AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Spectrum

| Computational Biology | Research Article

The distinct transcriptome of virulence-associated phylogenetic group B2 *Escherichia coli*

Jacob Hogins,[1] Zhenyu Xuan,[1] Philippe E. Zimmern,[2] Larry Reitzer[1]

AUTHOR AFFILIATIONS See affiliation list on p. 19.

ABSTRACT *Escherichia coli* strains of phylogenetic group B2 are often associated with urinary tract infections (UTIs) and several other diseases. Recent genomic and transcriptomic analyses have not suggested or identified specific genes required for virulence, but have instead suggested multiple virulence strategies and complex host-pathogen interactions. Previous analyses have not compared core gene expression between phylogenetic groups or between pathogens and nonpathogens within phylogenetic groups. We compared the core gene expression of 35 strains from three phylogenetic groups that included both pathogens and nonpathogens after growth in a medium that allowed comparable growth of both types of strains. K-means clustering suggested a B2 cluster with 17 group B2 strains and two group A strains; an AD cluster with six group A strains, five group D strains and one B2 strain; and four outliers which included the highly studied model uropathogenic *E. coli* strains UTI89 and CFT073. Half of the core genes were differentially expressed between B2 and AD cluster strains, including transcripts of genes for all aspects of macromolecular synthesis—replication, transcription, translation, and peptidoglycan synthesis—energy metabolism, and environmental-sensing transcriptional regulators. Notably, core gene expression between nonpathogenic and uropathogenic transcriptomes within phylogenetic groups did not differ. If differences between pathogens and nonpathogens exist, then the differences do not require transcriptional reprogramming. In summary, B2 cluster strains have a distinct transcription pattern that involves hundreds of genes. We propose that this transcription pattern is one factor that contributes to virulence.

IMPORTANCE *Escherichia coli* is a diverse species and an opportunistic pathogen that is associated with various diseases, such as urinary tract infections. When examined, phylogenetic group B2 strains are more often associated with these diseases, but the specific properties that contribute to their virulence are not known. From a comparative transcriptomic analysis, we found that group B2 strains grown in a nutrient-rich medium had a distinct transcription pattern, which is the first evidence that core gene expression differs between phylogenetic groups. Understanding the consequences of group B2 transcription pattern will provide important information on basic *E. coli* biology, the basis for *E. coli* virulence, and possibly for developing therapies for a majority of urinary tract infections and other group B2-associated diseases.

KEYWORDS *Escherichia coli*, pathogens, urinary tract infection, transcriptional regulation, transcription factors, phylogenetic analysis, metabolism

Urinary tract infections (UTIs) are common bacterial infections that often recur (1). A UTI requires rapid bacterial growth in the bladder; extensive interactions with urothelial cells which include attachment, entry, survival, and rapid intracellular replication; and release back into the bladder (2). Only one in 10^5–10^6 cells survive passage through urothelial cells, and the mechanism of intracellular survival is poorly

Editor Tino Polen, Forschungszentrum Jülich GmbH, Jülich, Germany

Address correspondence to Larry Reitzer, reitzer@utdallas.edu.

The authors declare no conflict of interest.

See the funding table on p. 19.

Received 18 May 2023
Accepted 28 July 2023
Published 19 September 2023

Copyright © 2023 Hogins et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

understood. Antibiotic treatment usually resolves UTIs but also increases the risk of bacteria developing multidrug resistance (3, 4). Recurrent UTIs (RUTIs) can last for years and become hard to control, in part, because of antibiotic recalcitrance (5). The WHO recognizes UTIs as loci for development of multidrug-resistant bacteria which can impact the outcome of other clinically important diseases that are treated with the same antibiotics (6). Alternate or adjunct therapies are urgently needed (2, 3, 7), but potential targets of therapy have been elusive, in part because of an incomplete understanding of virulence.

Escherichia coli, which causes most UTIs and RUTIs (3, 8), is a diverse species that can be divided into phylogenetic groups (9, 10). Although strains from several phylogenetic groups are pathogenic (11, 12), group B2 strains are more prevalent in UTIs and RUTIs (8, 11, 12), as well as in *E. coli*-caused neonatal meningitis (13, 14), inflammatory bowel diseases (Crohn's disease and ulcerative colitis) (15–17), and eye infections (18), are more often associated with colorectal cancer (19), and have more characteristics associated with multiple antibiotic resistance and extended spectrum β-lactamase production (20). Paradoxically, a phylogenetic evolutionary analysis found that group B2 strains have lost several environmental-adaptive genes, including genes for survival in extreme conditions of low pH, high temperature, and low osmolarity (21). This phylogenetic study proposed that different lineages had distinct adaptive trajectories, but provided no explanation for the advantage of the group B2 adaptive strategy.

A common strategy to understand virulence is to identify genes that are only present in pathogens, which assumes that acquisition of certain genes will endow nonpathogenic strains with pathogenic potential. This type of virulence gene is by definition an accessory gene, i.e., not present in all *E. coli* strains. However, virulence factor genes for uropathogenic *E. coli* (UPEC) have been frustratingly difficult to identify. More than two decades of searches for virulence genes culminated in a study that concluded that UPEC strains do not have a common set of virulence genes and that previous analyses could not distinguish virulence genes from group B2-specific accessory genes (22).

An alternate or, more accurately, a complementary hypothesis is that UPEC strains have a common set of more highly expressed genes. If virulence is a property of strains from different phylogenetic genetic groups, then this hypothesis assumes cores genes are more highly expressed in pathogenic strains. The first test of this hypothesis did not identify genes from urine-associated *E. coli* whose expression correlated with quantifiable properties of murine infections (22). The authors then proposed multiple virulence strategies based on a complex interplay between hosts and UPEC strains (22). This conclusion could serve as a definition of virulence for an opportunistic pathogen. A second transcriptomic analysis postulated that expressed virulence genes would only be apparent from bacteria isolated directly from the human host. UPEC RNA from symptomatic patients was extracted immediately after urine collection (*in vivo*) and compared to RNA extracted after growth in pooled urine (*ex vivo*) and a rich nutrient broth. The results were interpreted to suggest a transcriptional program for UPEC strains which had upregulated genes for rapid growth, such as genes for the translational machinery (23). Rapid growth has been proposed to be a virulence factor, but slow-growing strains have also been associated with UTIs (24–26), which suggests that rapid growth may not be sufficient for UTIs.

Our primary goal was to determine whether core gene expression differs between strains from group B2 and other groups because of the association of group B2 strains with a variety of diseases. Our secondary goal was to test the hypothesis that intracellular survival during an infection selects for strains that differ from their nonpathogenic relatives, and to this end, we compared core gene transcriptomes from UPEC and nonpathogenic *E. coli* (NPEC) strains within phylogenetic groups. Our primary conclusions are that group B2 strains generally had a distinct transcriptomic pattern, and that UPEC and NPEC transcriptomes were generally indistinguishable. In other words, our results show that B2 strains have a distinct transcription program which we propose results in a greater virulence potential than strains from other groups.

RESULTS

Comparisons between and within phylogenetic groups

We compared the transcriptomes from 35 strains: 14 NPEC and 9 UPEC strains from the ECOR collection, nine recently isolated strains causing RUTIs, the well-studied model UPEC strains CFT073 and UTI89, and the nonpathogenic lab strain W3110. Groups A, B2, and D were represented by 9, 20, and 6 strains, respectively. We employed ECOR strains because (i) they are readily available to the research community, and (ii) they contain UPEC and NPEC strains from different phylogenetic groups.

We assumed that major differences of core gene expression would be readily detectable regardless of the growth medium. Because previous studies have shown that UPEC strains grow faster in urine and other media (27–29), an important consideration for selecting a growth medium was for comparable growth rates for NPEC and UPEC strains. The medium was 0.5% glucose-1% tryptone-0.25% NaCl, and with few exceptions, all strains grew at the same rate. We grew cultures aerobically because direct measurement suggests 4%–7% oxygen content in the bladder, and because aerobic growth is easier to reproduce between labs (30, 31).

Comparison across phylogenetic groups requires analysis of shared genes, i.e., core genes. Two thousand six hundred sixty-seven genes were present in at least 34 of the 35 strains and were defined as core genes (Table S1). A multidimensional scaling (MDS) plot of core gene expression shows a B2 cluster with 17 B2 strains and two group A NPEC strains, and an AD cluster with 6, 5, and 1 strains from groups A, D, and B2, respectively (Fig. 1A). We confirmed the phylotypes of the two group A strains in the B2 cluster and the B2 strain in the AD cluster. Four UPEC strains—CFT073, PNK004, PNK006, and UTI89—mapped outside of the B2 and AD clusters, but K-mean clustering did not define these strains as a group. A heatmap of the 1,000 most variably expressed genes visually shows the differences between the AD and B2 clusters (Fig. 1B). Patterns of gene expression were similar between ECOR and recently isolated UPEC strains, which suggests that the ECOR strains have not undergone extensive lab evolution since their isolation. We conclude that (i) transcriptomic and phylogenetic groupings are similar, but not identical, and (ii) most B2 strains not only differ from most group A and D strains in accessory genes, which define the phylogenetic groups, but also in the pattern of core gene expression.

One thousand three hundred forty (50.2%) of core genes were differentially expressed genes (DEGs) [false discovery rate (FDR) <0.05] between the B2 and AD cluster strains (Table S1, DEG tab). Gene ontology analysis indicated that B2 cluster strains had higher expression of genes for translation (e.g., ribosome protein genes), several biosynthetic pathways (purines, glutamine, glycine, and ribose), iron acquisition, carbohydrate transport, and transport processes that involve protons, but lower expression of several genes for environmentally responsive transcription factors (TFs) (Table S1, GOMF and GOBP tabs).

Differences in core gene expression did not distinguish UPEC from NPEC strains. Within the B2 cluster, 61 genes with known functions (excluding y genes whose products have no known function) were differentially expressed (FDR < 0.05) between UPEC and NPEC strains (Table S2). Forty-two of these genes code for components of the flagella and chemotaxis machinery. The expression of motility genes is anomalous because the growth medium contained glucose which should prevent cyclic-AMP synthesis and flagella expression. Only 3 of the 19 B2 cluster strains expressed motility genes, and when these strains were removed from the analysis, UPEC and NPEC strains did not differentially express these 42 genes. Similarly, 10 of the 61 genes code for cyclic-AMP-regulated catabolic enzymes, and in this case, one strain (ECOR51) greatly distorted the statistics. When ECOR51 results were excluded, these 10 genes were not only not differentially expressed, but they were not expressed. The remaining nine genes were expressed in the recently isolated UPEC strains, but not in the ECOR UPEC strains. These genes are worthy of further analysis but have not been previously associated with

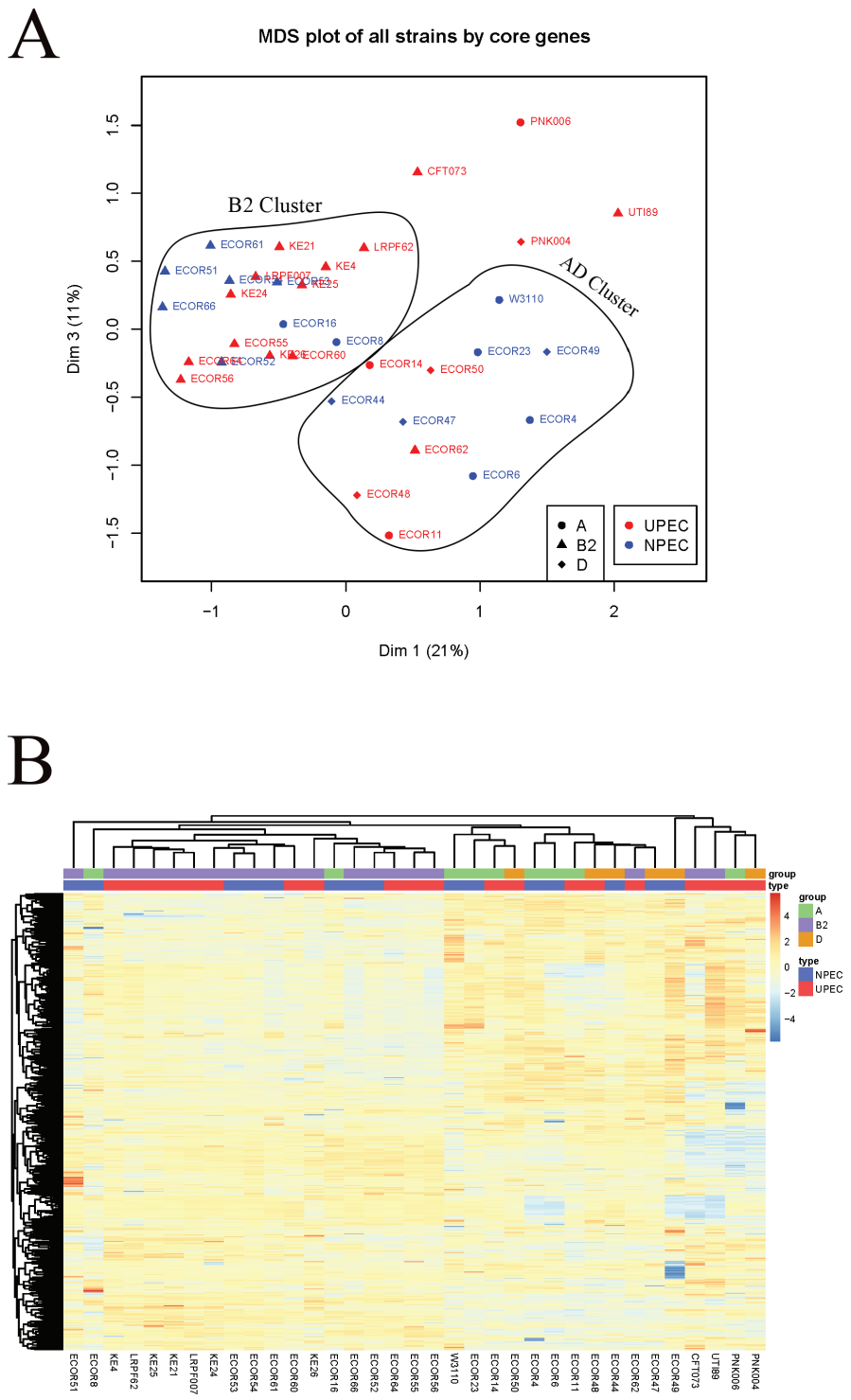


FIG 1 Visualizations of the total transcriptomes of 35 strains of *E. coli*. (A) Dimensional reductive analysis visualized by a MDS plot of these transcriptomes identifies two transcriptomic clusters labeled here as the B2 and AD clusters and four strains that do not form a separate cluster as determined by K-mean analysis. (B) Visualization of the top 1,000 most variable genes identified three branches corresponding to the clusters previously described. Two strains (ECOR 8 and 16) are genomically group A, but transcriptomically cluster B2, and one strain (ECOR62) is genomically group B2, but transcriptomically cluster AD.

virulence. We conclude that identification of a DEG not only requires a statistical analysis, but also an examination of transcripts from individual strains to ensure against false positives. For this reason, when we identify DEGs, we show the transcript levels for genes

in individual B2 and AD cluster strains. Within the AD cluster, no genes were differentially expressed between UPEC and NPEC strains (Table S3).

The transcriptomes of four outlier UPEC strains—CFT073, PNK004, PNK006, and UTI89—are from three different phylogenetic groups. A major difference between CFT073 and other B2 strains was 47-fold more *phoB* transcripts. PNK004's transcriptome is distorted by extraordinarily high expression of *rcsA*, which codes for a regulator of capsule synthesis, and genes of colanic acid. UTI89 and PNK006 did not have major expression distortions that were easily identified. In any case, these strains are not representative of most UPEC strains.

In summary, the transcriptomes of most strains clustered into two major groups which generally, but not always, followed the phylogenetic grouping. No obvious differences in core gene expression existed between UPEC and NPEC strains of the same phylogenetic group. The remaining sections describe the expression differences between B2 and AD cluster strains with a focus on the genes of macromolecular synthesis, macromolecular structure, and energy metabolism.

Macromolecular synthesis and structure

Translation

Gene ontology indicated that B2 cluster strains had greater expression of genes for the translational machinery and nucleotide precursor synthesis. The ratio of the sum of ribosomal protein transcripts to the total number of transcripts ranged from 0.067 to 0.436 and is statistically different between B2 and AD cluster strains (Fig. 2). Expression of translational machinery genes (ribosomal proteins, tRNA, RNA modifying enzymes, etc.) is controlled by the alarmones guanosine tetra- and penta-phosphate (collectively refered to as (p)ppGpp) which are synthesized in response to several stresses (32). (p)ppGpp is synthesized by RelA, binds to DksA, and is degraded by SpoT. Both (p)ppGpp and the (p)ppGpp-DksA complex destabilize weak open promoter complexes such as those for genes of many ribosomal proteins. B2 cluster strains had more *spoT* transcripts, fewer *dksA* transcripts, and about the same number of *relA* transcripts (Fig. 2), which is consistent with lower (p)ppGpp and a higher ribosomal content in B2 cluster strains compared to AD cluster strains.

Ribosome synthesis is dependent on metabolic pathways that generate nucleotides which require purines, pyrimidines, glutamine, glycine, and N^5-N^{10}-methylene tetrahydrofolate. All but one gene for adenine nucleotide synthesis were more highly expressed in B2 cluster strains (Fig. 3A and B). Curiously, genes for synthesis of guanine nucleotides (Fig. 3A) and pyrimidine nucleotides are not differentially expressed (Table S1). The ranges for *purE* and *glnA* normalized transcripts from individual strains do not overlap between the transcriptomic clusters and is particularly striking for *glnA* (Fig. 3B). The parallel expression of ribosomal proteins and anabolic pathways links translation and metabolism. In summary, B2 cluster strains have higher transcripts for components of the translational machinery.

DNA replication and structure

The DEGs for translational machinery suggest that B2 strains have a greater capacity for rapid growth, which implies that DEGs for other macromolecular processes should also be affected. Consistent with this prediction, the B2 strains had more transcripts for many, but not all, genes of the core DNA replisome and deoxyribonucleotide synthesis (Fig. 4A through C). Replication requires resolution of replication blocks, such as those caused by transcription complexes, which are resolved by the PriC, Rep, and UvrD helicases (33). B2 cluster strains had higher levels of *rep* and *uvrD* transcripts, and AD cluster strains had more *priC* transcripts (Fig. 4A). These results suggest that B2 and AD cluster strains either use different mechanisms to overcome replication blocks or encounter different types of replication blocks.

Supercoiling and nucleoid-binding proteins control DNA condensation (34). B2 strains had more transcripts for the gyrase genes (*gyrA* and *gyrB*) whose products introduce

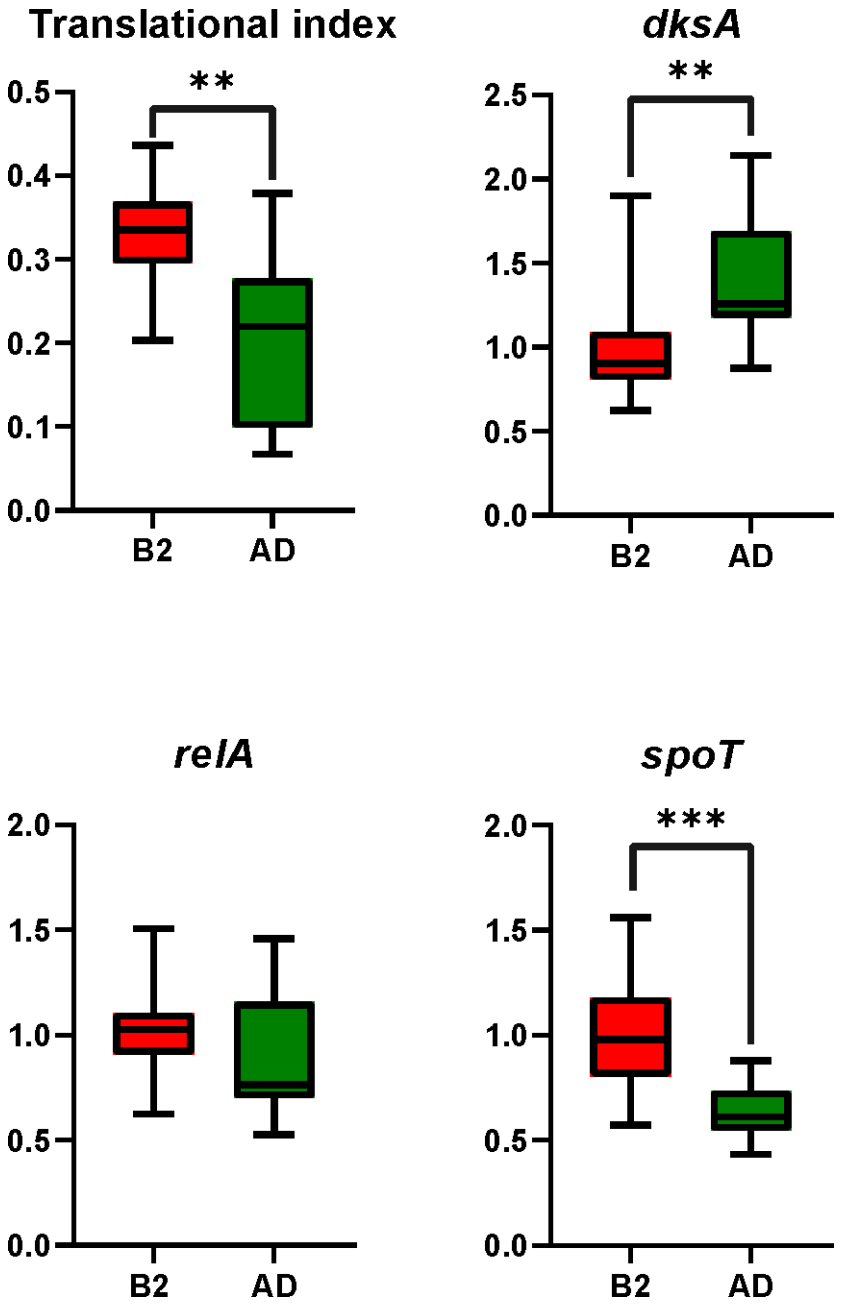


FIG 2 Comparison of transcripts from genes for ribosomal proteins and *dksA*, *relA*, and *spoT*. The translational index is the normalized sum of transcripts for proteins of the ribosome small and large subunits. For *dksA*, *relA*, and *spoT*, the transcripts per million counts are normalized to the average transcripts for B2 cluster strains. **P* < 0.05; ***P* < 0.01; ****P* < 0.001. Statistical significance was determined by Mann-Whitney test because we assumed a non-Gaussian distribution and unequal variance. Transcript values for all B2 and AD cluster strains were used with no exclusions.

negative supercoiling and fewer transcripts for many genes of nucleoid-associated proteins (*hns*, *ihfA*, *matP*, *dps*, and *hfq*) (Fig. 5A and C, results for only the first three genes are shown). The low molecular weight polyamines—putrescine and spermidine—stabilize negatively supercoiled DNA structure and stimulate gyrase activity (35, 36). B2 strains had more transcripts for the polyamine synthetic genes *speA*, *speB*, *speC*, and *speD*, and fewer transcripts for *speG* whose product modifies spermidine (Fig. 5B and C). Furthermore, phylogenetic group B2 strains lack genes of the two *puu* operons whose

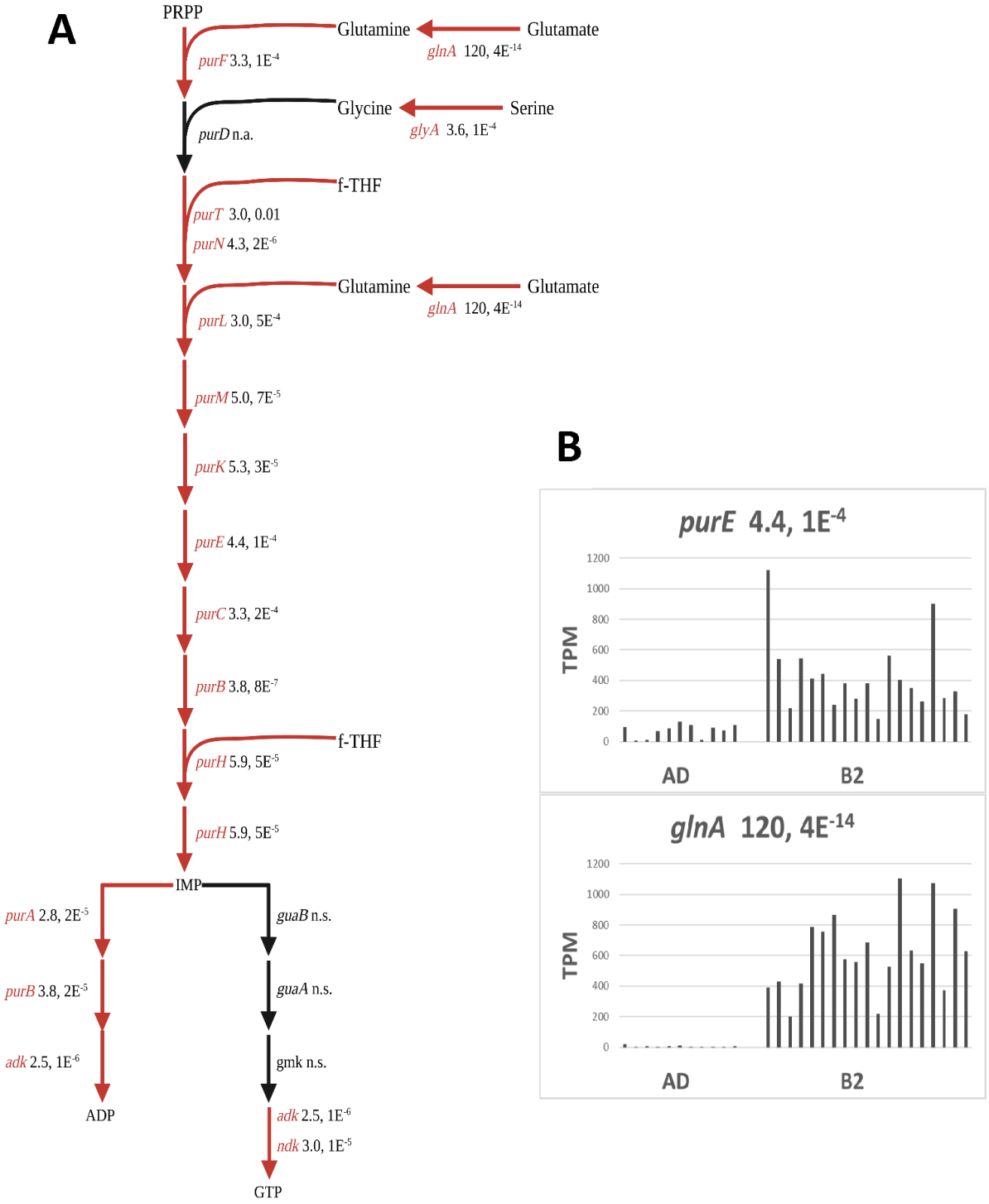


FIG 3 Comparison of transcripts from genes for enzymes that synthesize nucleotides. The first number next to a gene designation is the average number of normalized transcripts from B2 cluster strains divided by the average number of normalized transcripts from AD cluster strains. The second number is the FDR which was determined using the *EdgeR* package. The designation "n.s." means not significant, i.e., FDR <0.05. The designation "n.a." means not available: the gene did not appear as a conserved gene. Such genes could be essential but may have two distinct alleles that are scored as different genes.(A) Genes for proteins that synthesize purines are shown. The red arrow and red gene designation indicate a B2/AD ratio >1 (usually much greater) and FDR <0.05. The black arrow and gene designation indicate a ratio with an FDR >0.05 or that the information was not available because the gene was not considered conserved. f-THF is N^{10}-formyltetrahydrofolate (B) Normalized transcript levels for *glnA* and *purE* are shown from 11 AD cluster strains (the lab strain W3110 was excluded) and all 19 B2 cluster strains. In each histogram, values for individual AD and B2 cluster strains are on the left and right, respectively.

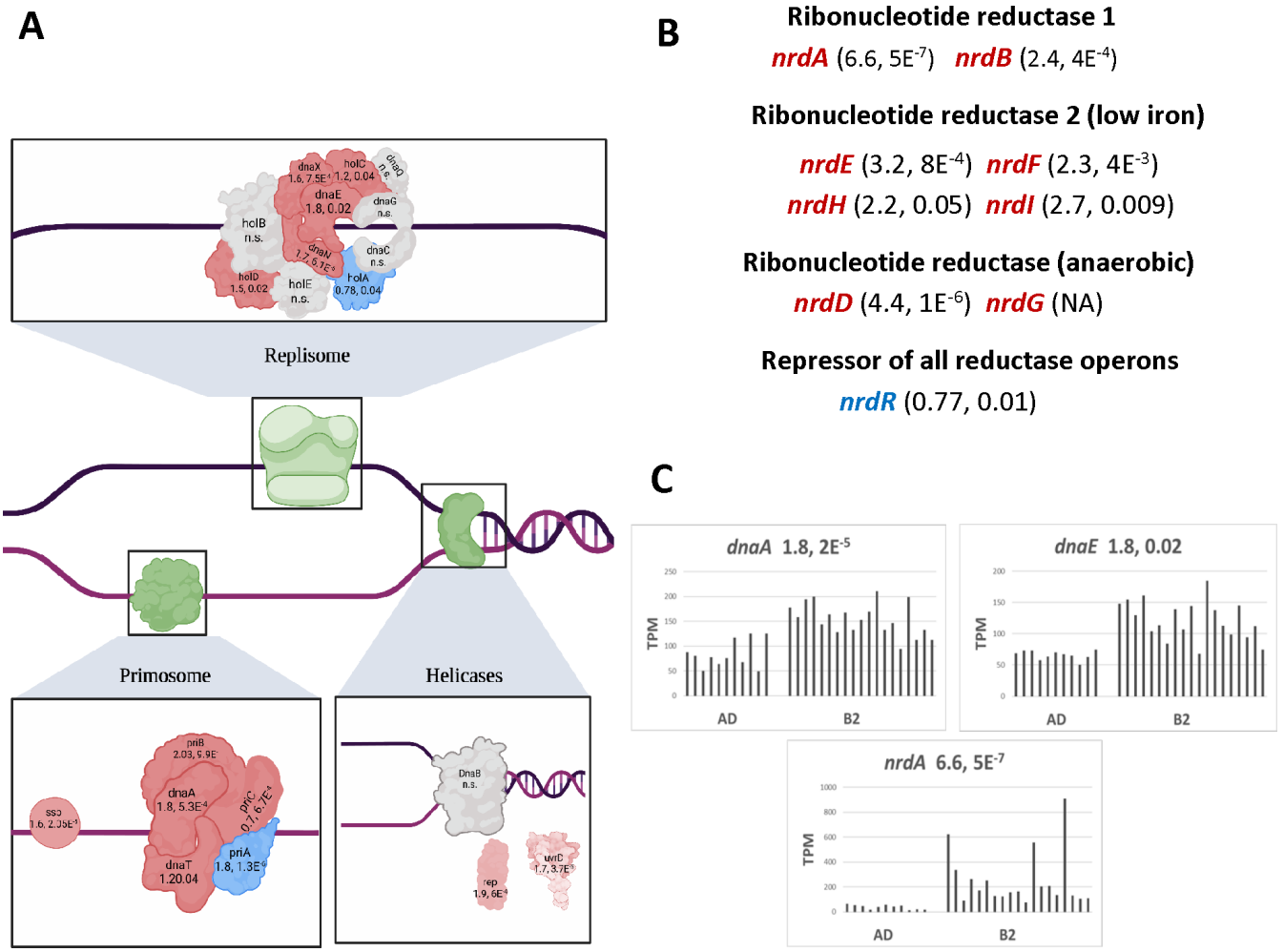


FIG 4 Comparison of transcripts from DNA replication genes. Genes more highly expressed in B2 and AD cluster strains are shown diagrammatically in red and blue, respectively. See the Fig. 3 legend for an explanation of the numbers and abbreviations. (A) Expression of genes whose products are part of the primosome and replisome is shown. The structures in the diagram are not meant to be structurally accurate. (B) The differentially expressed genes for the three ribonucleotide reductases are all higher in the B2 strains. Expression of the repressor, NrdR, is lower in B2 strains. (C) The normalized transcript levels for *dnaA*, *dnaE*, and *nrdA* are shown from individual AD and B2 cluster strains.

products degrade putrescine (37, 38). The normalized transcripts from AD and B2 cluster strains for *gyrA*, *hns*, and *speA* show little overlap for the expression ranges even when the expression ratio is only about twofold (Fig. 5C). The net result of these differences is to increase putrescine and spermidine synthesis and prevent their removal by SpeG and putrescine catabolism. In summary, these results suggest that DNA condensation in B2 cluster strains is more reliant on supercoiling—gyrase expression and increased polyamines which stimulate gyrase activity—whereas DNA condensation in AD cluster strains is more reliant on nucleoid-associated proteins.

Transcription

Three major factors that control transcription—core RNA polymerase, the promoter-recognizing σ factors, and the DNA-binding TFs—were differentially expressed between B2 and AD cluster strains. B2 strains had more transcripts for two core RNA polymerase genes (*rpoB* and *rpoC*) and two σ subunits genes (*rpoH* and *rpoN*) (Fig. 6A and B). The AD cluster strains had more transcripts for *rpoE* whose product responds to envelope stress (40) (Fig. 6A and B). Seventy-four TFs genes were differentially expressed between the AD and B2 cluster strains (FDR <0.05) (see Table S1, Transcription factor tab for expression of

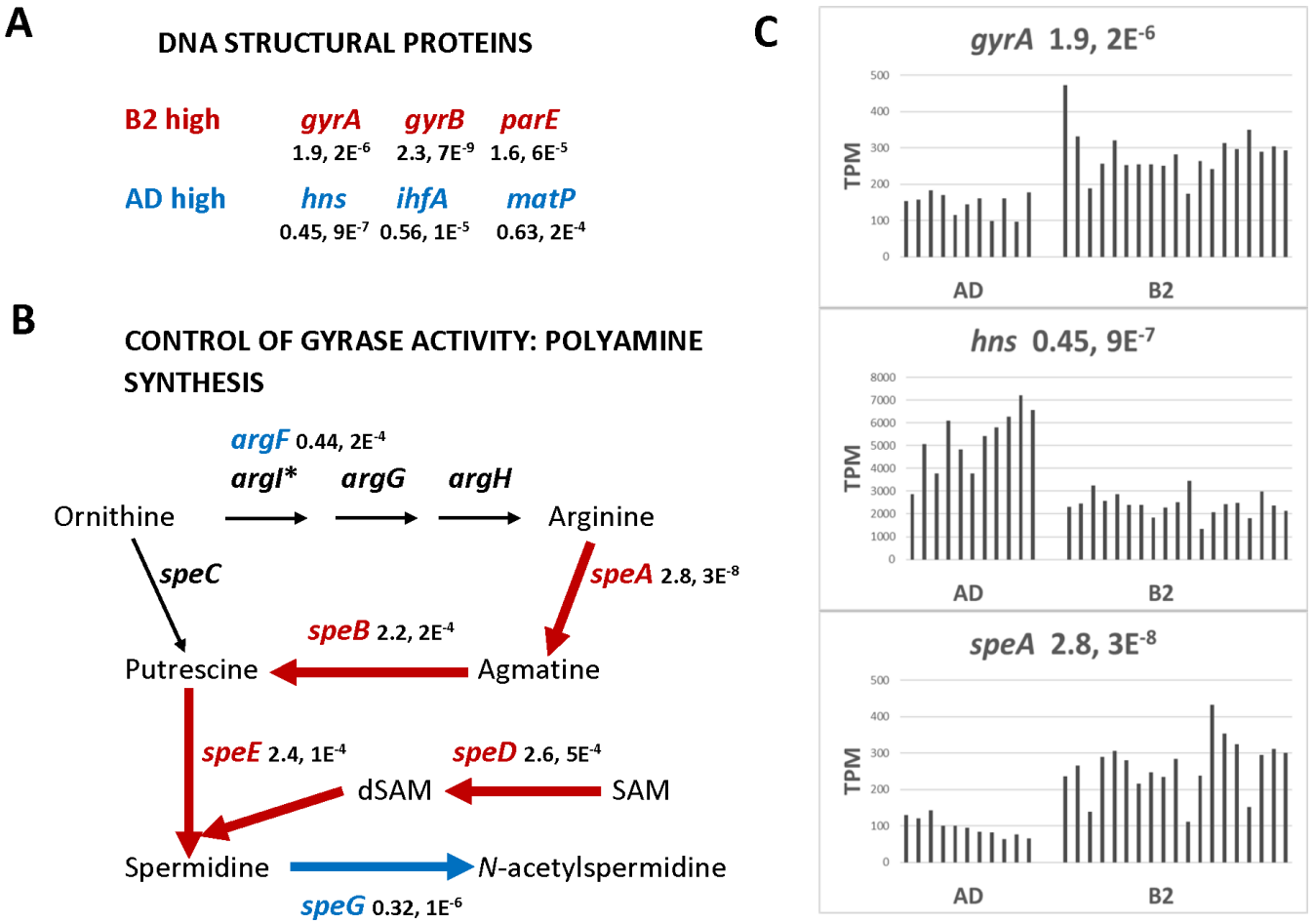


FIG 5 Transcripts for genes of DNA structure. Genes more highly expressed in B2 and AD cluster strains are shown diagrammatically in red and blue, respectively. See the Fig. 3 legend for an explanation of the numbers and abbreviations. (A) Genes for nucleoid-associated proteins and proteins that affect supercoiling are shown if the differential expression had FDR <0.001. The differentially expressed genes with FDRs between 0.05 and 0.001 that were higher in B2 cluster strains were *mukB* and *mukE*, and those higher in AD cluster strains were *dps*, *hfq*, and *hupB* (Table S1, DEG.B2vsAD tab). Several genes that affect DNA structure were not differentially expressed, i.e., FDR >0.05: *fis*, *hupA*, *ihfB*, *mukF*, *parC*, *topB*, and *sbmC* (Table S1, DEG.B2vsAD tab). (B) The pathway and genes for spermidine synthesis which stimulates DNA gyrase activity. The net effect is to increase putrescine and spermidine in B2 cluster strains. *argI* is a B2-specific gene (39). (C) The normalized transcript levels for *gyrA*, *hns*, and *speA* are shown from individual AD and B2 cluster strains.

all TF genes and Fig. S1 for a volcano plot of these genes). The products of 12 of these TF genes control more than 10 operons and >13% of regulon genes were differentially expressed in the growth conditions used. Ten of these 12 TF genes were more highly expressed in AD cluster strains (Table 1). The normalized transcripts from AD and B2 cluster strains for *rpoB* and *rpoE* showed some but not extensive overlap, whereas the normalized transcripts for *phoP* and *glnG* showed little or no overlap (Fig. 6B).

In addition to differentially expressed TF genes, we could deduce increased activity in B2 strains of the CpxAR two-component regulatory system that has been implicated in uropathogenesis and responds to envelope stress (41, 42). CpxAR-regulated *cpxP* was more highly expressed in B2 cluster strains (B2 to AD ratio of 8 with an FDR = 0.002), although *cpxA* and *cpxR* were not differentially expressed (Table S1, DEG tab). The Cpx system counters the TFs that are more highly expressed in AD cluster strains: CpxA physically interacts with ArcA—one of the major transcriptional regulators that are more highly expressed in AD strains (Table 1)—and CpxR represses *rpoE* expression (40).

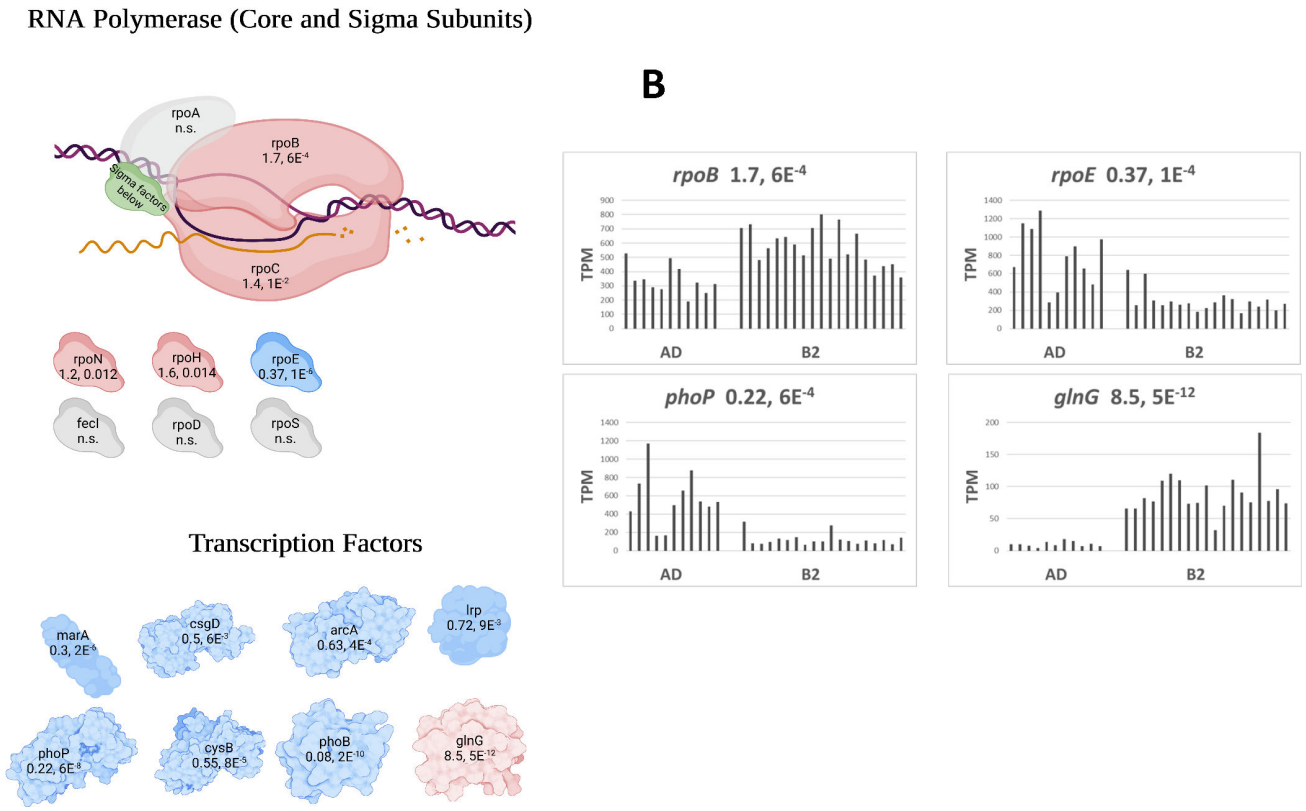


FIG 6 Transcripts for genes of the transcriptional machinery. See the Fig. 3 legend for an explanation of the numbers and abbreviations. (A) Expression of genes for core RNA polymerase, the sigma subunits, and stress-responsive transcription factors that are differentially expressed is shown. Genes more highly expressed in B2 and AD cluster strains are shown diagrammatically in red and blue, respectively. The diagrams are not meant to be structurally accurate. (B) The normalized transcript levels for *rpoB*, *rpoE*, *phoP*, and *glnG* are shown from individual AD and B2 cluster strains.

Peptidoglycan synthesis and structure

The peptidoglycan layer consists of long-chained polysaccharides crosslinked with short peptides which, when initially synthesized, contain L-alanine, D-glutamate, *meso*-diaminopimelate (an intermediate in lysine synthesis), and two molecules of D-alanine. The three phases of peptidoglycan synthesis are precursor formation, synthesis, and maturation [summarized in reference (40)]. All but one gene of the synthetic phase were statistically more highly expressed in B2 strains, even though the B2 to AD transcript ratios were only 20%–90% higher (Fig. 7A and B). Maturation involves processing of the peptide linkage: one or both D-alanine residues may be removed, and glycine can replace one D-alanine. The maturation genes *dacA*, *mrcA*, *mrcB*, and *mrdA* were more highly expressed in B2 strains, while *ftsI*, *ldtD*, and *ldtE* were more highly expressed in AD strains (Fig. 7A and B). The *ldtD* and *ldtE* genes are associated with stress (43). These results suggest possible differences in peptidoglycan structure between the transcriptomic groups. Such differences could partially explain the difficulty in viral-mediated transduction into group B2 UPEC strains (44, 45).

Energy metabolism and iron acquisition

Energy metabolism

B2 cluster strains had more transcripts for several genes of the tricarboxylic acid (TCA) cycle, electron transport (for NADH dehydrogenase I and II, ubiquinone, menaquinone,

TABLE 1 Summary of differentially expressed transcription factor genes[a]

Genes	Log_2FC	Log_2CPM	*P*-value	FDR	Regulon size	Regulon %
glnG	−3.08473	8.003008	5.80E-15	5.16E-12	15	41%
phoB	3.588802	7.522517	6.74E-13	1.96E-10	32	53%
phoP	2.183875	8.59083	1.36E-09	6.02E-08	36	32%
rpoE	1.419381	9.480466	5.61E-08	1.13E-06	75	18%
marA	1.757421	6.084431	1.01E-07	1.71E-06	32	30%
cysB	0.854713	7.764627	1.23E-05	8.38E-05	18	72%
nac	−1.85724	5.072371	6.41E-05	0.000333	134	13.3%
arcA	0.566224	9.862044	8.80E-05	0.000432	83	22%
metJ	0.547261	6.880234	0.001195	0.004014	10	13.3%
csgD	1.009655	2.039974	0.002189	0.006516	13	16.7%
lrp	0.461569	8.974365	0.003317	0.009233	44	21%
rpoN	−0.31829	9.852378	0.004851	0.012809	53	33.3%

[a]Genes presented in this table control more than 10 operons and >13% of the regulon is expressed. All differentially expressed transcription factor genes between B2 and AD cluster strains are presented in Table S1.

and cytochrome bo_3 and *bd* oxidase), ATP synthase (for the *atp* operon), acetate formation (*aceEF-lpdA* and *ackA*), and several gluconeogenic genes (*maeB*, *ppsA*, *fbp*, and *glpX*) (Fig. 8; Table S1). The genes that metabolize malate distinguish B2 from AD cluster strains (Fig. 8A and B): in B2 strains, genes that convert malate to pyruvate and of gluconeogenesis are more highly expressed, while in AD strains, genes that complete the TCA cycle are more highly expressed (Fig. 8A and B). This pattern of metabolism is consistent with a previous genetic analysis of energy metabolism in the UPEC strain CFT073: a reliance on amino acid degradation via the TCA cycle and gluconeogenesis instead of glycolysis (46). These results are remarkable because the glucose in the media should inhibit cyclic-AMP synthesis which is required for expression of most TCA cycle enzymes, the cytochrome bo_3 oxidase specified by the *cyoABCD* operon, the *aceEF-lpdA* operon, and *glpX* [summarized in reference (40)]. This apparent paradox is explained by a flux balance analysis of growth in a complex medium with glucose and amino acids that showed that the fastest growth is achieved with preferential use of amino acids which reduces glucose uptake (rapid amino acid degradation increases intracellular pyruvate which, in turn, inhibits glucose transport) and increases acetate overflow (47). The reduced glucose uptake could account for expression of genes that require cyclic-AMP. For glycolytic genes in either the Embden-Meyerhof-Parnas or the oxidative arm of the pentose-phosphate pathway, group B2 strains had more transcripts for *ptsG*, *gapA*, *tpi*, and *gndA* and fewer transcripts for *gpmM*, *eno*, *pykA*, and *zwf*. These DEGs do not suggest an obvious pattern. Despite these results, evidence exists that carbohydrate catabolism is important for UTI89 under certain circumstances, which suggests that our understanding of carbohydrate catabolism in UPEC strains is incomplete (48, 49). In summary, B2 cluster strains appears to have a pattern for the genes of energy metabolism that differs from that of AD cluster strains.

Iron acquisition

Iron is important for several components of energy-generating pathways and, not surprisingly, iron acquisition is a crucial virulence factor (50). B2 cluster strains had more transcripts for most of the core iron acquisition genes: (i) energy-transducing iron-chelate transport (*tonB*, *exbB*, and *exbD*), (ii) transport systems for ferric citrate (*fec* genes), ferrous iron (*feo* genes), ferric enterochelin transport (*fep* genes), and ferrichrome (*fhu* genes), and (iii) synthesis of enterochelin (*ent* genes) and enterochelin precursors (several *aro* genes) (Fig. 9A and B; Table S1, iron acquisition tab). Fur complexed to ferrous iron represses many of these genes, but *fur* was not differentially expressed between B2 and AD cluster strains. Therefore, greater expression of iron acquisition in B2 cluster strains implies lower intracellular iron. In summary, B2 cluster strains had higher expression of numerous iron acquisition genes.

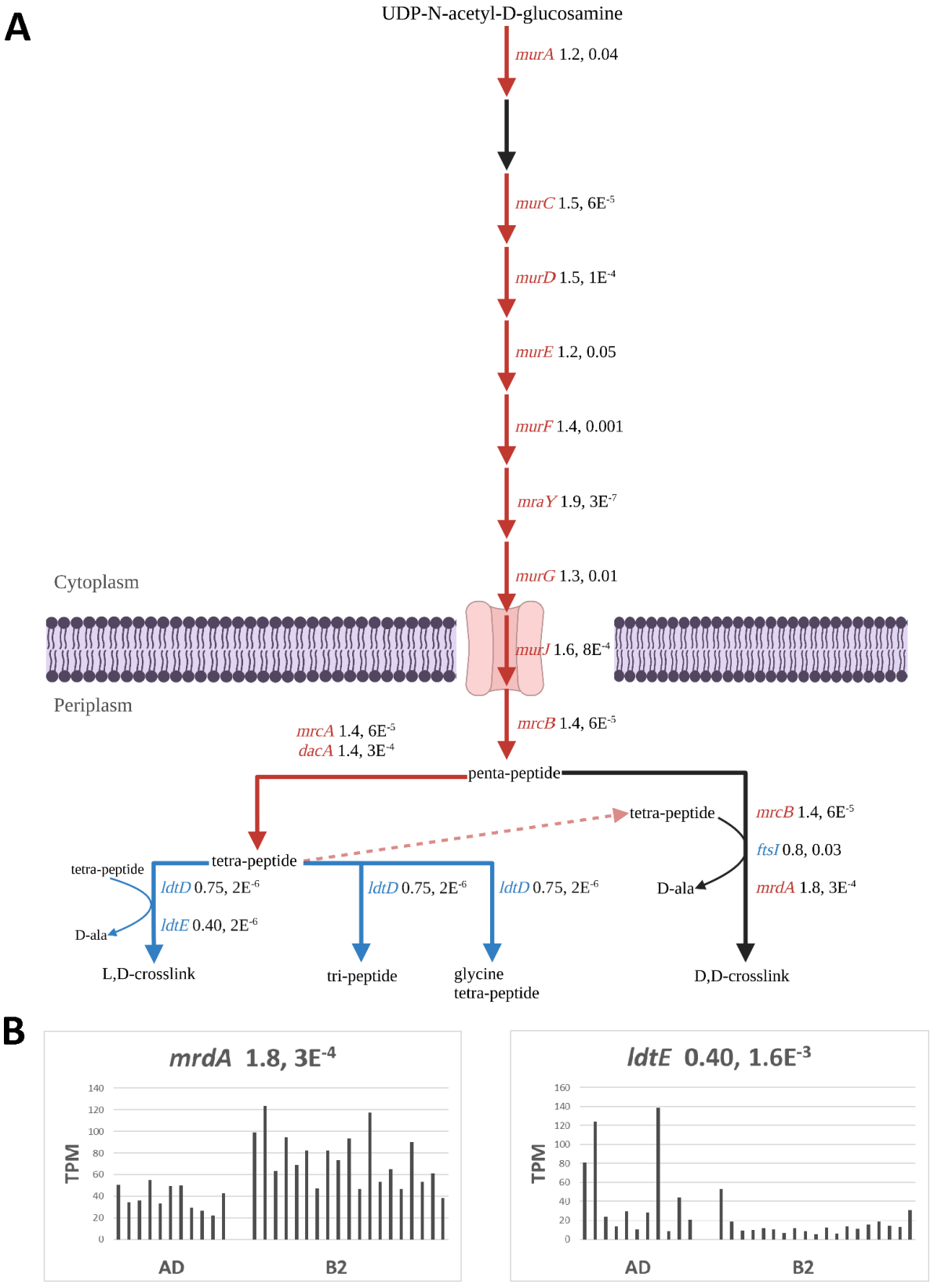


FIG 7 Transcripts for genes of peptidoglycan synthesis. (A) Expression of genes for peptidoglycan synthesis is shown starting from UDP-N-acetyl-D-glucosamine. Genes more highly expressed in B2 and AD cluster strains are shown diagrammatically in red and blue, respectively. See the Fig. 3 legend for an explanation of the numbers and abbreviations. Only genes with expression differences with an FDR <0.05 are shown. (B) The normalized transcript levels for *mrdA* and *ldtE* are shown from individual AD and B2 cluster strains.

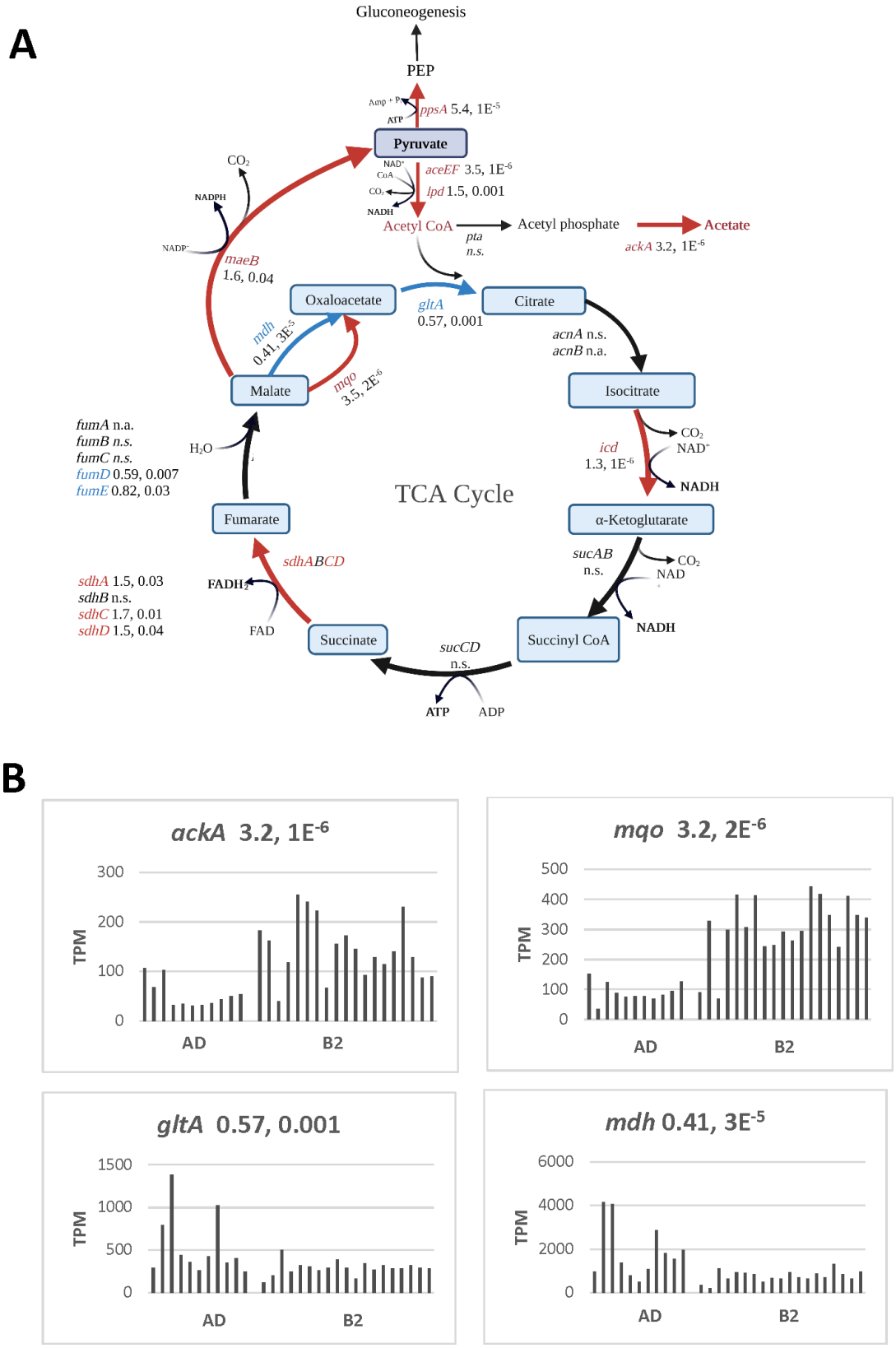


FIG 8 Transcripts from genes for TCA cycle enzymes. (A) TCA cycle genes that are more highly expressed in B2 and AD cluster strains are shown diagrammatically in red and blue, respectively. The black designations are for genes that are not differentially expressed between AD and B2 cluster strains. See the Fig. 3 legend for an explanation of the numbers and abbreviations. (B) The normalized transcript levels from individual AD and B2 cluster strains are shown for *ackA* and *mqo* which are more highly expressed in B2 cluster strains, and for *gltA* and *mdh* which are more highly expressed in AD cluster strains.

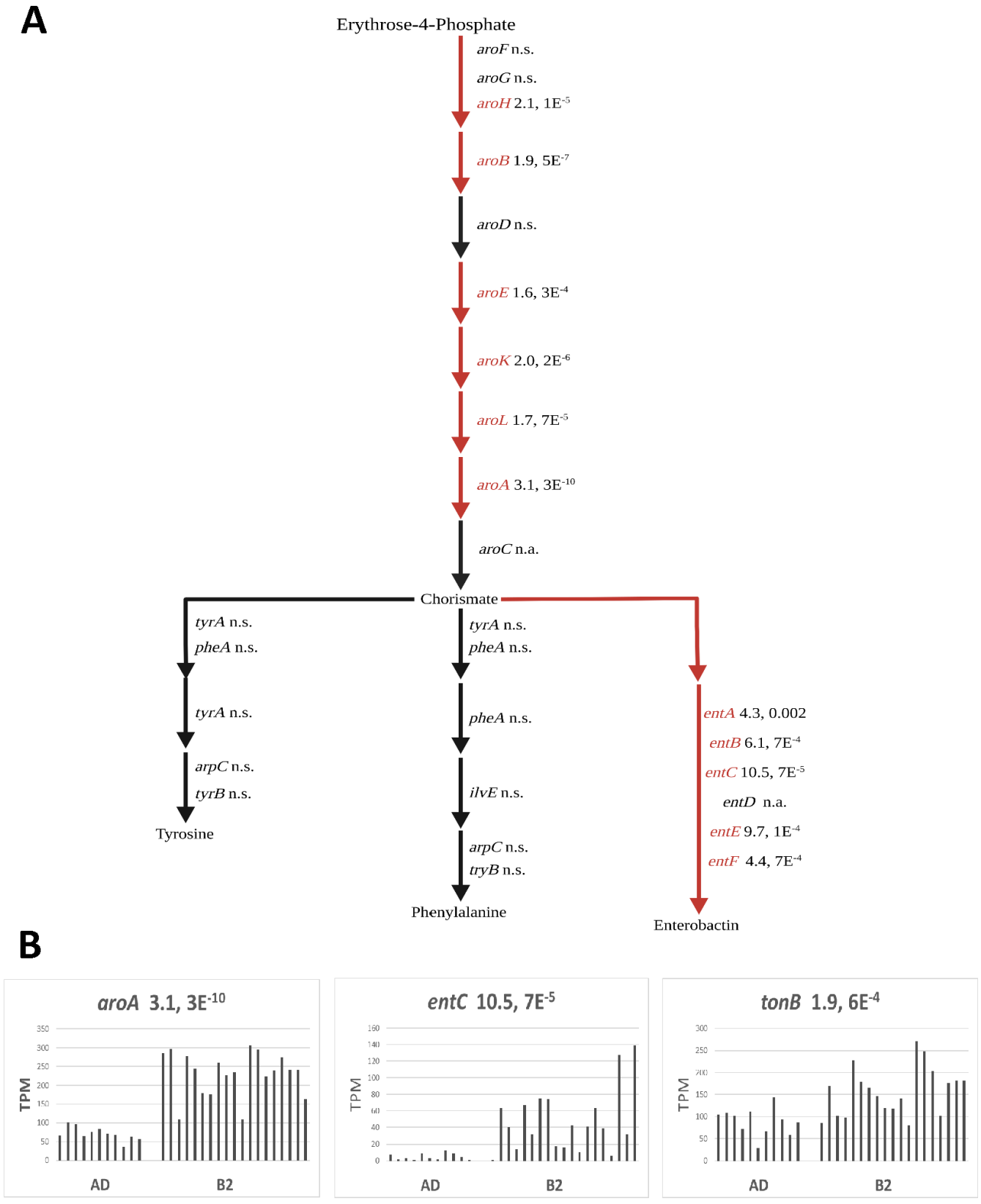


FIG 9 Transcripts for iron acquisition genes. (A) Expression of genes for synthesis of enterochelin and two other compounds, tyrosine and phenylalanine, derived from chorismate is shown. The red designations indicate genes that are more highly expressed in B2 cluster strains, and the black designations are for genes that are not differentially expressed between AD and B2 cluster strains. See the Fig. 3 legend for an explanation of the numbers and abbreviations. (B) The normalized transcript levels for *aroA*, *entC*, and *tonB* are shown from individual AD and B2 cluster strains. All are more highly expressed in B2 cluster strains.

DISCUSSION

The relation between phylogenetics and transcriptomics

Our first major conclusion is that *E. coli* phylogenetic group B2 strains, which are associated with several pathologies, had a distinct core gene transcription pattern when grown in a nutrient-rich environment. Almost half of the core genes were differentially expressed between B2 and AD cluster strains. The transcriptomic pattern generally followed phylogeny, which is based on the presence of specific chromosomal genes, although an occasional group A strain possessed the B2 transcriptome pattern, and vice versa (Fig. 1A). The malleability and discreteness of the transcriptome patterns suggest that the transition involves only a few genes, but our results do not suggest which genes are responsible for the transcriptional rewiring. One study in particular—the longest-running microbial evolution experiment—described a transcriptomic rewiring that affected some of the same genes that distinguish B2 and AD cluster strains: genes that control DNA supercoiling and the stringent response (51). The evolved changes were the result of daily cycles of starvation and growth which is likely to be relevant for pathogens that must survive and replicate in different, constantly changing, and nutritionally challenging environments. Recent results suggest that (p)ppGpp controls *gyrA* expression, which provides a mechanistic link between DNA supercoiling and the stringent response (52). A possible explanation for the distinct B2-type transcriptome is lower (p)ppGpp levels and more DNA supercoiling. Additional experiments are required to determine whether (p)ppGpp and supercoiling are sufficient to account for the distinct B2 transcriptome.

Evolutionary analysis has suggested that group B2 strains diverged early from other *E. coli* strains and have lost several stress-adaptive genes, which implied a different adaptive trajectory (21). In addition to loss of stress-adaptive genes, our results show that B2 strains have lower expression of stress-adaptive core genes in a relatively stress-free nutrient-rich environment (Table 1). Despite their lower expression in B2 cluster strains, several of these stress-responsive TFs are important during virulence of *E. coli* and other pathogens. The *phoP* gene is important for virulence of the uropathogen CFT073 and *Salmonella* and has been associated with control of the proton motive force, intracellular survival, and responses to low pH, low magnesium, and antimicrobial peptides (53) (54, 55). The phosphate-sensing PhoR-PhoB system has also been associated with *E. coli* virulence and is required for expression of a gene that regulates type I pili synthesis in CFT073 (56) (57, 58). RpoE has been implicated in virulence of *Salmonella* and *E. coli*, including UPEC strains (59–62). RpoE is stress-induced sigma subunit for RNA polymerase that responds to high temperature and osmolarity, misfolded proteins in the outer membrane, periplasm, and inner membrane, metal ion exposure, entry into stationary phase, and elevated (p)ppGpp (40). Finally, the stringent response has been associated with virulence and antibiotic resistance in a variety of pathogens (55, 63–68). Further experiments are required to determine whether B2 and AD strains differ in sensing and responding to a variety of stresses.

Little is known about the metabolic and physiological differences between the phylogenetic groups. Our results showed that B2 cluster strains had higher expression for several metabolic genes, including those for macromolecular synthesis, several components of the TCA cycle, electron transport, and acetogenesis. We have recently observed that two group B2 UPEC strains grow better in urine and dilute lab medium than a group A NPEC strain, which implies better nutrient acquisition (29). The higher levels of iron acquisition genes in group B2 strains are consistent with this observation (Fig. 9). Better nutrient acquisition will support faster growth in nutrient-poor environments, which is consistent with the faster growth of UPEC strains in urine compared to NPEC strains (27, 28). We propose that efficient nutrient acquisition and the capacity for faster growth—higher expression of TCA cycle enzymes, several electron transport components, ATP synthase, and iron acquisition in B2 strains—are coupled properties that enhance growth in urine.

Distinguishing UPEC and NPEC strains

Our second major conclusion is that transcriptomic analysis of core gene expression did not identify major differences between UPEC and NPEC strains within the B2 cluster. We cannot conclude that B2-specific properties are sufficient for virulence because our analysis was not capable of identifying strain-specific adaptations to individual hosts. S. Hultgren and colleagues have emphasized the importance of host-specific properties and host-pathogen interactions which limit generalizations about virulence (22). Such considerations would be expected for an opportunistic pathogen: the virulence mechanism may be dependent on the specific host, and B2 strains may have greater adaptability.

The problem with model UPEC strains

The vast majority of UPEC studies have focused on the model group B2 strains UTI89 and CFT073 with few studies of strains in other groups. Our analysis showed these two strains were among four UPEC strains that did not cluster within the B2 or AD transcriptome patterns. A distinctive property of CFT073 was 47-fold overexpression of *phoB* compared to other B2 strains. The only other strain with a similar level of *phoB* was W3110. These two strains have been passaged in labs for decades, and the *phoB* hyperexpression in both strains may be a property of laboratory evolution in the low phosphate amino acid-rich broths that are commonly used in labs, i.e., L-broth (LB). UTI89 has also been extensively passaged, which may have altered its transcriptome. The outliers' transcriptomes may be inherent to these strains when they were originally isolated, and given the diversity of host environments, these strains may provide information on some extreme environments. However, their properties may also result from lab evolution, and some of their properties are not necessarily shared with other UPEC strains.

Does the group B2-type transcription pattern contribute to virulence?

Group B2 strains are often associated with UTIs and a variety of pathologies. Virulence is necessarily a property of both the bacteria and the host response to the infection, which degrades with age and presumably shows substantial variation between individuals. With respect to the bacteria, we have established that B2 strains possess a unique transcription pattern when grown in a nutrient-rich medium. In addition, B2 strains also possess a unique set of non-core accessory genes that define the B2 group. Both properties of B2 strains are linked if the transcription factors encoded by core genes control expression of the accessory genes. Mobley and colleagues have proposed a distinct transcription pattern for UPEC strains isolated from a physiologically relevant environment (23, 69). Most, but not all, of these strains were group B2, and they argued that the transcription pattern is not a property of B2 strains (69). We can reconcile our results with this conclusion if the non-B2 strains exhibit a B2 transcription pattern, either by mutation or adaptation to a physiologically relevant environment. Nonetheless, several observations suggest that the B2 transcription pattern does contribute to virulence. First, the transcriptome of UPEC strains analyzed from the host shows high levels of the translational machinery (23, 69), which our analysis showed is a property of B2 cluster strains. Second, analysis of the pathway requirements for the model strain CFT073 indicates use of the TCA cycle and gluconeogenesis (46), and the same pattern is suggested by the B2 transcription pattern. Third, acetogenesis is a property of UPEC strains and acetate is a UTI biomarker (28, 70, 71). Acetogenesis is suggested by the B2-specific transcriptome. Fourth, CFT073 has been shown to have high levels of iron acquisition genes when grown in urine (28). Our results suggest that B2 strains have high endogenous levels of iron acquisition genes. The contribution of any single factor, such as a B2-type transcriptome, to virulence cannot be precisely determined. However, because UPEC gene expression in physiologically relevant environments parallels the gene expression pattern of B2 strains that we observed, we propose that the B2 transcription pattern is one, but not the only, factor that contributes to virulence. Further

studies will be required to determine where and when during an infection the transcription pattern described in this study is important.

Limitations of this study

The major limitations of our study are conceptual. Because of the need to grow UPEC and NPEC strains at similar rates, the bacteria were grown in a non-physiological environment. Understanding virulence will require analysis from bacteria grown in "more relevant" environments. Our results did not find differences between UPEC and NPEC strains, but such differences may exist. Our results suggest differences in the expression of stress-responsive transcription factor genes during growth in a nutrient-rich lab medium, but provide no information on differences in the responses to stress between the B2 and AD cluster strains. These topics require further study.

MATERIALS AND METHODS

Strains, bacterial growth, and RNA isolation

Strains and their sources are listed in Table 2. Strains were grown to single colonies overnight on LB. A culture was inoculated from a single colony into 10 mL of GT media (0.5% glucose, 1% tryptone, 0.25% NaCl) until mid-log growth phase, collected by a table top centrifugation at maximum speed for 10 min at 4°C, and the cell pellets were frozen at −80°C.

Frozen cell pellets were thawed on ice, suspended in 0.7 mL of buffer RLT (from Qiagen RNeasy Mini kit), and mechanically lysed using a bead beater (FastPrep-24 Classic from MP Biomedical) set at the highest setting for three 45-s cycles with a 5-min rest period on ice between cycles. Once lysed, a Qiagen RNeasy Mini Kit was used to collect RNA and treated with RNase-free DNase to remove DNA. Samples had to have at least 10 µg of RNA as determined by A260 (nanodrop). To remove the DNase and concentrate the RNA, a GeneJet RNA Isolation and Concentration Kit was used. We continued only if purified RNA was ≥1 µg/µL (quantified by A_{260} by nanodrop) and the A_{260}/A_{230} ratio was between 1.9 and 2.2. Then, RNA was quality checked using a Qubit 4 fluorometer (ThermoFisher) and we continued only with an RNA IQ score ≥8.0. To ensure the RNA was not degraded, 1 µg was added to a 1.0% agarose gel and electrophoresed for 1 h at 90V. If the rRNA bands were clear with no smearing, then 1 µg RNA was subjected to reverse transcription. cDNA was used as a template for qPCR using *arcA* as a target and *rpoD* as a library control. RNA was submitted if the no reverse transcriptase control (used to test for small amounts of contaminating DNA) returned at greater than 30 cycles, and the cDNA-containing reactions produced detectable amplification at ≤20 cycles. When the RNA passed all of the tests above, it was submitted to the genomics core facility at the University of Texas at Dallas for RNA sequencing. The core performed rRNA removal (RiboMinus Transcriptome Isolation Kit or RiboCop bacterial rRNA depletion—Lexogen), library preparation (Stranded Total RNA Prep—Illumina), and single-end Illumina sequencing.

Analysis of the RNA sequences

We compared the within-strain variances for five strains—CFT073, PNK004, PNK006, UTI89, and W3110—for which RNA was isolated three times to the between-strain variances. The within-strain gene variances were much smaller than the between-strain variances (Fig. S2). Therefore, we isolated RNA from most strains only once. The logic of this analysis was that analyzing more strains was more important than analyzing fewer strains with more replicates. We validated differences for *phoP* with quantitative reverse transcriptase PCR (Fig. S3), and for *glnA* with a *glnA-gfp* fusion as part of a separate study (77).

Annotations for KE and LRPF strains were performed by National Center for Biotechnology Information (NCBI) using pan-genome analysis pipeline (PGAP, version 6.4). The

TABLE 2 Strain list

Strain	Group	NPEC/UPEC	Source
ECOR 4	A	NPEC	(72)
ECOR 6	A	NPEC	(72)
ECOR 8	A	NPEC	(72)
ECOR 11	A	UPEC	(72)
ECOR 14	A	UPEC	(72)
ECOR 16	A	NPEC	(72)
ECOR 23	A	NPEC	(72)
ECOR 51	B2	NPEC	(72)
ECOR 52	B2	NPEC	(72)
ECOR 53	B2	NPEC	(72)
ECOR 54	B2	NPEC	(72)
ECOR 55	B2	UPEC	(72)
ECOR 56	B2	UPEC	(72)
ECOR 60	B2	UPEC	(72)
ECOR 61	B2	NPEC	(72)
ECOR 62	B2	UPEC	(72)
ECOR 64	B2	UPEC	(72)
ECOR 66	B2	NPEC	(72)
ECOR 44	D	NPEC	(72)
ECOR 47	D	NPEC	(72)
ECOR 48	D	UPEC	(72)
ECOR 49	D	NPEC	(72)
ECOR 50	D	UPEC	(72)
CFT073	B2	UPEC	(73)
KE4	B2	UPEC	(74)
KE21	B2	UPEC	(74)
KE24	B2	UPEC	(74)
KE25	B2	UPEC	(74)
KE26	B2	UPEC	(74)
LRPF007	B2	UPEC	(74)
LRPF62	B2	UPEC	(74)
PNK004	D	UPEC	(75)
PNK006	A	UPEC	(75)
UTI89	B2	UPEC	(76)
W3110	A	NPEC	Lab strain

gene annotation files of the other 28 strains were downloaded from FTP server in NCBI (https://ftp.ncbi.nlm.nih.gov/genomes/). All gff3 files were used to generate a pan-transcriptome by *Roary* (version 3.11.2) (78) with default parameter setting to assign all transcripts to 17,098 gene clusters. Among them, 2,689 clusters had ortholog genes in at least 34 strains and were defined as the core gene set. We used *Kallisto* (version 0.45.1) (79) to quantify gene expression from RNA-seq data of each strain by using annotated transcripts from all strains as reference. The gene expression of each cluster was defined as the sum of expression from all transcripts belonging to this cluster. To perform DEG analysis, we used *EdgeR* (version 3.40.2) (80, 81) to analyze read counts of clusters in core gene set across all strains. DEGs were detected with an adjusted *P*-value less than 0.05. *ClusterProfiler* (version 4.6.0) (82, 83) was used to perform function and pathway enrichment analysis. K-mean clustering ($n = 3$) was performed with all core gene-related clusters in *R* for grouping all strains.

ACKNOWLEDGMENTS

This work was supported by a UT Dallas Collaborative Biomedical Research Award grant to L.R. and P.E.Z., and Cain Foundation money to P.E.Z.

J.H. would like to thank Prof. Kelli Palmer, Ph.D., of The University of Texas at Dallas for her technical help, and use of space and resources. The guidance is forever appreciated. J.H., P.E.Z., and L.R. would like to thank the Genome Center at The University of Texas at Dallas for the services to support our research.

All authors contributed to the study design and coordination. J.H. performed all of the experiments and most of the statistical analysis. Z.X. directed and supervised the bioinformatics. P.E.Z. collected clinical bacterial isolates. P.E.Z. and L.R. obtained funding. L.R. supervised the wet lab work and wrote the manuscript. All authors contributed to the editorial process.

The authors declare no potential or realized conflict of interest.

AUTHOR AFFILIATIONS

[1]Department of Biological Sciences, The University of Texas at Dallas, Richardson, Texas, USA
[2]Department of Urology, The University of Texas Southwestern, Dallas, Texas, USA

AUTHOR ORCIDs

Jacob Hogins http://orcid.org/0000-0001-6041-7332
Philippe E. Zimmern http://orcid.org/0000-0002-7612-2042
Larry Reitzer http://orcid.org/0000-0002-4406-6090

FUNDING

Funder	Grant(s)	Author(s)
UT Dallas Collaborative Biomedical Research Award		Philippe E. Zimmern
		Larry Reitzer

AUTHOR CONTRIBUTIONS

Jacob Hogins, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – review and editing | Zhenyu Xuan, Data curation, Formal analysis, Software, Visualization, Writing – review and editing | Philippe E. Zimmern, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing | Larry Reitzer, Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

RNA sequencing data is available at GSE235964 through NCBI GEO.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Figures S1 to S3 (Spectrum02085-23-s0001.docx). S1: volcano plot of transcription factor genes; S2: Variation between strains versus between replicates; S3: RNAseq validation of phoP by quantitative reverse transcriptase PCR.
Table S1-S3 legends (Spectrum02085-23-s0002.docx). Overall description of contents of the supplemental Excel files.

Table S1 (Spectrum02085-23-s0003.xlsx). Differentially expressed genes between AD and B2 cluster genes.

Table S2 (Spectrum02085-23-s0004.xlsx). Gene expression between pathogens and nonpathogens within group B2 cluster strains.

Table S3 (Spectrum02085-23-s0005.xlsx). Differentially expressed genes between pathogenic and nonpathogenic strains of AD cluster strains.

REFERENCES

1. Rowe TA, Juthani-Mehta M. 2013. Urinary tract infection in older adults. Aging health 9. https://doi.org/10.2217/ahe.13.38
2. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 13:269–284. https://doi.org/10.1038/nrmicro3432
3. Foxman B. 2010. The epidemiology of urinary tract infection. Nat Rev Urol 7:653–660. https://doi.org/10.1038/nrurol.2010.190
4. Selekman RE, Shapiro DJ, Boscardin J, Williams G, Craig JC, Brandström P, Pennesi M, Roussey-Kesler G, Hari P, Copp HL. 2018. Uropathogen resistance and antibiotic prophylaxis: a meta-analysis. Pediatrics 142:e20180119. https://doi.org/10.1542/peds.2018-0119
5. Glover M, Moreira CG, Sperandio V, Zimmern P. 2014. Recurrent urinary tract infections in healthy and nonpregnant women. Urol Sci 25:1–8. https://doi.org/10.1016/j.urols.2013.11.007
6. Anonymous. 2021. Global antimicrobial resistance and use surveillance system (GLASS) report: 2021. World Health Organization, Geneva. Available from: https://apps.who.int/iris/handle/10665/341666
7. Malik RD, Wu YR, Christie AL, Alhalabi F, Zimmern PE. 2018. Impact of allergy and resistance on antibiotic selection for recurrent urinary tract infections in older women. Urology 113:26–33. https://doi.org/10.1016/j.urology.2017.08.070
8. Ejrnæs K. 2011. Bacterial characteristics of importance for recurrent urinary tract infections caused by *Escherichia coli*. Dan Med Bull 58:B4187.
9. Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep 5:58–65. https://doi.org/10.1111/1758-2229.12019
10. Clermont O, Gordon D, Denamur E. 2015. Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. Microbiology (Reading) 161:980–988. https://doi.org/10.1099/mic.0.000063
11. Lo Y, Zhang L, Foxman B, Zöllner S. 2015. Whole-genome sequencing of uropathogenic *Escherichia coli* reveals long evolutionary history of diversity and virulence. Infect Genet Evol 34:244–250. https://doi.org/10.1016/j.meegid.2015.06.023
12. Chen SL, Wu M, Henderson JP, Hooton TM, Hibbing ME, Hultgren SJ, Gordon JI. 2013. Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. Sci Transl Med 5:184ra60. https://doi.org/10.1126/scitranslmed.3005497
13. Basmaci R, Bonacorsi S, Bidet P, Biran V, Aujard Y, Bingen E, Béchet S, Cohen R, Levy C. 2015. *Escherichia coli* meningitis features in 325 children from 2001 to 2013 in France. Clin Infect Dis 61:779–786. https://doi.org/10.1093/cid/civ367
14. Wijetunge DSS, Gongati S, DebRoy C, Kim KS, Couraud PO, Romero IA, Weksler B, Kariyawasam S. 2015. Characterizing the pathotype of neonatal meningitis causing *Escherichia coli* (NMEC). BMC Microbiol 15:211. https://doi.org/10.1186/s12866-015-0547-9
15. Mirsepasi-Lauridsen HC, Vallance BA, Krogfelt KA, Petersen AM. 2019. *Escherichia coli* pathobionts associated with inflammatory bowel disease. Clin Microbiol Rev 32:e00060-18. https://doi.org/10.1128/CMR.00060-18
16. Kamali Dolatabadi R, Feizi A, Halaji M, Fazeli H, Adibi P. 2021. The prevalence of adherent-invasive *Escherichia coli* and its association with inflammatory bowel diseases: a systematic review and meta-analysis. Front Med (Lausanne) 8:730243. https://doi.org/10.3389/fmed.2021.730243
17. Kotlowski R, Bernstein CN, Sepehri S, Krause DO. 2007. High prevalence of *Escherichia coli* belonging to the B2+D phylogenetic group in inflammatory bowel disease. Gut 56:669–675. https://doi.org/10.1136/gut.2006.099796
18. Nunes PHS, Valiatti TB, Santos A de M, Nascimento J da S, Santos-Neto JF, Rocchetti TT, Yu MCZ, Hofling-Lima AL, Gomes TAT. 2022. Evaluation of the pathogenic potential of *Escherichia coli* strains isolated from eye infections. Microorganisms 10:1084. https://doi.org/10.3390/microorganisms10061084
19. Kohoutova D, Smajs D, Moravkova P, Cyrany J, Moravkova M, Forstlova M, Cihak M, Rejchrt S, Bures J. 2014. *Escherichia coli* strains of phylogenetic group B2 and D and bacteriocin production are associated with advanced colorectal neoplasia. BMC Infect Dis 14:733. https://doi.org/10.1186/s12879-014-0733-7
20. Badger-Emeka LI, Kausar N, Estrella E, Angeles GB. 2022. A three-year look at the phylogenetic profile, antimicrobial resistance, and associated virulence genes of uropathogenic *Escherichia coli*. Pathogens 11:631. https://doi.org/10.3390/pathogens11060631
21. Gonzalez-Alba JM, Baquero F, Cantón R, Galán JC. 2019. Stratified reconstruction of ancestral *Escherichia coli* diversification. BMC Genomics 20:936. https://doi.org/10.1186/s12864-019-6346-1
22. Schreiber HL, Conover MS, Chou W-C, Hibbing ME, Manson AL, Dodson KW, Hannan TJ, Roberts PL, Stapleton AE, Hooton TM, Livny J, Earl AM, Hultgren SJ. 2017. Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. Sci Transl Med 9:382. https://doi.org/10.1126/scitranslmed.aaf1283
23. Sintsova A, Frick-Cheng AE, Smith S, Pirani A, Subashchandrabose S, Snitkin ES, Mobley H. 2019. Genetically diverse uropathogenic *Escherichia coli* adopt a common transcriptional program in patients with UTIs. Elife 8:e49748. https://doi.org/10.7554/eLife.49748
24. Borderon E, Horodniceanu T. 1978. Metabolically deficient dwarf-colony mutants of *Escherichia coli*: deficiency and resistance to antibiotics of strains isolated from urine culture. J Clin Microbiol 8:629–634. https://doi.org/10.1128/jcm.8.6.629-634.1978
25. McIver CJ, Tapsall JW. 1995. Virulence and other phenotypic characteristics of urinary isolates of cysteine-requiring *Escherichia coli*. J Med Microbiol 42:39–42. https://doi.org/10.1099/00222615-42-1-39
26. Trülzsch K, Hoffmann H, Keller C, Schubert S, Bader L, Heesemann J, Roggenkamp A. 2003. Highly resistant metabolically deficient dwarf mutant of *Escherichia coli* is the cause of a chronic urinary tract infection. J Clin Microbiol 41:5689–5694. https://doi.org/10.1128/JCM.41.12.5689-5694.2003
27. Gordon DM, Riley MA. 1992. A theoretical and experimental analysis of bacterial growth in the bladder. Mol Microbiol 6:555–562. https://doi.org/10.1111/j.1365-2958.1992.tb01500.x
28. Reitzer L, Zimmern P. 2019. Rapid growth and metabolism of uropathogenic *Escherichia coli* in relation to urine composition. Clin Microbiol Rev 33:e00101-19. https://doi.org/10.1128/CMR.00101-19
29. Hogins J, Fan E, Seyan Z, Kusin S, Christie AL, Zimmern PE, Reitzer L. 2022. Bacterial growth of uropathogenic *Escherichia coli* in pooled urine is much higher than predicted from the average growth in individual urine samples. Microbiol Spectr 10:e0201622. https://doi.org/10.1128/spectrum.02016-22
30. Wang ZJ, Joe BN, Coakley FV, Zaharchuk G, Busse R, Yeh BM. 2008. Urinary oxygen tension measurement in humans using magnetic resonance imaging. Acad Radiol 15:1467–1473. https://doi.org/10.1016/j.acra.2008.04.013
31. Zaharchuk G, Busse RF, Rosenthal G, Manley GT, Glenn OA, Dillon WP. 2006. Noninvasive oxygen partial pressure measurement of human

body fluids *in vivo* using magnetic resonance imaging. Acad Radiol 13:1016–1024. https://doi.org/10.1016/j.acra.2006.04.016
32. Gourse RL, Chen AY, Gopalkrishnan S, Sanchez-Vazquez P, Myers A, Ross W. 2018. Transcriptional responses to ppGpp and DksA. Annu Rev Microbiol 72:163–184. https://doi.org/10.1146/annurev-micro-090817-062444
33. Guy CP, Atkinson J, Gupta MK, Mahdi AA, Gwynn EJ, Rudolph CJ, Moon PB, van Knippenberg IC, Cadman CJ, Dillingham MS, Lloyd RG, McGlynn P. 2009. Rep provides a second motor at the replisome to promote duplication of protein-bound DNA. Mol Cell 36:654–666. https://doi.org/10.1016/j.molcel.2009.11.009
34. Macvanin M, Adhya S. 2012. Architectural organization in *E. coli* nucleoid. Biochim Biophys Acta 1819:830–835. https://doi.org/10.1016/j.bbagrm.2012.02.012
35. Shao Q, Goyal S, Finzi L, Dunlap D. 2012. Physiological levels of salt and polyamines favor writhe and limit twist in DNA. Macromolecules 45:3188–3196. https://doi.org/10.1021/ma300211t
36. Duprey A, Groisman EA. 2020. DNA supercoiling differences in bacteria result from disparate DNA gyrase activation by polyamines. PLoS Genet 16:e1009085. https://doi.org/10.1371/journal.pgen.1009085
37. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguénec C, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Médigue C, Rocha EPC, Denamur E. 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet 5:e1000344. https://doi.org/10.1371/journal.pgen.1000344
38. Schneider BL, Reitzer L. 2012. Pathway and enzyme redundancy in putrescine catabolism in *Escherichia coli*. J Bacteriol 194:4080–4088. https://doi.org/10.1128/JB.05063-11
39. Hibbing ME, Dodson KW, Kalas V, Chen SL, Hultgren SJ. 2020. Adaptation of arginine synthesis among uropathogenic branches of the *Escherichia coli* phylogeny reveals adjustment to the urinary tract habitat. mBio 11:e02318-20. https://doi.org/10.1128/mBio.02318-20
40. Karp PD, Ong WK, Paley S, Billington R, Caspi R, Fulcher C, Kothari A, Krummenacker M, Latendresse M, Midford PE, Subhraveti P, Gama-Castro S, Muñiz-Rascado L, Bonavides-Martinez C, Santos-Zavaleta A, Mackie A, Collado-Vides J, Keseler IM, Paulsen I. 2018. The EcoCyc database. EcoSal Plus 8. https://doi.org/10.1128/ecosalplus.ESP-0006-2018
41. Dbeibo L, van Rensburg JJ, Smith SN, Fortney KR, Gangaiah D, Gao H, Marzoa J, Liu Y, Mobley HLT, Spinola SM. 2018. Evaluation of CpxRA as a therapeutic target for uropathogenic *Escherichia coli* infections. Infect Immun 86:e00798-17. https://doi.org/10.1128/IAI.00798-17
42. Fortney KR, Smith SN, van Rensburg JJ, Brothwell JA, Gardner JJ, Katz BP, Ahsan N, Duerfeldt AS, Mobley HLT, Spinola SM. 2022. CpxA phosphatase inhibitor activates CpxRA and is a potential treatment for uropathogenic *Escherichia coli* in a murine model of infection. Microbiol Spectr 10:e0243021. https://doi.org/10.1128/spectrum.02430-21
43. Nallamotu KC, Bahadur R, Kaul M, Reddy M. 2023. Peptidoglycan remodeling by an L,D-transpeptidase, LdtD during cold shock in *Escherichia coli*. J Bacteriol 205:e0038222. https://doi.org/10.1128/jb.00382-22
44. Hogins J, Sudarshan S, Zimmern P, Reitzer L. 2023. Facile transduction with P1 phage in *Escherichia coli* associated with urinary tract infections. J Microbiol Methods 208:106722. https://doi.org/10.1016/j.mimet.2023.106722
45. Battaglioli EJ, Baisa GA, Weeks AE, Schroll RA, Hryckowian AJ, Welch RA. 2011. Isolation of generalized transducing bacteriophages for uropathogenic strains of *Escherichia coli*. Appl Environ Microbiol 77:6630–6635. https://doi.org/10.1128/AEM.05307-11
46. Alteri CJ, Smith SN, Mobley HLT, Galán JE. 2009. Fitness of *Escherichia coli* during urinary tract infection requires gluconeogenesis and the TCA cycle. PLoS Pathog 5:e1000448. https://doi.org/10.1371/journal.ppat.1000448
47. Zampieri M, Hörl M, Hotz F, Müller NF, Sauer U. 2019. Regulatory mechanisms underlying coordination of amino acid and glucose catabolism in *Escherichia coli*. Nat Commun 10:3354. https://doi.org/10.1038/s41467-019-11331-5
48. Conover MS, Hadjifrangiskou M, Palermo JJ, Hibbing ME, Dodson KW, Hultgren SJ. 2016. Metabolic requirements of *Escherichia coli* in intracellular bacterial communities during urinary tract infection pathogenesis. mBio 7:e00104–e00116. https://doi.org/10.1128/mBio.00104-16
49. Sudarshan S, Hogins J, Ambagaspitiye S, Zimmern P, Reitzer L. 2021. The nutrient and energy pathway requirements for surface motility of nonpathogenic and uropathogenic *Escherichia coli*. J Bacteriol 203:e00467-20. https://doi.org/10.1128/JB.00467-20
50. Garcia EC, Brumbaugh AR, Mobley HLT. 2011. Redundancy and specificity of *Escherichia coli* iron acquisition systems during urinary tract infection. Infect Immun 79:1225–1235. https://doi.org/10.1128/IAI.01222-10
51. Philippe N, Crozat E, Lenski RE, Schneider D. 2007. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. Bioessays 29:846–860. https://doi.org/10.1002/bies.20629
52. Fernández-Coll L, Maciag-Dorszynska M, Tailor K, Vadia S, Levin PA, Szalewska-Palasz A, Cashel M. 2020. The absence of (p)ppGpp renders initiation of *Escherichia coli* chromosomal DNA synthesis independent of growth rates. mBio 11:e03223-19. https://doi.org/10.1128/mBio.03223-19
53. Alteri CJ, Lindner JR, Reiss DJ, Smith SN, Mobley HLT. 2011. The broadly conserved regulator Phop links pathogen virulence and membrane potential in *Escherichia coli*. Mol Microbiol 82:145–163. https://doi.org/10.1111/j.1365-2958.2011.07804.x
54. Groisman EA, Duprey A, Choi J. 2021. How the PhoP/PhoQ system controls virulence and Mg^{2+} homeostasis: lessons in signal transduction, pathogenesis, physiology, and evolution. Microbiol Mol Biol Rev 85:e0017620. https://doi.org/10.1128/MMBR.00176-20
55. Pontes MH, Sevostyanova A, Groisman EA. 2015. When too much ATP is bad for protein synthesis. J Mol Biol 427:2586–2594. https://doi.org/10.1016/j.jmb.2015.06.021
56. Crépin S, Chekabab S-M, Le Bihan G, Bertrand N, Dozois CM, Harel J. 2011. The Pho regulon and the pathogenesis of *Escherichia coli*. Vet Microbiol 153:82–88. https://doi.org/10.1016/j.vetmic.2011.05.043
57. Gardner SG, McCleary WR. 2019. Control of the *phoBR* regulon in *Escherichia coli*. EcoSal Plus 8. https://doi.org/10.1128/ecosalplus.ESP-0006-2019
58. Crépin S, Porcheron G, Houle S, Harel J, Dozois CM. 2017. Altered regulation of the diguanylate cyclase YaiC reduces production of type 1 fimbriae in a Pst mutant of uropathogenic *Escherichia coli* CFT073. J Bacteriol 199:e00168-17. https://doi.org/10.1128/JB.00168-17
59. Raivio TL. 2005. Envelope stress responses and Gram-negative bacterial pathogenesis. Mol Microbiol 56:1119–1128. https://doi.org/10.1111/j.1365-2958.2005.04625.x
60. Testerman TL, Vazquez-Torres A, Xu Y, Jones-Carson J, Libby SJ, Fang FC. 2002. The alternative sigma factor σ^E controls antioxidant defences required for *Salmonella* virulence and stationary-phase survival. Mol Microbiol 43:771–782. https://doi.org/10.1046/j.1365-2958.2002.02787.x
61. Humphreys S, Stevenson A, Bacon A, Weinhardt AB, Roberts M. 1999. The alternative sigma factor, σ^E, is critically important for the virulence of *Salmonella typhimurium*. Infect Immun 67:1560–1568. https://doi.org/10.1128/IAI.67.4.1560-1568.1999
62. Fan Y, Bai J, Xi D, Yang B. 2022. RpoE facilitates stress-resistance, invasion, and pathogenicity of *Escherichia coli* K1. Microorganisms 10:879. https://doi.org/10.3390/microorganisms10050879
63. Cabello FC, Godfrey HP, Bugrysheva JV, Newman SA. 2017. Sleeper cells: the stringent response and persistence in the *Borreliella (Borrelia) burgdorferi* enzootic cycle. Environ Microbiol 19:3846–3862. https://doi.org/10.1111/1462-2920.13897
64. Das B, Bhadra RK. 2020. (p)ppGpp metabolism and antimicrobial resistance in bacterial pathogens. Front Microbiol 11:563944. https://doi.org/10.3389/fmicb.2020.563944
65. Hobbs JK, Boraston AB. 2019. (p)ppGpp and the stringent response: an emerging threat to antibiotic therapy. ACS Infect Dis 5:1505–1517. https://doi.org/10.1021/acsinfecdis.9b00204
66. Irving SE, Choudhury NR, Corrigan RM. 2021. The stringent response and physiological roles of (pp)pGpp in bacteria. Nat Rev Microbiol 19:256–271. https://doi.org/10.1038/s41579-020-00470-y

67. Kundra S, Colomer-Winter C, Lemos JA. 2020. Survival of the fittest: the relationship of (p)ppGpp with bacterial virulence. Front Microbiol 11:601417. https://doi.org/10.3389/fmicb.2020.601417
68. Prusa J, Zhu DX, Stallings CL. 2018. The stringent response and *Mycobacterium tuberculosis* pathogenesis. Pathog Dis 76:fty054. https://doi.org/10.1093/femspd/fty054
69. Frick-Cheng AE, Sintsova A, Smith SN, Krauthammer M, Eaton KA, Mobley HLT, Sperandio V. 2020. The gene expression profile of uropathogenic *Escherichia coli* in women with uncomplicated urinary tract infections is recapitulated in the mouse model. mBio 11:e01412-20. https://doi.org/10.1128/mBio.01412-20
70. Anfora AT, Halladin DK, Haugen BJ, Welch RA. 2008. Uropathogenic *Escherichia coli* CFT073 is adapted to acetatogenic growth but does not require acetate during murine urinary tract infection. Infect Immun 76:5760–5767. https://doi.org/10.1128/IAI.00618-08
71. Lussu M, Camboni T, Piras C, Serra C, Del Carratore F, Griffin J, Atzori L, Manzin A. 2017. ^{1}H NMR spectroscopy-based metabolomics analysis for the diagnosis of symptomatic *E. coli*-associated urinary tract infection (UTI). BMC Microbiol 17:201. https://doi.org/10.1186/s12866-017-1108-1
72. Ochman H, Selander RK. 1984. Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol 157:690–693. https://doi.org/10.1128/jb.157.2.690-693.1984
73. Mobley HL, Green DM, Trifillis AL, Johnson DE, Chippendale GR, Lockatell CV, Jones BD, Warren JW. 1990. Pyelonephritogenic *Escherichia coli* and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. Infect Immun 58:1281–1289. https://doi.org/10.1128/iai.58.5.1281-1289.1990
74. Hogins J, Zimmern PE, Reitzer L. 2023. Genome sequences of seven clade B2 *Escherichia coli* strains associated with recurrent urinary tract infections in postmenopausal women. Microbiol Resour Announc 12:e0003523. https://doi.org/10.1128/mra.00035-23
75. De Nisco NJ, Neugent M, Mull J, Chen L, Kuprasertkul A, de Souza Santos M, Palmer KL, Zimmern P, Orth K. 2019. Direct detection of tissue-resident bacteria and chronic inflammation in the bladder wall of postmenopausal women with recurrent urinary tract infection. J Mol Biol 431:4368–4379. https://doi.org/10.1016/j.jmb.2019.04.008
76. Mulvey MA, Schilling JD, Hultgren SJ. 2001. Establishment of a persistent *Escherichia coli* reservoir during the acute phase of a bladder infection. Infect Immun 69:4572–4579. https://doi.org/10.1128/IAI.69.7.4572-4579.2001
77. Urs K, Zimmern PE, Reitzer L. 2023. Nitrogen-limitation independent control of *glnA* (glutamine synthetase) expression in *Escherichia coli* by urea, several amino acids, and post-transcriptional regulation. Physiology. https://doi.org/10.1101/2023.01.17.524496
78. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421
79. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527. https://doi.org/10.1038/nbt0816-888d
80. McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40:4288–4297. https://doi.org/10.1093/nar/gks042
81. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616
82. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb) 2:100141. https://doi.org/10.1016/j.xinn.2021.100141
83. Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16:284–287. https://doi.org/10.1089/omi.2011.0118