

# RNA-Seq analysis of isolate- and growth phase-specific differences in the global transcriptomes of enteropathogenic *Escherichia coli* prototype isolates

## OPEN ACCESS

### Edited by:

Dongsheng Zhou,  
Beijing Institute of Microbiology and  
Epidemiology, China

### Reviewed by:

Learn-Han Lee,  
Monash University Malaysia, Malaysia  
Dmitriy Ignatov,  
Shemyakin and Ovchinnikov Institute  
of Bioorganic Chemistry RAS, Russia

### \*Correspondence:

David A. Rasko,  
Department of Microbiology and  
Immunology, Institute for Genome  
Sciences, University of Maryland  
School of Medicine, 801 W. Baltimore  
Street, Suite 619, Baltimore, MD  
21201, USA  
drasko@som.umaryland.edu

### Specialty section:

This article was submitted to  
Food Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 25 March 2015

**Accepted:** 24 May 2015

**Published:** 12 June 2015

### Citation:

Hazen TH, Daugherty SC, Shetty A,  
Mahurkar AA, White O, Kaper JB and  
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.  
doi: 10.3389/fmicb.2015.00569

Tracy H. Hazen<sup>1,2</sup>, Sean C. Daugherty<sup>1</sup>, Amol Shetty<sup>1</sup>, Anup A. Mahurkar<sup>1</sup>,  
Owen White<sup>1</sup>, James B. Kaper<sup>2</sup> and David A. Rasko<sup>1,2\*</sup>

<sup>1</sup> Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA, <sup>2</sup> Department of  
Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA

Enteropathogenic *Escherichia coli* (EPEC) are a leading cause of diarrheal illness among infants in developing countries. *E. coli* isolates classified as typical EPEC are identified by the presence of the locus of enterocyte effacement (LEE) and the bundle-forming pilus (BFP), and absence of the Shiga-toxin genes, while the atypical EPEC also encode LEE but do not encode BFP or Shiga-toxin. Comparative genomic analyses have demonstrated that EPEC isolates belong to diverse evolutionary lineages and possess lineage- and isolate-specific genomic content. To investigate whether this genomic diversity results in significant differences in global gene expression, we used an RNA sequencing (RNA-Seq) approach to characterize the global transcriptomes of the prototype typical EPEC isolates E2348/69, B171, C581-05, and the prototype atypical EPEC isolate E110019. The global transcriptomes were characterized during laboratory growth in two different media and three different growth phases, as well as during adherence of the EPEC isolates to human cells using *in vitro* tissue culture assays. Comparison of the global transcriptomes during these conditions was used to identify isolate- and growth phase-specific differences in EPEC gene expression. These analyses resulted in the identification of genes that encode proteins involved in survival and metabolism that were coordinately expressed with virulence factors. These findings demonstrate there are isolate- and growth phase-specific differences in the global transcriptomes of EPEC prototype isolates, and highlight the utility of comparative transcriptomics for identifying additional factors that are directly or indirectly involved in EPEC pathogenesis.

**Keywords:** enteropathogenic *Escherichia coli*, RNA-sequencing, diversity, regulation, pathogenesis

## Introduction

Enteropathogenic *Escherichia coli* (EPEC) have been associated with moderate to severe cases of diarrhea and are a leading cause of lethal diarrheal illness among young children in developing countries (Ochoa and Contreras, 2011; Kotloff et al., 2013). EPEC are identified by the presence of the locus of enterocyte effacement (LEE), which encodes a type III secretion system (T3SS) and the intimin adherence factor, which are involved in translocation of bacterial factors into host cells, and adherence to the surface of host cells. EPEC are also characterized by the absence of the Shiga-toxin genes, which is typically present in the enterohemorrhagic *E. coli* (EHEC) (Nataro and Kaper, 1998; Kaper et al., 2004). EPEC are further identified as typical EPEC by the presence of genes encoding the bundle-forming pilus (BFP), which is a type IV pilus typically carried by the EPEC adherence factor (EAF) plasmid. Meanwhile *E. coli* isolates that contain the LEE and do not carry the Shiga-toxin phage or the BFP genes are considered atypical EPEC (Nataro and Kaper, 1998). However, we have previously demonstrated that the atypical EPEC includes isolates with genomic similarity to typical EPEC and EHEC (Hazen et al., 2013).

The major components of EPEC pathogenesis that have been characterized to date are the T3SS encoded by the LEE region, additional T3SS effectors that are encoded by insertion elements or phages inserted within the genome, and the BFP (McDaniel et al., 1995; Nataro and Kaper, 1998; Kaper et al., 2004; Mellies et al., 2007; Nisa et al., 2013). Intimin, encoded by the *eae* gene, of the LEE region confers intimate attachment to host cells while the T3SS translocates effector proteins across the host cell membrane that result in the formation of the attaching and effacing lesions (Jerse et al., 1990; Donnenberg et al., 1993; McDaniel et al., 1995; McDaniel and Kaper, 1997; Garmendia et al., 2005). Once inside the host cell, the type III secreted effectors induce changes including rearrangement of the actin cytoskeleton to form pedestals that further facilitate adherence of EPEC to host cells (Sperandio et al., 1999; Wong et al., 2011; Clements et al., 2012; Nisa et al., 2013). The transcriptional regulation of the LEE and BFP regions have been extensively studied (Gomez-Duarte and Kaper, 1995; Puente et al., 1996; Mellies et al., 1999, 2007; Sperandio et al., 1999, 2000; Elliott et al., 2000; Bustamante et al., 2001; Shin et al., 2001; Haack et al., 2003; Porter et al., 2004; Iyoda and Watanabe, 2005; Leverton and Kaper, 2005; Kendall et al., 2011). Regulation of EPEC virulence genes has been demonstrated to involve numerous transcriptional factors and to be influenced by environmental conditions and cell density (Puente et al., 1996; Mellies et al., 1999; Sperandio et al., 1999; Elliott et al., 2000; Shin et al., 2001; Deng et al., 2004; Kaper et al., 2004; Garmendia et al., 2005; Bhatt et al., 2011; Wong et al., 2011).

Investigations of EPEC pathogenesis and virulence factor regulation have primarily used a select few prototype isolates (E2348/69, B171, E22, E110019) (Viljanen et al., 1990; Rasko et al., 2008; Iguchi et al., 2009). Each of the previously used prototype isolates was examined in detail for specific reasons; E2348/69 is an EPEC1 isolate as defined by multi-locus sequence typing (MLST) and has been used in human trials (Levine et al.,

1978, 1985; Levine and Rennels, 1978), B171 is an EPEC2 isolate as defined by MLST and has been characterized for plasmid content and infection studies (Riley et al., 1987, 1990; Tobe et al., 1999), E22 is a rabbit adapted EPEC strain that has significant utility in the infection of animal models (Cheney et al., 1980) and E110019 is the prototype isolate of atypical EPEC that had a significantly different clinical presentation in a large Finnish outbreak that appeared to spread within households (Viljanen et al., 1990). The characterization of EPEC virulence-associated genes has also typically involved investigating a limited number of genes at a time, which are usually within virulence gene regions, or are known global regulators in other bacteria (Sperandio et al., 1999, 2000; Bustamante et al., 2001; Haack et al., 2003; Iyoda and Watanabe, 2005; Kendall et al., 2011). In contrast, global transcriptional analysis of bacterial pathogens has provided additional insight into genome-wide transcription during conditions that promote pathogenesis, as well as the identification of novel virulence associated factors (Bergholz et al., 2007; Camarena et al., 2010; Mandlik et al., 2011; Chugani et al., 2012; Sahl and Rasko, 2012). Microarrays have previously been used to identify global gene expression of EHEC in response to nutrient limitation, exposure to host cells, and during multiple growth phases (Bergholz et al., 2007; Jandu et al., 2009; Abu-Ali et al., 2010; Bingle et al., 2014). The regulons of the LEE-encoded regulator (Ler) in EHEC and EPEC prototype isolates was also recently described using microarray analyses (Bingle et al., 2014). Overall, these studies have not provided a global view of the transcription in EPEC.

A new method of investigating global transcriptomes is RNA sequencing (RNA-Seq), which is an unbiased high-throughput sequencing approach used to capture the global transcriptional response of an organism during particular conditions (Mortazavi et al., 2008; Wang et al., 2009; Martin and Wang, 2011; Oszolak and Milos, 2011). This method allows the simultaneous analysis of all regions of the genome, unlike methods such as microarray or quantitative reverse-transcription PCR (qRT-PCR), which are limited to analyzing known genomic regions as targets. Also, RNA-Seq can be used to analyze isolates that have diverse or unknown genomic content, unlike microarray analysis, which requires that samples exhibit sequence similarity to known targets that were used to develop the probes of the microarray. To date, RNA-Seq has been used to characterize the global transcriptomes of multiple human disease-associated bacteria (Perkins et al., 2009; Camarena et al., 2010; Sharma et al., 2010; Mandlik et al., 2011; Chugani et al., 2012), including enterohemorrhagic *E. coli* (EHEC) (Landstorfer et al., 2014) and enterotoxigenic *E. coli* (ETEC) (Sahl and Rasko, 2012). These studies used RNA-Seq to demonstrate that distinct isolates of *Pseudomonas aeruginosa* have differences in their quorum sensing regulons (Chugani et al., 2012), or to identify genes that are likely acting as global regulators in ETEC (Sahl and Rasko, 2012). However, to our knowledge there have been no prior studies that have used RNA-Seq to describe the global transcriptional response of EPEC. Furthermore, it is not known how much variability exists among the global virulence regulons of EPEC that have emerged in divergent phylogenomic lineages of *E. coli*, each with unique genomic content.

To investigate whether the genomic diversity of four EPEC prototype isolates results in unique transcriptional profiles, an RNA-Seq approach was used to identify the global transcriptomes of these isolates during standard laboratory growth conditions that have been used to study EPEC pathogenesis (Kenny et al., 1997; Sperandio et al., 1999; Leverton and Kaper, 2005). Using a combination of RNA-Seq analysis, comparative transcriptomics and comparative genomics, the current studies demonstrate there are isolate-specific transcriptional responses for different growth conditions including three growth phases and two media types. Differential expression analysis RNA-Seq data from conditions that promote virulence factor expression compared to conditions during which virulence factors are not typically expressed was used to identify all genes that are simultaneously expressed in association with virulence factors. Many of these additional genes encode proteins involved in central metabolism or survival that may indirectly contribute to EPEC pathogenesis. Additional studies in this manuscript examine the bacterial interaction with host cells and identified a transcriptional pattern that is unique to each of the prototype isolates, as well as a core transcriptome under this condition. These studies highlight the diversity in the transcriptional program of these prototype isolates and provide a glimpse into the core transcriptome, which may provide information on genes and gene products that can be targeted as therapeutics and diagnostics for this group of pathogens. The findings of these studies demonstrate the utility of comparative transcriptomics for investigating EPEC. However, a similar approach could be undertaken for any genomically diverse pathogen or species and provide vital information into the transcriptional patterns of pathogens under any number of growth conditions.

## Materials and Methods

### Bacterial Isolates and Media

The bacterial isolates examined in this study were previously characterized and their genome sequences are publicly available (Rasko et al., 2008; Iguchi et al., 2009; Hazen et al., 2013). The genomes were completely sequenced (E2348/69, NC\_011601.1-NC\_0116013.1) (Iguchi et al., 2009) or sequenced to high quality draft genomes (B171, AAJX00000000; C581-05, AIBE00000000; E110019, AAJW00000000) (Rasko et al., 2008; Hazen et al., 2013). The EPEC isolates were grown in Luria-Bertani (LB) broth medium (Difco) or in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 4.5 g/L of glucose (Gibco).

### Large Scale-BLAST Score Ratio (LS-BSR)

The four EPEC isolate (E2348/69, B171, C581-05, and E110019) genomes were compared with LS-BSR as previously described (Hazen et al., 2013; Sahl et al., 2014) using TBLASTN (Gertz et al., 2006). The predicted protein-encoding genes of each genome that had  $\geq 90\%$  nucleotide identity to each other were assigned to gene clusters using UCLUST (Edgar, 2010). Representative sequences of each gene cluster were then compared to each genome using TBLASTN (Gertz et al., 2006), and the TBLASTN scores were used to generate a BSR value indicating the detection

of each gene cluster in each of the eight genomes. The BSR value was generated by dividing the score of a gene compared to a genome by the score of the gene compared to its own sequence. The representative nucleotide sequence of each gene cluster is provided in Supplemental Data Set 2.

### RNA Isolation and Sequencing

The EPEC isolates were grown overnight in LB and were inoculated 1:100 into 50 ml of LB, or DMEM supplemented with 4.5 g/L glucose in a 250 ml flask. The flasks were grown at 37°C with shaking (225 rpm) to a final optical density of approximately 0.2, 0.5, or 1.0. Two biological replicates were generated for each EPEC isolate and media type. The cells were concentrated by centrifugation at 3500 rpm for 5 min. Total RNA was isolated using the Ribopure bacteria kit (Ambion) and treated with the Ribopure DNase I to remove contaminating DNA. The samples were then treated with the Turbo DNA-free kit (Ambion) to ensure all contaminating DNA was removed, and the RNA was verified to be DNA free by quantitative PCR (qPCR) analysis using primers for *rpoA*, which encodes the alpha subunit of the RNA polymerase using primers listed in Supplemental Table 1.

RNA-Seq of EPEC cells that have adhered to HeLa was performed as follows. HeLa cells were grown to a 90% confluence in a 6-well culture dish in DMEM supplemented with 4.5 g/L of glucose and 10% fetal bovine serum (FBS). The HeLa cells were washed with fresh DMEM supplemented with 4.5 g/L of glucose lacking FBS, and then 2 ml of the media was added before inoculating overnight cultures of each EPEC isolate 1:100 into the fresh media containing the HeLa cells. The EPEC isolates were allowed to adhere to the HeLa cells for 3 h at 37°C with no shaking. The media containing non-adhered EPEC cells was removed and the HeLa and adhered EPEC cells were rinsed twice with fresh media, then all cells (HeLa and EPEC) were removed by scraping and were resuspended in 1 ml of fresh DMEM. The cells were concentrated via centrifugation at 3500 rpm for 1 min, and total RNA was isolated from both the HeLa and EPEC cells using the MasterPure kit (Epicenter). Contaminating DNA was removed using the Turbo DNA-free kit (Ambion). The polyadenylated HeLa mRNA was depleted from the samples using the GenElute mRNA Miniprep kit (Sigma), which binds poly(A)<sup>+</sup> labeled mRNA to a spin column. The poly(A)<sup>+</sup> depleted fraction as well as the eluted poly(A)<sup>+</sup>-enriched fraction were both saved at -80°C for further processing. The poly(A)<sup>+</sup>-depleted fraction was treated again for contaminating DNA using the Turbo DNA-free kit, and was verified to be DNA-free using qPCR as described above. This DNA-free poly(A)<sup>+</sup>-depleted fraction was then submitted for library construction and 100 bp paired-end Illumina sequencing. All of the RNA samples were converted into cDNA libraries using the Ovation Prokaryotic RNA-Seq System (NuGen) that were sequenced on the Illumina HiSeq 2000 to generate 100 bp paired-end reads at the Institute for Genome Sciences Genome Resource Center. The Illumina sequencing reads generated for each RNA sample have been deposited in the short reads archive (SRA), and the accession numbers are listed in Supplemental Table 2.

## RNA-Seq Analyses

The Illumina reads generated for each RNA sample were analyzed and compared using an Ergatis-based (Orvis et al., 2010) RNA-Seq analysis pipeline. The completed genome and annotation that is publicly available for EPEC isolate E2348/69 was used for the RNA-Seq analysis of this isolate. The contigs of the draft genome assemblies of B171, C581-05, and E110019 were ordered relative to the completed genome and plasmids of E2348/69 (Iguchi et al., 2009). The contigs of each genome were then combined into several large pseudomolecule sequences, and the protein-encoding regions and other features were predicted using Prodigal (Hyatt et al., 2010) using an in-house annotation pipeline (Galens et al., 2011). The reads for each isolate were aligned to their corresponding genome using Bowtie (Langmead et al., 2009) and the number of reads that aligned to the predicted coding regions were determined using HTSeq (Anders et al., 2014). The differential expression of each gene under two different conditions was determined using DESeq v. 1.5.24 (Anders and Huber, 2010). The reads aligned to each gene were normalized then averaged for each of the two biological replicates. The fold-change and the  $\log_2$  of the fold-change (LFC) were calculated for each of the comparisons. The expression data were then filtered for further analysis and genes that were determined to be transcriptionally altered met the following criteria:  $LFC \geq 2$ ,  $\leq -2$ , minimum normalized read count = 10, false discovery rate (FDR)  $\leq 0.05$  that was determined using DESeq v. 1.5.24 and R-2.15.2. The circular displays of the genes that exhibited significant differential expression were generated using Circos 0.65 (Krzywinski et al., 2009). Heatmaps of the significant LFC values for the LEE and BFP genes were constructed using MeV (Saeed et al., 2006).

The genes in these four isolates are represented by 7493 LS-BSR gene clusters, which were then utilized to examine the expression values for each of these LS-BSR clusters that were conserved among all of the isolates included in this analysis. The analysis was performed using in-house Perl scripts and heatmaps were generated using R statistical package v2.15.2 that in turn utilized the DESeq v1.10.1 library for normalization and the gplots v2.11.0 library for generating the heat maps. The expression values were normalized using the DESeq method (Anders and Huber, 2010). Only 3302 conserved clusters (i.e., represented in all strains) were used to compute the eigenvectors using principal component analysis methods. The first and second principal components were utilized in a scatter plot to visualize the clustering of the strains by gene content and gene expression. The normalized gene expression values were also used to compute the standard deviation for each LS-BSR cluster across all samples and 2000 out of 3302 LS-BSR clusters showing the greatest standard deviations of expression values were utilized to generate a heatmap of the samples using the LS-BSR cluster expressions.

## Quantitative Reverse Transcriptase PCR (qRT-PCR)

The differences in expression during growth in DMEM compared to LB were determined for genes listed in Supplemental Table 1 using qRT-PCR. RNA that was used for RNA-Seq was reverse

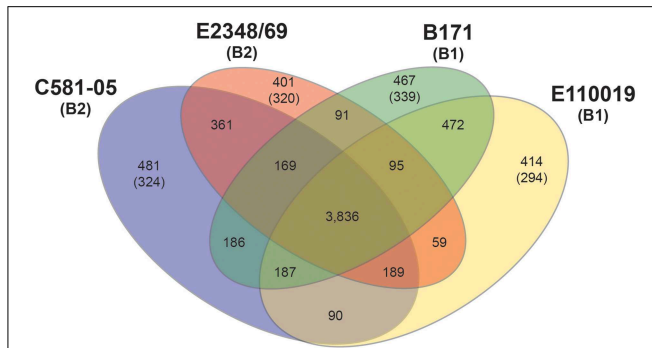
transcribed and primed with random hexamers to generate cDNA using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). The cDNA was diluted 1:20 into nuclease free water (Ambion) before analysis using qPCR. The qPCR on the reverse transcribed RNA samples was performed using SYBR Green master mix (Life Technologies) with 10  $\mu$ l reactions comprised of the following: 5  $\mu$ l of 2X SYBR master mix, 1  $\mu$ l of each of the 5  $\mu$ M forward and reverse primers (Supplemental Table 1), 1  $\mu$ l of nuclease free water (Ambion), and 2  $\mu$ l of cDNA diluted 1:20. Triplicate reactions were performed for each cDNA template and primer combination. The reactions were cycled in a 384-well plate on the 7900HT Fast Real-Time PCR System (Applied Biosystems) using a two-step reaction with an initial incubation of 50°C for 2 min, 95°C for 10 min, then 40 cycles of 95°C for 15 s and 60°C for 1 min, followed by a dissociation stage. The cycle threshold (Ct) values were calculated using the Applied Biosystems software. The Ct values of the biological replicates were averaged and the standard deviation was calculated. The Ct values of target genes of each sample were normalized by subtracting from it the Ct value of the constitutively-expressed RNA polymerase alpha subunit, *rpoA*, resulting in the  $\Delta$ Ct value of a particular gene for each sample. The difference in expression of a target gene ( $\Delta\Delta$ Ct) in the DMEM compared to the LB samples was then calculated by subtracting the  $\Delta$ Ct of the LB sample from the  $\Delta$ Ct of the DMEM sample. The fold difference of the expression of a particular gene in DMEM compared to LB was determined by calculating  $2^{-\Delta\Delta Ct}$ . The difference in expression is represented in the figures as the  $\log_2$  of the fold-difference ( $2^{-\Delta\Delta Ct}$ ) for each gene in DMEM compared to LB. The error bars indicate the standard deviation of the  $\Delta\Delta$ Ct values.

## Results and Discussion

### Comparative Genomics of the Four EPEC Prototype Isolates

The EPEC prototype isolates selected for global transcriptional analysis belong to four different EPEC phylogenomic lineages as defined in our recent study as well as being from the *E. coli* phylogroups B1 and B2 (Hazen et al., 2013). Three of these isolates (E2348/69, B171, E110019) have been frequently studied as EPEC prototype isolates in research investigating the virulence mechanisms of typical and atypical EPEC (Viljanen et al., 1990; Rasko et al., 2008; Iguchi et al., 2009). The fourth isolate, C581-05, was selected as a representative of the EPEC4 phylogenomic lineage, which is also in phylogroup B2 (Hazen et al., 2013).

Comparison of the genomic content of the four EPEC prototype isolates (E2348/69, B171, C581-05, E110019) using large-scale BLAST score ratio (LS-BSR) demonstrated there are 3836 genes that are present in all four isolates with significant similarity (LS-BSR  $\geq 0.8$ ) (Figure 1). There were a greater number of genes shared between EPEC prototype isolates of the same phylogroup (C581-05 and E2348/69 in phylogroup B2, vs. B171 and E110019 in phylogroup B1) than for isolates of different phylogroups (for example E2348/69 of phylogroup B2, and B171 of phylogroup B1) (Figure 1). In each isolate there was genome content that was not exclusive, but divergent in one or more of



**FIGURE 1 | Comparison of the genomic content of four EPEC prototype isolates.** Venn diagram illustrating the number of shared and unique genes identified among the four EPEC prototype isolates using LS-BSR (Sahl et al., 2014). The *E. coli* phylogroup (Tenaillon et al., 2010; Hazen et al., 2013) of each of the prototype isolates is indicated in parentheses below the isolate name. The number of core genes is indicated, which are genes that are present with significant similarity (LS-BSR  $\geq 0.8$ ) in all four of the EPEC isolates. The number of shared genes is indicated in the overlapping regions, which are the genes that have significant similarity (LS-BSR  $\geq 0.8$ ) in two or three of the EPEC isolates, but are divergent (LS-BSR  $< 0.8$ ) or absent (BSR  $< 0.4$ ) in the other EPEC isolate or isolates. The number of unique genes identified in each EPEC isolate genome are the genes that have LS-BSR  $\geq 0.8$  in the one genome and  $< 0.8$  in the other three genomes. The number of genes in parentheses represents the number of unique genes of each EPEC isolate, which have LS-BSR  $\geq 0.8$  in the one genome and  $< 0.4$  in the other three genomes.

the other EPEC isolates. The number of genes that were identified in a single EPEC isolate (LS-BSR  $\geq 0.8$ ) that were divergent (LS-BSR  $< 0.8$  but  $\geq 0.4$ ) or absent (LS-BSR  $< 0.4$ ) in all the other isolates, ranged from 401 to 481 (Figure 1). Each genome also contained isolate-specific content with the number of genes that were unique to a particular isolate and absent from all the other isolates (LS-BSR  $\geq 0.8$ , and  $< 0.4$  in all other isolates), ranging from 294 to 339 (Figure 1). These genomic comparisons allowed the determination of the phylogroup- and isolate-specific gene content of the EPEC prototype isolates. These findings were then used to compare the global transcriptomes of these isolates during growth in laboratory conditions that stimulate virulence factor expression.

### RNA-Seq Analysis of the Global Transcriptomes of EPEC Prototype Isolates

To investigate whether the unique genomic content of each of the EPEC prototype isolates resulted in phylogroup- or isolate-specific differences in the global transcriptomes an RNA-Seq approach was used to analyze each of the prototype isolates grown during standard laboratory conditions. The global transcriptomes were determined for each of the EPEC isolates during growth in rich media (LB), and minimal media (DMEM), which are laboratory conditions that have been previously used to investigate EPEC virulence factor expression. Known virulence factors of EPEC are typically expressed during growth in DMEM but not in LB (Puente et al., 1996; Rosenshine et al., 1996; Leverton and Kaper, 2005). RNA-Seq was also used to investigate the differences in the global transcriptomes during three growth phases (early exponential, late exponential, and

stationary phase) corresponding to the optical densities (OD<sub>600</sub>): 0.2, 0.5, and 1.0, respectively (Supplemental Table 2, Table 1, Figure 3).

The total number of Illumina HiSeq reads generated for the 56 RNA samples analyzed was approximately 5.25 billion (Supplemental Table 2). The total number of reads generated for each sample ranged from approximately 73.4 to 153 million, and the percentage of reads that mapped to each corresponding genome sequence ranging from 31 to 83% for the LB and DMEM samples (Supplemental Table 2). The percentage of reads that mapped to each EPEC isolate genome from the HeLa adherence assay samples was decreased, ranging from 28 to 35% (Supplemental Table 2), as would be expected for a sample containing both pathogen and host.

### Principal Component Analysis Reveals Isolate- and Media-Specific Patterns of Gene Expression

Principal component analysis of the normalized LS-BSR cluster expression patterns of each RNA-Seq sample demonstrated that there were isolate-specific trends among the samples (Supplemental Figure 1). The samples of B171 and E110019 of phylogroup B1, clustered together while the samples of E2348/69 and C581-05 of phylogroup B2 exhibited less similarity to each other (Supplemental Figure 1). Clustering of samples based on the normalized expression of the conserved LS-BSR clusters that were present in all four of the EPEC isolates also demonstrated similarity among the expression patterns for samples of the same media type with the exception of samples of C581-05 (LB, DMEM, or HeLa) (Figure 2). The samples generated for E2348/69, B171, and E110019 clustered into groups based on growth in the LB or DMEM media, with the exception of the E110019 LB OD<sub>600</sub> = 1 samples (Figure 2). This suggests that there was a similar transcriptional response in these isolates; however, there were significant differences in the E110019 gene expression during stationary phase in the nutrient rich LB media (Figure 2). The cluster analysis also demonstrated concordance among the biological replicates for all samples with the exception of the replicates of E2348/69 grown in DMEM to early stationary phase (OD<sub>600</sub> = 1) (Figure 2). We observed that E2348/69 grew slower in DMEM than the other EPEC isolates, which suggests it may have a reduced ability to grow for extended periods of time in the nutrient limiting conditions of DMEM. Interestingly, the C581-05 RNA-Seq samples exhibited divergence from the other EPEC isolate samples, as is demonstrated by the lack of media-specific clustering in the analysis of the expression of conserved genes (Figure 2). The C581-05 LB and DMEM samples grouped together rather than with the samples of the same media type (Figure 2). This suggests that while the EPEC isolates E2348/69, B171, and E110019 exhibited similar overall transcriptional responses to growth in LB and DMEM, C581-05 exhibited a different transcriptional response to growth in the same conditions. The differences in the transcriptional responses of these four EPEC prototype isolates highlights the issues with using a single isolate as the prototype to study the pathogenicity of a pathovar or pathogenic species.

**TABLE 1 | Numbers of genes that are differentially-expressed in the EPEC prototype isolates.**

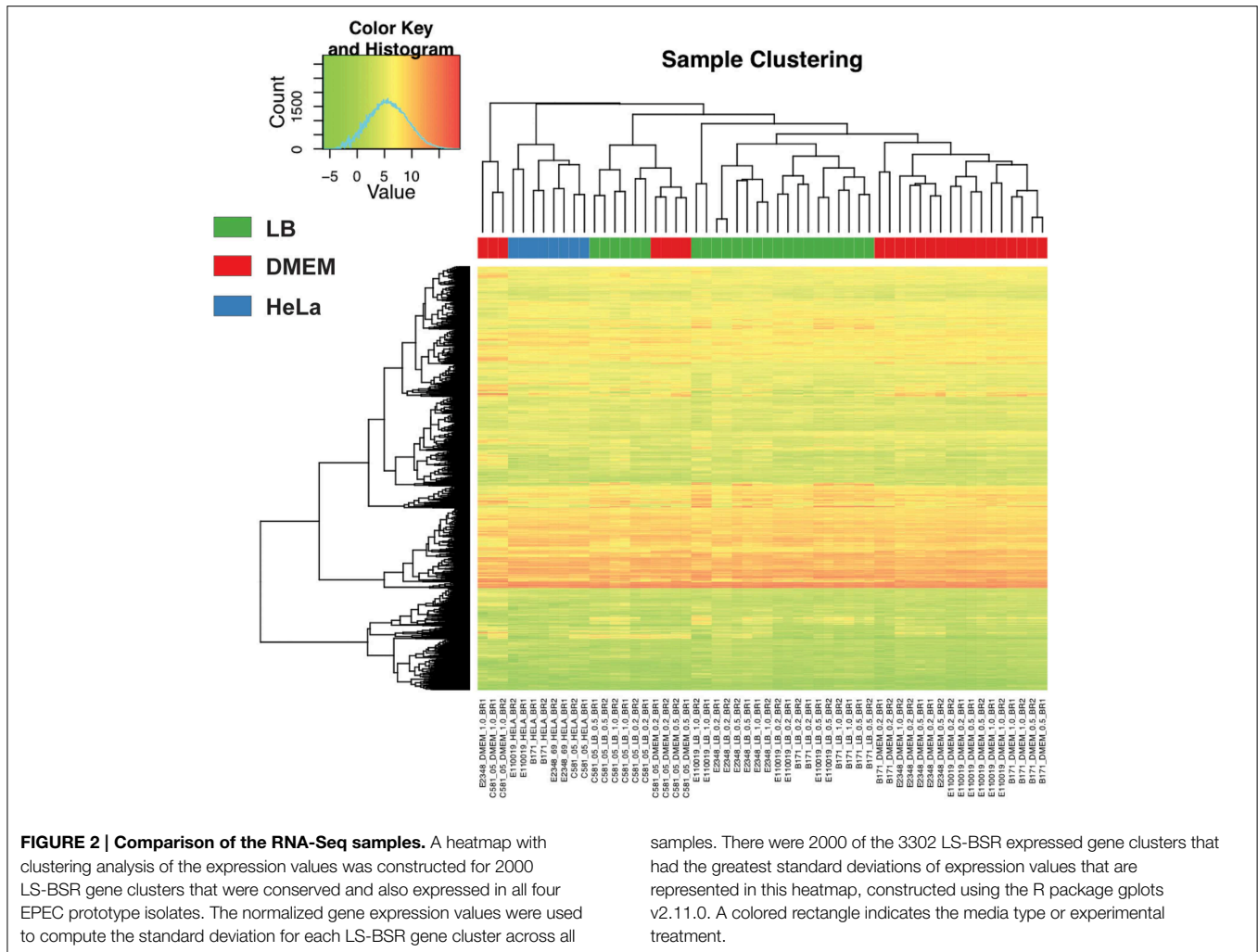
EPEC isolate	BFP content	Phylogenomic lineage	DE comparisons	OD <sub>600</sub>	Increased (LFC ≥ 2) <sup>a</sup>	Decreased (LFC ≤ -2) <sup>a</sup>	Total DE genes <sup>b</sup>	No. of DE genes of core clusters <sup>c</sup>	No. of DE genes of exclusive clusters <sup>d</sup>
E2348/69	BFP+	EPEC1	LB vs. LB	0.5 vs. 0.2	312	92	404	347	14
				1.0 vs. 0.5	113	103	216	184	4
				1.0 vs. 0.2	358	184	542	459	18
			DMEM vs. DMEM	0.5 vs. 0.2	26	26	52	42	0
				1.0 vs. 0.5	1	6	7	6	0
				1.0 vs. 0.2	5	3	8	8	0
			DMEM vs. LB	0.2 vs. 0.2	350	438	788	567	49
				0.5 vs. 0.5	203	272	475	356	21
				1.0 vs. 1.0	23	13	36	17	9
			HeLa vs. DMEM	H vs. 0.2	134	247	381	246	56
				H vs. 0.5	76	264	340	204	56
				H vs. 1.0	8	211	219	96	56
B171	BFP+	EPEC2	LB vs. LB	0.5 vs. 0.2	138	49	187	148	2
				1.0 vs. 0.5	75	122	197	146	3
				1.0 vs. 0.2	205	101	306	221	5
			DMEM vs. DMEM	0.5 vs. 0.2	307	115	422	253	15
				1.0 vs. 0.5	72	30	102	76	4
				1.0 vs. 0.2	287	116	403	243	20
			DMEM vs. LB	0.2 vs. 0.2	153	118	271	233	5
				0.5 vs. 0.5	245	276	521	363	18
				1.0 vs. 1.0	379	261	640	418	33
			HeLa vs. DMEM	H vs. 0.2	321	260	581	404	18
				H vs. 0.5	179	262	441	324	18
				H vs. 1.0	268	414	682	460	44
C581-05	BFP+	EPEC4	LB vs. LB	0.5 vs. 0.2	208	65	273	215	10
				1.0 vs. 0.5	207	194	401	300	17
				1.0 vs. 0.2	495	359	854	653	41
			DMEM vs. DMEM	0.5 vs. 0.2	125	33	158	100	12
				1.0 vs. 0.5	728	617	1345	900	116
				1.0 vs. 0.2	885	703	1588	1051	130
			DMEM vs. LB	0.2 vs. 0.2	138	75	213	192	1
				0.5 vs. 0.5	172	265	437	369	5
				1.0 vs. 1.0	473	362	835	520	85
			HeLa vs. DMEM	H vs. 0.2	294	239	533	340	45
				H vs. 0.5	295	333	628	421	50
				H vs. 1.0	754	1055	1809	1196	140
E110019	BFP-	None	LB vs. LB	0.5 vs. 0.2	356	237	593	464	10
				1.0 vs. 0.5	333	214	547	434	10
				1.0 vs. 0.2	639	365	1004	782	35
			DMEM vs. DMEM	0.5 vs. 0.2	108	40	148	118	11
				1.0 vs. 0.5	64	53	117	90	11
				1.0 vs. 0.2	209	97	306	204	32
			DMEM vs. LB	0.2 vs. 0.2	208	130	338	273	25
				0.5 vs. 0.5	223	317	540	427	16
				1.0 vs. 1.0	216	395	611	497	24
			HeLa vs. DMEM	H vs. 0.2	295	333	628	433	65
				H vs. 0.5	310	404	714	555	49
				H vs. 1.0	469	552	1021	749	95

<sup>a</sup>LFC is the Log<sub>2</sub> fold change of the genes that exhibit significant (LFC ≥ 2 or ≤ -2 and FDR ≤ 0.05) differential expression (DE).

<sup>b</sup>The total number of genes that exhibit significant (LFC ≥ 2 or ≤ -2 and FDR ≤ 0.05) DE.

<sup>c</sup>The number of DE genes that belong to core gene clusters (LS-BSR ≥ 0.8 in all four genomes).

<sup>d</sup>The number of DE genes that belong to gene clusters identified in one genome with an LS-BSR ≥ 0.8 and <0.4 in the other three genomes.

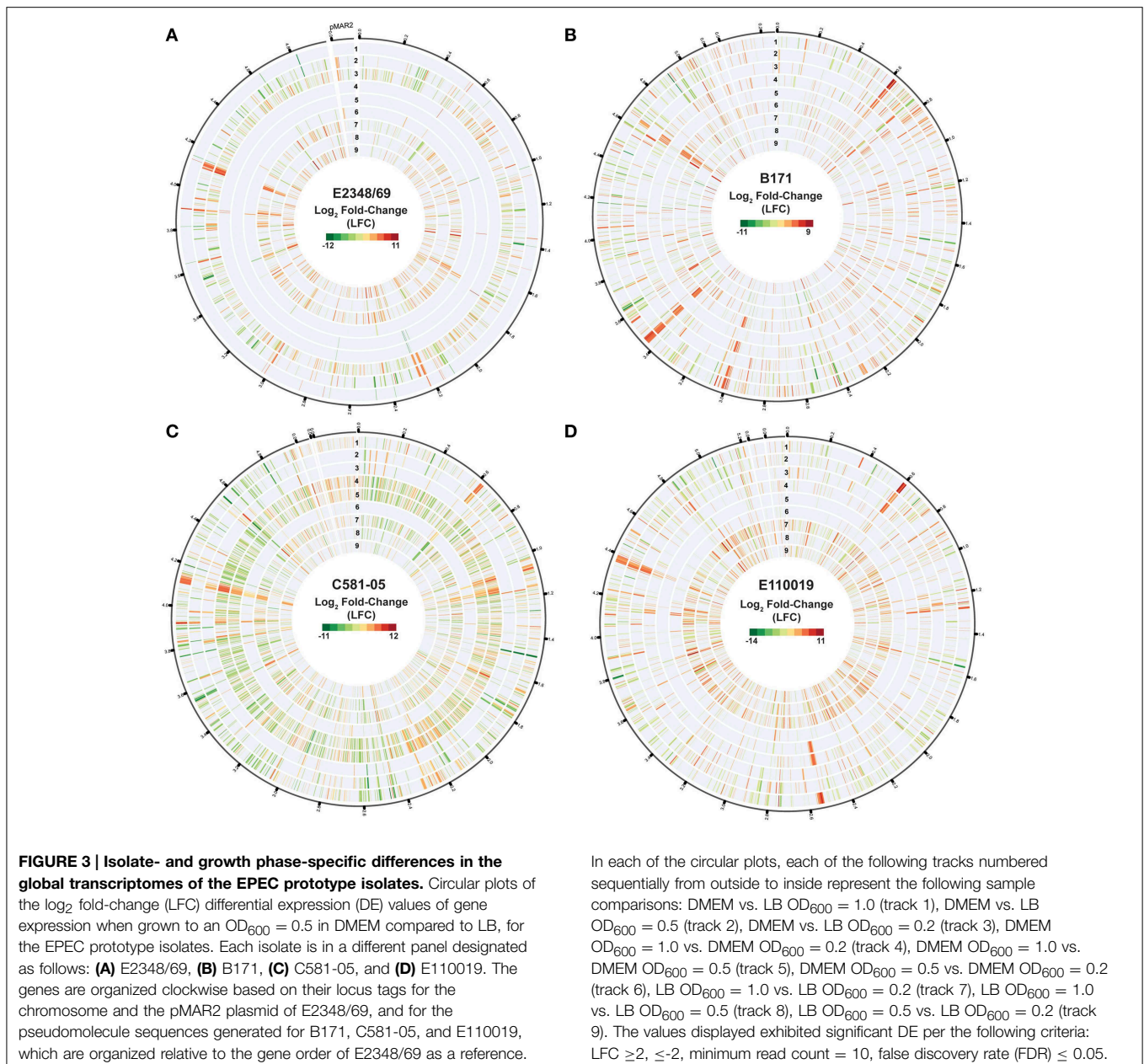


### Isolate-Specific Changes in Gene Expression

Differential expression analysis of the RNA-Seq samples further demonstrated there were identifiable isolate-specific differences in gene expression during growth in DMEM compared to LB (Table 1, Figure 3 tracks 1–3, Supplemental Data Sets 3–6). The total number of genes that exhibited significant differential expression during growth in DMEM compared to LB ranged from 36 to 835 depending on the growth phase and the EPEC isolate (Table 1). Overall, the EPEC isolates had a greater number of genes that were differentially-expressed in DMEM compared to LB during stationary phase ( $OD_{600} = 1$ ) than there were during exponential ( $OD_{600} = 0.5$ ) or early exponential growth ( $OD_{600} = 0.2$ ) (Table 1, Figure 3 tracks 1–3). However, E2348/69 had the opposite trend with fewer genes that had significant differential expression in DMEM compared to LB during exponential ( $OD_{600} = 0.5$ ) compared to early exponential growth ( $OD_{600} = 0.2$ ) (Table 1). These findings demonstrated that each of the four EPEC prototype isolates had isolate-specific responses to growth in the nutrient rich LB broth compared to the nutrient limiting conditions of DMEM.

### Growth Phase-Specific Changes in Gene Expression

The number of differentially-expressed genes also was altered for each of the EPEC isolates when comparing the different growth phases during growth in a single media type (Table 1, Figure 3 tracks 4–9). During growth in DMEM the number of genes that had significant differential expression was greatest for the stationary phase ( $OD_{600} = 1$ ) samples compared to the early exponential samples ( $OD_{600} = 0.2$ ), or for the late exponential ( $OD_{600} = 0.5$ ) compared to the early exponential ( $OD_{600} = 0.2$ ) samples than observed for the stationary phase ( $OD_{600} = 1$ ) compared to the late exponential ( $OD_{600} = 0.5$ ) (Table 1). Overall, E2348/69 had far fewer genes that exhibited significant differential expression when comparing the three different growth phases during growth in DMEM (7–52 genes) than was observed for the other three EPEC isolates during growth in DMEM (range from 102 to 1588 genes) (Table 1, Figure 3 tracks 4–6). The reduced number of genes that exhibited significant differential expression compared to E2348/69 DMEM grown to an  $OD_{600} = 1$  may in part be attributed to the observed



decrease in growth of this isolate. The number of differentially-expressed genes when comparing these two growth phases was also reduced when compared to the three other EPEC isolates under the same conditions (Table 1). This indicates that along with a reduced ability to grow for extended periods of time in the DMEM, E2348/69 also had a different transcriptional response to growth in DMEM than was observed for the other EPEC isolates.

### Media-Specific Changes in Gene Expression

This data set also allowed the examination of the differences in gene expression in different media. The majority of the genes that had significant differential expression when comparing growth phases were not identified in the comparison of

growth in differing media types (Table 2). Overall, there were fewer genes (range 0–501) that exhibited significant altered expression in both LB and DMEM when comparing between the different growth phases (Table 2). This finding highlights the considerable differences in the global transcriptomes of EPEC prototype isolates during growth in nutrient-rich compared to nutrient-limiting media. The greatest number of genes (501) that were differentially-expressed when comparing the different growth phases in both LB and DMEM was identified in the EPEC4 isolate C581-05 (Table 2). In most cases the genes that were differentially-expressed for the different growth phase comparisons of both LB and DMEM had similar trends of increased or decreased expression, and there were fewer genes



**TABLE 2 | Number of genes that are differentially-expressed during different growth phases.**

EPEC isolate	OD <sub>600</sub>	No. of differentially-expressed genes <sup>a</sup>											
		Total DE genes		In LB only			In DMEM only			In both LB and DMEM			
		LB	DMEM	Total	Increased	Decreased	Total	Increased	Decreased	Total	Increased	Decreased	Different trend
E2348/69	0.5 vs. 0.2	404	52	382	290	92	30	24	6	22	2	0	20
	1.0 vs. 0.5	216	7	216	113	103	7	1	6	0	0	0	0
	1.0 vs. 0.2	542	8	542	358	184	8	5	3	0	0	0	0
B171	0.5 vs. 0.2	187	422	151	113	38	386	298	88	36	5	7	24
	1.0 vs. 0.5	197	102	185	68	117	90	64	26	12	6	3	3
	1.0 vs. 0.2	306	403	224	137	87	321	239	82	82	40	6	36
C581-05	0.5 vs. 0.2	273	158	221	160	61	106	93	13	52	30	2	20
	1.0 vs. 0.5	401	1345	200	122	78	1144	649	495	201	78	115	8
	1.0 vs. 0.2	854	1588	353	224	129	1087	635	452	501	241	221	39
E110019	0.5 vs. 0.2	593	148	558	330	228	113	81	32	35	25	7	3
	1.0 vs. 0.5	547	117	511	309	202	81	53	28	36	3	4	29
	1.0 vs. 0.2	1004	306	873	535	338	175	108	67	131	96	22	13

<sup>a</sup>The number of genes that exhibited significant ( $LFC \geq 2$  or  $\leq -2$ ,  $FDR \leq 0.05$ ,  $\geq 10$  read counts) differential expression at two different growth phases in a single media type (LB or DMEM).

that were increased during growth in one media type and decreased in the other media type (Table 2). For example, there were only 22 genes of E2348/69 that exhibited significant differential expression in both DMEM and LB when comparing the early exponential ( $OD_{600} = 0.2$ ) and exponential ( $OD_{600} = 0.5$ ) growth phases (Table 2). This limited gene set can be partially attributed to the low number of E2348/69 genes (52) that had significant differential expression in DMEM when comparing growth phases (Table 1). Interestingly, nearly all (20 out of 22) of these genes exhibited a different expression trend in LB compared to DMEM, which is in contrast to the other EPEC isolates, which had a greater number of genes with the same trend in expression (Table 2). For each of the E2348/69 genes that had a different trend in expression in the DMEM and LB samples, the expression was increased in the LB samples and decreased in the DMEM samples.

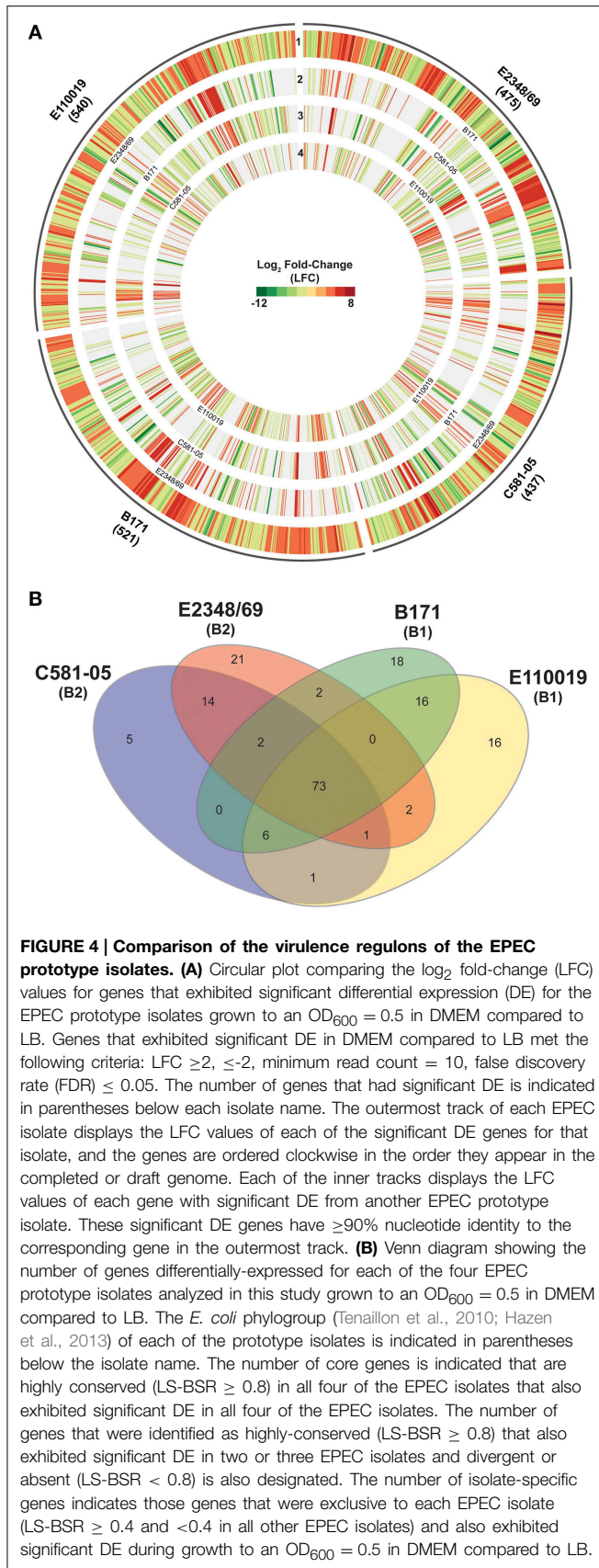
For most of the growth phase comparisons, the majority of the differentially-expressed genes were altered only in LB or DMEM and did not exhibit differential expression in both media types (Table 2). The number of genes that were altered only in LB ranged from 151 to 873, while the number of genes that were altered only in DMEM ranged from 7 to 1144 (Table 2). Among the genes that were differentially-expressed when comparing the different growth phases during growth in LB only, were genes involved in metabolism (Supplemental Data Sets 3–6). Meanwhile, genes that were differentially-expressed during different growth phases in DMEM included virulence-associated genes, plasmid-associated genes involved in conjugal transfer, putative phage genes, lipoproteins, and global regulators such as the histone-like protein H-NS, which regulates expression of the LEE (Bustamante et al., 2001; Mellies et al., 2007; Bhatt et al., 2011) (Supplemental Data Sets 3–6). Overall, these studies highlight the transcriptional differences among these

four prototype isolates in multiple growth phases and media types.

### Comparison of the Global Virulence Regulons of EPEC Prototype Isolates During Growth in Laboratory Conditions

To investigate isolate-specific differences in the expression of known virulence factors, and to identify additional genes that are coordinately-expressed with the known virulence factors, we identified the differences in expression of all protein-encoding genes of the four EPEC prototype isolates during exponential growth in DMEM compared to LB during exponential growth ( $OD_{600} = 0.5$ ) (Table 1, Supplemental Data Sets 3–6). The total number of genes that exhibited significant differential expression for E2348/69, B171, C581-05, and E110019 were 475, 521, 437, and 540 respectively (Table 1, Figure 4A). These genes include known virulence factors such as those within the LEE and BFP regions (Table 3). However, there were also a number of conserved genes involved in metabolism and stress responses including transporters, and additional isolate-specific genes that represent a potentially unique contribution to the virulence regulon of a particular EPEC isolate (Table 3). A previous study demonstrated there were clade-specific differences in the expression levels of LEE genes encoded by different O157:H7 EHEC isolates during growth in DMEM (Abu-Ali et al., 2010).

To our knowledge there has been no extensive comparison to date of phylogroup- or lineage-specific differences in virulence factor expression among EPEC isolates. Comparison of the differentially-expressed genes of the four prototype isolates that exhibited similarity based on the LS-BSR analysis demonstrated there were isolate-specific differences in the global transcriptomes during exponential growth ( $OD_{600} = 0.5$ ) in DMEM compared to LB (Figure 4A). This highlights the need



to examine multiple isolates and not rely on a single prototype strain for any pathovar to describe an entire group of isolates. Overall, the differentially-expressed genes that were identified in more than one of the EPEC isolates exhibited similar trends of increased or decreased expression (**Figure 4A**); however, in some instances there were genes identified in multiple EPEC isolates that were increased in some of the isolates and decreased in other isolates (**Figure 4A, Table 3**). These included some genes identified by LS-BSR as highly-conserved in all four of the EPEC isolates (Supplemental Table 3). Interestingly, several genes (*napB*, *napG*, and *napH*) of the *nap* operon, which encodes a periplasmic nitrate reductase (Brondijk et al., 2002), exhibited increased expression in the EPEC isolates of phylogroup B1 (B171 and E110019), but decreased expression in the phylogroup B2 EPEC isolates (E2348/69 and C581-05) (Supplemental Table 3). There were a total of 73 genes that were present in the genomes of all four of the EPEC isolates that were also differentially-expressed in all four EPEC isolates (**Figure 4B**). Of these 73 genes, there were four that had different trends of expression in these isolates, 18 exhibited consistently increased expression, and 51 exhibited consistently decreased expression in the four EPEC isolates (Supplemental Table 3).

Although there were a number of similarities in the response of the four EPEC prototype isolates to growth in DMEM compared to LB, there were many more genes that exhibited differential expression in one, two or three of the isolates (**Figure 4A**). The number of genes that were identified by LS-BSR as being unique (LS-BSR  $\geq 0.8$  and  $< 0.4$  in all other isolates) to an EPEC isolate that were differentially-expressed during exponential growth in DMEM compared to LB ranged from only 5–21 genes (**Table 1, Figure 4B**). The number of genes that were identified in two or three of the EPEC isolates and also had significant differential expression was similarly low, ranging from 0 to 16 genes (**Figure 4B**). Thus, the majority of the genes that exhibited differential expression in each of the EPEC isolates were genes that were highly-conserved in more than one isolate (**Table 1**). However, these genes did not exhibit significant differential expression in all four of the isolates (**Figure 4B**). This may in part be attributed to the stringent cut-offs used in this study for identifying genes with significant differential expression, and with less stringent cut-offs there may be more of these highly-conserved genes with low-level differences identified with significant differential expression in all four isolates.

### Differential-Expression of the LEE and BFP Regions

The differential expression of known virulence factors present in all four of the EPEC isolates such as genes encoded by the LEE demonstrated there are isolate-specific differences in the timing and conditions during which the LEE genes are expressed (**Figure 5A, Table 3**). The majority of the LEE-encoded genes had significantly increased expression during growth in DMEM compared to LB for the exponential growth ( $OD_{600} = 0.5$ ) and early stationary phase ( $OD_{600} = 1$ ) samples (**Figure 5A**). This finding was true for all four of the EPEC isolates except the stationary phase comparison of E2348/69, which was previously

**TABLE 3 | Select genes that exhibited significant differential expression in DMEM compared to LB at an OD<sub>600</sub> = 0.5.**

Gene ID <sup>a</sup>	Predicted protein function	Protein accession no. <sup>a</sup>	Log <sub>2</sub> fold-change (LFC) <sup>b,c</sup>			
			E2348/69	C581-05	B171	E110019
<b>LEE-ENCODED EFFECTORS, REGULATORS, AND ADHESION</b>						
<i>eae</i>	Intimin adherence factor	YP_002331401.1	NS	NS	4.57	4.82
<i>ler</i>	Transcriptional regulator	YP_002331430.1	2.81	NS	2.08	NS
<i>grlR</i>	Negative regulator GrlR	YP_002331420.1	5.43	NS	3.22	2.33
<i>grlA</i>	Positive regulator GrlA	YP_002331419.1	4.89	NS	3.6	NS
<i>espF</i>	LEE-encoded effector EspF	YP_002331392.1	5.17	NS	4.7	3.93
<i>espG</i>	LEE-encoded effector EspG	YP_002331432.1	4.71	2.67	4.07	3.11
<i>espH</i>	LEE-encoded effector EspH	YP_002331406.1	5.64	NS	4.05	2.62
<i>espZ</i>	LEE-encoded effector EspZ	YP_002331413.1	4.36	NS	3.31	2.68
<i>map</i>	LEE-encoded effector Map	YP_002331404.1	3.87	NS	3.83	2.97
<i>tir</i>	Translocated intimin receptor Tir	YP_002331403.1	NS	NS	4.34	3.57
<b>NON-LEE-ENCODED EFFECTORS</b>						
<i>espG</i>	T3SS secreted effector EspG-like protein	YP_002330404.1	2.56	NS	NS	NS
<i>nleG</i>	T3SS secreted effector NleI/NleG-like protein	YP_002328601.1	2.66	NS	NS	NS
<i>nleH</i>	T3SS secreted effector NleH-like protein	YP_002328982.1	NS	NS	NS	-2.00
<i>nleF</i>	T3SS secreted effector NleF-like protein	YP_002328983.1	NS	NS	NS	-2.40
<b>EAF PLASMID-ENCODED</b>						
<i>bfpA</i>	Major pilin structural unit bundlin	YP_002332157.1	4.07	NS	3.11	NA
<i>perA</i>	Transcriptional activator of the bfp operon	YP_002332173.1	4.67	NS	4.01	NA
<i>perB</i>	Transcriptional regulator	YP_002332174.1	4.69	NS	3.69	NA
<i>perC</i>	Transcriptional regulator	YP_002332175.1	4.41	-2.04	3.22	NA
<b>DE GENES OF CORE GENE CLUSTERS</b>						
<i>napB</i>	Citrate reductase cytochrome c-type subunit	YP_002329852.1	-4.11	-3.93	2.14	2.15
<i>napH</i>	Quinol dehydrogenase membrane component	YP_002329853.1	-3.45	-3.56	2.48	3.02
<i>napG</i>	Quinol dehydrogenase periplasmic component	YP_002329854.1	-3.4	-3.66	2.27	2.69
<i>rbsD</i>	D-ribose pyranase	YP_002331517.1	-5.15	2.46	-5.4	-5.27
<i>cirA</i>	Colicin I receptor	YP_002329807.1	4.16	5.82	3.62	4.62
<i>dppF</i>	Dipeptide transporter ATP-binding subunit	YP_002331254.1	4.55	3.62	2.15	2.84
<i>espG</i>	LEE-encoded effector EspG	YP_002331432.1	4.71	2.67	4.07	3.11
<i>narK</i>	Nitrate/nitrite transporter	YP_002328887.1	-8.07	-6.81	-2.67	-4.32
<i>fucl</i>	L-fucose isomerase	YP_002330550.1	-4.09	-2.45	-3.22	-2.96
<i>ygeV</i>	DNA-binding transcriptional regulator	YP_002330601.1	-3.29	-2.42	-4.43	-2.95
<i>ygfJ</i>	Hypothetical protein	YP_002330609.1	-5.57	-3.17	-2.32	-2.53
<i>nanT</i>	Sialic acid transporter	YP_002330964.1	-3.39	-2.67	-2.12	-2.67
<i>yjfO</i>	Biofilm stress and motility protein A	YP_002331964.1	-4.79	-2.21	-4.28	-3.59
<b>DE GENES OF EXCLUSIVE GENE CLUSTERS</b>						
<b>E2348/69</b>						
<i>espC</i>	Serine protease	YP_002330403.1	5.4	NA	NA	NA
<i>wzx</i>	O-antigen flippase	YP_002329687.1	5.2	NA	NA	NA
E2348C_2104	Lipoprotein	YP_002329618.1	2.68	NA	NA	NA
<b>C581-05</b>						
<i>idnD</i>	L-isonate 5-dehydrogenase	WP_024223218.1	NA	-2.02	NA	NA
<i>tctA</i>	Tripartite tricarboxylate transporter TctA family protein	YP_543360.1	NA	-2.05	NA	NA
<i>purR</i>	Cytochrome C peroxidase	WP_024223889.1	NA	-2.09	NA	NA

(Continued)

TABLE 3 | Continued

Gene ID <sup>a</sup>	Predicted protein function	Protein accession no. <sup>a</sup>	Log <sub>2</sub> fold-change (LFC) <sup>b,c</sup>			
			E2348/69	C581-05	B171	E110019
<b>B171</b>						
EcB171_5053	Glycosyl hydrolase 108 family protein	EDX28171.1	NA	NA	3.56	NA
<i>bor</i>	Lipoprotein Bor	EDX30251.1	NA	NA	-2.24	NA
EcB171_3359	Transcriptional regulator, GntR family	EDX31174.1	NA	NA	-2.3	NA
<b>E110019</b>						
<i>imm</i>	Colicin-E2 immunity protein	EDV85355.1	NA	NA	NA	2.04
EcE110019_3691	Transcriptional regulator, C terminal family protein	EDV86975.1	NA	NA	NA	2.03
<i>stbB</i>	Plasmid stability family protein	WP_000361389.1	NA	NA	NA	-2.17

<sup>a</sup>The gene symbol or locus id and the protein accession number are indicated for the top match protein. In some cases a protein match could not be identified for a cluster in a particular genome, which likely results from differences in the gene-calling that was used for LS-BSR compared to that used for the GenBank sequences. None indicates there was not a corresponding locus id for the particular genome.

<sup>b</sup>These are LFC values for samples that have been normalized for a single EPEC isolate, and have not been normalized across all EPEC isolates.

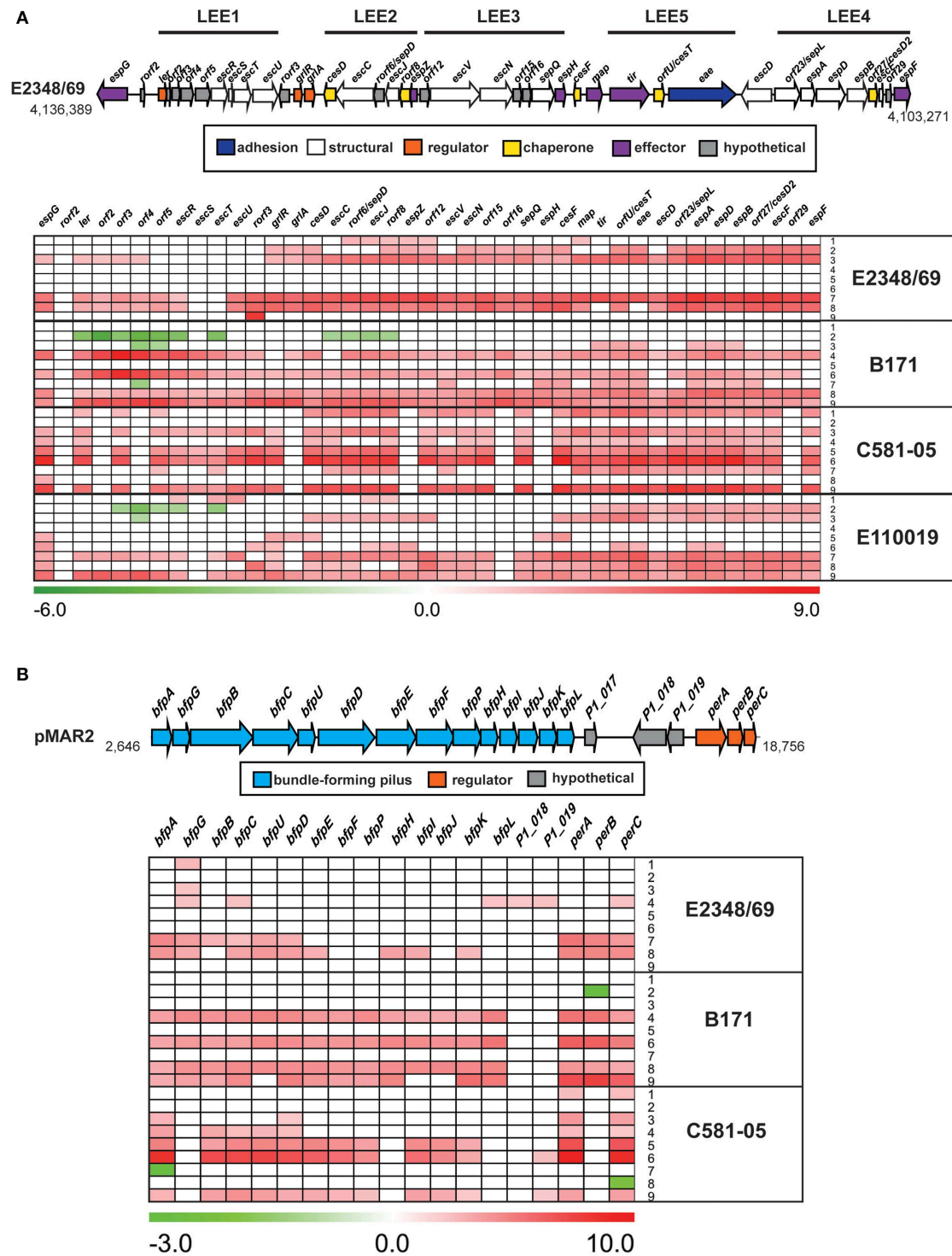
<sup>c</sup>NS indicates a value was not significant, while NA indicates a comparison was not applicable.

noted to have a different gene expression pattern (Figure 2). The increased expression of the intimin gene (*eae*) observed by RNA-Seq analysis was verified using qRT-PCR for these samples (Supplemental Figure 2). Interestingly, the trend of increased LEE gene expression in DMEM compared to LB was not observed for the C581-05 late exponential growth samples, which only exhibited significant expression for one gene, *espG*, in DMEM compared to LB during late exponential growth (OD<sub>600</sub> = 0.5) (Figure 5A). In contrast, the LEE genes of C581-05 exhibited significant differences in expression during late exponential (OD<sub>600</sub> = 0.5) compared to early exponential (OD<sub>600</sub> = 0.2) growth in LB (Figure 5A). This finding suggests that in contrast to other EPEC isolates such as B171, which did not exhibit significant increases in LEE gene expression during growth in LB or DMEM over time, the LEE genes of EPEC isolate C581-05 had increased expression over time during growth in LB (Figure 5A). This potentially suggests an additional role for quorum sensing in the regulation of LEE in C581-05 during growth in nutrient-rich media (LB). This type of variation in gene expression has been previously observed in *E. coli* as being quorum sensing regulated (Sperandio et al., 1999; Sircili et al., 2004) and it is possible that this is the case in these studies; however more detailed analysis of the contribution quorum sensing will require studies that are beyond the scope of the current manuscript. Another notable difference in the expression of LEE genes was identified for both B171 and E110019 of phylogroup B1, which exhibited decreased expression of LEE1 genes during stationary phase growth compared to exponential growth in LB (Figure 5A). This was in contrast to the LEE genes of both E2348/69 and C581-05 of phylogroup B2, which did not exhibit a significant difference in expression of the LEE genes during these conditions (Figure 5A). This further indicates there are phylogroup-specific differences in the transcriptional regulation of EPEC virulence factors.

The genes that were identified with significant genetic conservation and increased gene expression during the virulence-associated conditions in all EPEC isolates, represents the core

EPEC virulence regulon. The only LEE-encoded gene that exhibited significant differential expression during exponential growth in DMEM compared to LB in all four of the EPEC isolates was *espG*, which had increased expression in all of the EPEC isolates (Table 3). *EspG* is a type III secreted effector that alters the cytoskeleton of the host cell by interfering with the function of microtubules and golgi (Hardwidge et al., 2005; Shaw et al., 2005; Tomson et al., 2005; Wong et al., 2011). The increased expression of *espG* in all isolates was verified using qRT-PCR (Supplemental Figure 2). In contrast, few of the non-LEE-encoded T3SS effectors had increased expression during exponential growth in DMEM compared to LB (Table 3), suggesting a lack of global coordination of the regulation of these virulence factors in all isolates.

The differential expression of genes of the BFP region also differed for the three typical EPEC isolates (E2348/69, B171, and C581-05), which encoded the BFP operon (Figure 5B). The majority of the BFP genes exhibited significantly increased expression during exponential growth (OD<sub>600</sub> = 0.5) in DMEM compared to LB (Figure 5B), which is consistent with the findings of previous studies that EPEC virulence factors exhibit increased expression in DMEM (Puente et al., 1996; Rosenshine et al., 1996; Leverton and Kaper, 2005). Interestingly, EPEC isolate C581-05 did not have significantly increased expression during these conditions, and instead exhibited decreased expression of the transcriptional regulator, *perC* (Figure 5B). *PerC* has previously been demonstrated to activate transcription of LEE and BFP genes (Gomez-Duarte and Kaper, 1995; Tobe et al., 1996; Bustamante et al., 2011). This finding further suggests that virulence factors such as the BFP genes are constitutively expressed in C581-05 during exponential growth irrespective of media in some isolates. In contrast, the majority of the BFP genes of C581-05 exhibited increased expression during entry into stationary phase in DMEM compared to LB, indicating that BFP expression exhibits a greater increase during the nutrient-limiting conditions over time (Figure 5B).



**FIGURE 5 | Comparison of the isolate- and growth phase-specific differences in the expression of genes in the LEE and BFP regions.** A diagram and heatmap of the isolate- and growth phase-specific differences in expression of (A) genes encoded within the LEE region of E2348/69, B171, E110019, and C581-05, and (B) genes of the BFP operon encoded by the EAF plasmids of E2348/69, B171, and C581-05 (note E110019 naturally lacks the BFP encoding plasmid). Red indicates increased differential expression, green indicates decreased differential expression, and white indicates the difference in expression

was not significant or a gene was not present in the EPEC isolate. Each row represents the differential expression ( $\log_2$  fold-change; LFC) for the following different sample comparisons: LB OD<sub>600</sub> = 0.5 vs. LB OD<sub>600</sub> = 0.2 (row 1), LB OD<sub>600</sub> = 1.0 vs. LB OD<sub>600</sub> = 0.5 (row 2), LB OD<sub>600</sub> = 1.0 vs. LB OD<sub>600</sub> = 0.2 (row 3), DMEM OD<sub>600</sub> = 0.5 vs. DMEM OD<sub>600</sub> = 0.2 (row 4), DMEM OD<sub>600</sub> = 1.0 vs. DMEM OD<sub>600</sub> = 0.5 (row 5), DMEM OD<sub>600</sub> = 1.0 vs. DMEM OD<sub>600</sub> = 0.2 (row 6), DMEM vs. LB OD<sub>600</sub> = 0.2 (row 7), DMEM vs. LB OD<sub>600</sub> = 0.5 (row 8), and DMEM vs. LB OD<sub>600</sub> = 1.0 (row 9).

The decreased expression of the plasmid genes (*bfpA* and *perA*) of C581-05 during exponential growth in DMEM compared to LB, and increased expression of these genes during stationary phase was confirmed by qRT-PCR (Supplemental Figure 2). Again, these data suggest a role for quorum sensing in this isolate and highlight the strain specificity of virulence factor regulation in different isolates.

### Differential-Expression of EPEC Genes Not Previously Identified as Virulence Factors

Additional genes that exhibited significant differential expression during the virulence-inducing conditions included biotin synthesis genes, dipeptide transporters, and a gene encoding a putative colicin receptor protein (Table 3, Supplemental Table 3). The genes that were determined to have decreased expression in all four EPEC isolates included genes encoding transporters and many genes involved in metabolism (Table 3, Supplemental Table 3). The genes that were identified with significant differential expression in only one of the EPEC isolates included known virulence-associated genes, phage genes, a putative adhesin, lipoproteins, lipopolysaccharide biosynthesis genes, and numerous predicted transcriptional regulators (Table 3, Supplemental Table 3). For example, among the genes that were unique to the E2348/69 transcriptome was *espC*, which encodes a serine protease autotransporter that causes cytotoxicity to host cells and is translocated into the host cell by the type V secretion mechanism (Table 3) (Navarro-Garcia et al., 2004; Vidal and Navarro-Garcia, 2008). There were also several E2348/69 genes involved in O-antigen biosynthesis (*wzy*, *wzx*, *wbiO*) that exhibited the greatest increases in expression in DMEM compared to LB (Table 3, Supplemental Table 3). The capsular polysaccharide and O-antigen modification was recently demonstrated to promote pathogenesis of the uropathogenic *E. coli* (UPEC) by increasing their ability to survive in the gut (Sarkar et al., 2014). However, to our knowledge the role of the O-antigen gene expression during EPEC virulence has not been established.

Many of the non-virulence factor genes identified that are co-regulated with the virulence regulons of the EPEC isolates are involved in general metabolism or survival and may enhance the ability of the EPEC isolates to colonize and survive in the human gastrointestinal tract (Table 3). Some of the genes of the *nap* operon (*napB* and *napH*), which encodes a periplasmic nitrate reductase (Brondijk et al., 2002), were previously determined for the EHEC O157:H7 Sakai strain, to exhibit increased expression during exponential growth in minimal media (Bergholz et al., 2007). Additionally, several metabolic transcriptional regulators have recently been linked to pathogenesis of EHEC and EPEC. For example, a transcriptional regulator, *FusR*, of O157:H7 EHEC was identified that regulates EHEC metabolism and the expression of LEE and promotes intestinal colonization in response to the presence of the metabolic by-product, fucose, which was identified as being abundant in the mucus layer of the intestine (Pacheco et al., 2012). Also, *Cra* and *KdpE*, two other global transcriptional regulators of sugar production and potassium transport, have been demonstrated to also either directly or indirectly regulate expression of LEE

genes of EHEC (Njoroge et al., 2012, 2013). Another regulator of glycogen biosynthesis metabolism (Romeo et al., 1993), *CsrA*, was also previously determined to regulate expression of EPEC virulence factors including the LEE (Bhatt et al., 2009, 2011). Overall, the transcriptional studies demonstrate the interconnectedness of general metabolism and virulence and only through global studies can we begin to understand these relationships.

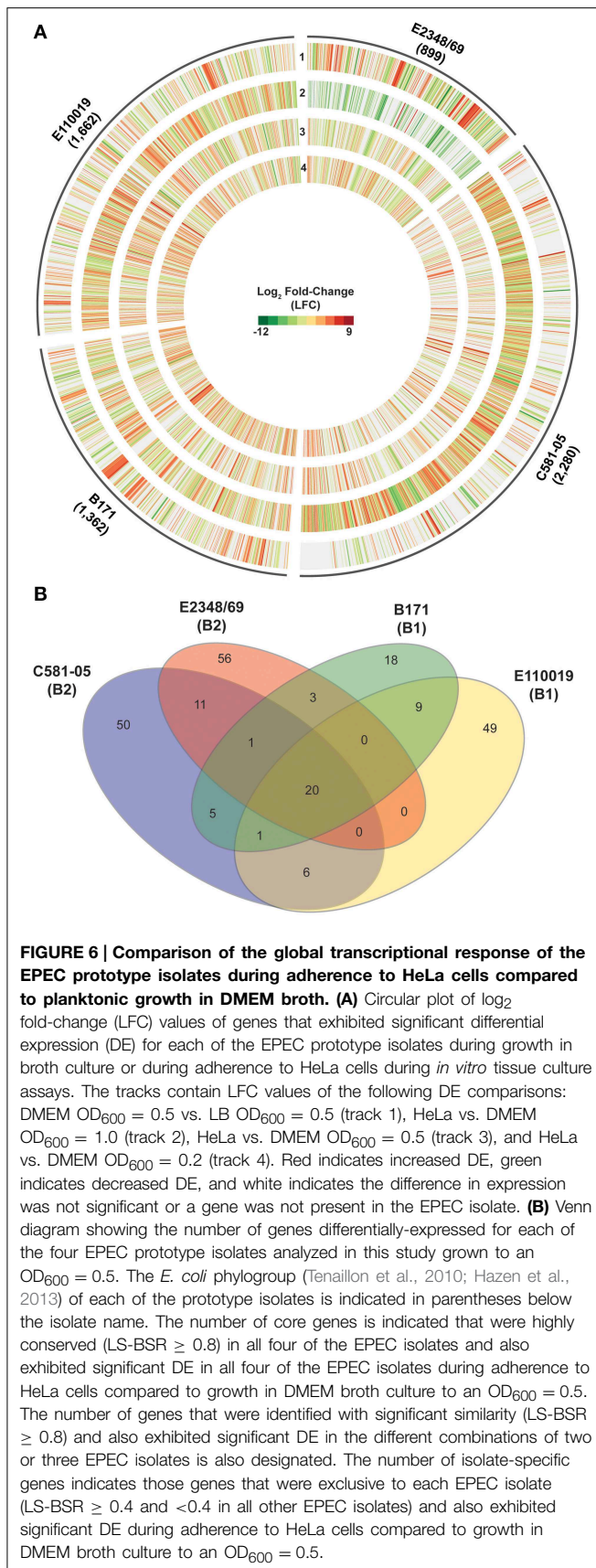
### Differences in the Global Transcriptomes of the Typical and Atypical EPEC Prototype Isolates

In addition to the phylogroup- and isolate-specific differences in gene expression, there were also differences when comparing the three typical EPEC isolates (E2348/69, B171, and C581-05) to the atypical EPEC isolate (E110019) (Supplemental Data Sets 3–6). The LS-BSR analysis of the EPEC isolate genomic content identified 169 genes that were highly-conserved (LS-BSR  $\geq 0.8$ ) among the typical EPEC isolate genomes that were divergent (LS-BSR  $< 0.8$ ,  $\geq 0.4$ ) or absent (LS-BSR  $< 0.4$ ) from the atypical EPEC isolate E110019 (Figure 1). Of these 169 genes, there were 20 that were also identified with significant differential expression during growth in DMEM compared to LB in two of the typical EPEC isolates and only two genes that were differentially-expressed in all three of the typical EPEC isolates (Supplemental Data Sets 3–5). The two genes that exhibited significant differential expression in all three of the typical EPEC isolates were *perC* and the lactose permease *lacY* (Supplemental Data Sets 3–5). Meanwhile, genes that were present and exhibited significant differential expression in E110019 that were absent from the typical EPEC isolates included a putative adhesin, a colicin immunity protein, a plasmid stability protein, a phage-associated gene, and numerous conserved hypothetical proteins (Table 3, Supplemental Table 3). These findings suggest that the plasmid maintained in E110019 is also regulated under these conditions, but is not similar to the EAF plasmid in the tEPEC isolates.

### Global Transcriptomes of EPEC During Adherence to Host Cells *In Vitro*

The mono culture studies described above highlight the variability of the EPEC transcriptome of isolates grown independently of external stimuli. To extend the studies above, the global transcriptome of each EPEC isolate during adherence to HeLa cells compared to growth density were examined (Figure 6A). The total number of genes that exhibited significant differential expression in one or more of these comparisons differed considerably for each EPEC isolate (ranged from 899 to 2280), as indicated by the size of the plot regions and numbers designated in parentheses for each EPEC isolate (Figure 6A, Supplemental Data Sets 3–6).

Comparison of the genes that had significant differential expression during adherence of the EPEC isolates to HeLa cells compared to exponential growth (OD<sub>600</sub> = 0.5) in DMEM broth demonstrated that most of the altered genes belonged to core gene clusters (Figure 6B). However, there were only 20 genes that were present and also were differentially expressed in all four of the EPEC isolates (Figure 6B, Supplemental Table 4).



Furthermore, all but two of these genes had similar trends in terms of increased or decreased expression in all four of the EPEC isolates (Table 4). Among the genes that exhibited increased expression in DMEM compared to LB for all four of the prototype isolates was *hscB*, which has been previously described as a co-chaperone involved in the formation of Fe-S cluster proteins of *E. coli* (Vickery et al., 1997; Hoff et al., 2000). One of the genes that exhibited different trends of expression in the four EPEC isolates was *nirC* (Table 4, Supplemental Table 4), which encodes a protein involved in nitrite uptake in *E. coli* (Clegg et al., 2002). The *nirC* gene was also identified as a virulence-associated factor that increased the survival of *Salmonella* in the presence of macrophages (Das et al., 2009). This finding demonstrates that there are phylogroup-specific responses of EPEC in the presence of host cells, suggesting EPEC isolates of different *E. coli* phylogroups have unique mechanisms for survival during infection, in addition to their differences in virulence factor content that have been previously described (Hazen et al., 2013).

The number of genes that were exclusive to one of the prototype isolates that also exhibited significant differential expression during adherence to HeLa compared to growth in DMEM media ranged from 18 to 56 genes (Figure 6B). Interestingly, of the 56 genes that were unique to E2348/69 that had significant differential expression, 41 of these genes were previously identified within integrative elements and phages of the E2348/69 chromosome (Iguchi et al., 2009) (Supplemental Table 4). Also, both of the EPEC isolates from phylogroup B2 (E2348/69 and C581-05) were identified with a greater number of genes that were differentially-expressed in HeLa compared to exponential growth ( $OD_{600} = 0.5$ ) in DMEM broth, than were differentially-expressed in DMEM compared to LB during exponential growth (Table 1). This is in contrast to the EPEC isolates from phylogroup B1 (B171 and E110019), which had similar numbers of genes that were differentially-expressed in the HeLa compared to DMEM, as there were in DMEM compared to LB (Table 1). This finding further indicates that EPEC prototype isolates exhibit phylogroup-specific differences and diverse transcriptional responses to conditions that facilitate pathogenesis.

## Conclusions

In this study, we have demonstrated that RNA-Seq and comparative genomics are powerful tools that can be used together to investigate the global transcriptome of an *E. coli* pathogen that has diverse genomic content. This approach can be used to provide insight into all genes that are simultaneously expressed with virulence factors and may directly or indirectly be involved in EPEC pathogenesis. By investigating differences in the global transcriptomes of EPEC isolates belonging to different evolutionary lineages of *E. coli*, we have demonstrated that genes comprising the unique genomic content of these isolates are in fact differentially-expressed during conditions that promote virulence factor expression. Thus, these studies highlight the need to investigate multiple genomically-diverse isolates rather than a

**TABLE 4 | Select genes that were differentially-expressed during adherence to HeLa compared to growth in DMEM broth.**

Gene ID <sup>a</sup>	Predicted protein	Protein accession no. <sup>a</sup>	Log <sub>2</sub> fold-change (LFC) in HeLa vs. DMEM (OD <sub>600</sub> = 0.5) <sup>b,c</sup>				LFC in DMEM vs. LB (OD <sub>600</sub> = 0.5) <sup>b,c</sup>
			E2348/69	C581-05	B171	E110019	
<b>DE genes of core gene clusters</b>							
<i>rbsD</i>	RbsD/FucU transport family protein	YP_002331517.1	-2.21	-4.52	-4.65	-2.96	Decreased in all, except increased in C581-05
<i>yliE</i>	Hypothetical protein	YP_002328350.1	-2.42	-6.14	-3.31	-4.27	NS
<i>slp</i>	Outer membrane protein Slp	YP_002331206.1	-4.60	-4.62	-3.67	2.73	NS
<i>yqjB</i>	SecD export N-terminal TM region family protein	YP_002330857.1	2.28	2.66	2.03	2.13	NS
<i>hscB</i>	Fe-S protein assembly co-chaperone HscB	YP_002330308.1	2.17	3.36	3.66	3.28	NS
<i>nirC</i>	Putative nitrite transporter	YP_002331085.1	2.60	2.06	-4.90	-7.24	NS
<b>DE GENES OF EXCLUSIVE GENE CLUSTERS</b>							
<b>E2348/69</b>							
<i>wzy</i>	O-antigen polymerase	YP_002329685.1	-5.92	NA	NA	NA	4.39
<i>rfaS</i>	Lipopolysaccharide core biosynthesis protein	YP_002331339.1	-6.63	NA	NA	NA	NS
<b>C581-05</b>							
<i>rfaG</i>	Glycosyl transferases group 1 family protein	WP_024235092.1	NA	-2.14	NA	NA	NS
<i>argR</i>	Arginine repressor, DNA binding domain protein	YP_002559154.1	NA	-2.37	NA	NA	NS
<i>yopT-like</i>	Cysteine protease, YopT-type domain protein	WP_024235005.1	NA	-2.75	NA	NA	NS
<b>B171</b>							
<i>mhpR</i>	Mhp operon transcriptional activator	EDX27960.1	NA	NA	2.77	NA	-4.25
<i>eamA</i>	EamA-like transporter family protein	WP_011251361.1	NA	NA	2.33	NA	NS
None	Transcriptional regulator, GntR family	EDX31174.1	NA	NA	-2.26	NA	-2.30
<b>E110019</b>							
<i>lutA</i>	Ferric aerobactin receptor LutA	EDV87696.1	NA	NA	NA	2.56	NS
<i>terA</i>	Tellurium resistance family protein	EDV86455.1	NA	NA	NA	-2.42	NS
EcE110019_4492	Sulfatase family protein	EDV86483.1	NA	NA	NA	-3.82	NS

<sup>a</sup>The gene symbol or locus id and the protein accession number are indicated for the top match protein. In some cases a protein match could not be identified for a gene cluster in a particular genome, which likely results from differences in the gene-calling that was used for LS-BSR compared to that used for the GenBank sequences. None indicates there was not a corresponding locus id for the particular genome.

<sup>b</sup>These are LFC values for samples that have been normalized for a single EPEC isolate, and have not been normalized across all EPEC isolates.

<sup>c</sup>NS indicates a value was not significant, while NA indicates a comparison was not applicable.

single prototype isolate when investigating virulence mechanisms of a pathovar. Identifying a global transcriptional response that is conserved among divergent EPEC isolates provides insight into the emergence of EPEC isolates in numerous *E. coli* lineages. The conserved transcriptional response of EPEC could be used to develop diagnostic tools to determine whether EPEC are contributing to an infection. This would be particularly useful for patients that carry multiple enteric pathogens, such as was demonstrated in the recent GEMS study (Kotloff et al., 2013) that identified the causative agents of diarrheal illness among young children and infants in numerous study sites in Africa and Asia.

## References

Abu-Ali, G. S., Ouellette, L. M., Henderson, S. T., Lacher, D. W., Riordan, J. T., Whittam, T. S., et al. (2010). Increased adherence and expression of virulence genes in a lineage of *Escherichia coli* O157:H7 commonly associated with human infections. *PLoS ONE* 5:e10167. doi: 10.1371/journal.pone.0010167

## Acknowledgments

This project was funded in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health grant numbers U19 AI090873, U19AI110820 and R37 AI021657.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.00569/abstract>

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P. T., and Huber, W. (2014). HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638



- Bergholz, T. M., Wick, L. M., Qi, W., Riordan, J. T., Ouellette, L. M., and Whittam, T. S. (2007). Global transcriptional response of *Escherichia coli* O157:H7 to growth transitions in glucose minimal medium. *BMC Microbiol.* 7:97. doi: 10.1186/1471-2180-7-97
- Bhatt, S., Edwards, A. N., Nguyen, H. T., Merlin, D., Romeo, T., and Kalman, D. (2009). The RNA binding protein CsrA is a pleiotropic regulator of the locus of enterocyte effacement pathogenicity island of enteropathogenic *Escherichia coli*. *Infect. Immun.* 77, 3552–3568. doi: 10.1128/IAI.00418-09
- Bhatt, S., Romeo, T., and Kalman, D. (2011). Honing the message: post-transcriptional and post-translational control in attaching and effacing pathogens. *Trends Microbiol.* 19, 217–224. doi: 10.1016/j.tim.2011.01.004
- Bingle, L. E., Constantinidou, C., Shaw, R. K., Islam, M. S., Patel, M., Snyder, L. A., et al. (2014). Microarray analysis of the Ler regulon in enteropathogenic and enterohaemorrhagic *Escherichia coli* strains. *PLoS ONE* 9:e80160. doi: 10.1371/journal.pone.0080160
- Brondijk, T. H., Fiegen, D., Richardson, D. J., and Cole, J. A. (2002). Roles of NapF, NapG and NapH, subunits of the *Escherichia coli* periplasmic nitrate reductase, in ubiquinol oxidation. *Mol. Microbiol.* 44, 245–255. doi: 10.1046/j.1365-2958.2002.02875.x
- Bustamante, V. H., Santana, F. J., Calva, E., and Puente, J. L. (2001). Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. *Mol. Microbiol.* 39, 664–678. doi: 10.1046/j.1365-2958.2001.02209.x
- Bustamante, V. H., Villalba, M. I., Garcia-Angulo, V. A., Vazquez, A., Martinez, L. C., Jimenez, R., et al. (2011). PerC and GrlA independently regulate Ler expression in enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 82, 398–415. doi: 10.1111/j.1365-2958.2011.07819.x
- Camarena, L., Bruno, V., Euskirchen, G., Poggio, S., and Snyder, M. (2010). Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA-sequencing. *PLoS Pathog.* 6:e1000834. doi: 10.1371/journal.ppat.1000834
- Cheney, C. P., Schad, P. A., Formal, S. B., and Boedeker, E. C. (1980). Species specificity of *in vitro* *Escherichia coli* adherence to host intestinal cell membranes and its correlation with *in vivo* colonization and infectivity. *Infect. Immun.* 28, 1019–1027.
- Chugani, S., Kim, B. S., Phattarasukol, S., Brittnacher, M. J., Choi, S. H., Harwood, C. S., et al. (2012). Strain-dependent diversity in the *Pseudomonas aeruginosa* quorum-sensing regulon. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2823–2831. doi: 10.1073/pnas.1214128109
- Clegg, S., Yu, F., Griffiths, L., and Cole, J. A. (2002). The roles of the polytopic membrane proteins NarK, NarU and NirC in *Escherichia coli* K-12: two nitrate and three nitrite transporters. *Mol. Microbiol.* 44, 143–155. doi: 10.1046/j.1365-2958.2002.02858.x
- Clements, A., Young, J. C., Constantinou, N., and Frankel, G. (2012). Infection strategies of enteric pathogenic *Escherichia coli*. *Gut Microbes* 3, 71–87. doi: 10.4161/gmic.19182
- Das, P., Lahiri, A., Lahiri, A., and Chakravorty, D. (2009). Novel role of the nitrite transporter NirC in *Salmonella* pathogenesis: SPI2-dependent suppression of inducible nitric oxide synthase in activated macrophages. *Microbiology* 155, 2476–2489. doi: 10.1099/mic.0.029611-0
- Deng, W., Puente, J. L., Gruenheid, S., Li, Y., Vallance, B. A., Vazquez, A., et al. (2004). Dissecting virulence: systematic and functional analyses of a pathogenicity island. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3597–3602. doi: 10.1073/pnas.0400326101
- Donnenberg, M. S., Tacket, C. O., James, S. P., Losonsky, G., Nataro, J. P., Wasserman, S. S., et al. (1993). Role of the *eaeA* gene in experimental enteropathogenic *Escherichia coli* infection. *J. Clin. Invest.* 92, 1412–1417. doi: 10.1172/JCI116717
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Elliott, S. J., Sperandio, V., Giron, J. A., Shin, S., Mellies, J. L., Wainwright, L., et al. (2000). The locus of enterocyte effacement (LEE)-encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infect. Immun.* 68, 6115–6126. doi: 10.1128/IAI.68.11.6115-6126.2000
- Galens, K., Orvis, J., Daugherty, S., Creasy, H. H., Angiuoli, S., White, O., et al. (2011). The IGS standard operating procedure for automated prokaryotic annotation. *Stand. Genomic Sci.* 4, 244–251. doi: 10.4056/signs.1223234
- Garmendia, J., Frankel, G., and Crepin, V. F. (2005). Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: translocation, translocation, translocation. *Infect. Immun.* 73, 2573–2585. doi: 10.1128/IAI.73.5.2573-2585.2005
- Gertz, E. M., Yu, Y. K., Agarwala, R., Schaffer, A. A., and Altschul, S. F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol.* 4:41. doi: 10.1186/1741-7007-4-41
- Gomez-Duarte, O. G., and Kaper, J. B. (1995). A plasmid-encoded regulatory region activates chromosomal *eaeA* expression in enteropathogenic *Escherichia coli*. *Infect. Immun.* 63, 1767–1776.
- Haack, K. R., Robinson, C. L., Miller, K. J., Fowlkes, J. W., and Mellies, J. L. (2003). Interaction of Ler at the LEE5 (*tir*) operon of enteropathogenic *Escherichia coli*. *Infect. Immun.* 71, 384–392. doi: 10.1128/IAI.71.1.384-392.2003
- Hardwidge, P. R., Deng, W., Vallance, B. A., Rodriguez-Escudero, I., Cid, V. J., Molina, M., et al. (2005). Modulation of host cytoskeleton function by the enteropathogenic *Escherichia coli* and *Citrobacter rodentium* effector protein EspG. *Infect. Immun.* 73, 2586–2594. doi: 10.1128/IAI.73.5.2586-2594.2005
- Hazen, T. H., Sahl, J. W., Fraser, C. M., Donnenberg, M. S., Scheutz, F., and Rasko, D. A. (2013). Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12810–12815. doi: 10.1073/pnas.1306836110
- Hoff, K. G., Silberg, J. J., and Vickery, L. E. (2000). Interaction of the iron-sulfur cluster assembly protein IscU with the Hsc66/Hsc20 molecular chaperone system of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7790–7795. doi: 10.1073/pnas.130201997
- Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Iguchi, A., Thomson, N. R., Ogura, Y., Saunders, D., Ooka, T., Henderson, I. R., et al. (2009). Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *J. Bacteriol.* 191, 347–354. doi: 10.1128/JB.01238-08
- Iyoda, S., and Watanabe, H. (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. doi: 10.1128/JB.187.12.4086-4094.2005
- Jandu, N., Ho, N. K., Donato, K. A., Karmali, M. A., Mascarenhas, M., Duffy, S. P., et al. (2009). Enterohemorrhagic *Escherichia coli* O157:H7 gene expression profiling in response to growth in the presence of host epithelia. *PLoS ONE* 4:e4889. doi: 10.1371/journal.pone.0004889
- Jerse, A. E., Yu, J., Tall, B. D., and Kaper, J. B. (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. doi: 10.1073/pnas.87.20.7839
- Kaper, J. B., Nataro, J. P., and Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2, 123–140. doi: 10.1038/nrmicro818
- Kendall, M. M., Gruber, C. C., Rasko, D. A., Hughes, D. T., and Sperandio, V. (2011). Hfq virulence regulation in enterohemorrhagic *Escherichia coli* O157:H7 strain 86-24. *J. Bacteriol.* 193, 6843–6851. doi: 10.1128/JB.06141-11
- Kenny, B., Abe, A., Stein, M., and Finlay, B. B. (1997). Enteropathogenic *Escherichia coli* protein secretion is induced in response to conditions similar to those in the gastrointestinal tract. *Infect. Immun.* 65, 2606–2612.
- Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Panchalingam, S., et al. (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. doi: 10.1016/S0140-6736(13)60844-2
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Landstorfer, R., Simon, S., Schober, S., Keim, D., Scherer, S., and Neuhaus, K. (2014). Comparison of strand-specific transcriptomes of enterohemorrhagic *Escherichia coli* O157:H7 EDL933 (EHEC) under eleven different environmental conditions including radish sprouts and cattle feces. *BMC Genomics* 15:353. doi: 10.1186/1471-2164-15-353
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

- Leverton, L. Q., and Kaper, J. B. (2005). Temporal expression of enteropathogenic *Escherichia coli* virulence genes in an *in vitro* model of infection. *Infect. Immun.* 73, 1034–1043. doi: 10.1128/IAI.73.2.1034-1043.2005
- Levine, M. M., Bergquist, E. J., Nalin, D. R., Waterman, D. H., Hornick, R. B., Young, C. R., et al. (1978). *Escherichia coli* strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. *Lancet* 1, 1119–1122. doi: 10.1016/S0140-6736(78)90299-4
- Levine, M. M., Nataro, J. P., Karch, H., Baldini, M. M., Kaper, J. B., Black, R. E., et al. (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. doi: 10.1093/infdis/152.3.550
- Levine, M. M., and Rennels, M. B. (1978). *E. coli* colonisation factor antigen in diarrhoea. *Lancet* 2, 534.
- Mandlik, A., Livny, J., Robins, W. P., Ritchie, J. M., Mekalanos, J. J., and Waldor, M. K. (2011). RNA-Seq-based monitoring of infection-linked changes in *Vibrio cholerae* gene expression. *Cell Host Microbe* 10, 165–174. doi: 10.1016/j.chom.2011.07.007
- Martin, J. A., and Wang, Z. (2011). Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12, 671–682. doi: 10.1038/nrg3068
- McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S., and Kaper, J. B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 92, 1664–1668. doi: 10.1073/pnas.92.5.1664
- McDaniel, T. K., and Kaper, J. B. (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. doi: 10.1046/j.1365-2958.1997.2311591.x
- Mellies, J. L., Barron, A. M., and Carmona, A. M. (2007). Enteropathogenic and enterohemorrhagic *Escherichia coli* virulence gene regulation. *Infect. Immun.* 75, 4199–4210. doi: 10.1128/IAI.01927-06
- Mellies, J. L., Elliott, S. J., Sperandio, V., Donnenberg, M. S., and Kaper, J. B. (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. doi: 10.1046/j.1365-2958.1999.01473.x
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Nataro, J. P., and Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11, 142–201.
- Navarro-Garcia, F., Canizalez-Roman, A., Sui, B. Q., Nataro, J. P., and Azamar, Y. (2004). The serine protease motif of EspC from enteropathogenic *Escherichia coli* produces epithelial damage by a mechanism different from that of Pet toxin from enteroaggregative *E. coli*. *Infect. Immun.* 72, 3609–3621. doi: 10.1128/IAI.72.6.3609-3621.2004
- Nisa, S., Scanlon, K. M., and Donnenberg, M. S. (2013). “Enteropathogenic *Escherichia coli*” in *Escherichia coli Pathotypes and Principles of Pathogenesis, 2nd Edn.*, ed M. S. Donnenberg (Oxford, UK: Academic Press).
- Njoroge, J. W., Gruber, C., and Sperandio, V. (2013). The interacting Cra and KdpE regulators are involved in the expression of multiple virulence factors in enterohemorrhagic *Escherichia coli*. *J. Bacteriol.* 195, 2499–2508. doi: 10.1128/JB.02252-12
- Njoroge, J. W., Nguyen, Y., Curtis, M. M., Moreira, C. G., and Sperandio, V. (2012). Virulence meets metabolism: Cra and KdpE gene regulation in enterohemorrhagic *Escherichia coli*. *MBio* 3, e00280–e00212. doi: 10.1128/mBio.00280-12
- Ochoa, T. J., and Contreras, C. A. (2011). Enteropathogenic *Escherichia coli* infection in children. *Curr. Opin. Infect. Dis.* 24, 478–483. doi: 10.1097/QCO.0b013e32834a8b8b
- Orvis, J., Crabtree, J., Galens, K., Gussman, A., Inman, J. M., Lee, E., et al. (2010). Ergatis: a web interface and scalable software system for bioinformatics workflows. *Bioinformatics* 26, 1488–1492. doi: 10.1093/bioinformatics/btq167
- Ozsolak, F., and Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12, 87–98. doi: 10.1038/nrg2934
- Pacheco, A. R., Curtis, M. M., Ritchie, J. M., Munera, D., Waldor, M. K., Moreira, C. G., et al. (2012). Fucose sensing regulates bacterial intestinal colonization. *Nature* 492, 113–117. doi: 10.1038/nature11623
- Perkins, T. T., Kingsley, R. A., Fookes, M. C., Gardner, P. P., James, K. D., Yu, L., et al. (2009). A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genet.* 5:e1000569. doi: 10.1371/journal.pgen.1000569
- Porter, M. E., Mitchell, P., Roe, A. J., Free, A., Smith, D. G., and Gally, D. L. (2004). Direct and indirect transcriptional activation of virulence genes by an AraC-like protein, PerA from enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 54, 1117–1133. doi: 10.1111/j.1365-2958.2004.04333.x
- Puente, J. L., Bieber, D., Ramer, S. W., Murray, W., and Schoolnik, G. K. (1996). The bundle-forming pili of enteropathogenic *Escherichia coli*: transcriptional regulation by environmental signals. *Mol. Microbiol.* 20, 87–100. doi: 10.1111/j.1365-2958.1996.tb02491.x
- Rasko, D. A., Rosovitz, M. J., Myers, G. S., Mongodin, E. F., Fricke, W. F., Gajer, P., et al. (2008). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* 190, 6881–6893. doi: 10.1128/JB.00619-08
- Riley, L. W., Junio, L. N., Libaek, L. B., and Schoolnik, G. K. (1987). Plasmid-encoded expression of lipopolysaccharide O-antigenic polysaccharide in enteropathogenic *Escherichia coli*. *Infect. Immun.* 55, 2052–2056.
- Riley, L. W., Junio, L. N., and Schoolnik, G. K. (1990). HeLa cell invasion by a strain of enteropathogenic *Escherichia coli* that lacks the O-antigenic polysaccharide. *Mol. Microbiol.* 4, 1661–1666. doi: 10.1111/j.1365-2958.1990.tb00543.x
- Romeo, T., Gong, M., Liu, M. Y., and Brun-Zinkernagel, A. M. (1993). Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J. Bacteriol.* 175, 4744–4755.
- Rosenshine, I., Ruschkowski, S., and Finlay, B. B. (1996). Expression of attaching/effacing activity by enteropathogenic *Escherichia coli* depends on growth phase, temperature, and protein synthesis upon contact with epithelial cells. *Infect. Immun.* 64, 966–973.
- Saeed, A. I., Bhagabati, N. K., Braisted, J. C., Liang, W., Sharov, V., Howe, E. A., et al. (2006). TM4 microarray software suite. *Meth. Enzymol.* 411, 134–193. doi: 10.1016/s0076-6879(06)11009-5
- Sahl, J. W., Caporaso, J. G., Rasko, D. A., and Keim, P. (2014). The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* 2:e332. doi: 10.7717/peerj.332
- Sahl, J. W., and Rasko, D. A. (2012). Analysis of Global Transcriptional Profiles of Enterotoxigenic *Escherichia coli* Isolate E24377A. *Infect. Immun.* 80, 1232–1242. doi: 10.1128/IAI.06138-11
- Sarkar, S., Ulett, G. C., Totsika, M., Phan, M. D., and Schembri, M. A. (2014). Role of capsule and O antigen in the virulence of uropathogenic *Escherichia coli*. *PLoS ONE* 9:e94786. doi: 10.1371/journal.pone.0094786
- Sharma, C. M., Hoffmann, S., Darfeuille, F., Reigier, J., Findeiss, S., Sittka, A., et al. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464, 250–255. doi: 10.1038/nature08756
- Shaw, R. K., Smollett, K., Cleary, J., Garmendia, J., Straatman-Iwanowska, A., Frankel, G., et al. (2005). Enteropathogenic *Escherichia coli* type III effectors EspG and EspG2 disrupt the microtubule network of intestinal epithelial cells. *Infect. Immun.* 73, 4385–4390. doi: 10.1128/IAI.73.7.4385-4390.2005
- Shin, S., Castanie-Cornet, M. P., Foster, J. W., Crawford, J. A., Brinkley, C., and Kaper, J. B. (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. doi: 10.1046/j.1365-2958.2001.02570.x
- Sircilli, M. P., Walters, M., Trabuasi, L. R., and Sperandio, V. (2004). Modulation of enteropathogenic *Escherichia coli* virulence by quorum sensing. *Infect. Immun.* 72, 2329–2337. doi: 10.1128/IAI.72.4.2329-2337.2004
- Sperandio, V., Mellies, J. L., Delahay, R. M., Frankel, G., Crawford, J. A., Nguyen, W., et al. (2000). Activation of enteropathogenic *Escherichia coli* (EPEC) LEE2 and LEE3 operons by Ler. *Mol. Microbiol.* 38, 781–793. doi: 10.1046/j.1365-2958.2000.02168.x
- Sperandio, V., Mellies, J. L., Nguyen, W., Shin, S., and Kaper, J. B. (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. doi: 10.1073/pnas.96.26.15196
- Tenailon, O., Skurnik, D., Picard, B., and Denamur, E. (2010). The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* 8, 207–217. doi: 10.1038/nrmicro2298

- Tobe, T., Hayashi, T., Han, C. G., Schoolnik, G. K., Ohtsubo, E., and Sasakawa, C. (1999). Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. *Infect. Immun.* 67, 5455–5462.
- Tobe, T., Schoolnik, G. K., Sohel, I., Bustamante, V. H., and Puente, J. L. (1996). Cloning and characterization of *bfpTVW*, genes required for the transcriptional activation of *bfpA* in enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 21, 963–975. doi: 10.1046/j.1365-2958.1996.531415.x
- Tomson, F. L., Viswanathan, V. K., Kanack, K. J., Kanteti, R. P., Straub, K. V., Menet, M., et al. (2005). Enteropathogenic *Escherichia coli* EspG disrupts microtubules and in conjunction with Orf3 enhances perturbation of the tight junction barrier. *Mol. Microbiol.* 56, 447–464. doi: 10.1111/j.1365-2958.2005.04571.x
- Vickery, L. E., Silberg, J. J., and Ta, D. T. (1997). Hsc66 and Hsc20, a new heat shock cognate molecular chaperone system from *Escherichia coli*. *Protein Sci.* 6, 1047–1056. doi: 10.1002/pro.5560060511
- Vidal, J. E., and Navarro-Garcia, F. (2008). EspC translocation into epithelial cells by enteropathogenic *Escherichia coli* requires a concerted participation of type V and III secretion systems. *Cell. Microbiol.* 10, 1975–1986. doi: 10.1111/j.1462-5822.2008.01181.x
- Viljanen, M. K., Peltola, T., Junnila, S. Y., Olkkonen, L., Jarvinen, H., Kuistila, M., et al. (1990). Outbreak of diarrhoea due to *Escherichia coli* O111:B4 in schoolchildren and adults: association of Vi antigen-like reactivity. *Lancet* 336, 831–834. doi: 10.1016/0140-6736(90)92337-H
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Wong, A. R., Pearson, J. S., Bright, M. D., Munera, D., Robinson, K. S., Lee, S. F., et al. (2011). Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. *Mol. Microbiol.* 80, 1420–1438. doi: 10.1111/j.1365-2958.2011.07661.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Hazen, Daugherty, Shetty, Mahurkar, White, Kaper and Rasko. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.