



Transcriptional Variation of Diverse Enteropathogenic *Escherichia coli* Isolates under Virulence-Inducing Conditions

Tracy H. Hazen,^{a,b} Sean C. Daugherty,^a Amol C. Shetty,^a James P. Nataro,^c
David A. Rasko^{a,b}

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

ABSTRACT Enteropathogenic *Escherichia coli* (EPEC) bacteria are a diverse group of pathogens that cause moderate to severe diarrhea in young children in developing countries. EPEC isolates can be further subclassified as typical EPEC (tEPEC) isolates that contain the bundle-forming pilus (BFP) or as atypical EPEC (aEPEC) isolates that do not contain BFP. Comparative genomics studies have recently highlighted the considerable genomic diversity among EPEC isolates. In the current study, we used RNA sequencing (RNA-Seq) to characterize the global transcriptomes of eight tEPEC isolates representing the identified genomic diversity, as well as one aEPEC isolate. The global transcriptomes were determined for the EPEC isolates under conditions of laboratory growth that are known to induce expression of virulence-associated genes. The findings demonstrate that unique genes of EPEC isolates from diverse phylogenomic lineages contribute to variation in their global transcriptomes. There were also phylogroup-specific differences in the global transcriptomes, including genes involved in iron acquisition, which had significant differential expression in the EPEC isolates belonging to phylogroup B2. Also, three EPEC isolates from the same phylogenomic lineage (EPEC8) had greater levels of similarity in their genomic content and exhibited greater similarities in their global transcriptomes than EPEC from other lineages; however, even among closely related isolates there were isolate-specific differences among their transcriptomes. These findings highlight the transcriptional variability that correlates with the previously unappreciated genomic diversity of EPEC.

IMPORTANCE Recent studies have demonstrated that there is considerable genomic diversity among EPEC isolates; however, it is unknown if this genomic diversity leads to differences in their global transcription. This study used RNA-Seq to compare the global transcriptomes of EPEC isolates from diverse phylogenomic lineages. We demonstrate that there are lineage- and isolate-specific differences in the transcriptomes of genomically diverse EPEC isolates during growth under *in vitro* virulence-inducing conditions. This study addressed biological variation among isolates of a single pathovar in an effort to demonstrate that while each of these isolates is considered an EPEC isolate, there is significant transcriptional diversity among members of this pathovar. Future studies should consider whether this previously undescribed transcriptional variation may play a significant role in isolate-specific variability of EPEC clinical presentations.

KEYWORDS EPEC, *Escherichia coli*, diversity, transcriptome

Received 28 March 2017 Accepted 6 May 2017 Published 25 July 2017

Citation Hazen TH, Daugherty SC, Shetty AC, Nataro JP, Rasko DA. 2017. Transcriptional variation of diverse enteropathogenic *Escherichia coli* isolates under virulence-inducing conditions. *mSystems* 2:e00024-17. <https://doi.org/10.1128/mSystems.00024-17>.

Editor Katrine L. Whiteson, University of California, Irvine

Copyright © 2017 Hazen et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to David A. Rasko, drasko@som.umaryland.edu.

For a commentary on this article, see <https://doi.org/10.1128/mSystems.00049-17>.

Enteropathogenic *Escherichia coli* (EPEC) is a leading cause of moderate to severe diarrhea among young children, particularly in developing countries (1). EPEC strains are characterized by the presence of the locus of enterocyte effacement (LEE) region and are subclassified as typical EPEC (tEPEC) or atypical EPEC (aEPEC) by the presence or absence of the bundle-forming pilus (BFP), respectively (2–4). The LEE region is a characteristic feature of both EPEC and enterohemorrhagic *E. coli* (EHEC), which includes the O157:H7 and non-O157 EHEC that are a significant cause of foodborne outbreaks in the United States (2, 5).

The LEE region encodes the intimin adherence protein, the translocated intimin receptor protein (Tir), and a type III secretion system (T3SS), which have been identified as major components of the EPEC and EHEC virulence mechanisms (2, 3, 6–8). The intimin and Tir proteins, as well as the T3SS, are involved in attachment to host cells and the translocation of effector proteins that confer changes in the host cell (6, 7). BFP is a type IV pilus that is involved in the localized adherence to the host cell, which is a unique feature that is found in tEPEC and not in aEPEC or EHEC (2, 3).

Global views of bacterial transcriptomes have provided insight into genome-wide virulence gene regulation, as well as the identification of novel virulence factors (9–12). RNA sequencing (RNA-Seq) provides an unbiased high-throughput sequencing approach that can capture the global transcriptional response of an organism under particular growth or environmental conditions (13–16). RNA-Seq has been used to investigate the global transcriptomes of numerous pathogenic bacteria (9, 17–21) and also to examine the transcriptome of different *E. coli* pathovars, specifically EHEC (21–23) and ETEC (9). RNA-Seq has also been used to investigate differences in the transcriptional responses of genomically diverse commensal and environmental *E. coli* isolates during growth under multiple laboratory conditions (24).

The EPEC virulence mechanisms have been extensively characterized for a limited number of EPEC isolates (25–30). However, considering the genomic diversity that has been described in recent years for EPEC isolates (31–33), this raised the issue of whether EPEC isolates have greater variability in the transcription of their virulence mechanisms. To investigate whether there are differences in the global transcriptomes of three frequently studied EPEC reference isolates, we previously used RNA-Seq to identify genes that were coordinately expressed under multiple laboratory growth conditions (34). Included among the isolates examined were the frequently studied E2348/69 and B171 archetype EPEC strains, which have been used in many previous studies to characterize EPEC virulence mechanisms (25, 26, 35–40). The transcriptional study by Hazen et al. included the global transcriptomes of the archetype aEPEC isolate E110019 (27) and of a recently described tEPEC isolate, C581-05, which belongs to a *E. coli* phylogenomic lineage that is different from that of the archetype strains (31, 34). The global transcriptomes of these four EPEC isolates were compared during growth in multiple medium types and three different growth phases (early exponential, late exponential, and early stationary phase) (34). The findings demonstrated that these EPEC isolates exhibited isolate-, medium-, and growth-phase-specific differences in their global transcriptomes (34). Surprisingly, there were also differences in the timing of expression of the LEE genes, the key defining virulence factor, among these four EPEC isolates (34).

The current report describes the application of RNA-Seq to investigate differences in the global transcriptomes of nine phylogenomically diverse EPEC isolates representing eight EPEC lineages from three *E. coli* phylogroups (B2, B1, and A) (34, 41, 42). Among the EPEC isolates analyzed in the current study were E2348/69, B171, C581-05, and E110019 (34). The additional five EPEC isolates in the current study originated from the Global Enteric Multicenter Study (GEMS) (43) and belong to unexplored EPEC phylogenomic lineages. By including GEMS isolates, we were able to consider the genomic variation of contemporary circulating EPEC isolates that have caused illness within the last decade in countries in Africa, where tEPEC is associated with increased lethality among infants with diarrheal illness (44). Inclusion of the more recent and genomically diverse human EPEC isolates is critical for our deeper understanding of the virulence

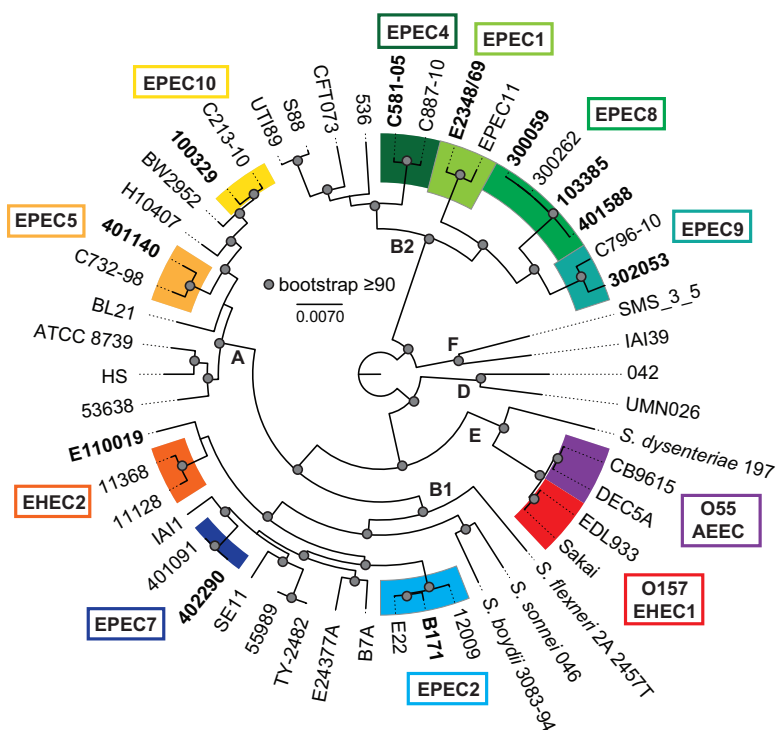


FIG 1 Phylogenomic analysis of representative EPEC isolates. The genome sequences of representative EPEC isolates were compared with those of a reference collection of diverse *E. coli* and *Shigella* isolates that had been sequenced previously and are available in the public domain. The genomes were aligned using Mugsy (95) as previously described (31, 94). A 1.9-Mb aligned region was used to generate a maximum-likelihood phylogeny with 100 bootstrap replicates using RAxML v.7.2.8 (97), and the results were visualized using FigTree v.1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). The representative EPEC isolates that were analyzed using RNA-Seq are indicated in bold.

mechanisms of modern isolates that have not been passaged many times in the laboratory, which could potentially introduce mutations that would alter virulence phenotypes. For instance, the EPEC archetype strain, E2348/69, was initially isolated in 1969 and, as we recently described, there are multiple versions of this strain in existence that exhibit genomic differences along with altered growth and virulence phenotypes (45).

This report identifies phylogroup-, lineage-, and isolate-specific differences in the global transcriptomes of genomically diverse EPEC isolates. Overall, this report addresses the biological variation among the isolates of a single *E. coli* pathovar and demonstrates that, while each of these isolates is considered an EPEC isolate, and a great deal is known about some of the isolates, very little is known about the transcriptional diversity of the members of this pathovar or of the *E. coli* species in general.

RESULTS

Phylogenomic analysis of the representative EPEC isolates. Phylogenomic analysis of a collection of *E. coli* isolates classified as EPEC based on the presence of the LEE and/or BFP regions has previously demonstrated that isolates from the EPEC pathovar occur in numerous phylogenomic lineages (31–33). The purpose of this study was to investigate the impact of genomic diversity on the global transcriptional regulons of isolates in diverse phylogenomic lineages during growth under laboratory conditions that promote virulence factor expression (46–48). RNA-Seq was used to analyze the global transcriptomes of nine EPEC isolates (E2348/69, B171, C581-05, 401140, 402290, 401588, 302053, 100329, and E110019) representing eight EPEC phylogenomic lineages and three *E. coli* phylogroups (A, B1, and B2) (Fig. 1; Table 1). The EPEC isolates analyzed

TABLE 1 Number of shared or unique genes identified using LS-BSR analysis^a

Isolate ID	Location of isolation	Date of isolation	Phylogenomic lineage ^b	Phylogroup ^c	No. of isolate-specific gene clusters ^d
E2348/69	England	1969	EPEC1	B2	206
B171	United States	1983	EPEC2	B1	131
C581-05	Africa	NK	EPEC4	B2	141
401140	Kenya	2008	EPEC5	A	90
402290	Kenya	2009	EPEC7	B1	169
401588	Kenya	2008	EPEC8	B2	212
302053	Mozambique	2009	EPEC9	B2	53
100329	The Gambia	2008	EPEC10	A	62
E110019	Finland	1987	None	B1	164

^aID, identifier; NK, date of isolation not known.

^bThe total number of core gene clusters (LS-BSR \geq 0.8) in all EPEC isolates.

^cThe numbers of gene clusters with significant similarity (LS-BSR \geq 0.8) in all genomes of one phylogroup that were divergent (LS-BSR $<$ 0.8, \geq 0.4) or absent (LS-BSR $<$ 0.4) from genomes of other phylogroups were 44 (phylogroup A), 62 (phylogroup B1), and 128 (phylogroup B2).

^dThe isolate-specific genes are those identified in one genome with an LS-BSR \geq 0.8 and in the other genomes with an LS-BSR $<$ 0.4.

in this study included the frequently studied E2348/69 and B171 EPEC reference isolates from the EPEC1 and EPEC2 phylogenomic lineages, respectively, as well as archetype α EPEC isolate E110019 (34) (Fig. 1). Also, we included a previously characterized isolate, C581-05 (31, 34), from the EPEC4 lineage (49). The nonarchetype EPEC isolates from the GEMS (32, 43) were selected as genomic representatives of other diverse EPEC phylogenomic lineages (32). Where available, we selected tEPEC isolates that contained a complete or nearly complete BFP region. These isolates belong to the previously identified EPEC5 phylogenomic lineage or to the newly defined EPEC7, EPEC8, EPEC9, and EPEC10 phylogenomic lineages, which were designated based on phylogenomic analysis of 70 EPEC isolates from the GEMS (32).

Comparative genomic analysis of the EPEC isolates. To understand the potential transcriptional variation among the genomically diverse EPEC isolates, we first performed a comparative genomic large-scale BLAST score ratio (LS-BSR) analysis (50). Genomic comparison of the nine EPEC genome sequences identified 2,989 gene clusters that had significant similarity (LS-BSR \geq 0.8) in all genomes, representing a conserved core EPEC genome (Table 1). The number of phylogroup-specific gene clusters identified among the EPEC isolates ranged from 44 to 128 (Table 1). The phylogroup-specific gene clusters were those identified with significant similarity (LS-BSR \geq 0.8) in all genomes of each of the *E. coli* phylogroups (A, B1, or B2) that were divergent (LS-BSR $<$ 0.8 and \geq 0.4) or absent (LS-BSR $<$ 0.4) from the genomes of the other phylogroups. The number of isolate-specific gene clusters that were specific to one of the genomes and were absent (LS-BSR $<$ 0.4) from all other genomes ranged from 53 to 212 (Table 1). These numbers of genes were similar to those identified for other comparisons of *E. coli* (31, 32, 51).

Characteristics of the RNA-Seq samples. The global transcriptomes of each of the nine EPEC isolates were characterized using RNA-Seq during growth in lysogeny broth (LB) or Dulbecco's modified Eagle's medium (DMEM) to determine the interisolate and interphylogenomic lineage variation. Growth in nutrient-limited DMEM has previously been demonstrated to induce virulence factor expression of EPEC compared to growth in nutrient-rich LB, which does not promote expression of the majority of the primary EPEC virulence factors (46–48). The global transcriptomes were also characterized for three EPEC8 phylogenomic lineage isolates under the same growth conditions (Fig. 1; see also Table S2 in the supplemental material). These comparisons provide insight into the variation of the global transcriptomes of isolates of the same phylogenomic lineage under the same growth conditions. RNA-Seq was performed on a total of 44 RNA samples, generating approximately 2.6 billion Illumina sequence reads across all samples. The number of total reads and number of mapped reads are provided in Table S2. The number of reads that mapped to protein-coding (genic) regions ranged from 5 to

15 million reads per sample (14% to 56% of the total mapped reads), with an average of 10 million reads per sample mapping to protein-coding genes (Table S2). The approximate sequence coverage of mapped reads to each genome across all the samples ranged from 104× to 303×, with an average level of coverage of 202× (Table S2).

Principal-component analysis of the RNA-Seq reads for each sample of the nine EPEC isolates demonstrated there was a correlation among the biological replicates for each medium type (Fig. 2A). This demonstrated that the medium type had a significant impact on the global transcriptomes of the EPEC, as all of the LB samples (triangles in Fig. 2A) and all of the DMEM samples (circles in Fig. 2A) grouped together (Fig. 2A). Hierarchical cluster analysis of the read counts for 674 gene clusters that were identified in all nine EPEC isolates and exhibited the greatest deviation of expression also demonstrated a similar correlation among the biological replicates and samples of the same medium type (Fig. 2B, red and blue labels).

Identification of the core EPEC transcriptional regulon. The total number of genes of each EPEC isolate that exhibited significant differential expression during growth in DMEM compared to LB ranged from 334 to 572 (Table 2), representing between ~6% and 10% of the genome. Among these, there were 145 to 278 genes that exhibited increased expression and 134 to 315 genes that exhibited decreased expression (Table 2). A total of 242 to 394 of the differentially expressed genes of each EPEC isolate were identified as being part of the conserved EPEC core gene set ($LS\text{-}BSR \geq 0.8$ in all nine of the EPEC isolates) (Table 2). Interestingly, the core genes represented more than half of the total differentially expressed genes of each of the EPEC isolates examined (Table 2); however, there were only 21 conserved core genes that exhibited significant differential expression in all nine of the EPEC isolates (Table S3). The majority of the remaining core genes that were differentially expressed during growth in DMEM compared to LB encoded proteins involved in central metabolism such as glycerol-3-phosphate and biotin biosynthesis (Table S3). This demonstrates that a small number of highly conserved genes comprise the core regulon of genomically diverse EPEC isolates under these virulence-inducing laboratory conditions. Furthermore, this core regulon does not include most of the virulence factors hypothesized to be regulated under these growth conditions.

Genes involved in biotin synthesis, including *bioD*, had increased expression in all nine EPEC isolates in both the RNA-Seq and quantitative reverse transcriptase PCR (qRT-PCR) analyses (Table S3; see also Fig. S1 in the supplemental material). Also, other genes involved in central metabolism (*nark*, *glpD*, and *treB*) exhibited decreased expression in all nine EPEC isolates during growth in DMEM compared to LB (Table S3; Fig. S1). Comparison of the LFC values of all differentially expressed genes in each of the EPEC isolates demonstrated there were regions within each genome that exhibited similar expression trends among multiple isolates but not all isolates (Fig. 3).

Phylogroup- and isolate-specific gene expression. Beyond the core conserved gene clusters, we wanted to identify the gene clusters that were present and transcriptionally active in the specific phylogroups, phylogenomic lineages, or only in individual isolates. In this study, the number of phylogroup-specific genes ranged from 44 to 128, depending on the EPEC isolate (Table 1); however, the number of genes that were phylogroup-specific and also exhibited altered expression ranged from 0 to 18 (Table 2), depending on the isolate. The phylogroup-specific gene clusters that were differentially expressed included genes encoding conserved hypothetical proteins and genes involved in iron acquisition (Table S3). We anticipated that we would find a greater number of genes that were identified as phylogroup specific, but this was not the case, suggesting a more significant contribution of isolate-specific genes to variations in the global transcriptomes.

The number of isolate-specific genes ranged from 53 to 212 in the nine representative EPEC isolates (Table 1) and included several previously characterized virulence genes (Table S3). The T3SS secreted effector gene, *espV*, was identified and found to be

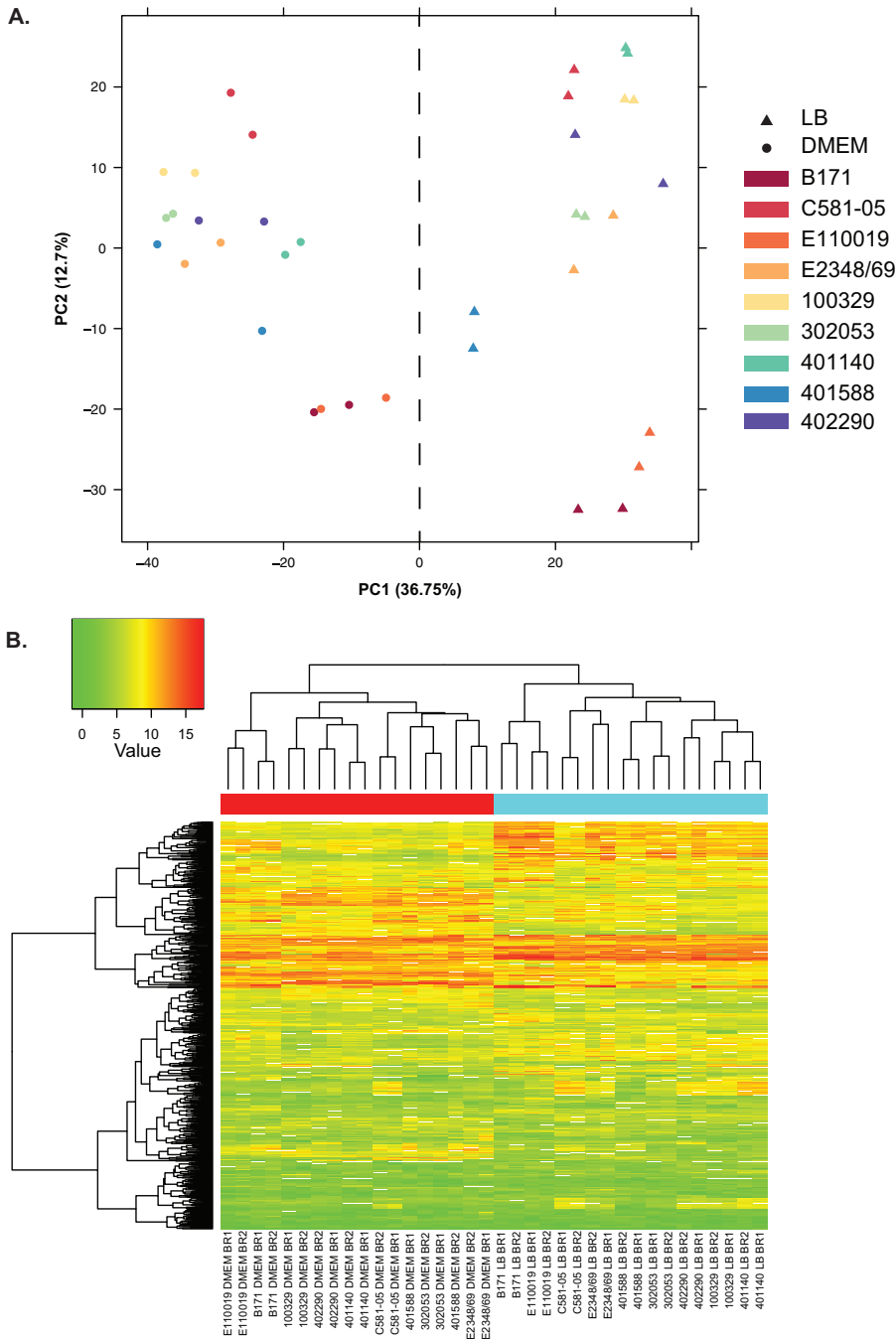


FIG 2 Principal-component analysis and hierarchical cluster analysis of the RNA-Seq samples examined in this study. (A) Principal-component analysis of the expression of gene clusters identified in the EPEC isolates, comparing all RNA-Seq samples analyzed. The first (PC1) and second (PC2) principal components were visualized in a scatterplot to demonstrate the clustering of the strains by gene content and gene expression. The comparative matrix is composed of the maximum normalized expression values for each gene cluster in each of the isolates and under each of the conditions examined for the gene clusters that were present among all isolates. The first component contains 36.75% of the variation, and the second component is responsible for 12.7% of the variation. The samples are colored by isolate, and symbols represent LB or DMEM as indicated in the legend. (B) A heatmap with clustering analysis of the expression values was constructed for the 674 LS-BSR gene clusters that were present in all of the EPEC isolates and had the greatest standard deviations of expression values. The normalized gene expression values were used to compute the standard deviation for each LS-BSR gene cluster across all samples. The heatmap was constructed using the R package gplots v2.11.0. A red label at the top of the heatmap designates the DMEM samples, while a blue label designates the LB samples.

TABLE 2 Number of genes that were differentially expressed in the EPEC isolates examined in this study

Isolate ID	Phylogenomic lineage ^a	Phylogroup ^a	LFC ≥ 2 ^b	LFC ≤ -2 ^b	Total DE genes ^c	No. of DE genes of core clusters ^d	No. of DE genes of phylogroup-specific clusters ^e	No. of DE genes of isolate-specific clusters ^f	Total DE sRNAs ^g
E2348/69	EPEC1	B2	180	251	431	253	18	10	6
B171	EPEC2	B1	220	255	475	249	2	4	10
C581-05	EPEC4	B2	162	235	397	291	12	1	17
401140	EPEC5	A	145	189	334	242	1	3	22
402290	EPEC7	B1	243	268	511	354	1	2	7
401588	EPEC8	B2	267	134	401	253	11	4	12
302053	EPEC9	B2	228	315	543	392	15	0	21
100329	EPEC10	A	278	294	572	394	1	2	30
E110019	None	B1	172	246	418	280	0	7	9

^aThe phylogenomic lineage and phylogroup are those that have been previously described (Hazen et al. [32], Jaureguy et al. [41], Tenailon et al. [42]).

^bLFC, log₂-fold change of the genes that exhibited significant (LFC ≥ 2 or ≤ -2 and FDR ≤ 0.05) differential expression (DE).

^cThe total number of genes that exhibited significant (LFC ≥ 2 or ≤ -2 and FDR ≤ 0.05) DE.

^dThe total number of core gene clusters (LS-BSR ≥ 0.8 in all genomes) was 2,989.

^eThe number of clusters in all EPEC genomes of one phylogroup that were divergent or absent (LS-BSR < 0.8) from EPEC genomes of the other phylogroups were 44 (phylogroup A), 62 (phylogroup B1), and 128 (phylogroup B2).

^fThe isolate-specific genes are those that were in one genome with an LS-BSR ≥ 0.8 and in the other genomes with an LS-BSR < 0.4.

^gThe total number of sRNA that were previously investigated in *E. coli* by Raghavan et al. (60) and exhibited significant (LFC ≥ 2 or ≤ -2 and FDR ≤ 0.05) DE in each of the nine EPEC isolates.

differentially expressed only in aEPEC isolate E110019, while the cytolethal distending toxin genes, *cdtAB*, were identified and found to be differentially expressed only in tEPEC isolate 401140 (Table S3). These results highlight the potential contribution of isolate-specific genes to the global transcriptomes of these genomically diverse EPEC isolates. Further investigation is necessary to determine whether particular isolate-specific genes are global transcriptional regulators that can explain differences in the transcription of highly conserved genes among the diverse EPEC isolates.

Within-lineage conservation and variability of the global transcriptomes of EPEC8 isolates. We also used RNA-Seq to examine two additional EPEC8 isolates (103385 and 300059) in order to investigate whether there was variation in the global transcriptomes of EPEC isolates within the same phylogenomic lineage (Table S2; Fig. 4). Comparison of the three EPEC8 isolates (103385, 300059, and 401588) using LS-BSR identified 4,359 core gene clusters (Fig. 4B). In contrast, there were only 49, 162, and 170 isolate-specific gene clusters that were present with significant similarity (LS-BSR ≥ 0.8) in only one of the three EPEC8 isolates and were divergent (LS-BSR < 0.8 and ≥ 0.4) or absent (LS-BSR < 0.4) from the other EPEC8 isolates (Fig. 4B). The number of gene clusters that were present with significant similarity (LS-BSR ≥ 0.8) in two of the EPEC8 isolates but divergent in (LS-BSR < 0.8 but ≥ 0.4) or absent from (LS-BSR < 0.4) the third isolate ranged from 50 to 229 (Fig. 4B). These data demonstrate that while the isolates were EPEC and were within the same phylogenomic lineage, they still contained considerable genomic variation. The total number of genes that exhibited significant differential expression in each of the three EPEC8 isolates ranged from 401 to 635 (Fig. 4A). Of these, there were 123 core gene clusters that exhibited significant differential expression in all three of the EPEC8 isolates (Fig. 4C). As anticipated, this number of common and expressed genes is greater than that determined for all of the EPEC isolates but were still relatively few. Among these core and expressed gene clusters were genes encoding predicted proteins involved in metabolism and iron uptake and genes encoding numerous hypothetical proteins (Table S4). There were 4, 6, and 20 gene clusters that were identified in only one of the EPEC8 isolates that also had significant differential expression (Fig. 4C). Among these isolate-specific genes were genes encoding putative proteins involved in iron uptake and also phage-associated genes (Table S4). These findings demonstrate that while there is greater similarity in the global transcriptomes of EPEC isolates belonging to the same phylogenomic lineage, the isolate-specific genes that are unique to each of the isolates contribute to differences in their global transcriptomes.

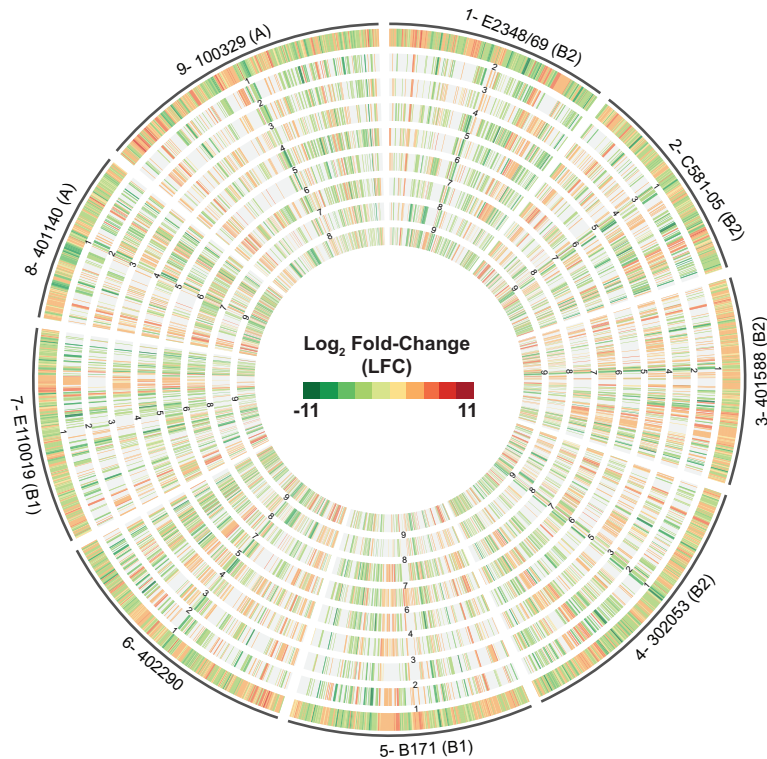


FIG 3 Comparison of the global transcriptomes of nine EPEC isolates. A circular plot of the \log_2 -fold-change (LFC) values for genes that exhibited significant differential expression during exponential growth in DMEM compared to LB is shown. The outermost track contains all of the significant LFC values for each of the indicated nine EPEC isolates. The inner tracks are numbered to correspond to the same number of each EPEC isolate in the outermost track. The phylogroup that each EPEC isolate belongs to is indicated in parentheses following the isolate designation. For example, 5-B171 (B1) indicates that data tracks containing B171 data are labeled track 5 in all of the comparisons, while B171 is the isolate designation and B1 the phylogroup designation. The inner tracks contain the LFC values of genes of another of the EPEC isolates belonging to the same LS-BSR gene cluster as the genes in the outermost reference track. The genes that were not identified in the other EPEC isolates or did not exhibit significant differential expression are absent from the inner tracks.

Differential expression of known EPEC virulence genes. To determine the transcriptional patterns of known EPEC virulence genes in these genomically diverse EPEC isolates, we compared the trends of expression of the LEE and BFP genes in each of the isolates (Fig. 5). Similar to what was observed with the four archetype EPEC isolates (34), there were differences in the expression levels of LEE genes among the nine EPEC isolates (Fig. 5A). As expected, the majority of the LEE genes exhibited significantly increased expression in DMEM compared to LB for all of the tEPEC isolates belonging to phylogroups B1 and B2 (Fig. 5A). An exception to this trend was EPEC4 isolate C581-05 of phylogroup B2, which showed significant differential expression of only 3 of the 41 LEE genes under these conditions (Fig. 5A). Also, there were only 4 to 5 of the 41 LEE genes that exhibited significant differential expression in the two EPEC isolates from phylogroup A, 401140 and 100329 (Fig. 5A). Interestingly, the *bfp* genes of four of the eight tEPEC isolates, three from phylogroup B2 and one from phylogroup A, did not exhibit significant differential expression under these conditions (Fig. 5B). In contrast, nearly all of the *bfp* and *per* genes exhibited increased expression in EPEC isolates B171 and 402290, both of *E. coli* phylogroup B1 (Fig. 5B). These findings highlight the phylogroup-specific regulation of these important virulence factors.

In addition to investigating differences in the expression of the LEE and BFP regions, we interrogated differences in expression of other characterized EPEC virulence factors such as autotransporters and other non-LEE secreted effectors (Fig. 6). Of the autotransporter-encoding genes that were detected with significant differential expres-

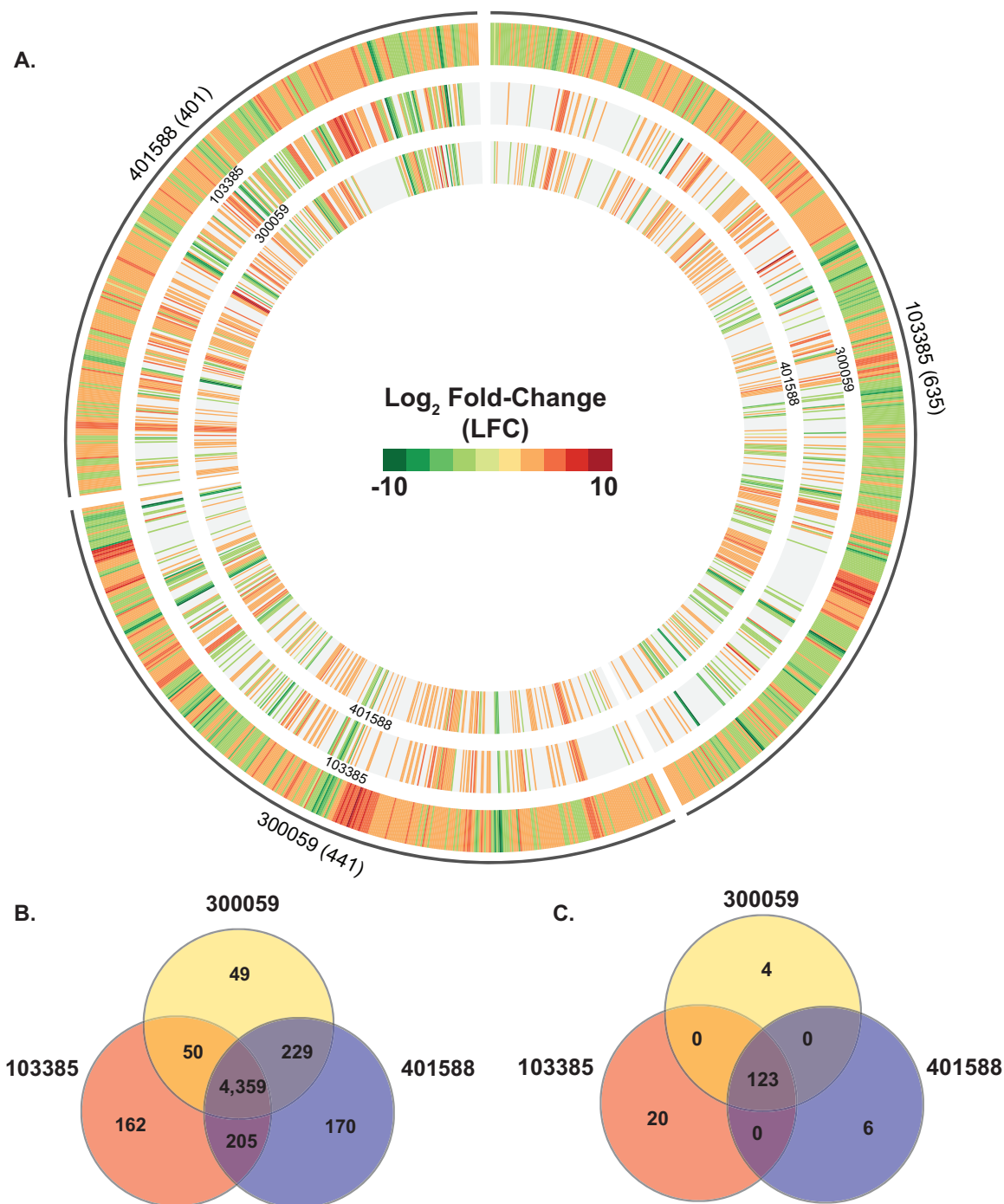


FIG 4 Comparison of the global transcriptomes of multiple EPEC isolates of the EPEC8 phylogenomic lineage. (A) Circular plot of the log₂-fold-change (LFC) values for genes that exhibited significant differential expression during exponential growth in DMEM compared to LB. The outermost track contains all of the significant LFC values for one of the three EPEC8 isolates. The inner tracks contain the LFC values of genes of another of the EPEC8 isolates belonging to the same LS-BSR gene cluster as the genes in the outermost reference track. The genes that were not identified in the other EPEC8 isolates and/or did not exhibit significant differential expression are absent from the inner tracks. The number of genes that had significant differential expression is indicated in parentheses next to each isolate name. (B) The number of genes that were highly conserved (LS-BSR \geq 0.8) in all four of the EPEC isolates is indicated in the center. The number of genes that were identified with significant similarity (LS-BSR \geq 0.8) that also exhibited significant differential expression in two or three EPEC isolates is also designated. The number of isolate-specific genes indicates the genes that were identified with significant similarity in one EPEC isolate and that were divergent or absent from the other two EPEC8 isolates. (C) Venn diagram showing the number of genes differentially expressed for each of the EPEC8 isolates analyzed in this study grown to an OD₆₀₀ of 0.5 in DMEM compared to LB. The number of core genes that were highly conserved (LS-BSR \geq 0.8) in all three of the EPEC8 isolates that also exhibited significant differential expression in all of the EPEC8 isolates is indicated in the center. There were no genes that were present with significant similarity and also differentially expressed in only two of the three EPEC8 isolates. The number of isolate-specific genes indicates those genes that exhibited significant similarity (LS-BSR \geq 0.8) that were divergent or absent (LS-BSR < 0.8) from the other two EPEC8 isolates and also exhibited significant differential expression during growth to an OD₆₀₀ of 0.5 in DMEM compared to LB.

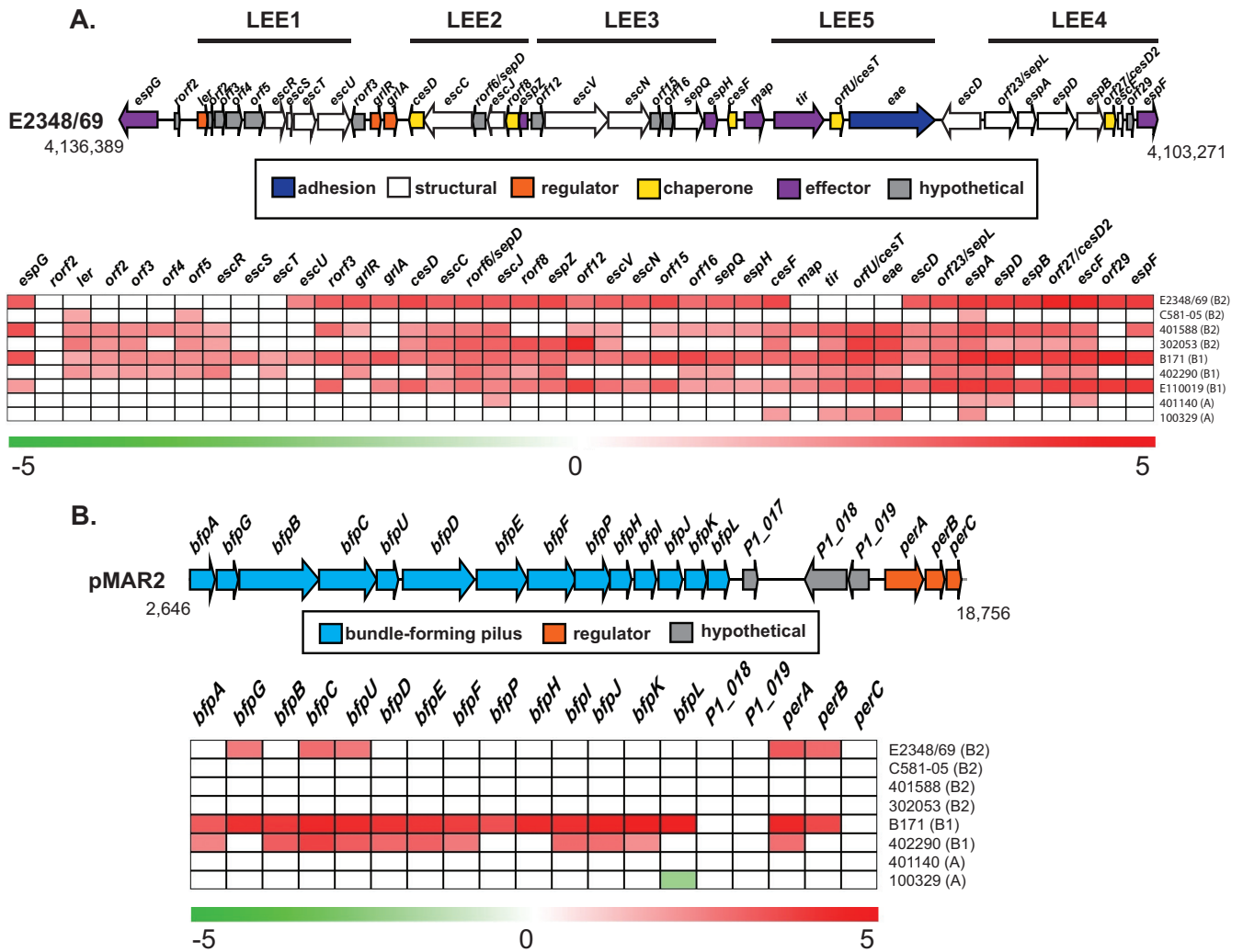


FIG 5 Differential expression analysis of genes carried by the LEE and BFP regions in each of the representative EPEC isolates analyzed. (A) Diagram of the genetic structure and results of differential expression analysis of LEE genes for the nine representative EPEC isolates examined (E2348/69, B171, C581-05, 401588, 401140, 402290, 302053, 100329, and E110019) during exponential growth ($OD_{600} = 0.5$) in DMEM compared to LB. The values in the heatmap are the significant \log_2 -fold-change (LFC) values for LEE genes. (B) Diagram of the gene organization and the RNA-Seq LFC values of genes carried by the BFP operon of the eight EPEC isolates that contained BFP (E2348/69, B171, C581-05, 401588, 401140, 402290, 302053, and 100329). The expression values in the heatmap are significant LFC values for BFP genes.

sion in the EPEC genomes, *espC* was identified as having significantly increased expression only in EPEC isolate 401588 (Fig. 6). Also, of the LEE and non-LEE T3SS effectors that were detected in the genomes, more than half had increased expression during growth in DMEM compared to LB (Fig. 6). Only two of the T3SS effectors exhibited discordant trends of differential expression among the EPEC isolates (Fig. 6). These included *espL*, which had increased expression in EPEC isolate 402290 but decreased expression in EPEC isolate 100329, and *nleF*, which had increased expression in EPEC isolates 402290 and 401588 but decreased expression in aEPEC isolate E110019 (Fig. 6). Interestingly, most of the T3SS effectors that had decreased expression (*espV*, *nleA*, *nleF*, *nleG*, *ospB*, and *tccP*) were identified in aEPEC isolate E110019 (Fig. 6). Further analysis is necessary to determine whether these effectors would also exhibit decreased expression in additional aEPEC isolates during growth under these same conditions.

Semiconserved genes of the global EPEC regulon. There were additional genes belonging to LS-BSR gene clusters that were identified among many but not all of the EPEC isolates and that were thus not identified as part of the conserved core regulon even though they exhibited similar expression trends in two or more of the EPEC

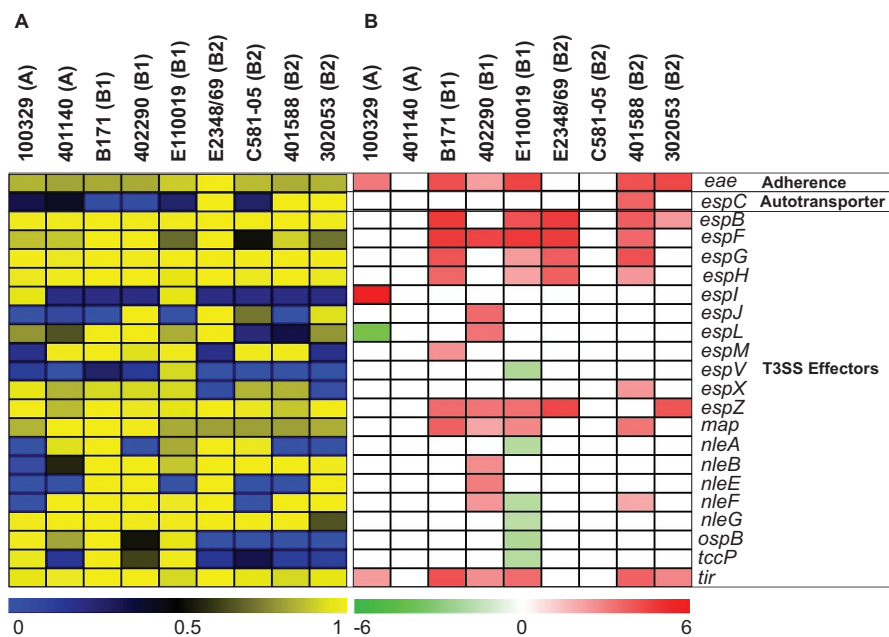


FIG 6 Differential expression analysis of known virulence genes of EPEC. The phylogroup that each EPEC isolate belongs to is indicated in parentheses. (A) Heatmap of LS-BSR values indicating the presence or absence of known EPEC virulence genes in the genomes of each of the EPEC isolates analyzed. Genes present with significant similarity are indicated by yellow, genes with divergent similarity are indicated by black, and genes that are absent are indicated by blue. (B) Heatmap of the log₂-fold-change (LFC) values for known virulence genes of EPEC that exhibited significant differential expression during exponential growth in DMEM compared to LB. The color gradient indicates decreased expression (green) or increased expression (red) of the virulence genes, while white indicates a gene that either was not present in the isolate or did not exhibit significant differential expression.

isolates (Table S3; see also Data Set S2 in the supplemental material). For example, there were proteins involved in biosynthesis of flagella that exhibited decreased expression in six of the nine EPEC isolates (Table S5). This is consistent with previous reports that genes encoding flagellar proteins have an expression pattern opposite that of virulence factors under these virulence gene-inducing growth conditions (52–55).

Similar to the differentially expressed conserved core genes, many of the differentially expressed semiconserved core genes encode proteins involved in metabolism, such as the *his* genes encoding predicted proteins for histidine synthesis, which had increased expression in nearly all of the EPEC isolates (Table S5). Among the genes that were conserved in all of the EPEC isolates but had significant differential expression in only a limited number of isolates were genes encoding predicted proteins involved in processes involving resistance to the host such as colanic acid biosynthesis (Table S5). The colanic acid biosynthesis genes had increased expression in EPEC isolates E2348/69 and 100329, which belong to two different phylogroups (B2 and A, respectively) (Table S5). Whether these genes have a unique role in pathogenesis of these EPEC isolates is not yet clear.

There were a total of 2,382 LS-BSR gene clusters that were identified with significant similarity (LS-BSR ≥ 0.8) in one or more of the EPEC isolates that contained a predicted protein domain of secreted or membrane-associated proteins. Among these gene clusters, there were 595 that had significant differential expression in one or more of the EPEC isolates (Data Set S2). Of these, 32 gene clusters were also differentially expressed in one or more of the EPEC isolates analyzed (Table S6). There were 26 gene clusters that had increased expression in DMEM compared to LB, while only six had decreased expression (Table S6). Included among these were genes encoding predicted proteins previously known to be secreted or membrane-associated proteins such as type IV pilus, a sucrose porin, and T3SS proteins (Table S6). There were also nine gene clusters encoding hypothetical proteins that could be further investigated for their role in the virulence mechanism of EPEC (Table S6).

Differential expression of known sRNAs of *E. coli*. To determine whether there was significant differential expression among small RNAs (sRNAs) in the genomically diverse EPEC isolates, the RNA-Seq reads were mapped to the previously identified *E. coli* sRNAs (56–60) that could be identified in each EPEC genome. The total number of sRNAs that exhibited significant differential expression in the EPEC isolates ranged from 6 to 30 sRNAs per isolate (Table 2). Unlike what we observed for some of the protein-coding genes that exhibited significant differential expression, the sRNAs examined did not have any phylogroup-specific expression trends.

Interestingly, the sRNA DsrA exhibited significant differential expression in only three of the EPEC isolates, all of which were part of phylogroup B2 (Table S7). Similar to some of the other sRNAs, DsrA had an expression trend in EPEC isolate C581-05 that was different from that of other isolates from phylogroup B2 (Table S7). The expression of *dsrA* was decreased in C581-05 in DMEM compared to LB, whereas it was increased in isolates 103385 and 300059 under the same conditions (Table S7). DsrA has been described for its role in regulating a number of global transcriptional pathways, including the RpoS-mediated stress response mechanism (61–63), which then potentially alters the expression of other pathways, such as the LEE region of EHEC and EPEC (64).

The sRNAs GlmY and GlmZ (65) both had increased expression during growth in DMEM compared to LB (Table S7). GlmY had significant differential expression in all of the EPEC isolates, except aEPEC isolate E110019 (Table S7). In comparison, GlmZ had significant differential expression only in EPEC isolate 302053 (Table S7). In *E. coli* K-12, GlmY and GlmZ regulate expression of *glmS*, which encodes the enzyme glucosamine-6-phosphate required for hexosamine metabolism, which generates precursor molecules for the synthesis of amino sugars that are used to make peptidoglycan and lipopolysaccharides (65). Meanwhile, GlmY and GlmZ in EHEC were found to regulate acid resistance, tryptophan metabolism, adhesion, and the expression of non-LEE effectors (22, 66). Previous studies demonstrated that *glmZ* is constitutively expressed (67), whereas the expression of *glmY* is regulated by QseEF in EHEC (66, 68). Our findings demonstrate that *glmY* had increased expression during growth in DMEM compared to LB in EPEC and that *glmZ* expression was also increased, but it was not great enough to be considered significant by our criteria for all but one of the EPEC isolates, 302053 (Table S7).

Another sRNA that had similar expression trends in nearly all of the EPEC isolates was RyhB, which exhibited increased expression in DMEM versus LB in all but three of the EPEC isolates (Table S7). RyhB is an sRNA that regulates iron metabolism, is required for siderophore production by uropathogenic *E. coli* (69, 70), and was previously described as having increased expression during growth in minimal media (60, 71). We observed similar results of increased expression of *ryhB* in minimal media (DMEM) compared to nutrient-rich media (LB) for 8 of the total of 11 EPEC isolates examined (Table S7). Interestingly, genes involved in iron acquisition (*hmuV* and *shuA*) had increased expression in the EPEC isolates belonging to phylogroup B2 (Table S3). To our knowledge, the role of RyhB in regulation of iron metabolism in EPEC has yet to be determined. Future studies are needed to determine the role of these sRNAs in regulating gene expression of EPEC.

DISCUSSION

Previous studies investigating the transcriptional networks and regulators of EPEC pathogenesis have focused on a limited number of genes in a limited number of isolates (46, 72–76). These genes are typically within pathogenicity islands or within other regions containing known virulence factors or have been described as global regulators of virulence in other bacteria (46, 72–76). In particular, the transcriptional regulation of the LEE and BFP regions has been extensively studied (46, 47, 72–83). Regulation of EPEC virulence mechanisms is known to involve numerous transcription factors, which are influenced by environmental conditions, including cell density (2, 46, 78, 79, 82–86). No previous studies of EPEC virulence had used an unbiased global

sequencing approach such as RNA-Seq to identify all genes that are simultaneously expressed under virulence gene-inducing growth conditions.

In addition to investigating a small number of genes at a time, previous studies also investigated EPEC virulence mechanisms using primarily a select few isolates (E2348/69, B171, E22, E110019) (25, 26, 35–40), which represent only 2 EPEC phylogenomic lineages (EPEC1 and EPEC2) of the more than 10 that have been described (31, 32). The results of our previous study demonstrated that there are differences in the global transcriptomes of these frequently studied EPEC isolates during growth under virulence gene-inducing laboratory conditions (34), indicating there were significant differences associated with even the archetype isolates. However, the archetype isolates were isolated in the past and may not represent the modern EPEC isolates.

In the current study, we demonstrated that EPEC isolates representing genomically diverse lineages can have limited conservation with respect to their transcriptional responses under virulence-inducing laboratory growth conditions (see Table S3 in the supplemental material). Our study investigated the global transcriptional response of EPEC under a single growth condition (high-glucose DMEM) that induces the expression of virulence among EPEC isolates (46–48). However, other conditions have also been described that induce EPEC virulence, including growth in static cultures (87) or with different glucose concentrations (88). The focus of our study was on investigating variation in the transcriptional response of genomically diverse EPEC isolates under a single set of conditions known to induce virulence gene expression, and we would anticipate that there would also be significant variation observed if other growth conditions were investigated. Given the considerable genomic diversity identified among the EPEC isolates analyzed, it was not surprising that there were only 21 conserved gene clusters that were present that also exhibited significant differential expression in all nine of the EPEC isolates (Table S3). Nearly all of these genes had known functions associated with central metabolism. In contrast, many of the known EPEC virulence factors were not identified as part of the core EPEC regulon under the examined growth conditions due to genetic diversity or differences in the timing of transcription of these genes (Fig. 5 and 6). This included genes within the LEE and BFP regions that are critical to EPEC virulence (Fig. 5). Many of these genes exhibited phylogroup-specific differences in expression, a result that was also observed in our previous studies (34). Although all the isolates in the current study were identified as EPEC, this study highlighted the variability of the transcriptional responses of these isolates. The underlying regulatory mechanisms of EPEC virulence have been assumed to be similar in all isolates based on the study of a limited number of genes in a limited number of archetype isolates; however, the current study highlighted that this assumption represents an underestimation of isolate-specific genomic content that may contribute to transcriptional differences and variable clinical presentations.

The phylogenomically diverse EPEC isolates analyzed in this study contained unique isolate-specific genomic content that was differentially expressed under the virulence-inducing conditions, representing an isolate-specific transcriptional response (Table S3). The cytolethal distending toxin, identified only in EPEC isolate 401140 of phylogroup A (Table S3; see also Fig. S1 in the supplemental material), represents an accessory virulence factor that has been identified in multiple pathovars of *E. coli*, including some EPEC isolates (89–92), but is absent from many EPEC isolates and is not considered a major component of the EPEC virulence mechanism (4). Comparison of the regulons of multiple EPEC isolates belonging to the same phylogenomic lineage (EPEC8) demonstrated a greater number of conserved genes that were present and also differentially expressed in all three of these isolates (Fig. 4; Table S4). This finding is not surprising, considering that there is greater genomic similarity among these EPEC isolates than among EPEC isolates from different lineages or phylogroups. However, it was interesting that there was unique genomic content in each of these EPEC8 isolates that exhibited differential expression under the virulence-inducing conditions. This finding demonstrates that there was additional diversification of EPEC isolates within the same

phylogenomic lineage that is either directly or indirectly linked to EPEC transcriptional regulation.

Further investigation is needed to determine what role, if any, these genes have in coordinating gene expression of EPEC. This study characterized the biological variation among the isolates of a single pathovar and demonstrated that, while each of these isolates is considered an EPEC isolate based on genomic features, they exhibited differences in their global transcriptomes, and very little is known about how the diversity of their transcriptional responses can result in differing clinical presentations. The observed transcriptional variation in responses to host and other signals may be more biologically significant for the observed clinical presentation than the genome content alone. Continued investigation of differences in the global transcriptomes of EPEC isolates in the presence of other pathogenic *E. coli* strains, commensal *E. coli* strains, and other members of the gut microbiome is ongoing and will be required to fully comprehend the importance of this transcriptional variability.

MATERIALS AND METHODS

Bacterial isolates and media. The EPEC isolates examined in this study (Table 1) were previously characterized, and their genome sequences are publicly available (25, 26, 31, 32, 93). The EPEC isolates were grown in lysogeny broth (LB) media (Difco) or in Dulbecco's modified Eagle's medium (DMEM) supplemented with 4.5 g/liter of glucose (Gibco).

Phylogenomic analysis. The genomes of the EPEC isolates analyzed in this study were compared with 40 previously sequenced *E. coli* and *Shigella* genomes by whole-genome phylogenomic analysis as previously described (31, 94). The genomes were aligned using Mugsy (95), and homologous blocks were concatenated using the bx-python toolkit (https://bitbucket.org/james_taylor/bx-python). The columns that contained one or more gaps were removed using Mothur (96). The resulting 1.98-Mb aligned region from each of the genomes was used to construct a maximum-likelihood phylogeny with 100 bootstrap replicates using RAxML v7.2.8 (97). The phylogeny was constructed using the GTR model of nucleotide substitution with the gamma model of rate heterogeneity and 100 bootstrap replicates. The phylogeny was then visualized using FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Large-scale BLAST score ratio (LS-BSR) analysis. The nine EPEC isolates (E2348/69, B171, C581-05, 401140, 402290, 401588, 302053, 100329, and E110019) represent eight different EPEC phylogenomic lineages (Table 1) and were examined along with two additional EPEC8 isolates (300059 and 103385). All of these isolates were subjected to comparative genomics using LS-BSR as previously described (31, 50, 98). Briefly, the predicted protein-coding genes of each genome that had $\geq 80\%$ nucleotide identity to each other were assigned to gene clusters using uclust (99). The representative sequences of each gene cluster were translated and the amino acid sequences were compared to those of each genome using TBLASTN (100) with composition-based adjustment turned off. The bit scores were used to generate a BSR value indicating the detection of each gene cluster in each of the genomes by dividing the score of a gene compared to a genome by the score of the gene compared to its own sequence. The genomes of ED1a, SE11, HS, and K-12 strain MG1655 were also included in the LS-BSR analysis to provide nonpathogenic reference isolates for the identification of conserved *E. coli* genes, as well as pathogen-specific genes. The nucleotide sequences of the gene clusters are included in Data Set S1 in the supplemental material.

The functions of the proteins encoded by each gene cluster were predicted using the Institute for Genome Sciences (IGS) prokaryotic annotation pipeline (101). The presence of protein and functional domains characteristic of membrane-associated or secreted proteins (analyzed using TMHMM, SignalP, and an in-house script that searches for motifs common in outer membrane proteins) or of lipoproteins (analyzed using TMH, SignalP I, and SignalP II) among the proteins encoded by each gene cluster was predicted using the indicated algorithms and the IGS prokaryotic annotation pipeline (101).

RNA isolation and sequencing. The EPEC isolates were grown overnight in LB and were inoculated at 1:100 into 50 ml of LB, or of DMEM supplemented with 4.5 g/liter glucose, in a 250-ml flask. The cultures were grown at 37°C with shaking (225 rpm) to an optical density at 600 nm (OD_{600}) of approximately 0.5, corresponding to the exponential-growth phase. Total RNA was isolated and prepared for sequencing from the cell pellet using a Ribopure bacteria kit (Ambion) and treated with Ribopure DNase I to remove contaminating DNA. The samples were then treated with a Turbo DNA-free kit (Ambion) to ensure that all contaminating DNA was removed. RNA samples were verified to be DNA-free by quantitative PCR (qPCR) analysis for the conserved *rp0A* gene as previously described (34). The DNA-free RNA samples were submitted for library construction using an Ovation prokaryotic RNA-Seq system (NuGen) and were sequenced using 100-bp paired-end sequencing at the Institute for Genome Sciences Genome Resource Center on an Illumina HiSeq platform.

RNA-Seq analyses. The Illumina reads generated for each RNA sample were analyzed and compared using an in-house Ergatis-based (102) RNA-Seq analysis pipeline as previously described (34). Briefly, the RNA-Seq reads were aligned to the corresponding genome sequences using Bowtie (103), and the number of reads that aligned to the protein-coding regions and intergenic regions was determined using HTSeq (104). The differential expression of each gene across the biological replicates for DMEM compared to LB was determined using DESeq (105). The \log_2 -fold-change (LFC) values were calculated

for the DMEM samples compared to the LB samples of each EPEC isolate. The gene expression data were then filtered for further analysis using the following criteria: $LFC \geq 2$ or ≤ -2 and false-discovery rate (FDR) ≤ 0.05 . Genes that met these criteria were identified as having significant differential expression during growth in DMEM compared to LB. The differential expression of the sRNAs was examined by detecting all of the previously identified sRNAs (56–60) in each of the EPEC genomes using BLAST analysis. The differential expression of the identified sRNA regions was then determined as described above for the protein-coding genes.

Circular displays of the differential expression data were generated using Circos v. 0.67-6 (106). The genes that exhibited significant LFC values and belonged to the same LS-BSR gene cluster were aligned in the plots for each of the EPEC isolates. Heat maps of the significant LFC values for the LEE and BFP genes and other virulence factors of each EPEC isolate were generated using MeV (107).

Correlation of the read counts for all of the RNA-Seq samples, including each of the two biological replicates, was examined as previously described (34). Expression values were normalized using DESeq v1.10.1 (105). The conserved LS-BSR gene clusters were used to compute the eigenvectors by principal-component analysis. The first and second principal components were displayed in a scatterplot to visualize the clustering of the RNA-Seq samples by isolate and medium type. The analysis was performed using R statistical package v2.15.2, and the data were visualized using gplots v2.11.0.

The normalized gene expression values were also used to compute the standard deviation for each LS-BSR cluster across all samples, and the 674 LS-BSR gene clusters showing the greatest standard deviations of expression values were used to generate a heatmap of the samples. The heatmap was generated using R statistical package v2.15.2, which in turn used DESeq v1.10.1 for normalization and gplots v2.11.0 for visualization.

Quantitative reverse transcriptase PCR (qRT-PCR). The trends of differential expression of selected genes during growth in DMEM compared to LB were verified for the genes listed in Table S1 in the supplemental material using previously described qRT-PCR methods (34, 108). Briefly, RNA was reverse transcribed and primed with random hexamers to generate cDNA using a SuperScript III first-strand synthesis system for RT-PCR (Invitrogen). The cDNA was diluted 1:20 into nuclease-free water (Ambion) before analysis with qPCR was performed. The qPCR was performed on the reverse-transcribed RNA samples by the use of SYBR green master mix (Life Technologies, Inc.) with 10- μ l reaction mixtures comprised of the following: 5 μ l of 2 \times SYBR master mix, 1 μ l of each of the 5 μ M forward and reverse primers (Table S1), 1 μ l of nuclease-free water (Ambion), and 2 μ l of cDNA diluted 1:20 with nuclease-free water. Triplicate reactions were performed for each cDNA template and primer combination. The reactions were cycled using a 384-well plate on a 7900HT Fast real-time PCR system (Applied Biosystems) and a two-step reaction with an initial incubation performed at 50°C for 2 min and 95°C for 10 min and then 40 cycles of 95°C for 15 s and 60°C for 1 min, followed by a dissociation stage. The cycle threshold (C_T) values were calculated using Applied Biosystems software. The C_T values of the biological replicates were averaged, and the standard deviation was calculated. The C_T value for each of the target genes of each sample was normalized by subtracting from it the C_T value of the constitutively expressed RNA polymerase alpha subunit gene, *rpoA*, resulting in the ΔC_T value of a particular gene for each sample. The difference between the expression level of a target gene ($\Delta \Delta C_T$) in the DMEM samples and its expression level in the LB samples was then calculated by subtracting the ΔC_T of the LB sample from the ΔC_T of the DMEM sample. The fold difference in the levels of expression of a particular gene in DMEM and in LB was determined by calculating the $2^{-\Delta \Delta C_T}$ value. The difference in expression is represented in the figures as the $\log_2(2^{-\Delta \Delta C_T})$ value for each gene in DMEM compared to LB. The error bars indicate the standard deviations of the $\Delta \Delta C_T$ values. Differences in the exact fold change values between the RNA-Seq and the qRT-PCR data were observed; however, this result was most likely due to the differences in the methodology required for the amplification, library construction, and sequencing involved in RNA-Seq versus a focused and optimized qRT-PCR assay.

Data availability. All raw data generated by RNA-Seq analysis have been deposited in the Short Read Archive (SRA) under the accession numbers listed in Table S2, and the expression data were deposited in GEO under accession number [GSE73885](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73885).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00024-17>.

DATA SET S1, PDF file, 5.6 MB.

DATA SET S2, XLSX file, 0.7 MB.

FIG S1, PDF file, 0.1 MB.

TABLE S1, PDF file, 0.1 MB.

TABLE S2, PDF file, 0.05 MB.

TABLE S3, PDF file, 0.1 MB.

TABLE S4, PDF file, 0.1 MB.

TABLE S5, PDF file, 0.1 MB.

TABLE S6, PDF file, 0.05 MB.

TABLE S7, PDF file, 0.05 MB.

ACKNOWLEDGMENTS

This project was funded in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract no. HHSN272200900009C and grant numbers U19 AI090873 and U19 AI110820 and by start-up funds from the state of Maryland.

REFERENCES

- Ochoa TJ, Contreras CA. 2011. Enteropathogenic *Escherichia coli* infection in children. *Curr Opin Infect Dis* 24:478–483. <https://doi.org/10.1097/QCO.0b013e32834a8b8b>.
- Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2:123–140. <https://doi.org/10.1038/nrmicro818>.
- Nataro JP, Kaper JB. 1998. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 11:142–201.
- Nisa S, Scanlon KM, Donnenberg MS. 2013. Enteropathogenic *Escherichia coli*. In Donnenberg MS (ed), *Escherichia coli* pathotypes and principles of pathogenesis, 2nd ed. Academic Press, New York, NY.
- Tarr PI, Gordon CA, Chandler WL. 2005. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* 365:1073–1086. [https://doi.org/10.1016/S0140-6736\(05\)71144-2](https://doi.org/10.1016/S0140-6736(05)71144-2).
- McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB. 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci U S A* 92:1664–1668. <https://doi.org/10.1073/pnas.92.5.1664>.
- McDaniel TK, Kaper JB. 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. <https://doi.org/10.1046/j.1365-2958.1997.2311591.x>.
- Jerse AE, Yu J, Tall BD, Kaper JB. 1990. A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. *Proc Natl Acad Sci U S A* 87:7839–7843. <https://doi.org/10.1073/pnas.87.20.7839>.
- Sahl JW, Rasko DA. 2012. Analysis of global transcriptional profiles of enterotoxigenic *Escherichia coli* isolate E24377A. *Infect Immun* 80:1232–1242. <https://doi.org/10.1128/IAI.06138-11>.
- Chugani S, Kim BS, Phattarasakul S, Brittnacher MJ, Choi SH, Harwood CS, Greenberg EP. 2012. Strain-dependent diversity in the *Pseudomonas aeruginosa* quorum-sensing regulon. *Proc Natl Acad Sci U S A* 109:E2823–E2831. <https://doi.org/10.1073/pnas.1214128109>.
- Bergholz TM, Wick LM, Qi W, Riordan JT, Ouellette LM, Whittam TS. 2007. Global transcriptional response of *Escherichia coli* O157:H7 to growth transitions in glucose minimal medium. *BMC Microbiol* 7:97. <https://doi.org/10.1186/1471-2180-7-97>.
- Camarena L, Bruno V, Euskirchen G, Poggio S, Snyder M. 2010. Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA-sequencing. *PLoS Pathog* 6:e1000834. <https://doi.org/10.1371/journal.ppat.1000834>.
- Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63. <https://doi.org/10.1038/nrg2484>.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621–628. <https://doi.org/10.1038/nmeth.1226>.
- Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12:87–98. <https://doi.org/10.1038/nrg2934>.
- Martin JA, Wang Z. 2011. Next-generation transcriptome assembly. *Nat Rev Genet* 12:671–682. <https://doi.org/10.1038/nrg3068>.
- Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hacker Müller J, Reinhardt R, Stadler PF, Vogel J. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464:250–255. <https://doi.org/10.1038/nature08756>.
- Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomson NR, Dougan G. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genet* 5:e1000569. <https://doi.org/10.1371/journal.pgen.1000569>.
- Oliver HF, Orsi RH, Ponnala L, Keich U, Wang W, Sun Q, Cartinhour SW, Filiatrault MJ, Wiedmann M, Boor KJ. 2009. Deep RNA sequencing of *monocytogenes* reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs. *BMC Genomics* 10:641. <https://doi.org/10.1186/1471-2164-10-641>.
- Mandlik A, Livny J, Robins WP, Ritchie JM, Mekalanos JJ, Waldor MK. 2011. RNA-Seq-based monitoring of infection-linked changes in *Vibrio cholerae* gene expression. *Cell Host Microbe* 10:165–174. <https://doi.org/10.1016/j.chom.2011.07.007>.
- Zhao H, Chen C, Xiong Y, Xu X, Lan R, Wang H, Yao X, Bai X, Liu X, Meng Q, Zhang X, Sun H, Zhao A, Bai X, Cheng Y, Chen Q, Ye C, Xu J. 2013. Global transcriptional and phenotypic analyses of *Escherichia coli* O157:H7 strain Xuzhou21 and its pO157_Sal cured mutant. *PLoS One* 8:e65466. <https://doi.org/10.1371/journal.pone.0065466>.
- Gruber CC, Sperandio V. 2015. Global analysis of posttranscriptional regulation by GlmY and GlmZ in enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 83:1286–1295. <https://doi.org/10.1128/IAI.02918-14>.
- Landstorfer R, Simon S, Schober S, Keim D, Scherer S, 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. <https://doi.org/10.1186/1471-2164-15-353>.
- Vital M, Chai B, Østman B, Cole J, Konstantinidis KT, Tiedje JM. 2015. Gene expression analysis of *E. coli* strains provides insights into the role of gene regulation in diversification. *ISME J* 9:1130–1140. <https://doi.org/10.1038/ismej.2014.204>.
- Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D, Asadulghani M, Kurokawa K, Dean P, Kenny B, Quail MA, Thurston S, Dougan G, Hayashi T, Parkhill J, Frankel G. 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *J Bacteriol* 191:347–354.
- Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebahia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J. 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* 190:6881–6893. <https://doi.org/10.1128/JB.00619-08>.
- Viljanen MK, Peltola T, Junnila SY, Oikkonen L, Järvinen H, Kuistila M, Huovinen P. 1990. Outbreak of diarrhoea due to *Escherichia coli* O111:B4 in schoolchildren and adults: association of Vi antigen-like reactivity. *Lancet* 336:831–834. [https://doi.org/10.1016/0140-6736\(90\)92337-H](https://doi.org/10.1016/0140-6736(90)92337-H).
- Cheney CP, Schad PA, Formal SB, Boedeker EC. 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.
- Srikhanta YN, Hocking DM, Wakefield MJ, Higginson E, Robins-Browne RM, Yang J, Tauschek M. 2013. Control of bacterial virulence by the RalR regulator of the rabbit-specific enteropathogenic *Escherichia coli* strain E22. *Infect Immun* 81:4232–4243. <https://doi.org/10.1128/IAI.00710-13>.
- Srikhanta YN, Hocking DM, Praszkiar J, Wakefield MJ, Robins-Browne RM, Yang J, Tauschek M. 2013. RegR virulence regulon of rabbit-specific enteropathogenic *Escherichia coli* strain E22. *Infect Immun* 81:1078–1089. <https://doi.org/10.1128/IAI.01325-12>.
- Hazen TH, Sahl JW, Fraser CM, Donnenberg MS, Scheutz F, Rasko DA. 2013. Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. *Proc Natl Acad Sci U S A* 110:12810–12815. <https://doi.org/10.1073/pnas.1306836110>.
- Hazen TH, Donnenberg MS, Panchalingam S, Antonio M, Hossain A, Mandomando I, Ochieng JB, Ramamurthy T, Tamboura B, Qureshi S, Quadri F, Zaidi A, Kotloff KL, Levine MM, Barry EM, Kaper JB, Rasko DA, Nataro JP. 2016. Genomic diversity of EPEC associated with clinical

- presentations of differing severity. *Nat Microbiol* 1:15014. <https://doi.org/10.1038/nmicrobiol.2015.14>.
33. Ingle DJ, Tauschek M, Edwards DJ, Hocking DM, Pickard DJ, Azzopardi KI, Amarasena T, Bennett-Wood V, Pearson JS, Tamboura B, Antonio M, Ochieng JB, Oundo J, Mandomando I, Qureshi S, Ramamurthy T, Hossain A, Kotloff KL, Nataro JP, Dougan G, Levine MM, Robins-Browne RM, Holt KE. 2016. Evolution of atypical enteropathogenic *E. coli* by repeated acquisition of LEE pathogenicity island variants. *Nat Microbiol* 1:15010. <https://doi.org/10.1038/nmicrobiol.2015.10>.
 34. Hazen TH, Daugherty SC, Shetty A, Mahurkar AA, White O, Kaper JB, Rasko DA. 2015. RNA-Seq analysis of isolate- and growth phase-specific differences in the global transcriptomes of enteropathogenic *Escherichia coli* prototype isolates. *Front Microbiol* 6:569. <https://doi.org/10.3389/fmicb.2015.00569>.
 35. Levine MM, Bergquist EJ, Nalin DR, Waterman DH, Hornick RB, Young CR, Sotman S. 1978. *Escherichia coli* strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. *Lancet* i:1119–1122.
 36. Levine MM, Rennels MB. 1978. *E. coli* colonisation factor antigen in diarrhoea. *Lancet* ii:534.
 37. Levine MM, Nataro JP, Karch H, Baldini MM, Kaper JB, Black RE, Clements ML, O'Brien AD. 1985. The diarrheal response of humans to some classic serotypes of enteropathogenic *Escherichia coli* is dependent on a plasmid encoding an enteroadhesiveness factor. *J Infect Dis* 152: 550–559. <https://doi.org/10.1093/infdis/152.3.550>.
 38. Tobe T, Hayashi T, Han CG, Schoolnik GK, Ohtsubo E, Sasakawa C. 1999. Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. *Infect Immun* 67: 5455–5462.
 39. Riley LW, Junio LN, Libaek LB, Schoolnik GK. 1987. Plasmid-encoded expression of lipopolysaccharide O-antigenic polysaccharide in enteropathogenic *Escherichia coli*. *Infect Immun* 55:2052–2056.
 40. Riley LW, Junio LN, Schoolnik GK. 1990. HeLa cell invasion by a strain of enteropathogenic *Escherichia coli* that lacks the O-antigenic polysaccharide. *Mol Microbiol* 4:1661–1666. <https://doi.org/10.1111/j.1365-2958.1990.tb00543.x>.
 41. Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbone E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S. 2008. Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* 9:560. <https://doi.org/10.1186/1471-2164-9-560>.
 42. Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 8:207–217. <https://doi.org/10.1038/nrmicro2298>.
 43. Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, Farag TH, van Eijk A, Adegbola RA, Alonso PL, Breiman RF, Golam Faruque AS, Saha D, Sow SO, Sur D, Zaidi AK, Biswas K, Panchalingam S, Clemens JD, Cohen D, Glass RI, Mintz ED, Sommerfelt H, Levine MM. 2012. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. *Clin Infect Dis* 55(Suppl 4):S232–S245. <https://doi.org/10.1093/cid/cis753>.
 44. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omoro R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acácio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 382:209–222. [https://doi.org/10.1016/S0140-6736\(13\)60844-2](https://doi.org/10.1016/S0140-6736(13)60844-2).
 45. Nisa S, Hazen TH, Assatourian L, Nougayrède JP, Rasko DA, Donnenberg MS. 2013. *In vitro* evolution of an archetypal enteropathogenic *Escherichia coli* strain. *J Bacteriol* 195:4476–4483. <https://doi.org/10.1128/JB.00704-13>.
 46. Sperandio V, Mellies JL, Nguyen W, Shin S, Kaper JB. 1999. Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 96:15196–15201. <https://doi.org/10.1073/pnas.96.26.15196>.
 47. Leverton LQ, Kaper JB. 2005. Temporal expression of enteropathogenic *Escherichia coli* virulence genes in an *in vitro* model of infection. *Infect Immun* 73:1034–1043. <https://doi.org/10.1128/IAI.73.2.1034-1043.2005>.
 48. Kenny B, Abe A, Stein M, Finlay BB. 1997. Enteropathogenic *Escherichia coli* protein secretion is induced in response to conditions similar to those in the gastrointestinal tract. *Infect Immun* 65:2606–2612.
 49. Lacher DW, Steinsland H, Blank TE, Donnenberg MS, Whittam TS. 2007. Molecular evolution of typical enteropathogenic *Escherichia coli*: clonal analysis by multilocus sequence typing and virulence gene allelic profiling. *J Bacteriol* 189:342–350. <https://doi.org/10.1128/JB.01472-06>.
 50. Sahl JW, Caporaso JG, Rasko DA, 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. <https://doi.org/10.7717/peerj.332>.
 51. Sahl JW, Sistrunk JR, Fraser CM, Hine E, Baby N, Begum Y, Luo Q, Sheikh A, Qadri F, Fleckenstein JM, Rasko DA. 2015. Examination of the enterotoxigenic *Escherichia coli* population structure during human infection. *mBio* 6:e00501. <https://doi.org/10.1128/mBio.00501-15>.
 52. Girón JA, Torres AG, Freer E, Kaper JB. 2002. The flagella of enteropathogenic *Escherichia coli* mediate adherence to epithelial cells. *Mol Microbiol* 44:361–379. <https://doi.org/10.1046/j.1365-2958.2002.02899.x>.
 53. Walters M, Sperandio V. 2006. Quorum sensing in *Escherichia coli* and *Salmonella*. *Int J Med Microbiol* 296:125–131. <https://doi.org/10.1016/j.ijmm.2006.01.041>.
 54. Sircili MP, Walters M, Trabulsi LR, Sperandio V. 2004. Modulation of enteropathogenic *Escherichia coli* virulence by quorum sensing. *Infect Immun* 72:2329–2337. <https://doi.org/10.1128/IAI.72.4.2329-2337.2004>.
 55. Hughes DT, Clarke MB, Yamamoto K, Rasko DA, Sperandio V. 2009. The QseC adrenergic signaling cascade in enterohemorrhagic *E. coli* (EHEC). *PLoS Pathog* 5:e1000553. <https://doi.org/10.1371/journal.ppat.1000553>.
 56. Rivas E, Klein RJ, Jones TA, Eddy SR. 2001. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* 11:1369–1373. [https://doi.org/10.1016/S0960-9822\(01\)00401-8](https://doi.org/10.1016/S0960-9822(01)00401-8).
 57. Chen S, Lesnik EA, Hall TA, Sampath R, Griffey RH, Ecker DJ, Blyn LB. 2002. A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems* 65:157–177. [https://doi.org/10.1016/S0303-2647\(02\)00013-8](https://doi.org/10.1016/S0303-2647(02)00013-8).
 58. Yachie N, Numata K, Saito R, Kanai A, Tomita M. 2006. Prediction of non-coding and antisense RNA genes in *Escherichia coli* with gapped Markov model. *Gene* 372:171–181. <https://doi.org/10.1016/j.gene.2005.12.034>.
 59. Tran TT, Zhou F, Marshburn S, Stead M, Kushner SR, Xu Y. 2009. De novo computational prediction of non-coding RNA genes in prokaryotic genomes. *Bioinformatics* 25:2897–2905. <https://doi.org/10.1093/bioinformatics/btp537>.
 60. Raghavan R, Groisman EA, Ochman H. 2011. Genome-wide detection of novel regulatory RNAs in *E. coli*. *Genome Res* 21:1487–1497. <https://doi.org/10.1101/gr.119370.110>.
 61. Sledjeski D, Gottesman S. 1995. A small RNA acts as an antisilencer of the H-NS-silenced *rcaA* gene of *Escherichia coli*. *Proc Natl Acad Sci U S A* 92:2003–2007. <https://doi.org/10.1073/pnas.92.6.2003>.
 62. Majdalani N, Cunnig C, Sledjeski D, Elliott T, Gottesman S. 1998. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc Natl Acad Sci U S A* 95:12462–12467. <https://doi.org/10.1073/pnas.95.21.12462>.
 63. Lease RA, Belfort M. 2000. Riboregulation by DsrA RNA: trans-actions for global economy. *Mol Microbiol* 38:667–672. <https://doi.org/10.1046/j.1365-2958.2000.02162.x>.
 64. Laaberki MH, Janabi N, Oswald E, Repoila F. 2006. Concert of regulators to switch on LEE expression in enterohemorrhagic *Escherichia coli* O157:H7: interplay between Ler, GrlA, HNS and RpoS. *Int J Med Microbiol* 296:197–210. <https://doi.org/10.1016/j.ijmm.2006.02.017>.
 65. Urban JH, Vogel J. 2008. Two seemingly homologous noncoding RNAs act hierarchically to activate *glmS* mRNA translation. *PLoS Biol* 6:e64. <https://doi.org/10.1371/journal.pbio.0060064>.
 66. Gruber CC, Sperandio V. 2014. Posttranscriptional control of microbe-induced rearrangement of host cell actin. *mBio* 5:e01025-13. <https://doi.org/10.1128/mBio.01025-13>.
 67. Göpel Y, Lüttmann D, Heroven AK, Reichenbach B, Dersch P, Görke B. 2011. Common and divergent features in transcriptional control of the homologous small RNAs GlmY and GlmZ in *Enterobacteriaceae*. *Nucleic Acids Res* 39:1294–1309. <https://doi.org/10.1093/nar/gkq986>.
 68. Reichenbach B, Göpel Y, Görke B. 2009. Dual control by perfectly overlapping sigma 54- and sigma 70- promoters adjusts small RNA

- GlmY expression to different environmental signals. *Mol Microbiol* 74:1054–1070. <https://doi.org/10.1111/j.1365-2958.2009.06918.x>.
69. Porcheron G, Habib R, Houle S, Caza M, Lépine F, Daigle F, Massé E, Dozois CM. 2014. The small RNA RyhB contributes to siderophore production and virulence of uropathogenic *Escherichia coli*. *Infect Immun* 82:5056–5068. <https://doi.org/10.1128/IAI.02287-14>.
 70. Massé E, Gottesman S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* 99:4620–4625. <https://doi.org/10.1073/pnas.032066599>.
 71. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. 2001. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* 15:1637–1651. <https://doi.org/10.1101/gad.901001>.
 72. Kendall MM, Gruber CC, Rasko DA, Hughes DT, Sperandio V. 2011. Hfq virulence regulation in enterohemorrhagic *Escherichia coli* O157:H7 strain 86-24. *J Bacteriol* 193:6843–6851. <https://doi.org/10.1128/JB.06141-11>.
 73. Iyoda S, 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. <https://doi.org/10.1128/JB.187.12.4086-4094.2005>.
 74. Haack KR, Robinson CL, Miller KJ, Fowlkes JW, Mellies JL. 2003. Interaction of Ler at the LEE5 (*tir*) operon of enteropathogenic *Escherichia coli*. *Infect Immun* 71:384–392. <https://doi.org/10.1128/IAI.71.1.384-392.2003>.
 75. Sperandio V, Mellies JL, Delahay RM, Frankel G, Crawford JA, Nguyen W, Kaper JB. 2000. Activation of enteropathogenic *Escherichia coli* (EPEC) LEE2 and LEE3 operons by Ler. *Mol Microbiol* 38:781–793. <https://doi.org/10.1046/j.1365-2958.2000.02168.x>.
 76. Bustamante VH, Santana FJ, Calva E, Puente JL. 2001. Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. *Mol Microbiol* 39:664–678. <https://doi.org/10.1046/j.1365-2958.2001.02209.x>.
 77. Elliott SJ, Sperandio V, Girón JA, Shin S, Mellies JL, Wainwright L, Hutcheson SW, McDaniel TK, Kaper JB. 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. <https://doi.org/10.1128/IAI.68.11.6115-6126.2000>.
 78. Mellies JL, Elliott SJ, Sperandio V, Donnenberg MS, Kaper JB. 1999. The per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). *Mol Microbiol* 33:296–306. <https://doi.org/10.1046/j.1365-2958.1999.01473.x>.
 79. Puente JL, Bieber D, Ramer SW, Murray W, Schoolnik GK. 1996. The bundle-forming pili of enteropathogenic *Escherichia coli*: transcriptional regulation by environmental signals. *Mol Microbiol* 20:87–100. <https://doi.org/10.1111/j.1365-2958.1996.tb02491.x>.
 80. Porter ME, Mitchell P, Roe AJ, Free A, Smith DG, Gally DL. 2004. Direct and indirect transcriptional activation of virulence genes by an AraC-like protein, PerA from enteropathogenic *Escherichia coli*. *Mol Microbiol* 54:1117–1133. <https://doi.org/10.1111/j.1365-2958.2004.04333.x>.
 81. Gómez-Duarte OG, Kaper JB. 1995. A plasmid-encoded regulatory region activates chromosomal *eaeA* expression in enteropathogenic *Escherichia coli*. *Infect Immun* 63:1767–1776.
 82. Shin S, Castanie-Cornet MP, Foster JW, Crawford JA, Brinkley C, Kaper JB. 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. <https://doi.org/10.1046/j.1365-2958.2001.02570.x>.
 83. Mellies JL, Barron AM, Carmona AM. 2007. Enteropathogenic and enterohemorrhagic *Escherichia coli* virulence gene regulation. *Infect Immun* 75:4199–4210. <https://doi.org/10.1128/IAI.01927-06>.
 84. Wong AR, Pearson JS, Bright MD, Munera D, Robinson KS, Lee SF, Frankel G, Hartland EL. 2011. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. *Mol Microbiol* 80:1420–1438. <https://doi.org/10.1111/j.1365-2958.2011.07661.x>.
 85. Deng W, Puente JL, Gruenheid S, Li Y, Vallance BA, Vázquez A, Barba J, Ibarra J, O'Donnell P, Metalnikov P, Ashman K, Lee S, Goode D, Pawson T, Finlay BB. 2004. Dissecting virulence: systematic and functional analyses of a pathogenicity island. *Proc Natl Acad Sci U S A* 101:3597–3602. <https://doi.org/10.1073/pnas.0400326101>.
 86. Garmendia J, Frankel G, Crepin VF. 2005. Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: translocation, translocation, translocation. *Infect Immun* 73:2573–2585. <https://doi.org/10.1128/IAI.73.5.2573-2585.2005>.
 87. Bustamante VH, Villalba MI, García-Angulo VA, Vázquez A, Martínez LC, Jiménez R, Puente JL. 2011. PerC and GrlA independently regulate Ler expression in enteropathogenic *Escherichia coli*. *Mol Microbiol* 82:398–415. <https://doi.org/10.1111/j.1365-2958.2011.07819.x>.
 88. Njoroge JW, Nguyen Y, Curtis MM, Moreira CG, Sperandio V. 2012. Virulence meets metabolism: *cra* and *KdpE* gene regulation in enterohemorrhagic *Escherichia coli*. *mBio* 3:e00280-12. <https://doi.org/10.1128/mBio.00280-12>.
 89. Scott DA, Kaper JB. 1994. Cloning and sequencing of the genes encoding *Escherichia coli* cytolethal distending toxin. *Infect Immun* 62:244–251.
 90. Pickett CL, Cottle DL, Pesci EC, Bikah G. 1994. Cloning, sequencing, and expression of the *Escherichia coli* cytolethal distending toxin genes. *Infect Immun* 62:1046–1051.
 91. Anderson JD, MacNab AJ, Gransden WR, Damm SM, Johnson WM, Lior H. 1987. Gastroenteritis and encephalopathy associated with a strain of *Escherichia coli* O55:K59:H4 that produced a cytolethal distending toxin. *Pediatr Infect Dis J* 6:1135–1136. <https://doi.org/10.1097/00006454-198712000-00015>.
 92. Tóth I, Nougayrède JP, Dobrindt U, Ledger TN, Boury M, Morabito S, Fujiwara T, Sugai M, Hacker J, Oswald E. 2009. Cytolethal distending toxin type I and type IV genes are framed with lambdoid prophage genes in extraintestinal pathogenic *Escherichia coli*. *Infect Immun* 77:492–500. <https://doi.org/10.1128/IAI.00962-08>.
 93. Donnenberg MS, Hazen TH, Farag TH, Panchalingam S, Antonio M, Hossain A, Mandomando I, Ochieng JB, Ramamurthy T, Tamboura B, Zaidi A, Levine MM, Kotloff K, Rasko DA, Nataro JP. 2015. Bacterial factors associated with lethal outcome of enteropathogenic *Escherichia coli* infection: genomic case-control studies. *PLoS Negl Trop Dis* 9:e0003791. <https://doi.org/10.1371/journal.pntd.0003791>.
 94. Sahl JW, Steinsland H, Redman JC, Angiuoli SV, Nataro JP, Sommerfelt H, Rasko DA. 2011. A comparative genomic analysis of diverse clonal types of enterotoxigenic *Escherichia coli* reveals pathovar-specific conservation. *Infect Immun* 79:950–960. <https://doi.org/10.1128/IAI.00932-10>.
 95. Angiuoli SV, Salzberg SL. 2011. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 27:334–342. <https://doi.org/10.1093/bioinformatics/btq665>.
 96. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
 97. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>.
 98. Sahl JW, Gillette JD, Schupp JM, Waddell VG, Driebe EM, Engelthaler DM, Keim P. 2013. Evolution of a pathogen: a comparative genomics analysis identifies a genetic pathway to pathogenesis in *Acinetobacter*. *PLoS One* 8:e54287. <https://doi.org/10.1371/journal.pone.0054287>.
 99. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
 100. Gertz EM, Yu YK, Agarwala R, Schäffer AA, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: improving the tBLASTn module of BLAST. *BMC Biol* 4:41. <https://doi.org/10.1186/1741-7007-4-41>.
 101. Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, White O, Wortman J, Mahurkar A, Giglio MG. 2011. The IGS standard operating procedure for automated prokaryotic annotation. *Stand Genomic Sci* 4:244–251. <https://doi.org/10.4056/signs.1223234>.
 102. Orvis J, Crabtree J, Galens K, Gussman A, Inman JM, Lee E, Nampally S, Riley D, Sundaram JP, Felix V, Whitty B, Mahurkar A, Wortman J, White O, Angiuoli SV. 2010. Ergatis: a web interface and scalable software system for bioinformatics workflows. *Bioinformatics* 26:1488–1492. <https://doi.org/10.1093/bioinformatics/btq167>.
 103. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
 104. Anders S, Pyl PT, Huber W. 2015. HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
 105. Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* 11:R106. <https://doi.org/10.1186/gb-2010-11-r106>.

106. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645. <https://doi.org/10.1101/gr.092759.109>.
107. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. 2006. TM4 microarray software suite. *Methods Enzymol* 411:134–193. [https://doi.org/10.1016/S0076-6879\(06\)11009-5](https://doi.org/10.1016/S0076-6879(06)11009-5).
108. Hazen TH, Kaper JB, Nataro JP, Rasko DA. 2015. Comparative genomics provides insight into the diversity of the attaching and effacing *Escherichia coli* virulence plasmids. *Infect Immun* 83:4103–4117. <https://doi.org/10.1128/IAI.00769-15>.