the number of differentially expressed chromosomal genes that are shared or exclusive to wild-type E2348/69, E2348/69 Δ*perABC* mutant, and JPN15 RNA-Seq samples. Red indicates the number of chromosomal genes that were differentially expressed in wild-type E2348/69, yellow indicates the genes differentially expressed in the Δ*perABC* mutant, and blue indicates the genes differentially expressed in JPN15. The regions of overlap indicate the number of differentially expressed genes that were shared between the different samples. The black numbers in the outer circle represent the total number of transcriptionally altered genes; in red are the number of exclusive genes in each isolate that had increased expression, and in green are the number of exclusive genes in each isolate that had decreased expression.

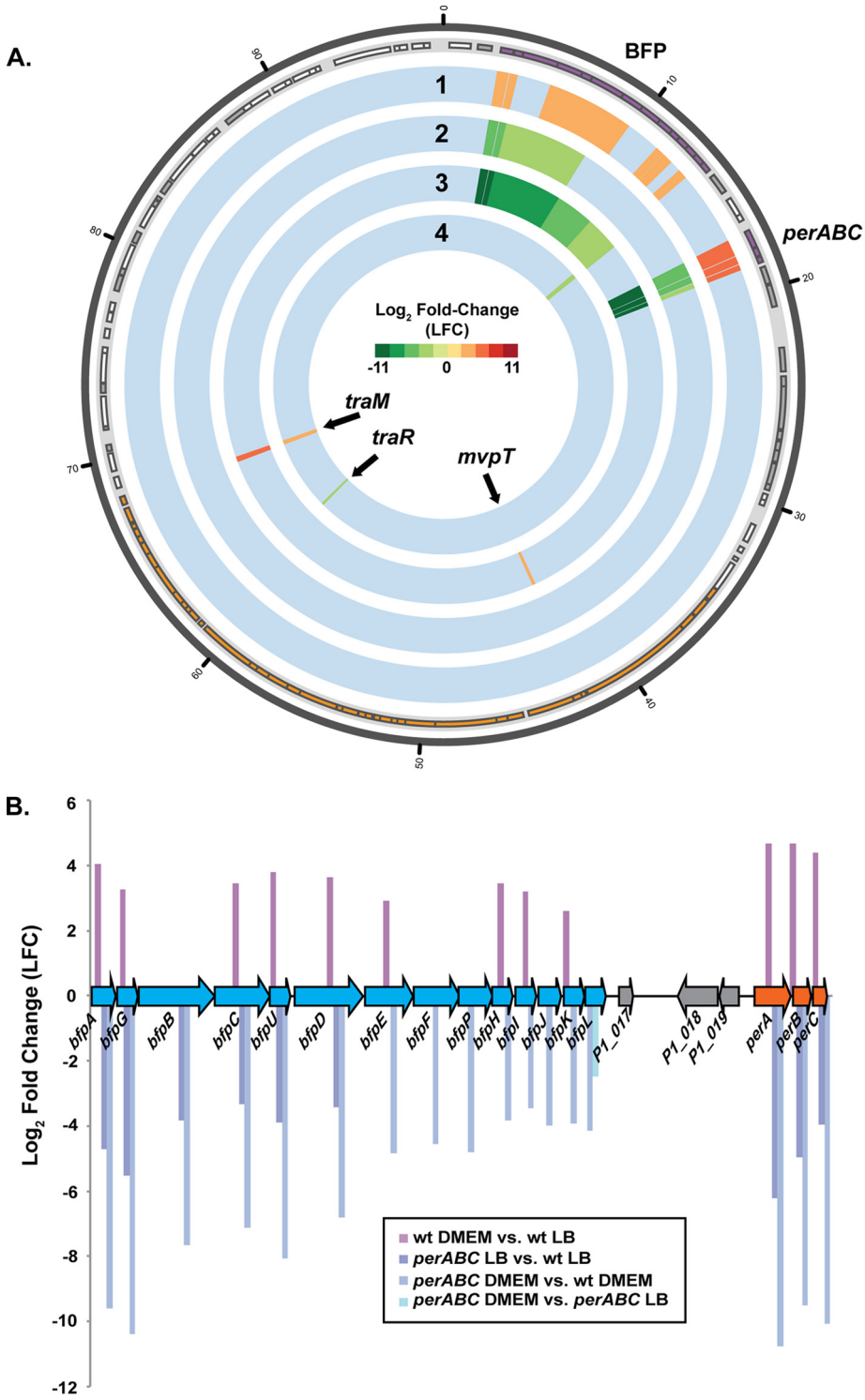

FIG 6 RNA-Seq analysis of the global transcription of the E2348/69 EAF plasmid pMAR2. (A) Circular display of the differential expression analysis of genes on the EAF plasmid pMAR2 (22). RNA-Seq was used to determine changes in the global transcription of wild-type E2348/69, an E2348/69 Δ*perABC* mutant, and JPN15. The color scale indicates $\log_2$ fold change (LFC) from −11 (green) to 11 (red). The outermost track denotes the location of the protein-encoding genes and pseudogenes (22). Genes are indicated in the light gray outer track of the plot as rectangles of different sizes according to their coordinates and are colored purple to indicate BFP-associated genes, orange to indicate the conjugal transfer region, white to indicate protein-encoding genes, and gray to indicate pseudogenes. The tracks of differential expression data are as follows, from the outer track to the innermost track: wild-type E2348/69 in DMEM versus wild-type E2348/69 in LB (track 1), the Δ*perABC* mutant in LB versus wild-type E2348/69 LB (track 2), the Δ*perABC* mutant in DMEM versus wild-type E2348/69 in DMEM (track 3), and the Δ*perABC* mutant in DMEM versus the Δ*perABC* mutant in LB (track 4). (B) Diagram of the pMAR2 BFP operon and the LFC values of each gene in the BFP operon. The approximate size of each gene and the coding strand location are indicated by the size and orientation of the arrows. Genes identified in blue are involved in pilus biogenesis, while orange indicates the plasmid-encoded regulator genes, and gray indicates genes that encode hypothetical proteins.

conjugation of pMAR2 and whether this is dependent on quorum-sensing activity of its own population or in response to the presence of other bacteria.

The second gene outside the BFP operon that had altered expression in the Δ*perABC* mutant compared to wild-type E2348/69 was the gene for the postsegregational killing (PSK) toxin, *mvpT*. MvpT is a toxin that is cotranscribed with its antitoxin protein MvpA; however, it is more stable than the antitoxin and degrades more slowly (71). PSK systems destroy cells that have lost the plasmid encoding the antitoxin, promoting the persistence of the plasmid in a bacterial population (72). This was demonstrated to increase the stability of the *Shigella flexneri* virulence plasmid (73). In the absence of *perABC*, and during growth in DMEM, *mvpT* exhibited increased expression compared to wild-type E2348/69 (Fig. 6A; also, see Data Set S3 in the supplemental material). This finding suggests that *perABC* regulates expression of *mvpT* and may contribute to the stability of pMAR2 under virulence conditions. Thus, in the absence of *perABC*, the increased expression of *mvpT* may be a mechanism to increase stability of the plasmid in order to be maintained by the population. However, further studies are necessary to determine the role of *perABC* genes in regulating *mvpT* and whether this contributes to differences in stability of the EAF plasmid.

To investigate whether any other pEAF genes contribute to the coordinated expression of the E2348/69 virulence regulon, we used RNA-Seq to identify variation in the differentially expressed genes for an isogenic plasmid-free mutant of E2348/69 (JPN15) compared to the wild-type E2348/69 during growth in LB or DMEM. The number of differentially expressed genes during growth in DMEM was greater for JPN15 (170 genes) compared to wild-type E2348/69 than for the Δ*perABC* mutant compared to E2348/69 (90 genes) (Table 2 and Fig. 5A; also, see Data Sets S3 and S8 in the supplemental material), demonstrating that other genes of the pMAR2 plasmid contribute to the virulence regulon of E2348/69. Among the genes that exhibited decreased expression in the JPN15 strain compared to wild-type E2348/69 that were not identified as differentially expressed in the Δ*perABC* mutant compared to the wild type was *chuA*, which was previously described to be involved in iron transport in the presence of heme (74) (see Data Set S8 in the supplemental material). This finding suggests that additional genes of the EAF plasmid, other than *perABC*, play a direct or indirect role in the regulation of these iron transport genes in wild-type E2348/69 grown in DMEM. There were 267 chromosomal genes that had altered expression in both the Δ*perABC* mutant and JPN15 grown in DMEM compared to LB, while they had 321 and 318 differentially expressed genes in common with wild-type E2348/69, respectively (Fig. 5B). There were 188 genes that were differentially expressed when grown in DMEM compared to LB in all three strains (wild-type E2348/69, the Δ*perABC* mutant, and JPN15) (Fig. 5B). These findings taken together suggest that changes to the EAF plasmid, whether deletion of *perABC* or loss of the entire plasmid, influence the dynamics of global transcription of E2348/69 during growth under virulence-promoting conditions.

Instability of the EAF plasmid. The EAF plasmid has previously been described as considerably stable, and in particular, pMAR2 of E2348/69 was stably maintained during growth in the laboratory (20). However, loss of pMAR2 was observed following the passage of E2348/69 through adults in clinical trials (20, 75). There were three LEE$^+$ *bfpA*$^+$ isolates that exhibited plasmid instability under standard laboratory growth conditions, resulting in the isolation of isogenic LEE$^+$, *bfpA*-negative colonies (Fig. 7). The *bfpA*$^+$ and *bfpA*-negative colonies from each of these EPEC isolates had identical ERIC PCR patterns, verifying that they are the same isolate (Fig. 7). The plasmid profiles of the isogenic *bfpA*$^+$ and *bfpA*-negative colonies verified the absence of a large plasmid band (≥60 kb) with a size similar to that of the EAF plasmid pMAR2 (21), from E2348/69 (Fig. 7). These EPEC isolates that exhibited EAF plasmid loss during growth under standard laboratory conditions were all in phylogroup B2 in the phylogenomic analysis (Fig. 1). The genomes of the three EPEC isolates that exhibited plasmid instability (401588, 302312, and 302053) did not contain genes with significant similarity to the PSK toxin and antitoxin genes *mvpT* and *mvpA* of pMAR2. Further research is required to determine whether the stability of the EAF plasmid depends on the functionality of the *mvp* addiction system and whether disruption or reduced expression of these

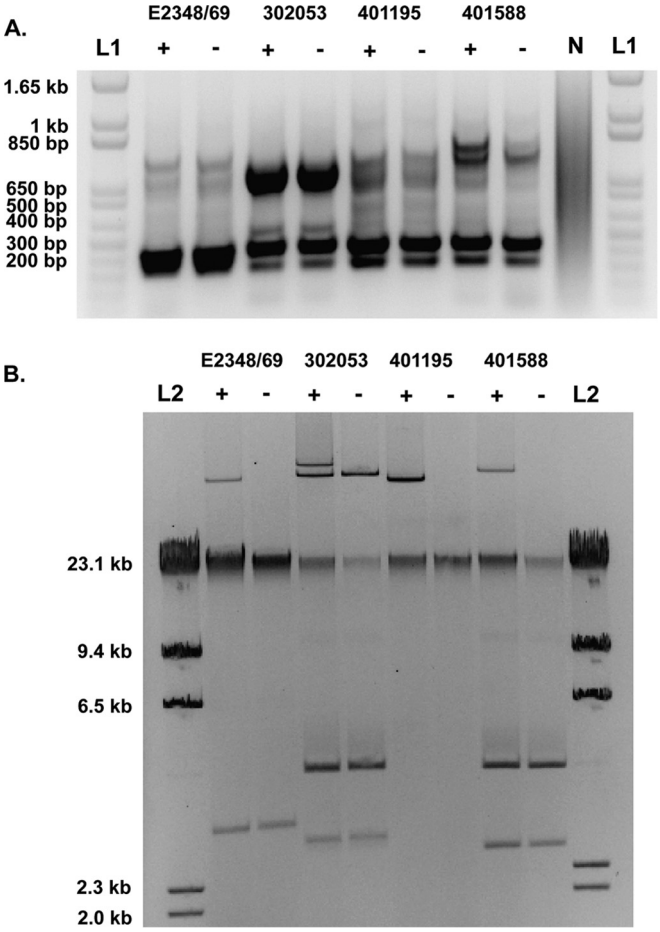

FIG 7 Gel image of the plasmid profiles of EPEC isolates. (A) ERIC PCR profiles of the EPEC prototype isolate E2348/69 and three EPEC isolates (302053, 401195, and 401588). "L1" indicates the 1-kb Plus ladder (Invitrogen). "N" indicates a no-template PCR control. (B) Plasmid profiles of E2348/69, 302053, 401195, and 401588. L2 indicates the lambda HindIII ladder (New England BioLabs), which is a linear DNA ladder that was included for reference comparison to the linearized chromosomal DNA in each plasmid isolation that corresponds to the top band of the ladder. Plus and minus signs indicate the presence and absence of the EAF plasmid in each of the isolates in panels A and B.

genes contributes to the EAF plasmid loss we observed under standard laboratory growth conditions.

Conclusions. In the current study, we used comparative genomics to investigate the genetic diversity of virulence plasmids in a collection of 210 AEEC isolates. While the genetic content of the O157:H7 virulence plasmid was highly conserved, there was greater sequence diversity of plasmid-associated genes identified in the non-O157 EHEC and the EPEC isolates. However, genes with similarity to genes of the EAF plasmids were identified in phylogenomically diverse AEEC isolates that lacked *bfp*, indicating that there are EAF plasmid genes that are conserved with other AEEC plasmids, including some in non-O157 EHEC isolates. Further investigation of the EAF plasmids demonstrated that the BFP operon was likely acquired by multiple ancestral plasmids. Alternately, the EAF plasmids may have undergone genetic modifications, such as rearrangement, loss, or acquisition of insertion elements, or the entire EAF plasmid may have been lost and acquired multiple times by isolates within the same phylogenomic lineage. In addition to the considerable genetic diversity we have described for the EAF plasmids from diverse EPEC phylogenomic lineages, RNA-Seq analysis of E2348/69 and its EAF plasmid-free (JPN15) and Δ*perABC* mutants demonstrated that plasmid genes other than *perABC* influence the expression of both plasmid and chromosomal genes under virulence-inducing conditions. Further investigation of the genetic diversity of the EAF plasmid and the contribution of this diversity to the virulence regulon of EPEC will yield insight into whether there are lineage- or isolate-specific differences that influence the clinical severity of EPEC-associated infections.

ACKNOWLEDGMENTS

This project was funded in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under NIH grant numbers U19 AI090873 and U19 AI110820 and by start-up funds from the State of Maryland.

REFERENCES

1. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol **2:**123–140. http://dx.doi.org/10.1038/nrmicro818.
2. **Pennington H.** 2010. *Escherichia coli* O157. Lancet **376:**1428–1435. http://dx.doi.org/10.1016/S0140-6736(10)60963-4.
3. **Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acacio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM.** 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet **382:**209–222. http://dx.doi.org/10.1016/S0140-6736(13)60844-2.
4. **Ochoa TJ, Contreras CA.** 2011. Enteropathogenic *Escherichia coli* infection in children. Curr Opin Infect Dis **24:**478–483. http://dx.doi.org/10.1097/QCO.0b013e32834a8b8b.
5. **Jerse AE, Yu J, Tall BD, Kaper JB.** 1990. A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. Proc Natl Acad Sci U S A **87:**7839–7843. http://dx.doi.org/10.1073/pnas.87.20.7839.
6. **Elliott SJ, Wainwright LA, McDaniel TK, Jarvis KG, Deng YK, Lai LC, McNamara BP, Donnenberg MS, Kaper JB.** 1998. The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. Mol Microbiol **28:**1–4.
7. **McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc Natl Acad Sci U S A **92:**1664–1668. http://dx.doi.org/10.1073/pnas.92.5.1664.
8. **Donnenberg MS, Giron JA, Nataro JP, Kaper JB.** 1992. A plasmid-encoded type IV fimbrial gene of enteropathogenic *Escherichia coli* associated with localized adherence. Mol Microbiol **6:**3427–3437. http://dx.doi.org/10.1111/j.1365-2958.1992.tb02210.x.
9. **Tobe T, Beatson SA, Taniguchi H, Abe H, Bailey CM, Fivian A, Younis R, Matthews S, Marches O, Frankel G, Hayashi T, Pallen MJ.** 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 U S A **103:**14941–14946. http://dx.doi.org/10.1073/pnas.0604891103.
10. **Burgos Y, Beutin L.** 2010. Common origin of plasmid encoded alpha-hemolysin genes in Escherichia coli. BMC Microbiol **10:**193. http://dx.doi.org/10.1186/1471-2180-10-193.
11. **Welch RA.** 1991. Pore-forming cytolysins of gram-negative bacteria. Mol Microbiol **5:**521–528. http://dx.doi.org/10.1111/j.1365-2958.1991.tb00723.x.
12. **Menestrina G, Moser C, Pellet S, Welch R.** 1994. Pore-formation by *Escherichia coli* hemolysin (HlyA) and other members of the RTX toxins family. Toxicology **87:**249–267. http://dx.doi.org/10.1016/0300-483X(94)90254-2.
13. **Makino K, Ishii K, Yasunaga T, Hattori M, Yokoyama K, Yutsudo CH, Kubota Y, Yamaichi Y, Iida T, Yamamoto K, Honda T, Han CG, Ohtsubo E, Kasamatsu M, Hayashi T, Kuhara S, Shinagawa H.** 1998. Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. DNA Res **5:**1–9. http://dx.doi.org/10.1093/dnares/5.1.1.
14. **Burland V, Shao Y, Perna NT, Plunkett G, Sofia HJ, Blattner FR.** 1998. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. Nucleic Acids Res **26:**4196–4204. http://dx.doi.org/10.1093/nar/26.18.4196.
15. **Sohel I, Puente JL, Ramer SW, Bieber D, Wu CY, Schoolnik GK.** 1996. Enteropathogenic *Escherichia coli*: identification of a gene cluster coding for bundle-forming pilus morphogenesis. J Bacteriol **178:**2613–2628.
16. **Stone KD, Zhang HZ, Carlson LK, Donnenberg MS.** 1996. A cluster of fourteen genes from enteropathogenic *Escherichia coli* is sufficient for the biogenesis of a type IV pilus. Mol Microbiol **20:**325–337. http://dx.doi.org/10.1111/j.1365-2958.1996.tb02620.x.
17. **Baldini MM, Kaper JB, Levine MM, Candy DC, Moon HW.** 1983. Plasmid-mediated adhesion in enteropathogenic *Escherichia coli*. J Ped Gastroent Nutr **2:**534–538. http://dx.doi.org/10.1097/00005176-198302030-00023.
18. **Donnenberg MS, Kaper JB.** 1992. Enteropathogenic *Escherichia coli*. Infect Immun **60:**3953–3961.
19. **Nisa S, Scanlon KM, Donnenberg MS.** 2013. Enteropathogenic *Escherichia coli*, p 75–119. *In* Donnenberg MS (ed), *Escherichia coli* pathotypes and principles of pathogenesis, 2nd ed. Academic Press, New York, NY.
20. **Levine MM, Nataro JP, Karch H, Baldini MM, Kaper JB, Black RE, Clements ML, O'Brien AD.** 1985. The diarrheal response of humans to some classic serotypes of enteropathogenic *Escherichia coli* is dependent on a plasmid encoding an enteroadhesiveness factor. J Infect Dis **152:**550–559. http://dx.doi.org/10.1093/infdis/152.3.550.
21. **Brinkley C, Burland V, Keller R, Rose DJ, Boutin AT, Klink SA, Blattner FR, Kaper JB.** 2006. Nucleotide sequence analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid pMAR7. Infect Immun **74:**5408–5413. http://dx.doi.org/10.1128/IAI.01840-05.
22. **Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D, Asadulghani M, Kurokawa K, Dean P, Kenny B, Quail MA, Thurston S, Dougan G, Hayashi T, Parkhill J, Frankel G.** 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J Bacteriol **191:**347–354. http://dx.doi.org/10.1128/JB.01238-08.
23. **Tobe T, Hayashi T, Han CG, Schoolnik GK, Ohtsubo E, Sasakawa C.** 1999. Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. Infect Immun **67:**5455–5462.
24. **Giron JA, Ho AS, Schoolnik GK.** 1991. An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. Science **254:**710–713. http://dx.doi.org/10.1126/science.1683004.
25. **Nataro JP, Maher KO, Mackie P, Kaper JB.** 1987. Characterization of plasmids encoding the adherence factor of enteropathogenic *Escherichia coli*. Infect Immun **55:**2370–2377.
26. **Okeke IN, Borneman JA, Shin S, Mellies JL, Quinn LE, Kaper JB.** 2001. Comparative sequence analysis of the plasmid-encoded regulator of en-

teropathogenic *Escherichia coli* strains. Infect Immun **69:**5553–5564. http://dx.doi.org/10.1128/IAI.69.9.5553-5564.2001.
27. **Lacher DW, Steinsland H, Blank TE, Donnenberg MS, Whittam TS.** 2007. Molecular evolution of typical enteropathogenic *Escherichia coli*: clonal analysis by multilocus sequence typing and virulence gene allelic profiling. J Bacteriol **189:**342–350. http://dx.doi.org/10.1128/JB.01472-06.
28. **Blank TE, Zhong H, Bell AL, Whittam TS, Donnenberg MS.** 2000. Molecular variation among type IV pilin (*bfpA*) genes from diverse enteropathogenic *Escherichia coli* strains. Infect Immun **68:**7028–7038. http://dx.doi.org/10.1128/IAI.68.12.7028-7038.2000.
29. **Contreras CA, Ochoa TJ, Lacher DW, DebRoy C, Navarro A, Talledo M, Donnenberg MS, Ecker L, Gil AI, Lanata CF, Cleary TG.** 2010. Allelic variability of critical virulence genes (*eae*, *bfpA* and *perA*) in typical and atypical enteropathogenic *Escherichia coli* in Peruvian children. J Med Microbiol **59:**25–31. http://dx.doi.org/10.1099/jmm.0.013706-0.
30. **Iida M, Okamura N, Yamazaki M, Yatsuyanagi J, Kurazono T, Suzuki R, Hiruta N, Isobe J, Seto K, Kawano K, Narimatsu H, Ratchtrachenchai OA, Okabe N, Ito K.** 2010. Classification of *perA* sequences and their correlation with autoaggregation in typical enteropathogenic *Escherichia coli* isolates collected in Japan and Thailand. Microbiol Immunol **54:**184–195. http://dx.doi.org/10.1111/j.1348-0421.2010.00212.x.
31. **Hazen TH, Sahl JW, Fraser CM, Donnenberg MS, Scheutz F, Rasko DA.** 2013. Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. Proc Natl Acad Sci U S A **110:**12810–12815. http://dx.doi.org/10.1073/pnas.1306836110.
32. **Leverton LQ.** 2004. An examination of enteropathogenic *Escherichia coli* virulence gene regulation. Ph.D. thesis. University of Maryland, Baltimore, MD.
33. **Sahl JW, Morris CR, Emberger J, Fraser CM, Ochieng JB, Juma J, Fields B, Breiman RF, Gilmour M, Nataro JP, Rasko DA.** 2015. Defining the phylogenomics of *Shigella* species: a pathway to diagnostics. J Clin Microbiol **53:**951–960. http://dx.doi.org/10.1128/JCM.03527-14.
34. **Sahl JW, Beckstrom-Sternberg SM, Babic-Sternberg JS, Gillece JD, Hepp CM, Auerbach RK, Tembe W, Wagner DM, Keim PS, Pearson T.** 2015. The *in silico* genotyper (ISG): an open-source pipeline to rapidly identify and annotate nucleotide variants for comparative genomics applications. bioRxiv.
35. **Delcher AL, Salzberg SL, Phillippy AM.** 2003. Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinformatics Chapter 10:Unit 10.13, http://dx.doi.org/10.1002/0471250953.bi1003s00.
36. **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22:**2688–2690. http://dx.doi.org/10.1093/bioinformatics/btl446.
37. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
38. **Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:**4673–4680. http://dx.doi.org/10.1093/nar/22.22.4673.
39. **Kimura M.** 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol **16:**111–120. http://dx.doi.org/10.1007/BF01731581.
40. **Rasko DA, Myers GS, Ravel J.** 2005. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics **6:**2. http://dx.doi.org/10.1186/1471-2105-6-2.
41. **Gertz EM, Yu YK, Agarwala R, Schaffer AA, Altschul SF.** 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol **4:**41. http://dx.doi.org/10.1186/1741-7007-4-41.
42. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
43. **Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA.** 2009. Circos: an information aesthetic for comparative genomics. Genome Res **19:**1639–1645. http://dx.doi.org/10.1101/gr.092759.109.
44. **Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ.** 2005. Identification of plasmids by PCR-based replicon typing. J Microbiol Methods **63:**219–228. http://dx.doi.org/10.1016/j.mimet.2005.03.018.
45. **Johnson TJ, Wannemuehler YM, Johnson SJ, Logue CM, White DG, Doetkott C, Nolan LK.** 2007. Plasmid replicon typing of commensal and pathogenic *Escherichia coli* isolates. Appl Environ Microbiol **73:**1976–1983. http://dx.doi.org/10.1128/AEM.02171-06.
46. **Sobecky PA, Mincer TJ, Chang MC, Helinski DR.** 1997. Plasmids isolated from marine sediment microbial communities contain replication and incompatibility regions unrelated to those of known plasmid groups. Appl Environ Microbiol **63:**888–895.
47. **Kieser T.** 1984. Factors affecting the isolation of CCC DNA from *Streptomyces lividans* and *Escherichia coli*. Plasmid **12:**19–36. http://dx.doi.org/10.1016/0147-619X(84)90063-5.
48. **Johnson JR, Clabots C.** 2000. Improved repetitive-element PCR fingerprinting of *Salmonella enterica* with the use of extremely elevated annealing temperatures. Clin Diagn Lab Immunol **7:**258–264.
49. **Versalovic J, Koeuth T, Lupski JR.** 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res **19:**6823–6831. http://dx.doi.org/10.1093/nar/19.24.6823.
50. **Hazen TH, Daugherty SC, Shetty A, Mahurkar AA, White O, Kaper JB, Rasko DA.** 2015. RNA-Seq analysis of isolate- and growth phase-specific differences in the global transcriptomes of enteropathogenic *Escherichia coli* prototype isolates. Front Microbiol **6:**569. http://dx.doi.org/10.3389/fmicb.2015.00569.
51. **Anders S, Huber W.** 2010. Differential expression analysis for sequence count data. Genome Biol **11:**R106. http://dx.doi.org/10.1186/gb-2010-11-10-r106.
52. **Sahl JW, Rasko DA.** 2012. Analysis of global transcriptional profiles of enterotoxigenic *Escherichia coli* isolate E24377A. Infect Immun **80:**1232–1242. http://dx.doi.org/10.1128/IAI.06138-11.
53. **Tenaillon O, Skurnik D, Picard B, Denamur E.** 2010. The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol **8:**207–217. http://dx.doi.org/10.1038/nrmicro2298.
54. **Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbonnelle E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S.** 2008. Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. BMC Genomics **9:**560. http://dx.doi.org/10.1186/1471-2164-9-560.
55. **Hazen TH, Sahl JW, Redman JC, Morris CR, Daugherty SC, Chibucos MC, Sengamalay NA, Fraser-Liggett CM, Steinsland H, Whittam TS, Whittam B, Manning SD, Rasko DA.** 2012. Draft genome sequences of the diarrheagenic *Escherichia coli* collection. J Bacteriol **194:**3026–3027. http://dx.doi.org/10.1128/JB.00426-12.
56. **Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, Farag TH, van Eijk A, Adegbola RA, Alonso PL, Breiman RF, Golam Faruque AS, Saha D, Sow SO, Sur D, Zaidi AK, Biswas K, Panchalingam S, Clemens JD, Cohen D, Glass RI, Mintz ED, Sommerfelt H, Levine MM.** 2012. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. Clin Infect Dis **55**(Suppl 4)**:**S232–S245. http://dx.doi.org/10.1093/cid/cis753.
57. **Johnson TJ, Nolan LK.** 2009. Pathogenomics of the virulence plasmids of *Escherichia coli*. Microbiol Mol Biol Rev **73:**750–774. http://dx.doi.org/10.1128/MMBR.00015-09.
58. **Scaletsky IC, Souza TB, Aranda KR, Okeke IN.** 2010. Genetic elements associated with antimicrobial resistance in enteropathogenic *Escherichia coli* (EPEC) from Brazil. BMC Microbiol **10:**25. http://dx.doi.org/10.1186/1471-2180-10-25.
59. **Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T.** 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc Natl Acad Sci U S A **106:**17939–17944. http://dx.doi.org/10.1073/pnas.0903585106.
60. **Bortolini MR, Trabulsi LR, Keller R, Frankel G, Sperandio V.** 1999. Lack of expression of bundle-forming pili in some clinical isolates of enteropathogenic *Escherichia coli* (EPEC) is due to a conserved large deletion in the *bfp* operon. FEMS Microbiol Lett **179:**169–174. http://dx.doi.org/10.1111/j.1574-6968.1999.tb08723.x.
61. **Leverton LQ, Kaper JB.** 2005. Temporal expression of enteropathogenic *Escherichia coli* virulence genes in an *in vitro* model of infection. Infect Immun **73:**1034–1043. http://dx.doi.org/10.1128/IAI.73.2.1034-1043.2005.
62. **Sperandio V, Mellies JL, Nguyen W, Shin S, Kaper JB.** 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 U S A **96:**15196–15201. http://dx.doi.org/10.1073/pnas.96.26.15196.
63. **Raetz CR, Whitfield C.** 2002. Lipopolysaccharide endotoxins. Annu Rev Biochem **71:**635–700. http://dx.doi.org/10.1146/annurev.biochem.71.110601.135414.
64. **Tobe T, Schoolnik GK, Sohel I, Bustamante VH, Puente JL.** 1996. Cloning and characterization of *bfpTVW*, genes required for the transcriptional activation of *bfpA* in enteropathogenic *Escherichia coli*. Mol Microbiol **21:**963–975. http://dx.doi.org/10.1046/j.1365-2958.1996.531415.x.
65. **Gomez-Duarte OG, Kaper JB.** 1995. A plasmid-encoded regulatory region activates chromosomal *eaeA* expression in enteropathogenic *Escherichia coli*. Infect Immun **63:**1767–1776.
66. **Swiderska A, Berndtson AK, Cha MR, Li L, Beaudoin GM III, Zhu J, Fuqua C.** 2001. Inhibition of the *Agrobacterium tumefaciens* TraR quorum-sensing regulator; interactions with the TraM anti-activator. J Biol Chem **276:**49449–49458. http://dx.doi.org/10.1074/jbc.M107881200.
67. **Fuqua C, Burbea M, Winans SC.** 1995. Activity of the *Agrobacterium* Ti plasmid conjugal transfer regulator TraR is inhibited by the product of the *traM* gene. J Bacteriol **177:**1367–1373.
68. **Hwang I, Cook DM, Farrand SK.** 1995. A new regulatory element modulates homoserine lactone-mediated autoinduction of Ti plasmid conjugal transfer. J Bacteriol **177:**449–458.
69. **He X, Chang W, Pierce DL, Seib LO, Wagner J, Fuqua C.** 2003. Quorum sensing in *Rhizobium* sp. strain NGR234 regulates conjugal transfer (*tra*) gene expression and influences growth rate. J Bacteriol **185:**809–822. http://dx.doi.org/10.1128/JB.185.3.809-822.2003.
70. **Mellies JL, Elliott SJ, Sperandio V, Donnenberg MS, Kaper JB.** 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. http://dx.doi.org/10.1046/j.1365-2958.1999.01473.x.
71. **Sayeed S, Reaves L, Radnedge L, Austin S.** 2000. The stability region of the large virulence plasmid of *Shigella flexneri* encodes an efficient postsegregational killing system. J Bacteriol **182:**2416–2421. http://dx.doi.org/10.1128/JB.182.9.2416-2421.2000.
72. **Hayes F.** 2003. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science **301:**1496–1499. http://dx.doi.org/10.1126/science.1088157.
73. **Sayeed S, Brendler T, Davis M, Reaves L, Austin S.** 2005. Surprising dependence on postsegregational killing of host cells for maintenance of the large virulence plasmid of *Shigella flexneri*. J Bacteriol **187:**2768–2773. http://dx.doi.org/10.1128/JB.187.8.2768-2773.2005.
74. **Torres AG, Payne SM.** 1997. Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7. Mol Microbiol **23:**825–833. http://dx.doi.org/10.1046/j.1365-2958.1997.2641628.x.
75. **Donnenberg MS, Tacket CO, James SP, Losonsky G, Nataro JP, Wasserman SS, Kaper JB, Levine MM.** 1993. Role of the *eaeA* gene in experimental enteropathogenic *Escherichia coli* infection. J Clin Invest **92:**1412–1417. http://dx.doi.org/10.1172/JCI116717.
76. **Eisen MB, Spellman PT, Brown PO, Botstein D.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A **95:**14863–14868. http://dx.doi.org/10.1073/pnas.95.25.14863.
77. **Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J.** 2006. TM4 microarray software suite. Methods Enzymol **411:**134–193. http://dx.doi.org/10.1016/S0076-6879(06)11009-5.
78. **Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J.** 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques **34:**374–378.