Medical Research Archives Volume 4 Issue 6.

Insights from the Whole Genome Sequence of *Escherichia coli* RS218 in Identifying Potential Virulence Traits Common to *E. coli* Causing Neonatal Meningitis and Other Extra-intestinal Diseases

Authors:
Dona Saumya S. Wijetunge (dwijetunge@hmc.psu.edu)[1],
Robab Katani (rxk104@psu.edu)[1],
Vivek Kapur (vkapur@psu.edu),
Subhashinie Kariyawasam (sxk91@psu.edu)[1, 2*]

Authors Note:
[1]Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, University Park, PA 16802, USA
[2]Center for Molecular Immunology and Infectious Disease, University Park, PA 16802, USA

*__Corresponding Author:__
Subhashinie Kariyawasam
Email: (sxk91@psu.edu)

Abstract

Escherichia coli RS218 is the prototypic strain of neonatal meningitis causing *E. coli* (NMEC) that has been used in many studies relevant to NMEC pathogenesis. In the present study, the whole genome of *E. coli* RS218 was sequenced and compared with the genomes of eight other extraintestinal pathogenic *E. coli* (ExPEC) and the laboratory strain of *E. coli* K-12. Analysis of *E. coli* RS218 genome revealed that it consists of a circular chromosome of 5.087 Mb in size and a 114-Kbp plasmid with an average G+C content of 50.6%. The chromosome contains 4,658 coding sequences, 88 transfer RNAs, 22 ribosomal RNAs, one clustered regularly interspaced short palindromic repeats array, and five noncoding RNAs. *Escherichia coli* RS218 genome demonstrated 98% nucleotide similarity to cystitis causing *E. coli* strain UTI89. Comparative genomic analysis identified a total of 51 genomic islands (GIs) in *E. coli* RS218 which were not present in the *E. coli* K-12 genome. Of these GIs, 16 were common to all NMEC strains studied whereas two GIs were common to all ExPEC. The GIs common to all NMEC encode for several sugar uptake pathways, arginine/ornithine metabolism, iron uptake systems, and putative adhesins and invasins, which may contribute to NMEC pathogenesis. This study also identified NMEC-specific traits that might play a role in initial colonization of mucosal epithelia and penetration of the intestinal and blood brain barriers by NMEC. Overall, these data will facilitate a better understanding of the genetic bases of virulence and adaptation mechanisms of ExPEC.

Keywords: Comparative genomics, *Escherichia coli*, ExPEC, Neonatal meningitis, Whole genome sequencing, Virulence traits.

Insights from the Whole Genome Sequence of *Escherichia coli* RS218 in Identifying Potential Virulence Traits Common to *E. coli* Causing Neonatal Meningitis and Other Extra-intestinal Diseases

1. Introduction

The whole genome sequencing of pathogens has opened up a new era of understanding the genetic basis of bacterial pathogens in terms of their clinical, pathological, epidemiological, diagnostic, and evolutionary perspectives (Feist, 2008; Flint, 2008; Handelsman, 2004; Morozova & Marra, 2008). The first bacterial genome that was sequenced was the genome of the Gram-negative coccobacilllus, *Haemophilus influenzae* in 1995 (Fleischmann et al., 1995). Since then sequencing of various commensal and pathogenic strains of bacteria has enabled the scientists to develop testable hypotheses on genetic events that drive evolution of bacterial virulence and adaptation (Fleischmann et al., 1995).

To date, more than 1000 *E. coli* genome sequencing projects have been reported to GenBank, of which 63 genomes have been completed and available to the public (http://www.ncbi.nlm.nih.gov/genome/167). Genome comparison studies have revealed that the *E. coli* genome is open in the sense that it acquires novel genes from other bacterial species, and consists of a core genome which is conserved across all *E. coli* strains, and an accessory genome which contains specific genetic regions responsible for genotypic and phenotypic diversity among different *E. coli* pathovars (Rasko, 2008; Touchon, 2009). It has also been reported that the *E. coli* pangenome (both core and accessory genomes) consists of a reservoir of more than 13,000 genes (Rasko, 2008). Interestingly, out of these 13,000 genes, only 2,200 belong to the core genome, providing the genetic basis for marked diversity observed among pathogenic *E. coli* (Rasko, 2008). Similarly, comparative genomics of *E. coli* with other closely related enteric bacterial genera revealed a stepwise acquisition of foreign DNA by *E. coli* from *Salmonella*, *Yersinia* and *Vibrio* species (Dobrindt et al., 2002). It has also been proven that genes that are horizontally transferred can be used to define a pathovar of *E. coli* as indicated for intestinal pathogenic (a.k.a. diarrheagenic) *E. coli* (Croxen & Finlay, 2010). Previous studies directed towards *E. coli* pathogenomics have focused mainly on intestinal pathogenic and uropathogenic *E. coli* (UPEC) with only a few studies have included genomes of non-UPEC extra-intestinal pathogenic *E. coli* (ExPEC) (Brzuszkiewicz et al., 2006; Rasko, 2008). The main reason for this is the unavailability of adequate numbers of complete genome sequences of ExPEC pathotypes other than UPEC. For instance, at present, only three genomes of neonatal meningitis-causing *E. coli* (NMEC) and two genomes of avian pathogenic *E. coli* (APEC) have been completely sequenced and are available in the public domain (Johnson et al., 2012; Lu et al., 2011; Mangiamele et al., 2013; Peigne et al., 2009).

Even though NMEC has been considered as one of the major pathogens associated with meningitis during early period of human life, its pathogenesis is not fully defined (Croxen & Finlay, 2010). *Escherichia coli* RS218 (hereafter referred to as RS218) has been isolated from cerebrospinal fluid of a neonatal patient diagnosed with meningitis in the early 1980s (Silver, Aaronson, Sutton, & Schneerson, 1980). This strain is considered as a reference strain of NMEC and many studies on NMEC pathogenesis have used this strain of *E. coli* (Maruvada & Kim, 2012; Teng et al., 2005; Teng et al., 2010; Zhou et al., 2012). Although sequencing of RS218 genome was initialed in 2006 by a team of researchers at University of Wisconsin (Madison, WI) under a project funded by the National Institutes of Health, its complete genome sequence is still not available to the public (http://www.genome.wisc.edu/sequencing/rs218.htm). Therefore, the objective of the present study was to sequence the whole genome of RS218 to identify potential virulence traits of NMEC that might aid in filling the knowledge gaps pertaining to NMEC pathogenesis.

2. Materials and Methods

2.1 *Escherichia coli* RS218 strain

Escherichia coli RS218 strain (O18:H7:K1) was kindly provided by Dr. James Johnson, VA Medical Center, University of Minnesota, MN, USA (Silver et al., 1980). It was isolated from the cerebrospinal fluid of an infant with meningitis in the 1980's (Silver et al., 1980). This strain belongs to the B2 phylogenetic group and sequence type 95 (ST95) (Silver et al., 1980).

2.2 Genomic DNA extraction

Genomic DNA was isolated using Promega Genomic Wizard Kit (Promega Corporation, Madison, WI). Briefly, 1 mL of overnight culture was centrifuged for 5 min at 15,000 x g in a microcentrifuge. The pellet was resuspended in 480 µl of 50 mM ethylenediaminetetraacetic acid (EDTA) and subjected to DNA extraction according to manufacturer's instructions. The genomic DNA pellet was dissolved in sterile double distilled H_2O, quantified using a Nanodrop (Thermo Scientific, Wilmington, DE) and stored at -20°C.

2.3 Whole genome sequencing

Genome sequencing was performed by Ion Torrent PGM Sequencing Technology (Life Technologies, Grand Island, NY) at the Genomics Core Facility of the Pennsylvania State University (University Park, PA) using a 318 chip to provide over 100-fold coverage of the genome. Approximately, 100 ng of high purity genomic DNA was submitted for sequencing.

2.4 Construction of the whole genome optical map (OpMap)

The whole genome restriction optical map was generated using *Nco*I digestion by OpGen, Inc. (Gaithersburg, MD) (21). In brief, high molecular weight DNA was extracted, linearized, immobilized on to a Mapcard containing micro channels which hold single chromosomes and subsequently digested with the *Nco*I restriction enzyme. Resultant genomic DNA fragments were stained with a fluorescence dye, lengths were measured using fluorescent microscopy and assembled by overlapping fragments to generate a map of restriction cut sites. Restriction map was visualized using the MapSolver software version 3.0 (OpGen).

2.5 Genome assembly and gap closure

Ion Torrent whole genome sequencing generated approximately six million reads with an average length of 250 ± 31 bp. These short reads were assembled using both *de novo* and reference guided assembly by SeqMan NGen 11.0 software and visualized using SeqMan Pro 11.0 software (DNASTAR, Madison, WI). *Escherichia coli* UTI89 (NCBI Accession CP000243) was selected as the reference genome for reference guided assembly based on its closest similarity to *Nco*I restriction map of RS218. The genomic contigs obtained by both assemblies were ordered according to the OpMap to acquire correct orientation in the chromosome. Plasmid contigs were identified using Basic Local Alignment Search Tool for Nucleotides (BLASTn) algorithm from National Center for Biotechnology Information (NCBI, http://blast.ncbi.nlm.nih.gov/Blast.cgi) and assembled using the reference plasmid sequence, pUTI89 (NCBI Accession CP000243.1). Ambiguous or low coverage areas and gaps in both the chromosome and the plasmid were closed by primer walking method using Universal Primer Walking Kit (Clontech Laboratories, Mountain View, CA) followed by Sanger sequencing at the Penn State University's Genomics Core Facility.

2.6 Genome annotation

Annotation was performed using automated annotation pipelines, the Rapid Annotation Subsystem Technology (RAST, http://rast.nmpdr.org) and NCBI Prokaryotic Annotation Pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_prok)/ (Aziz et al., 2008). Initial annotations were curated manually using Artemis genome browser and annotation tool (https://www.sanger.ac.uk/resources/software/artemis) and protein-protein blast (BLASTp) algorithm (http://www.ncbi.nlm.nih.gov).

2.7 Identification of repetitive elements and prophage regions

Initially predicted mobile elements through annotation pipeline were blasted using ISFinder web tool (https://www-is.biotoul.fr/) to identify the types of IS within the RS218 genome (Siguier, Perochon, Lestrade, Mahillon, & Chandler, 2006). Phage related sequences were searched using PHAge Search Tool (PHAST, http://phast.wishartlab.com) (Zhou, Liang, Lynch, Dennis, & Wishart, 2011).

2.8 Comparative genomics

The complete sequence of RS218 chromosome was aligned to other ExPEC pathogens namely, IHE3034 (NCBI Accession CP001969), S88 (NCBI Accession CU928161), CE10 (NCBI Accession NC_017646.1), CFT073 (NCBI Accession AE014075), 536 (NCBI Accession CP000247), UTI89 (NCBI Accession CP000243), APECO1 (NCBI Accession CP000468), O78 (NCBI Accession CP004009) and the laboratory strain *E. coli* K-12 MG155 (NCBI Accession U00096) using BLASTn algorithm and Mauve alignment tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Alignments were visualized using Blast Ring Image Generator (BRIG) (http://brig.sourceforge.net/) and Circular Genome Viewer (http://wishart.biology.ualberta.ca/cgview/). Each GI identified in the RS218 genome in comparison to *E. coli* K-12 was blasted against these selected ExPEC genomes with the cutoff values of 90% coverage and 90% identity.

3. Results and Discussion

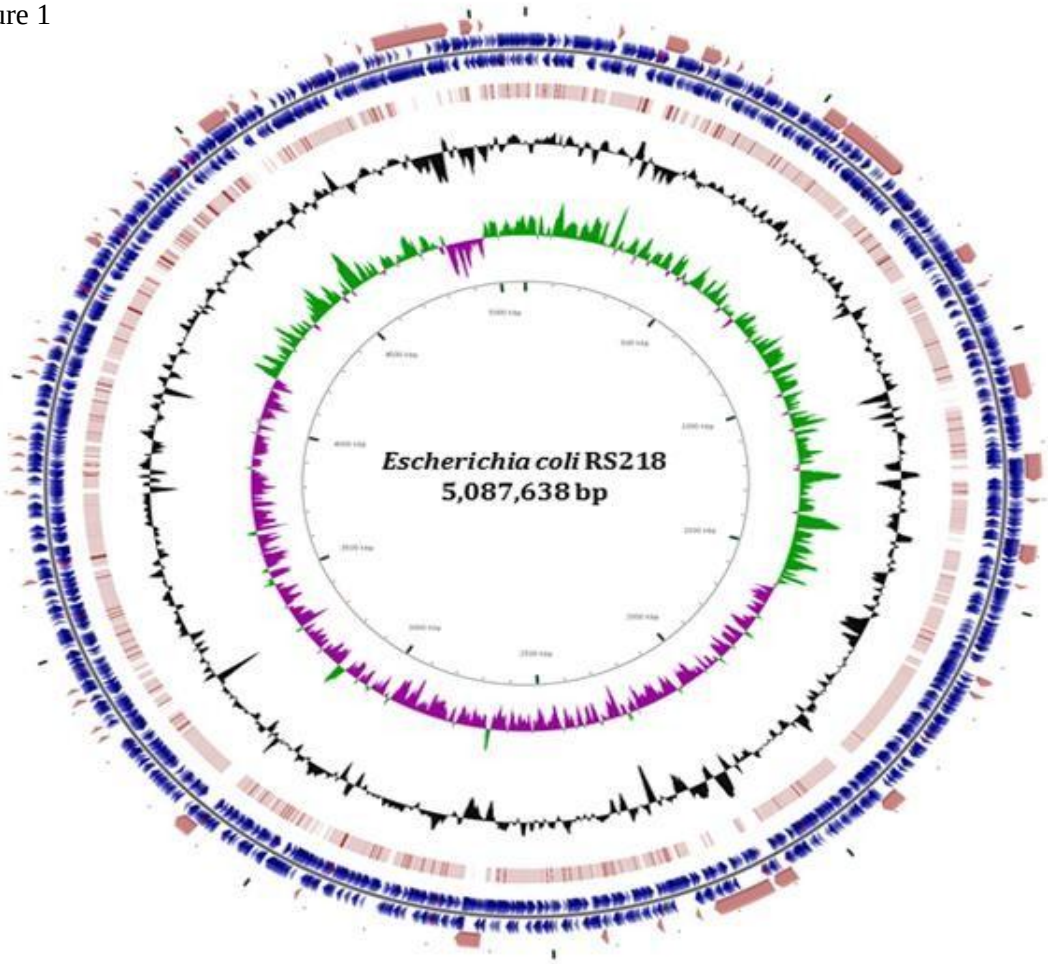
3.1 General characteristics of *E. coli* RS218 genome

The RS218 genome consists of a single circular chromosome (approximately 5.087 Mb in size) and a 114,231-bp plasmid (pRS218). These findings are consistent with the observations of the RS218 genome project initiated by researchers at University of Wisconsin. The RS218 chromosome is considerably larger than the *E. coli* K-12 chromosome (4.6 Mb) as observed for other ExPEC chromosomes, such as UPEC strains CFT073 (5.23 Mb) and UTI89 (5.21 Mb) 536 (4.94 Mb) and APEC O1 (5.5 Mb). The RS218 chromosome contains 4,658 coding sequences (CDS), 88 transfer RNAs (tRNAs), 22 ribosomal RNAs (rRNAs), one clustered regularly interspaced short palindromic repeats (CRISPR) array and five noncoding RNAs. The average GC content of the genome is 50.6% similar to most of the other *E. coli* that have been sequenced (Rasko, 2008; Welch et al., 2002). However, within the genome, a marked variation in the GC content of some coding sequences (ranged from 46% to 54%) was observed suggesting an acquisition of DNA through horizontal gene transfer (Figure 1).

Figure 1. **Circular map of the chromosome of *E. coli* RS218**. The innermost ring represents the *E. coli* K-12 chromosome used as a reference and its coordinates. The second inner ring (in purple/green) plots the GC skew of the reference, followed by the black ring which plots the G+C content. Pink ring indicates the BLASTn comparison between the chromosomes of *E. coli* RS218 and *E. coli* K-12. Blue arrows depict the coding sequences of RS218 in forward and reverse strands (inner and outer rings, respectively). The outermost red arrows indicate the genomic islands (GIs) identified in RS218 which were absent in *E. coli* K-12.

Insights from the Whole Genome Sequence of *Escherichia coli* RS218 in Identifying Potential Virulence Traits Common to *E. coli* Causing Neonatal Meningitis and Other Extra-intestinal Diseases

Figure 1



Escherichia coli RS218 harbors a large conjugative plasmid (pRS218) with a similar GC content (51.02%) to that of the chromosome which has already been sequenced and reported (Wijetunge, 2014; Wijetunge et al., 2014). The complete annotated genome of *E. coli* RS218 can be found in NCBI GenBank (http://www.ncbi.nlm.nih.gov/bioprojec) under the accession number CP007149. General features of RS218 genome and the plasmid are shown in Table 1. A schematic representation of the chromosome is illustrated in Figure 1.

Table 1. General characteristics of *E. coli* RS218 genome.

Chromosome	
Size	5,087,638 bp
GC content	50.06%
No. of genes	5,043
No. of coding sequences	4,658
rRNA (5S,16S, 23S) genes	22
tRNA genes	88
Plasmid	
Size	114,231 bp
GC content	51.06%
No. of genes	160
No. of coding sequences	130

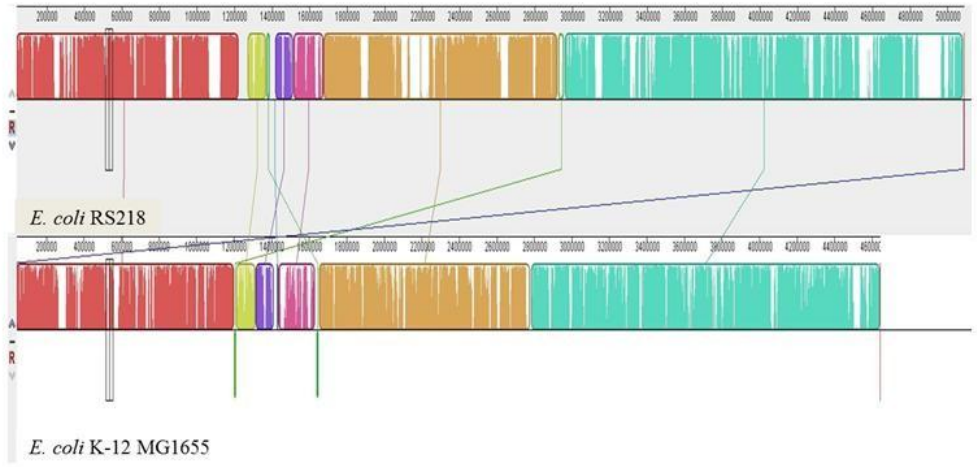
3.2 Genomic islands (GIs) of RS218

The overall sequence architecture of the RS218 genome was compared to the genome of the *E. coli* K-12 MG1655. The *E. coli* K-12 was originally isolated in 1922 from a stool sample taken from a convalescent diphtheria patient in Palo Alto, CA. Due to its noninvasive phenotype in the intestinal epithelium, this strain has been used in many studies as a nonpathogenic reference genome to compare genomes of pathogenic *E. coli* (Bachmann, 2006; Flint,2008; Rasko, 2008).

Alignment of RS218 genome with *E. coli* K-12 showed ten locally collinear blocks revealing an overall conservation of the core genome and the architecture of *E. coli* species in general (Figure 2). However, some of these blocks are inverted, and /or rearranged in terms of orientation compared to each other whereas the strain-specific regions are located within or outside of the collinear blocks. Both of these observations indicate genomic differences between the two strains that might be responsible for pathogenic and nonpathogenic phenotypes of RS218 and *E. coli* K-12, respectively. In a previous study, Xie et al. (2006) identified 22 GIs in RS218 chromosome, some which were shown to be involved in bacteremia and invasion of the blood brain barrier (BBB) (Xie et al., 2006). In particular, RS218 mutant strains lacking RS218-derived islands (RDI) containing the genes for certain adhesins (S fimbriae, antigen 43, type 1-like fimbriae Fim2, P fimbriae, F17-like fimbriae, antigen 43, Hek, and hemagglutinin), protein secretion systems (T5SS for antigen 43, type II/III secretin, and T1SS for hemolysin), iron uptake systems (Iro and Hmu), toxins (alpha-hemolysin), and some other virulence factors (cytotoxic necrotizing factor 1 or CNF-1, PuvA, and K1 capsule biosynthesis) exhibited reduced levels of bacteremia as compared to their wild-type counterparts. Similarly, RS218 mutants lacking the RDIs containing the genes that encode proteins, such as IcmF and IcmH, type 1-like fimbriae

Fim2, type II/III secretin, and IbeA as well as *sia* and *hip*-like operons demonstrated impaired ability to bind to and/or invade human brain microvascular endothelial cells (HBMEC). Although these 21 GIs provided some useful information regarding RS218 pathogenesis, the full genome sequence is necessary for complete understanding of molecular mechanisms involved in NMEC pathogenesis.

Figure 2. Pairwise alignment of *E. coli* RS218 genome to the laboratory strain *E. coli* K-12 using Mauve. Locally collinear blocks (LCBs) of DNA are depicted in the same color and connected via corresponding connection lines. Inverted regions of *E. coli* K-12 genome compared to the orientation of *E. coli* RS218 genome are drawn below the axis. White and grey regions in each genome indicate strain-specific regions.



In the present study, RS218-specific CDS were analyzed based on the orthology explained by cutoff values of >90% of identity and >90% of coverage. This analysis identified 1,249 unique CDS in RS218 which were absent in *E. coli* K-12. For further analysis, these ortholog genes were clustered according to their locations in the RS218 genome and the regions larger than 4 kb were designated as GIs. Accordingly, 51 GIs ranging from 4.07 Kbp to 123.08 Kbp were identified in the RS218 chromosome, which are flanked by tRNA sites, phage-related sequences or mobile genetic elements (S1 Table). These islands encoded potential virulence proteins, such as fimbrial adhesins, invasins, cytotoxins, and iron acquisitions systems (S1 Table). Some of these virulence factors (capsular polysaccharides, invasin of brain endothelium or Ibe, S fimbriae, SitABCD, etc.) have previously been described as essential virulence factors of NMEC, in particular of those possessing a K1 capsule (Xie et al., 2006). Also, some other proteins, such as Vat, salmochelin, and invasins of Peyer's patches have been recognized as virulence factors of other pathogenic *E. coli* but their involvement in NMEC pathogenesis has yet to be elucidated (Maruvada & Kim, 2012; Robbins et al., 1974; Teng et al., 2010; Zhou et al., 2012).

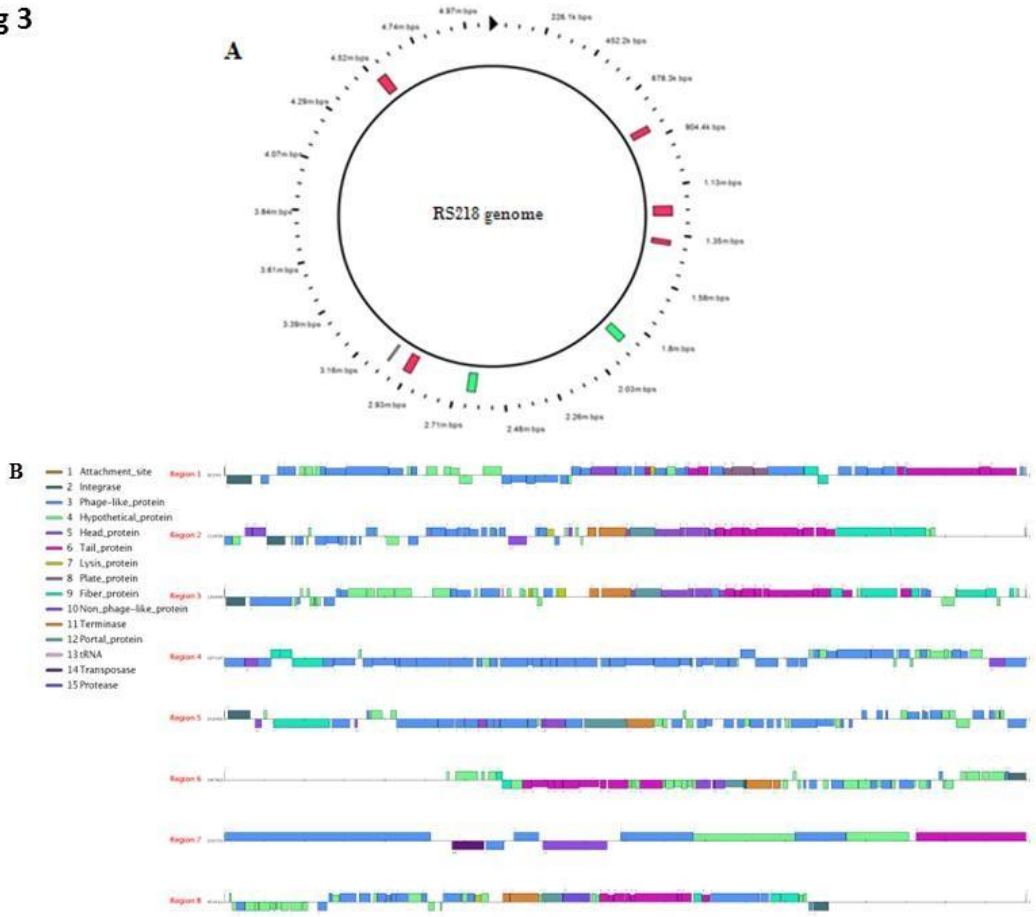
3.3 Bacteriophages in RS218

Phages in bacterial genomes provide information regarding pathogen evolution where some phage-associated genes are regarded as essential elements of pathogenesis of certain bacterial diseases. For example, major virulence determinants of enterohemorrhagic *E. coli*, the Shiga toxins and type III secretory apparatus are thought to be acquired by *E. coli* from *Shigella* via phages by means of horizontal gene transfer (Jarvis, 1996). Therefore, RS218 genome was analyzed for the presence of phages and phage-associated genes using PHAST (Zhou et al., 2011).

With this analysis, a total of eight phage regions were identified in the RS218 genome. These included five intact phages which encode all essential proteins for a complete functional phage as well as incomplete or partial phage areas that lack some of the essential elements of phages. The relative locations and the details of each phage region are depicted in Figure 3. Interestingly, three of these phage regions indicated horizontally-acquired virulence genes: *sit*ABCD from *Salmonella*, HPI from *Yersinia* and the capsular polysaccharide operon (S1 Table). Some of the proteins associated with these regions, such as capsular polysaccharides have already been implicated in the pathogenesis of NMEC (Robbins et al., 1974). It has also been demonstrated that the O-acetyltransferase gene *neuO* which mediates acetylation of polysialic acid O of the capsule is located on the *E. coli* K1-specific phage island CUS-3 (King et al., 2007). This phage acts as an active mobile contingency locus and is responsible for population-wide capsule variation seen among *E. coli* K1 (King et al., 2007). A mutant strain of RS218 strain lacking the CUS-3 prophage island was impaired in its ability to bind to HBMEC (Xie et al., 2006).

Figure 3. Characteristics of the bacteriophages in *E. coli* RS218 genome. A) Schematic diagram of relative positions of phages inserted in the *E. coli* RS218 chromosome. Red, intact phage; Green, phages lacking few essential genes; Grey, incomplete phage. B) Open reading frames present in each phage region with relevance to phage-associated functions.

Fig 3



3.4 Virulence and fitness-associated genes in RS218

3.4.1 Iron acquisition systems

Iron is a vital element for survival and multiplication of bacteria because of its role in respiration, DNA replication, and protection against oxidative stress (Miethke, 2007; Schaible, 2004; Skaar, 2010). Since iron is tightly bound to proteins, such as hemoglobin, transferrin, lactoferrin, and ferritin, the scarcity of free iron is one of the major obstacles encountered by bacterial pathogens inside the mammalian host (Miethke, 2007). Therefore, acquisition of iron that is already bound to iron-binding proteins is a prerequisite for pathogens particularly at extra-intestinal locations (Gao et al., 2012). In ExPEC, multiple iron acquisition systems encoded by the chromosome and/or plasmids have been identified (Mangiamele et al., 2013; Skaar, 2010). These uptake systems can be divided into three main categories: siderophore-based systems, heme acquisition systems, and transferrin/lactoferrin receptors (Gao et al., 2012). Recent studies have shown that harboring multiple siderophore systems are more common in ExPEC than in commensal

strains of *E. coli* (Garénaux et al., 2011; Johnson et al., 2005; Watts et al., 2012). The chromosome of RS218 also encodes six iron acquisition systems (salmochelin, enterobactin, yersiniabactins, FhuBCD, SitABCD, and EfeU) and several hemin, ferritin, and ferrochrome receptors (eg. ChuA, PiuC, CorA, etc.). Interestingly, most of these proteins are encoded by genes located on GIs and are discussed in detail below (S1 Table).

Salmochelins are glucosylated derivatives of enterobactin which were initially identified in *Salmonella* sp. (Müller, 2009). Salmochelin operon encodes five (IroBCDEN) proteins: IroB, which glycosylates a portion of enterobactin to produce salmochelin that enhances hydrophobicity to siderophore and the inability to be bound by the host proteins; IroC, which is an ABC transporter that exports salmochelin; IroN, which is a receptor for iron-bound salmochelin that transports iron back to the periplasm; IroE, which is a periplasmic esterase that degrades salmochelin to the linear trimeric form; and IroD, which degrades salmochelin (Müller, 2009). Besides its role in iron acquisition, IroN, the siderophore receptor of salmochelin operon has also been shown to facilitate internalization of bacteria into the uroepithelial cells *in vitro* (Feldmann, Sorsa, Hildinger, & Schubert, 2007). Recent studies have shown that salmochelin is prevalent in all subpathotypes of ExPEC and it is recognized as a urovirulence factor in UPEC (Feldmann et al., 2007;Gao et al., 2012; Müller, 2009). A high prevalence of salmochelin (64%) has been perceived in NMEC as well (Nègre et al., 2004). In consistent with these observations, *E. coli* RS218 carries a complete *iroBCDEN* operon on GI 9 together with the S fimbrial/Antigen 43 cluster. Nègre et al. (2004) demonstrated that *iroN* plays a more significant role in the virulence of the NMEC strain C5 (O18:K1:H7) compared to two other iron acquisition systems, yersiniabactin and ChuA, and was associated with high levels of bacteremia, which is a prerequisite for meningitis (Nègre et al., 2004). Also, the involvement of salmochelin encoded by the virulence plasmid of NMEC strain S88 (O45:K1:H7) in bacteremia has been demonstrated by Peigne et al. (2009). Taken together, high prevalence of salmochelin in NMEC strains and previous observations relating to its key role in high-level bacteremia indicate that salmochelin is not only a urovirulent trait but also might be a meningitic trait.

Yersiniabactins are encoded by the high pathogenicity island (HPI) of highly pathogenic *Yersinia sp.* (Carniel, 1996). The same HPI was found frequently in *E. coli* belonging to the phylogenetic group B2, which mostly contains pathogenic *E. coli* (Bach, 2000; Schubert, 2002). The complete and partial operons of yersiniabactin are located in the GIs 17 and 18 of RS218, which are flanked by $tRNA^{Asn}$-GTT, which is similar to the HPI in pathogenic *Yersinia* sp. (S1Table). In a mouse model of ExPEC infection, the HPI had a higher correlation with ExPEC virulence than other virulence traits, such as *pap*, *sfa/foc*, *aer*, *hly*, K1 antigen, *afa*, and *ibeA*, providing a strong evidence for its contribution to ExPEC pathogenesis (Schubert, 2002). In contrast, a study by Xie et al. (2006) suggested the HPI is not involved in *E. coli* meningitis pathogenesis (Xie et al., 2006).

Additionally, the RS218 chromosome contains the *Salmonella* Pathogenicity Island 1-associated iron acquisition operon *sitABCD* in the GI 10 (S1 Table). On this GI, the *sitABCD* operon is located just downstream to a lambda family intact phage indicating phage-associated introduction of foreign DNA into the RS218 genome. The functional significance of *sitABCD* operon has been well studied in avian pathogenic *E. coli* (APEC) (Sabri et al., 2008; Sabri et al., 2006) . Unlike RS218, *sitABCD* operon of APEC is located on the virulence plasmid, pAPEC-1 (Sabri et al., 2006). Functional analysis using isotope uptake experiments has shown that *sitABCD* in APEC strain χ7122 mediates transport of iron and manganese (Sabri et al., 2006). Both of these elements are essential for bacterial growth and to resist the detrimental effects of hydrogen peroxide which is a major bactericidal molecule present in the phagocytes (Sabri et al., 2006). It has been demonstrated that SitABCD is involved in the resistance to hydrogen peroxide in *Salmonella*, *Shigella*, and APEC (Kehres, Janakiraman, Slauch, & Maguire, 2002; Runyen-Janecky, Dazenski, Hawkins, & Warner, 2006). However, no such experiments have been carried out to identify the functional correlation of SitABCD in human ExPEC despite the high prevalence of *sit*A in NMEC and UPEC.

While the iron acquisition systems discussed above are present in RS218 but not in *E. coli* K-12, some other iron acquisition systems are common to both RS218 and *E. coli* K-12. These common systems may have a role in providing fitness to the bacterium for adaptation to its microenvironment. For example, siderophore enterobactin is a triscatechol derivative which enables bacteria to recover iron from the environment (Raymond, Dertz, & Kim, 2003). Enterobactin is synthesized under iron-limiting conditions and excreted into the environment where it binds Fe (III) with a high affinity and specificity (Raymond et al., 2003). The ferric siderophore complexes are then taken up into the cells by a specific ATP-binding cassette transporter complex (Chen, Wu, Bryan, & Dunaway-Mariano, 2009). This siderophore system is common to many Gram-negative bacteria including both pathogenic and commensal *E. coli* and was recently detected in *Streptomyces* species (Gram-positive bacteria) as well (Raymond et al., 2003). The RS218 genome also encodes for FhuB, C and D, which are required for the uptake of iron from aerobactin and other siderophores. However, RS218 does not encode genes of the aerobactin synthesis system, such as IucA, B, C and D or IutA indicating that FhuCDB system is involved in iron acquisition in place of aerobactin. In *E. coli*, it has been demonstrated that FhuCDB transporter accepts at least four different siderophores except aerobactin and it also provides binding sites for colicin and phages (Köster, 1991). However, more studies are required to implicate the relevance of FhuCDB in NMEC pathogenesis or overall ExPEC virulence. Moreover, RS218 genome contains genes that encode EfeU iron acquisition proteins which are specifically induced under low pH and anaerobic or microaerophilic environmental conditions (Große et al., 2006). It is believed that neonates acquire pathogenic *E. coli* during the delivery process from the vaginal microflora of the mother (Kim, 2003). It is also known that the human vagina is acidic

due to the presence of lactobacilli. At low pH, iron exists as the ferrous ion (Fe^{2+}), and therefore, the ferric iron (Fe^{3+}) availability is very low. Hence, harboring the EfeU iron acquisition system which binds to Fe^{2+} and is induced by acidic environmental conditions might be an important strategy exploited by RS218 in vaginal colonization to enhance its fitness in this particular niche. Other than these well studied iron acquisition receptors, RS218 also harbors several putative hemin receptors, iron-binding proteins and ferritin-like proteins that might be involved in enhancing the fitness of RS218 inside the eukaryotic host.

3.4.2 Adhesins

Bacterial adherence to host cells via adhesins represents a crucial step during the establishment of an infection (Klemm & Schembri, 2000). Most of these structures found in pathogenic bacteria have a dual role as invasins and adhesins while some act only as either adhesins or invasins (Klemm & Schembri, 2000). There are two major types of adhesins present in pathogenic bacteria, fimbrial and nonfimbrial adhesins (Klemm & Schembri, 2000). Fimbrial adhesins provide a receptor-mediated contact with host tissues which facilitates microbial colonization at mucosal surfaces, biofilm formation as well as initiation of internalization of some pathogenic bacteria by host cells (Klemm & Schembri, 2000). Regulatory cross-talk between fimbrial operons has been described for different fimbrial determinants of pathogenic *E. coli* (Hacker, 1992). In the RS218 genome, we have identified eight fimbrial clusters, namely, mannose-specific type I fimbriae, sialic acid-specific fimbriae (S fimbriae), pyelonephritis-associated fimbriae (P fimbriae), meningitis-associated and temperature-regulated fimbriae (Mat fimbriae), curli fimbriae, Yad fimbriae, type IV fimbriae and YgiL fimbriae. Except for type I fimbriae, curli and type IV fimbriae, the other fimbrial gene clusters of RS218 genome are located within the GIs (S1 Table). Interestingly, all these fimbriae have been implicated in pathogenesis of various infections caused by *E. coli* (Kulkarni et al., 2009; Lehti et al., 2010; Spurbeck et al., 2011; Stins et al., 1994; Teng et al., 2005).

In neonatal meningitis, type I fimbriae have been implicated in binding to and invasion of human brain microvascular endothelial cells (HBMEC) (Kim, 2003; Teng et al., 2005). The tip adhesin of type I fimbriae, the FimH, binds with CD48 which increases the Rho-dependent changes in the cytoskeleton that facilitates internalization of bacteria by HBMEC (Kim, 2003). Similarly, S fimbriae have also been implicated in NMEC virulence (Stins et al., 1994). Saukkonen et al. (1988) observed that *E. coli* clones harboring S fimbriae were more virulent in a rat pup model of neonatal meningitis than those containing type I fimbrial genes (Saukkonen et al., 1988). Later, Yang et al. (2004) reported that S fimbriae did not play a major role in *E. coli* K1 binding to HBMEC *in vitro* and crossing the BBB *in vivo* suggesting an alternative role for S fimbriae in NMEC virulence other than acting as an adhesin (Wang, Wen, & Kim, 2004).

Three other fimbrial clusters present in RS218; Yad, Pap, and YgiL, all of which have previously been identified as urovirulence factors (Spurbeck et al., 2011). For example, Yad fimbriae adhere bacteria to bladder epithelial cells, whereas YgiL and Pap are involved in bacterial adherence to renal tubular cells (Spurbeck et al., 2011).

Possession of these adhesins implies that RS218 possesses the ability to adhere to uroepithelial cells, which might act as a fitness factor for the survival of bacteria in the urogenital tract (Spurbeck et al., 2011). In contrast, Mat and curli fimbriae are specifically involved in the adherence of bacteria to the intestinal epithelium and formation of biofilms which might be another strategy used by RS218 for environmental adaptation as seen in many other pathogenic bacteria. Thus, these adhesins might be involved in RS218 colonization of maternal intestinal and uroepithelial cells as well as initial colonization of neonates.

In addition to fimbrial adhesins, RS218 also possesses several non-fimbrial adhesins. Unlike fimbrial adhesins, these are embedded in the bacterial cell wall and often are associated with self-recognition (autoaggregation) and biofilm formation (Antao, Wieler, & Ewers, 2009). *Escherichia coli* RS218 encodes two well characterized non-fimbrial adhesins, adhesin involved in diffuse adherence (AIDA-I) and antigen 43 (Ag43), and several other putative adhesins. Both AIDA-I and Ag43 are autotransporter (AT) proteins which are transported to target sites by type V secretion systems (Hasman, Chakraborty, & Klemm, 1999; Laarmann & Schmidt, 2003). The AT protein, Ag43, has been identified in many pathogenic and nonpathogenic bacteria and is involved in biofilm formation and adherence to and internalization of *E. coli* to intestinal epithelial cells (Hasman et al., 1999). Several variant phenotypes of Ag43 and their functional differences have also been observed in *E. coli* (Klemm et al., 2004; Klemm & Schembri, 2000). For example, Ag43 of *E. coli* K-12 exhibited an autoaggregatory phenotype whereas Ag43 of UPEC CFT073 and RS218 lacked this functionality but maintained its ability to form biofilms (Klemm et al., 2004). The AT protein, AIDA-I was first identified in *E. coli* strain 2787 which was isolated from a patient having infantile diarrhea (Benz & Schmidt, 1992). Subsequently, it was identified in many swine enterotoxigenic *E. coli* (ETEC) strains causing edema and post-weaning diarrhea and recognized as a virulence factor that facilitates adherence of ETEC to the intestinal epithelium (Ravi, 2007). Interestingly, in the RS218 genome, the gene encoding AIDA-I is located as a single insertion just adjacent to $tRNA^{Pro}$ gene without associated mobile elements or phage sequences. This gene is not present in the *E. coli* K-12 genome. However, the hepatosylation gene (*aah*) which is essential for the formation of functional AIDA-I adhesin was not observed in the RS218 genome suggesting that AIDA-I encoded by RS218 genome is not functional and has no significant role in RS218 adherence to epithelia (Inga Benz & Schmidt, 2001). Nevertheless, AIDA-like adhesins have been linked to APEC and UPEC pathogenesis (Allsopp et al., 2012). For instance, AIDA-I like adhesin, UpaH in uropathogenic *E. coli* is involved in biofilm formation and colonization of the bladder epithelial cells (Allsopp et al., 2012). In addition to AIDA-I, the RS218 genome also encodes for two AIDA-like adhesins, which may play a role in RS218 virulence.

The GI 1 of RS218 contains a gene that encodes another adhesin/invasin-like protein belonging to the AT family. The BlastP program identified a conserved domain at the C terminus similar to adhesins/invasins such as YadA in

pathogenic *Yersinia*, Saa in LEE-negative Shiga toxin-producing *E. coli* (STEC), and *E. coli* immunoglobulin-binding (Eib) adhesin of STEC (Kirjavainen et al., 2008; Leo & Goldman, 2009; Paton et al., 2001). Regardless of the structural differences at the N termini of these AT proteins, they all have been implicated in enterocyte attaching and effacing phenotype as demonstrated on HeLa cells (Leo & Goldman, 2009). To date, no virulence trait of NMEC has been linked to initial colonization of NMEC in the infant intestinal epithelium and these AT proteins seem to be promising targets to look for such mechanisms.

3.4.3 Invasins

Invasins are surface proteins of bacteria which promote bacterial internalization into host cells. In neonatal meningitis, there are two protective epithelial barriers encountered by NMEC in the process of bacterial entry into the meninges: mucosal epithelial cells and HBMEC or BBB. In NMEC, quite a few invasins have been identified with regard to penetration of BBB. These include Ibe proteins, outer membrane protein A (OmpA), arylsulfatase (AslA), conjugal transfer protein J (TraJ) encoded by plasmid genes, and new lipoprotein 1 (NlpI) (Hill, 2004; Hoffman, 2000; Maruvada & Kim, 2012; Teng et al., 2010). The functions of all of these invasins have been studied in depth to understand their relevance to NMEC pathogenesis (Badger, Wass, & Kim, 2000; Badger, Wass, Weissman, & Kim, 2000; Bloch, Huang, Rode, & Kim, 1996; Hill, 2004; Hoffman, 2000; Maruvada & Kim, 2012; Teng et al., 2010). However, virulence traits involved in the penetration of maternal intestinal mucosal barrier have yet to be identified. In RS218 genome, several putative adhesins, particularly related to intestinal epithelial adherence and invasion, were detected. For example, YadA has been shown to have both adhesive and invasive properties. Other putative invasins encoded by RS218 at locus numbers W817_11190 (coordinates 1329217 to 1330611) and W817_07205 (coordinates 1329217 to 1330611), and invasin-like protein encoded by GI 17 with yersiniabactin cluster contain bacterial Ig-like domains, which have been implicated in bacterial adherence to extracellular matrix and internalization into enterocytes (Bodelón, Palomino, & Fernández, 2013) . These bacterial Ig-like domains are found in many invasins and adhesins, such as intimin of attaching and effacing *E. coli*, invasins of *Yersinia* species which mediate internalization of bacteria into M cells, and adhesins of *Yesrsinia* and *Salmonella* which are involved in bacterial colonization of Peyer's patches (Bodelón et al., 2013). However, the exact mechanistic roles of these putative invasins of RS218 are yet to be elucidated.

3.4.4 Toxins

Pathogenic *E. coli* produce several exotoxins that are imperative for their pathogenesis. The RS218 genome also encodes several toxins or putative toxins that have already been demonstrated in other NMEC and/or other pathogenic *E. coli*. Among the toxins that are well characterized, the CNF1 belongs to the RTX family toxins that have been implicated in the penetration of BBB by *E. coli* K1(Khan et al., 2002). In RS218, this toxin is encoded by a gene operon located on GI 49. This gene operon is flanked by an IS110 family transposase and contains a complete set of genes including those that encode the

proteins required for RTX toxin transport and activation. Apart from CNF1, RS218 also contains an alpha hemolysin (HlyA) and its activator HlyC and the transporter HlyD. The RS218 genome also encodes another hemolysin, HlyE which is a novel pore-forming toxin found in *E. coli*, *Salmonella* Typhi, and *Shigella flexneri* (Hunt et al., 2010; Wyborn et al., 2004). It has been shown that these HlyE family hemolysins are structurally different from HlyA and do not require activation or RTX transporters; instead, they are secreted in the active form and the expression is induced under anaerobic conditions such as in the intestinal environment (Hunt et al., 2010). The HlyE has also been demonstrated in APEC, which does not contain a classical RTX operon, and in some enterohemorrhagic *E. coli* (Wyborn et al., 2004).

Another cytotoxin present in RS218 is vacuolating cytotoxin (Vat) which is located on GI 3. It belongs to the serine protease ATs of *Enterobacteriaceae* (SPATE) family which includes several virulence-associated proteins of bacteria, such as *E. coli* (e. g. Sat and PicU), *Shigella* (PicU), and *Neisseria* and *Haemophilus* (e. g. IgA1 proteases and Hap) (Parreira & Gyles, 2003). It is known that *vat* is more prevalent in ExPEC than in commensal *E. coli* indicating its pathogenic involvement in extra-intestinal infections (Parreira & Gyles, 2003). Although many studies have investigated the prevalence of *vat* in ExPEC, the mechanistic role of Vat in relation to pathogenesis has only been demonstrated in APEC (Oh et al., 2012; Parreira & Gyles, 2003). In the APEC strain Ec222, Vat was responsible for vacuolating cytotoxicity observed in chicken embryo fibroblast cells, and severe septicemia and increased mortality observed in a chicken respiratory infection model of colibacillosis (Parreira & Gyles, 2003). Interestingly, *vat* was more prevalent in NMEC than in commensal fecal *E. coli*, suggesting Vat can be an important target to unravel the pathogenic mechanisms of NMEC (Logue et al., 2012) . The RS218 genome also encodes two bacteriocins, namely, colicin E (ColE) and ColV, and their immunity proteins (Gérard et al., 2005; Jakes & Cramer, 2012). These toxic proteins act as antibacterial peptides that kill or inhibit the growth of colicin-negative bacteria while maintaining immunity for colicinogenic bacteria (Jakes & Cramer, 2012).

Apart from these cytotoxins, RS218 contains several toxin-antitoxin operons (TA) which are not present in *E. coli* K-12. These TA systems are proteins that are encoded by the chromosome and /or plasmids and are involved in enhancing the fitness of bacteria (Yamaguchi, Park, & Inouye, 2011). For example, plasmid-encoded TA systems endorse the existence of plasmid and plasmid-encoded fitness by selectively killing plasmid-free daughter cells (Yamaguchi et al., 2011). The exact mechanism of chromosome-encoded TA systems was not fully understood until Norton et al., (2012) described their role in pathoadaptaion of *E. coli* (Norton & Mulvey, 2012). Since a single bacterium could encode several TA systems (RS218 encodes for 7 TA systems), these investigators observed that in the UPEC strain CFT073, each of these TA systems is niche-specific and behaves in a manner that some TA systems are required for intestinal colonization while others are required for colonization of bladder epithelial and renal

tubular cells (Norton & Mulvey, 2012). These observations indicated that PasTI TA system is involved in the formation of persister cells, which makes ExPEC resistant to antibiotics, to nutrient deprivation, and to reactive oxygen species (Norton & Mulvey, 2012). The chromosome of RS218 also carries all these TA systems except PasTI. Therefore, the role of each of these TA systems of NMEC in different host tissues will be an exciting area for future studies to address current pitfalls of NMEC pathogenesis.

3.5 Secretion Systems

Protein secretion plays a pivotal role in bacterial communication, environmental adaptation, and survival (Kostakioti et al., 2005). There are many specialized secretion systems identified in Gram-negative bacteria (Kostakioti et al., 2005). The RS218 genome encodes all these secretion systems, except a type III secretion systems (T3SS). It also carries genes for SecYEG and the twin arginine pathway that are functional components of common secretion systems as well as some of the specialized secretion systems (Kostakioti et al., 2005). Although T3SS was absent in NMEC belonging to B2 phylogroup (IHE3034 and S88), it was present in the group D NMEC strain CE10 (Lu et al., 2011).

Type I secretion systems (T1SS) or ABC transporters are important secretory components of pathogenic *E. coli* which produce toxins, such as hemolysins and colicins (Delepelaire, 2004). As discussed earlier, RS218 genome encodes hemolysins as well as ColV and E which might be some of the proteins transported by T1SS. Type II secretion systems (T2SS) also have been identified in pathogenic *E. coli* which transport A-B toxins, such as heat labile enterotoxin of ETEC (Nivaskumar & Francetic). However, the proteins secreted through T2SS in RS218 have yet to be identified. Type IV secretion systems (T4SS) which share a structural similarity with T2SS are involved in the transportation of large proteins and DNA molecules across the cell membranes of Gram-negative bacteria (Kostakioti et al., 2005; Nivaskumar & Francetic). It is well known that *E. coli* conjugative pili use a T4SS (Kostakioti et al., 2005; Nivaskumar & Francetic). The RS218 genome contains a complete operon for a T4SS and a large conjugative plasmid indicating the functional relevancy of a T4SS. It has been observed that both T2SS and T4SS are involved in the virulence of UPEC, particularly for the bacterial persistence in the urinary tract and renal tissues (Kulkarni et al., 2009). Therefore, studies identifying the involvement of these secretion systems in the penetration of BBB by NMEC and their persistence in brain tissues might open up new opportunities to understand NMEC virulence. Type V secretion systems (T5SS) or autotransporters are Sec-dependent apparatus involved in toxin secretion in pathogenic *E. coli* (Gawarzewski et al., 2013). As aforementioned, RS218 encodes many AT proteins which might be secreted by the T5SS. Additionally, a virulence-associated type VI secretion system is also present in the RS218 genome. Zhou et al., (2012) demonstrated that RS218 genome contains T6SS in a GI which is flanked by $tRNA^{Asp}$ and its effector molecules, hemolysin-coregulated protein (Hcp) and valine glycine repeat (VgrG) located on the same island (Zhou et al., 2012). Also, they noticed that Hcp proteins facilitate invasion of BBB by NMEC (Zhou et al., 2012). Our

study identified four copies of *hcp* in the RS218 genome, of which two are located on the T6SS pathogenicity island (PAI) and the other two on a different PAI. Similarly, RS218 harbors two copies of *vgrG* clusters with and without a T6SS. Some studies have shown that Hcp and VgrG present in the same gene cluster along with T6SS are structural components rather than the effectors, and real effector molecules of T6SS are clustered elsewhere on the genome (Gawarzewski et al., 2013). Further studies are essential to clarify if this observation is valid for NMEC as well. The RS218 also encodes for the extracellular nucleation-precipitation (ENP) pathway or type VIII secretion system (T8SS). This secretion system is responsible for the secretion and assembly of curli fimbriae which are involved in biofilm formation, host cell adhesion, invasion and activation of the host immune system (Hammer et al., 2007). Other than these protein secretion systems, RS218 contains genes that encode ion transporters, such as cation transporters, tripartite ATP-independent periplasmic (TRAP) transporters, and antiporters.

3.6 Comparative genomics

Comparison of the complete genome of *E. coli* RS218 with other sequenced ExPEC genomes deposited in the public databases (Table 2 and Figure 4) revealed that 79% of RS218 genome was similar to the genome of *E. coli* laboratory strain K-12 MG1655, with a 98% identity at the nucleotide level. Among ExPEC genomes, the UPEC strain UTI89 isolated from the urine of a patient with acute cystitis demonstrated the highest similarity to the genome of RS218 (Table 2). Among NMEC strains, the genome of IHE3034 strain, which was isolated from a neonate diagnosed with meningitis in Finland in 1976 showed the greatest similarity to RS218. Despite the geographical disparity in terms of the origin of *E. coli* RS218 and IHE3034 strains, they both belong to the same serogroup (O18), phylogropup (B2) and the sequence type (ST95) suggesting a close genetic relatedness between the two strains (Moriel et al., 2010).

Figure 4. Pangenome analysis of ExPEC genomes with *E. coli* RS218 genome. The innermost ring represents the *E. coli* RS218 chromosome used as a reference and its coordinates. The black plot indicates the G+C content followed by the plot (in purple/green) indicating GC skew of the reference and the other rings indicate the BLASTn comparison between the pangenomes of ExPEC and *E. coli* K12. The NMEC pangenome included whole genome sequences of *E. coli* strains, IHE3034, ECS88 and CE10; UPEC pangenome included *E. coli* strains, CFT073, UPEC536 and UTI89; APEC pangenome included *E. coli* strains, APECO1 and APECO78.

Insights from the Whole Genome Sequence of *Escherichia coli* RS218 in Identifying Potential Virulence Traits Common to *E. coli* Causing Neonatal Meningitis and Other Extra-intestinal Diseases

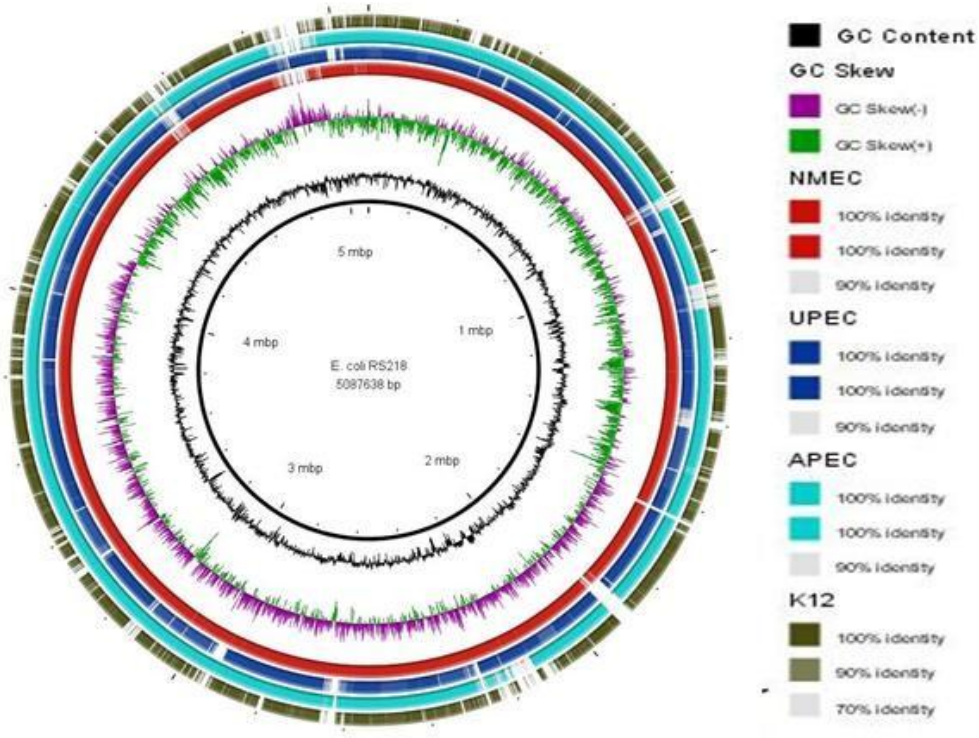


Table 2. Comparison of general characteristics of *E. coli* RS218 genome with other sequenced ExPEC genomes used in this study.

E.coli strain	Pathotype	Phylogroup	Serotype	ST type	Size (Mb)	GC%	# of Genes	# of Proteins	Plasmids	Coverage %
RS218	NMEC	B2	O18:K1:H7	95	5.08	50.06	5,043	4658	1	-
IHE3034	NMEC	B2	O18:K1:H7	95	5.11	50.7	4,966	4,753	1	96
S88	NMEC	B2	O45:K1:H7	95	5.17	50.7	5,187	4,823	1	93
CE10	NMEC	D	O7:K1:H7	62	5.38	50.6	5,269	5,080	4	86
CFT073	UPEC	B2	O6:K2:H1	73	5.23	50.5	5,574	5,364	0	92
UPEC536	UPEC	B2	O6:K15:H31	92	4.94	50.5	4,779	4,619	0	91
UTI89	UPEC	B2	O18:K1:H7	95	5.18	50.6	5,272	5,162	0	98
APECO1	APEC	B2	O1:K1:H7	95	5.50	50.3	4,968	4,853	2	93
APECO78	APEC	B1	O78:K80	23	4.80	50.7	4,695	4,588	0	80
K-12	Lab	A	OR:H48:K-	08	4.64	50.8	4,288	4,141	-	79

We also analyzed the distribution of RS218 GIs among the sequenced ExPEC strains to identify pathotype-specific GIs. Of the 51 GIs possessed by RS218, 48 islands were present at least in one other ExPEC. All three GIs (7, 16, and 44) which were unique to RS218 were bacteriophage regions (S1 Table). Two GIs (29 and 38) were shared by all ExPEC revealing that these islands might play an important role in ExPEC virulence regardless of their specific pathotype. Interestingly, one of these islands, the GI 38 encodes for core oligosaccharide (core OS) of bacterial lipopolysaccharide (LPS) which is crucial for barrier function of the outer membrane and is regarded as a virulence factor (Amor, 2000). Five distinct core OS structures namely, K-12 and R1–R4 have been identified in *E. coli* (Amor, 2000). The genes present in the GI 38 resembles the R1 core OS type suggesting that all ExPEC may belong to this core OS type (Whitfield et al., 1999). The R1 core OS has also been the most frequently found core OS type in pathogenic *E. coli* (Amor, 2000). Another GI of RS218, the GI 29 encodes for putative polysaccharide biosynthesis proteins containing tetratricopeptide repeats (TPRs). These types of proteins which include new lipoprotein I (Nlp I) also have been found in pathogenic bacteria. Such genes might be crucial for *E. coli* to cause extraintestinal disease or to survive in the extraintestinal environments since these genes particularly mediate functions, such as membrane barrier, virulence and protein transport (Wilson et al., 2005).

When considering the NMEC pathotype, we identified 16 GIs including previously discussed two islands which were common to all NMEC used for comparison purposes (Table 3). The GI 5 encodes for a sugar transport system for multiple sugars belonging to lignocelluloses which are not fermentable by *Enterobactericeae* but by many anaerobic bacteria that inhabit the intestinal tract and produce sugar end products, such as ribose, xylose, arabinose, and galactoside (Saraoui et al., 2013). Therefore, harboring a transport system for these sugars may enhance the survival of NMEC in nutrient-limited and competitive environments like in the host tissues. The GI 11 encodes DNA methylase enzyme (Ade-MTase) along with a putative ABC membrane transport system and hemin receptor (S1 Table). In pathogenic bacteria, such as *Salmonella* and *Yersinia*, DNA methylation is required for expression of several virulence traits and is considered as a virulence property (Honma et al., 2004; Low, 2001). Since all NMEC analyzed in this study contained this GI, it will be interesting to see if Ade-MTase is involved in virulence gene expression in NMEC as is the case with other pathogenic bacteria. The GI 13 encodes for a putative integral membrane transport system which is similar to multidrug resistance efflux systems. However, we observed that the RS218 strain is susceptible to almost all commonly used antimicrobials (unpublished data). Therefore, functional aspects of these efflux systems other than their role in antimicrobial resistance might open up new avenues to identify novel mechanistic advantages for NMEC pathogens that produce these proteins. Another GI common to all NMEC is GI 15 which encodes for a chitobiose-specific phosphotransferase system (PTS-Chitobiose). Chitobiose is a glycoprotein which is abundant in BBB (Yoshida et al., 2007). It has previously been demonstrated

that OmpA of *E. coli* K1 interaction with chitobiose present in BBB is crucial for the internalization of bacteria into the brain endothelial cells (Datta, 2003). Shifting of sugar fermentation pathways have already been characterized as fitness traits in pathogenic bacteria (Fabich et al., 2008; Njoroge et al., 2012). Taken together, PTS-chitobiose might be involved in the survival of NMEC in cytoplasmic vacuoles and bound chitobiose may act as an energy source within the vacuole. Also GIs 30, 33, 37, and 42 encode for ribose transport systems, a putative carbohydrate PTS system, a fructose-specific PTS, and a sorbose-specific PTS which may enhance the fitness of NMEC allowing them to survive under nutrient-limited conditions. Additionally, GI 39 encodes for a tricarboxylate (TCA) transport operon (*tctCBA*) similar to *Salmonella* Typhimurium, which enables bacteria to use metabolites like citrate (Widenhorn et al., 1989). Possession of numerous carbohydrate transport systems might enhance metabolic fitness of NMEC. Therefore, studies on functional and mechanistic properties of these systems, availability of corresponding sugars in the host environment and their involvement in metabolic shifting and gene regulation of NMEC inside the host are important to understand the niche-specific pathoadaptation of NMEC.

Table 3. Genomic islands (GIs) of *E. coli* RS218 that are shared among all other NMEC with >90% coverage and >90% similarity.

GI #	Size (bp)	Description
5	5,127	Ribose/xylose/arabinose/galactoside periplasmic ABC transporter
11	8,620	DNA methylation (Ade-MTase),Ton B dependent OMP, hemin ABC transporter, trehalase
13	7,492	RND efflux system
15	10,248	PTS system-chitobiose-specific components, moltoprotein
17	38,272	Yersiniabactin, adhesin and invasin (Peyer's patches colonization)
24	28,480	Capsular polysaccharide biosynthesis and secretion
26	5,535	Prepillin, accessory colonization factor (AcfD) precursor
30	8,495	Ribose ABC transport system
33	5,600	Putative carbohydrate PTS system
34	8,951	Hemin transport proteins, ChuA
37	6,943	PTS system, fructose-specific components
39	5,439	Tricarboxylate transport system
42	8,368	PTS system, sorbose-specific components
48	6,383	Arginine/ornithine metabolism
27	7,322	Conserved tetratricopeptide repeats protein
36	9,097	O antigen glycosylation proteins

The GIs 14 and 24 encode for yersiniabactin, adhesin/invasin and the K1 capsular polysaccharide biosynthesis pathway proteins. The GI 26, which is located just downstream to the capsular polysaccharide island, encodes for prepillin of type IV pili and accessory colonization factor D, and putative YghG proteins similar to ETEC and *Vibrio cholerae* T2SS structural components, which were required for colonization of the mouse intestines and toxin production (Strozen et al., 2012). The GI 48 encodes for arginine/ornithine antiporters and deaminases, which are considered as acid resistance mechanisms in bacteria (Poolman et al., 1987). It has been observed that pathogens which enter through the oral route have evolved several acid resistance mechanisms to survive the stomach acidity. The major route of transmission of NMEC is believed to be the oral route (Casiano-Colón & Marquis, 1988). Moreover, the human vaginal environment is also acidic; therefore, bearing an acid resistance operon in NMEC may enhance the survival and fitness of NMEC in the vaginal environment of the mother and the gastric environment of both the mother and the infant.

Comparative genomics with other ExPEC revealed a greater distribution of RS218 GIs (range 28-48) in ExPEC belonging to B2 phylogroup (*E. coli* strains UTI89, IHE3034, APECO1, S88, CFT073, and UPEC536), compared to ExPEC belonging to groups A and D (range 3-16) (Figure 5). Clustering of ExPEC strains based on the presence or absence of RS218 GIs revealed two major clusters which separate phylogroup B2, which is considered as the most common phylogroup among pathogenic *E. coli*, from groups A and D, suggesting a higher incidence of horizontal gene transfer in ExPEC belonging to B2 phylogroup compared to ExPEC belonging to other phylogroups (Figure 6). Although RS218 harbors three unique islands, which contain mostly phage regions, none of these GIs were universally present in NMEC that have been sequenced to date. We further observed that GIs clustered the ST95 strains together and ExPEC belonging to this cluster contained the highest number of GIs (Figure 6). This subcluster included all B2 NMEC, the cystitis-causing UPEC strain UTI89, and APECO1, which was originally isolated from the lung tissue of a turkey with colisepticemia (Johnson et al., 2012). *Escherichia coli* strains belonging to this subcluster shared eight RS218 GIs (5, 8, 12, 23, 24, 25, 32, and 35), which are not present in ExPEC belonging to other sequence types. These GIs contain genes for T6SS, putative lipoproteins, capsular associated polysaccharides, and putative fimbrial adhesins, which might be involved in ST95 *E. coli*-specific virulence. Furthermore, the presence of RS218 GIs in APECO1 in high numbers indicates that APEC strains may have the potential to cause disease in humans as suggested by previous investigators (Johnson et al., 2012).

Figure 5. Distribution of RS218 genomic islands (GIs) among other ExPEC strains used in the study.

Blue, presence of a GI; Grey, absence of a GI; GI designations correspond to Table S1.

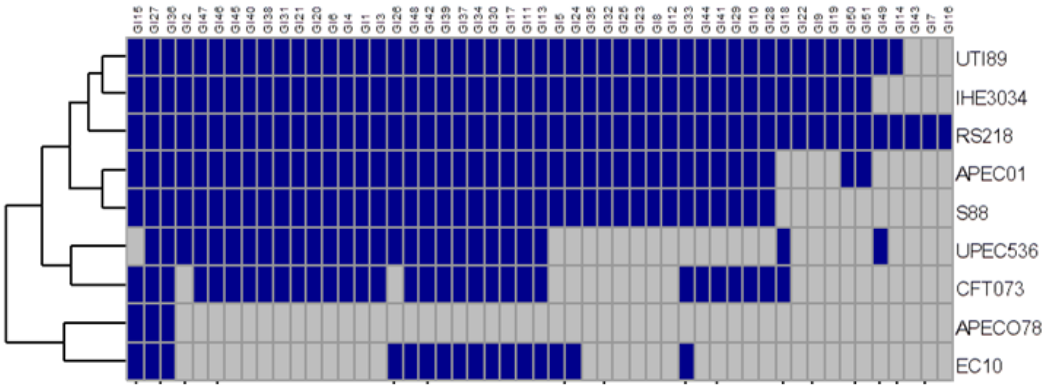
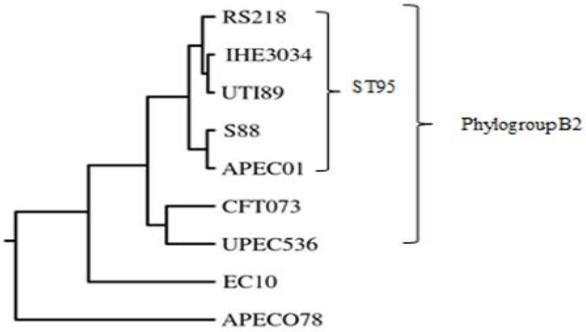


Figure 6. Clustering of ExPEC strains based on the presence or absence of RS218-associated genomic islands (GIs). The ExPEC strains containing more than 24 GIs belonged to the B2 phylogentic group and clustered together. Strains possessing eight of the RS218 GIs (5, 8, 12, 23, 24, 25, 32, and 35) formed a subcluster within the main cluster and they all belonged to ST95. These eight GIs were not present in EXPEC belonging to other STs.



4. Conclusions

Whole genome sequence evaluation of RS218 identified many potential virulence traits that might broaden our understanding of NMEC pathogenesis, particularly in relation to initial colonization of mucosal epithelia in both the mother and the infant as well as penetration of the intestinal mucosal barrier and BBB which are not yet fully understood. These include metabolic pathways, putative adhesins, invasins, iron acquisition systems, and cytotoxins, some of which have already been studied in other pathogenic *E. coli*. Finally, the presence of RS218 GIs in other ExPEC strains suggests that acquisition of GIs might be associated with phylogeny although the horizontal gene transfer is thought to be a random event.

Acknowledgements
We are grateful to Pennsylvania State University's Genomics Core Facility for performing the whole genome sequencing.

We are thankful to Dr. Prabhani Kuruppumullage Don for her assistance with generation of heat maps using R code.

References

1. Allsopp, L. P., Beloin, C., Moriel, D. G., Totsika, M., Ghigo, J.-M., & Schembri, M. A. (2012). Functional heterogeneity of the UpaH autotransporter protein from uropathogenic *Escherichia coli*. *J bacteriol, 194*(21), 5769-5782. doi:10.1128/jb.01264-12

2. Amor, K. (2000). Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. *Infect Immun, 68*(3), 1116-1124.

3. Antao, E.-M., Wieler, L., & Ewers, C. (2009). Adhesive threads of extraintestinal pathogenic *Escherichia coli*. *Gut Pathogens, 1*(1), 22.

4. Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., . . . Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics, 9*(1), 1-15. doi:10.1186/1471-2164-9-75

5. Bach, S. (2000). The high-pathogenicity island is present in different members of the family. *FEMS Microbiol Lett, 183*(2), 289-294. doi:10.1111/j.1574-6968.2000.tb08973.x

6. Bachmann, B. (2006). *Escherichia coli* and *Salmonella: Cellular and Molecular Biology*: ASM Press.

7. Badger, J. L., Wass, C. A., & Kim, K. S. (2000). Identification of *Escherichia coli* K1 genes contributing to human brain microvascular endothelial cell invasion by differential fluorescence induction. *Mol Microbiol, 36*(1), 174-182. doi:10.1046/j.1365-2958.2000.01840.x

8. Badger, J. L., Wass, C. A., Weissman, S. J., & Kim, K. S. (2000). Application of signature-tagged mutagenesis for identification of *Escherichia coli* K1 genes that contribute to invasion of human brain microvascular endothelial cells. *Infect Immun, 68*(9), 5056-5061. doi:10.1128/iai.68.9.5056-5061.2000

9. Benz, I., & Schmidt, M. A. (1992). Isolation and serologic characterization of AIDA-I, the adhesin mediating the diffuse adherence phenotype of the diarrhea-associated *Escherichia coli* strain 2787 (O126:H27). *Infect Immun, 60*(1), 13-18.

10. Benz, I., & Schmidt, M. A. (2001). Glycosylation with heptose residues mediated by the *aah* gene product is essential for adherence of the AIDA-I adhesin. *Mol Microbiol, 40*(6), 1403-1413. doi:10.1046/j.1365-2958.2001.02487.x

11. Bloch, C. A., Huang, S.-H., Rode, C. K., & Kim, K. S. (1996). Mapping of noninvasion Tn*phoA* mutations on the *Escherichia coli* O18: K1: H7 chromosome. *FEMS Microbiol Lett, 144*(2-3), 171-176.

12. Bodelón, G., Palomino, C., & Fernández, L. Á. (2013). Immunoglobulin domains in *Escherichia coli* and other enterobacteria: from pathogenesis to applications in antibody technologies. *FEMS Microbiol Rev, 37*(2), 204-250. doi:10.1111/j.1574-6976.2012.00347.x

13. Brzuszkiewicz, E., Brüggemann, H., Liesegang, H., Emmerth, M., Ölschläger, T., Nagy, G., . . . Dobrindt, U. (2006). How to become a uropathogen: Comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proc Natl Acad Sci, 103*(34), 12879-12884. doi:10.1073/pnas.0603038103

14. Carniel, E. (1996). Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia*

enterocolitica. *J Bacteriol, 178*(23), 6743-6751.
15. Casiano-Colón, A., & Marquis, R. E. (1988). Role of the arginine deiminase system in protecting oral bacteria and an enzymatic basis for acid tolerance. *Appl Environ Microbiol, 54*(6), 1318-1324.
16. Chen, D., Wu, R., Bryan, T. L., & Dunaway-Mariano, D. (2009). In vitro kinetic analysis of substrate specificity in enterobactin biosynthetic lower pathway enzymes provides insight into the biochemical function of the hot dog-fold thioesterase EntH. *Biochemistry, 48*(3), 511-513. doi:10.1021/bi802207t
17. Croxen, M. A., & Finlay, B. B. (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol, 8*(1), 26-38. doi:http://www.nature.com/nrmicro/journal/v8/n1/suppinfo/nrmicro2265_S1.html
18. Datta, D. (2003). Interaction of *E. coli* outer-membrane protein A with sugars on the receptors of the brain microvascular endothelial cells. *Proteins, 50*(2), 213-221. doi:10.1002/prot.10257
19. Delepelaire, P. (2004). Type I secretion in gram-negative bacteria. *Biochim Biophys Acta, 1694*(1–3), 149-161. doi:http://dx.doi.org/10.1016/j.bbamcr.2004.05.001
20. Dobrindt, U., Hentschel, U., Kaper, J. B., & Hacker, J. (2002). Genome plasticity in pathogenic and nonpathogenic enterobacteria. In J. Hacker & J. Kaper (Eds.), *Pathogenicity Islands and the Evolution of Pathogenic Microbes* (Vol. 264/1, pp. 157-175): Springer Berlin Heidelberg.
21. Fabich, A. J., Jones, S. A., Chowdhury, F. Z., Cernosek, A., Anderson, A., Smalley, D., . . . Conway, T. (2008). Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infect Immun, 76*(3), 1143-1152. doi:10.1128/iai.01386-07
22. Feist, A. M. (2008). The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol, 26*(6), 659-667. doi:10.1038/nbt1401
23. Feldmann, F., Sorsa, L. J., Hildinger, K., & Schubert, S. (2007). The salmochelin siderophore receptor iron contributes to invasion of urothelial cells by extraintestinal pathogenic *Escherichia coli* in vitro. *Infect Immun, 75*(6), 3183-3187. doi:10.1128/iai.00656-06
24. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., . . . Venter, J. C. (1995). Whole-genome random sequencing and assembly of *Haemophilus Influenzae* Rd. *Science, 269*(5223), 496-512. doi:10.2307/2887657
25. Flint, H. J. (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol, 6*(2), 121-131. doi:10.1038/nrmicro1817
26. Gao, Q., Wang, X., Xu, H., Xu, Y., Ling, J., Zhang, D., . . . Liu, X. (2012). Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiol, 12*(1), 143.
27. Garénaux, A., Caza, M., & Dozois, C. M. (2011). The ins and outs of siderophore mediated iron uptake by extra-intestinal pathogenic *Escherichia coli*. *Vet Microbiol, 153*(1–2), 89-98.

doi:http://dx.doi.org/10.1016/j.vetmic.2011.05.023
28. Gawarzewski, I., Smits, S. H., Schmitt, L., & Jose, J. (2013). Structural comparison of the transport units of type V secretion systems. *Biol Chem, 394*(11), 1385-1398.
29. Gérard, F., Pradel, N., & Wu, L.-F. (2005). Bactericidal activity of Colicin V is mediated by an inner membrane protein, SdaC, of *Escherichia coli*. *J Bacteriol, 187*(6), 1945-1950. doi:10.1128/jb.187.6.1945-1950.2005
30. Große, C., Scherer, J., Koch, D., Otto, M., Taudte, N., & Grass, G. (2006). A new ferrous iron-uptake transporter, EfeU (YcdN), from *Escherichia coli*. *Mol Microbiol, 62*(1), 120-131. doi:10.1111/j.1365-2958.2006.05326.x
31. Hacker, J. (1992). Role of fimbrial adhesins in the pathogenesis of *Escherichia coli* infections. *Can J Microbiol, 38*(7), 720-727.
32. Hammer, N. D., Schmidt, J. C., & Chapman, M. R. (2007). The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *Proc Nat Acad Sci, 104*(30), 12494-12499. doi:10.1073/pnas.0703310104
33. Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev, 68*(4), 669-685. doi:10.1128/mmbr.68.4.669-685.2004
34. Hasman, H., Chakraborty, T., & Klemm, P. (1999). Antigen-43-mediated autoaggregation of *Escherichia coli* is blocked by fimbriation. *J Bacteriol, 181*(16), 4834-4841.
35. Hill, V. T. (2004). TraJ-dependent *Escherichia coli* K1 interactions with professional phagocytes are important for early systemic dissemination of infection in the neonatal rat. *Infect Immun, 72*(1), 478-488.
36. Hoffman, J. A. (2000). *Escherichia coli* K1 *aslA* contributes to invasion of brain microvascular endothelial cells in vitro and in vivo. *Infect Immun, 68*(9), 5062-5067.
37. Honma, Y., Fernández, R. E., & Maurelli, A. T. (2004). A DNA adenine methylase mutant of *Shigella flexneri* shows no significant attenuation of virulence. *Microbiol, 150*(4), 1073-1078. doi:10.1099/mic.0.26781-0
38. Hunt, S., Green, J., & Artymiuk, P. (2010). Hemolysin E (HlyE, ClyA, SheA) and Related Toxins. In G. Anderluh & J. Lakey (Eds.), *Proteins Membrane Binding and Pore Formation* (Vol. 677, pp. 116-126): Springer New York.
39. Jakes, K. S., & Cramer, W. A. (2012). Border crossings: Colicins and transporters. *Ann Rev Genet, 46*(1), 209-231. doi:doi:10.1146/annurev-genet-110711-155427
40. Jarvis, K. G. (1996). Secretion of extracellular proteins by enterohemorrhagic *Escherichia coli* via a putative type III secretion system. *Infect Immun, 64*(11), 4826-4829.
41. Johnson, J. R., Kuskowski, M. A., Gajewski, A., Soto, S., Horcajada, J. P., de Anta, M. T. J., & Vila, J. (2005). Extended virulence genotypes and phylogenetic background of *Escherichia coli* isolates from patients with cystitis, pyelonephritis, or prostatitis. *J Infect Dis, 191*(1), 46-50. doi:10.1086/426450
42. Johnson, T. J., Wannemuehler, Y., Kariyawasam, S., Johnson, J. R., Logue, C. M., & Nolan, L. K. (2012). Prevalence of avian-pathogenic *Escherichia coli* strain o1 genomic islands among extraintestinal and

commensal *E. coli* isolates. *J Bacteriol, 194*(11), 2846-2853. doi:10.1128/jb.06375-11

43. Kehres, D. G., Janakiraman, A., Slauch, J. M., & Maguire, M. E. (2002). SitABCD is the alkaline Mn2+ transporter of *Salmonella enterica* serovar Typhimurium. *J Bacteriol, 184*(12), 3159-3166. doi:10.1128/jb.184.12.3159-3166.2002

44. Khan, N. A., Wang, Y., Kim, K. J., Chung, J. W., Wass, C. A., & Kim, K. S. (2002). Cytotoxic necrotizing factor-1 contributes to *Escherichia coli* K1 invasion of the central nervous system.*J BioloChem, 277*(18), 15607-15612. doi:10.1074/jbc.M112224200

45. Kim, K. S. (2003). Neurological diseases: Pathogenesis of bacterial meningitis: from bacteraemia to neuronal injury. *Nat Rev Neurosci, 4*(5), 376-385. doi:10.1038/nrn1103

46. King, M. R., Vimr, R. P., Steenbergen, S. M., Spanjaard, L., Plunkett, G., Blattner, F. R., & Vimr, E. R. (2007). *Escherichia coli* K1-specific bacteriophage CUS-3 distribution and function in phase-variable capsular polysialic acid O acetylation. *J Bacteriol, 189*(17), 6447-6456. doi:10.1128/jb.00657-07

47. Kirjavainen, V., Jarva, H., Biedzka-Sarek, M., Blom, A. M., Skurnik, M., & Meri, S. (2008). *Yersinia enterocolitica* serum resistance proteins YadA and Ail bind the complement regulator C4b-binding protein. *PLoS Pathog, 4*(8), e1000140. doi:10.1371/journal.ppat.1000140

48. Klemm, P., Hjerrild, L., Gjermansen, M., & Schembri, M. A. (2004). Structure-function analysis of the self-recognizing Antigen 43 autotransporter protein from *Escherichia coli*. *Mol Microbiol, 51*(1), 283-296. doi:10.1046/j.1365-2958.2003.03833.x

49. Klemm, P., & Schembri, M. A. (2000). Bacterial adhesins: function and structure. *Intl J Med Microbiol, 290*(1), 27-35.doi:http://dx.doi.org/10.1016/S1438-4221(00)80102-2

50. Kostakioti, M., Newman, C. L., Thanassi, D. G., & Stathopoulos, C. (2005). Mechanisms of protein export across the bacterial outer membrane. *J Bacteriol, 187*(13), 4306-4314. doi:10.1128/jb.187.13.4306-4314.2005

51. Köster, W. (1991). Iron(III) hydroxamate transport across the cytoplasmic membrane of *Escherichia coli*. *Biol Met, 4*(1), 23-32. doi:10.1007/bf01135553

52. Kulkarni, R., Dhakal, B. K., Slechta, E. S., Kurtz, Z., Mulvey, M. A., & Thanassi, D. G. (2009). Roles of putative type II secretion and type IV pilus systems in the virulence of uropathogenic *Escherichia coli*. *PLoS ONE, 4*(3), e4752. doi:10.1371/journal.pone.0004752

53. Laarmann, S., & Schmidt, M. A. (2003). The *Escherichia coli* AIDA autotransporter adhesin recognizes an integral membrane glycoprotein as receptor. *Microbiol, 149*(7), 1871-1882. doi:10.1099/mic.0.26264-0

54. Lehti, T. A., Bauchart, P., Heikkinen, J., Hacker, J., Korhonen, T. K., Dobrindt, U., & Westerlund-Wikström, B. (2010). Mat fimbriae promote biofilm formation by meningitis-associated *Escherichia coli*. *Microbiol, 156*(8), 2408-2417. doi:10.1099/mic.0.039610-0

55. Leo, J. C., & Goldman, A. (2009). The immunoglobulin-binding Eib proteins from *Escherichia coli* are receptors for IgG Fc. *Mol Immunol, 46*(8–9), 1860-1866.

doi:http://dx.doi.org/10.1016/j.molimm.2009.02.024
56. Logue, C. M., Doetkott, C., Mangiamele, P., Wannemuehler, Y. M., Johnson, T. J., Tivendale, K. A., . . . Nolan, L. K. (2012). Genotypic and phenotypic traits that distinguish neonatal meningitis-associated *Escherichia coli* from fecal *E. coli* isolates of healthy human hosts. *Appl Environ Microbiol, 78*(16), 5824-5830. doi:10.1128/aem.07869-11
57. Low, D. A. (2001). Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infect Immun, 69*(12),7197-7204. doi:10.1128/iai.69.12.7197-7204.2001
58. Lu, S., Zhang, X., Zhu, Y., Kim, K. S., Yang, J., & Jin, Q. (2011). Complete genome sequence of the neonatal-meningitis-associated *Escherichia coli* strain CE10. *J Bacteriol, 193*(24), 7005. doi:10.1128/jb.06284-11
59. Mangiamele, P., Nicholson, B., Wannemuehler, Y., Seemann, T., Logue, C. M., Li, G., . . . Nolan, L. K. (2013). Complete genome sequence of the avian pathogenic *Escherichia coli* strain APEC O78. *Genome Announc, 1*(2). doi:10.1128/genomeA.00026-13
60. Maruvada, R., & Kim, K. S. (2012). IbeA and OmpA of *Escherichia coli* K1 exploit Rac1 activation for invasion of human brain microvascular endothelial cells. *Infect Immun, 80*(6), 2035-2041. doi:10.1128/iai.06320-11
61. Miethke, M. (2007). Siderophore-based iron acquisition and pathogen control. *Microbiol Mol Biol Rev, 71*(3), 413-451. doi:10.1128/mmbr.00012-07
62. Moriel, D. G., Bertoldi, I., Spagnuolo, A., Marchi, S., Rosini, R., Nesta, B., . . . Serino, L. (2010). Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proc Nat Acad Sci, 107*(20), 9072-9077. doi:10.1073/pnas.0915077107
63. Morozova, O., & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics, 92*(5), 255-264. doi:10.1016/j.ygeno.2008.07.001
64. Müller, S. I. (2009). Salmochelin, the long-overlooked catecholate siderophore of *Salmonella*. *Biometals, 22*(4), 691-695. doi:10.1007/s10534-009-9217-4
65. Nègre, V. L., Bonacorsi, S., Schubert, S., Bidet, P., Nassif, X., & Bingen, E. (2004). The siderophore receptor IroN, but not the high-pathogenicity island or the hemin receptor ChuA, contributes to the bacteremic step of *Escherichia coli* neonatal meningitis. *Infect Immun, 72*(2), 1216-1220. doi:10.1128/iai.72.2.1216-1220.2004
66. Nivaskumar, M., & Francetic, O. Type II secretion system: A magic beanstalk or a protein escalator. *Biochim Biophys Acta.*(0).doi:http://dx.doi.org/10.1016/j.bbamcr.2013.12.020
67. Njoroge, J. W., Nguyen, Y., Curtis, M. M., Moreira, C. G., & Sperandio, V. (2012). Virulence meets metabolism: *cra* and *kdpE* gene regulation in enterohemorrhagic *Escherichia coli*. *mBio, 3*(5). doi:10.1128/mBio.00280-12
68. Norton, J. P., & Mulvey, M. A. (2012).Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic *Escherichia coli*. *PLoS Pathog, 8*(10), e1002954. doi:10.1371/journal.ppat.1002954

69. Oh, J. Y., Kang, M. S., Yoon, H., Choi, H. W., An, B. K., Shin, E. G., . . . Kwon, Y. K. (2012). The embryo lethality of *Escherichia coli* isolates and its relationship to the presence of virulence-associated genes. *Poult Sci, 91*(2), 370-375.
70. Parreira, V. R., & Gyles, C. L. (2003). A novel pathogenicity island integrated adjacent to the *thrW* tRNA gene of avian pathogenic *Escherichia coli* encodes a vacuolating autotransporter toxin. *Infect Immun, 71*(9), 5087-5096. doi:10.1128/iai.71.9.5087-5096.2003
71. Paton, A. W., Srimanote, P., Woodrow, M. C., & Paton, J. C. (2001). Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* strains that are virulent for humans. *Infect Immun, 69*(11), 6999-7009. doi:10.1128/iai.69.11.6999-7009.2001
72. Peigne, C., Bidet, P., Mahjoub-Messai, F., Plainvert, C., Barbe, V., Médigue, C., . . . Bonacorsi, S. (2009). The plasmid of *Escherichia coli* strain S88 (O45:K1:H7) that causes neonatal meningitis is closely related to avian pathogenic *E. coli* plasmids and is associated with high-level bacteremia in a neonatal rat meningitis model. *Infect Immun, 77*(6), 2272-2284. doi:10.1128/iai.01333-08
73. Poolman, B., Driessen, A. J., & Konings, W. N. (1987). Regulation of arginine-ornithine exchange and the arginine deiminase pathway in *Streptococcus lactis*. *J Bacteriol, 169*(12), 5597-5604.
74. Rasko, D. A. (2008). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol, 190*(20), 6881-6893. doi:10.1128/jb.00619-08
75. Ravi, M. (2007). Contribution of AIDA-I to the pathogenicity of a porcine diarrheagenic *Escherichia coli* and to intestinal colonization through biofilm formation in pigs. *Vet Microbiol, 120*(3-4), 308-319. doi:10.1016/j.vetmic.2006.10.035
76. Raymond, K. N., Dertz, E. A., & Kim, S. S. (2003). Enterobactin: An archetype for microbial iron transport. *Proc Nat Acad Sci, 100*(7), 3584-3588. doi:10.1073/pnas.0630018100
77. Robbins, J. B., McCracken, G. H., Gotschlich, E. C., Ørskov, F., Ørskov, I., & Hanson, L. A. (1974). *Escherichia coli* K1 capsular polysaccharide associated with neonatal meningitis. *N Engl J Med, 290*(22), 1216-1220. doi:doi:10.1056/NEJM197405302902202
78. Runyen-Janecky, L., Dazenski, E., Hawkins, S., & Warner, L. (2006). Role and regulation of the *Shigella flexneri* Sit and MntH systems. *Infect Immun, 74*(8), 4666-4672. doi:10.1128/iai.00562-06
79. Sabri, M., Caza, M., Proulx, J., Lymberopoulos, M. H., Brée, A., Moulin-Schouleur, M., . . . Dozois, C. M. (2008). Contribution of the SitABCD, MntH, and FeoB metal transporters to the virulence of avian pathogenic *Escherichia coli* O78 strain χ7122. *Infect Immun, 76*(2), 601-611. doi:10.1128/iai.00789-07
80. Sabri, M., Léveillé, S., & Dozois, C. M. (2006). A SitABCD homologue from an avian pathogenic *Escherichia coli* strain mediates transport of iron and manganese and resistance to hydrogen peroxide. *Microbiol, 152*(3), 745-758. doi:10.1099/mic.0.28682-0
81. Saraoui, T., Parayre, S., Guernec, G., Loux, V., Montfort, J., Cam, A., . . .

Falentin, H. (2013). A unique in vivo experimental approach reveals metabolic adaptation of the probiotic *Propionibacterium freudenreichii* to the colon environment. *BMC Genomics*, *14*(1), 911.

82. Saukkonen, K. M., Nowicki, B., & Leinonen, M. (1988). Role of type 1 and S fimbriae in the pathogenesis of *Escherichia coli* O18:K1 bacteremia and meningitis in the infant rat. *Infect Immun*, *56*(4), 892-897.

83. Schaible, U. E. (2004). Iron and microbial infection. *Nat Rev Microbiol*, *2*(12), 946-953. doi:10.1038/nrmicro1046

84. Schubert, S. (2002). Yersinia high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal infections. *Infect Immun*, *70*(9), 5335-5337.

85. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., & Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res*, *34*(suppl 1), D32-D36. doi:10.1093/nar/gkj014

86. Silver, R. P., Aaronson, W., Sutton, A., & Schneerson, R. (1980). Comparative analysis of plasmids and some metabolic characteristics of *Escherichia coli* K1 from diseased and healthy individuals. *Infect Immun*, *29*(1), 200-206.

87. Skaar, E. P. (2010). The battle for iron between bacterial pathogens and their vertebrate hosts. *PLoS Pathog*, *6*(8), e1000949. doi:10.1371/journal.ppat.1000949

88. Spurbeck, R. R., Stapleton, A. E., Johnson, J. R., Walk, S. T., Hooton, T. M., & Mobley, H. L. T. (2011). Fimbrial profiles predict virulence of uropathogenic *Escherichia coli* strains: Contribution of Ygi and Yad fimbriae. *Infect Immun*, *79*(12), 4753-4763. doi:10.1128/iai.05621-11

89. Stins, M. F., Prasadarao, N. V., Ibric, L., Wass, C. A., Luckett, P., & Kim, K. S. (1994). Binding characteristics of S fimbriated *Escherichia coli* to isolated brain microvascular endothelial cells. *Am J Pathol*, *145*(5), 1228-1236.

90. Strozen, T. G., Li, G., & Howard, S. P. (2012). YghG (GspSβ) is a novel pilot protein required for localization of the GspSβ type II secretion system secretin of enterotoxigenic *Escherichia coli*. *Infect Immun*, *80*(8), 2608-2622. doi:10.1128/iai.06394-11

91. Teng, C.-H., Cai, M., Shin, S., Xie, Y., Kim, K.-J., Khan, N. A., . . . Kim, K. S. (2005). *Escherichia coli* K1 RS218 interacts with human brain microvascular endothelial cells via type 1 fimbria bacteria in the fimbriated state. *Infect Immun*, *73*(5), 2923-2931. doi:10.1128/iai.73.5.2923-2931.2005

92. Teng, C.-H., Tseng, Y.-T., Maruvada, R., Pearce, D., Xie, Y., Paul-Satyaseela, M., & Kim, K. S. (2010). NlpI contributes to *Escherichia coli* K1 strain RS218 interaction with human brain microvascular endothelial cells. *Infect Immun*, *78*(7), 3090-3096. doi:10.1128/iai.00034-10

93. Touchon, M. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet*, *5*(1), e1000344. doi:10.1371/journal.pgen.1000344

94. Wang, Y., Wen, Z. G., & Kim, K. S. (2004). Role of S fimbriae in *Escherichia coli* K1 binding to brain microvascular endothelial cells in vitro and penetration into the central nervous system in vivo. *Microb Pathog*,*37*(6),287-293. doi:http://dx.doi.org/10.1016/j.micpath.2004.09.002

95. Watts, R. E., Totsika, M., Challinor, V. L., Mabbett, A. N., Ulett, G. C., De Voss, J. J., & Schembri, M. A. (2012). Contribution of siderophore systems to growth and urinary tract colonization of asymptomatic bacteriuria *Escherichia coli*. *Infect Immun, 80*(1), 333-344. doi:10.1128/iai.05594-11

96. Welch, R. A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., . . . Blattner, F. R. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci, 99*(26), 17020-17024.

97. Whitfield, C., Heinrichs, D. E., Yethon, J. A., Amor, K. L., Monteiro, M. A., & Perry, M. B. (1999). Assembly of the R1-type core oligosaccharide of *Escherichia coli* lipopolysaccharide. *J Endotox Res, 5*(3), 151-156. doi:10.1177/09680519990050030901

98. Widenhorn, K. A., Somers, J. M., & Kay, W. W. (1989). Genetic regulation of the tricarboxylate transport operon (tctI) of *Salmonella* Typhimurium. *J Bacteriol, 171*(8), 4436-4441.

99. Wijetunge, D. S. S. (2014). Pathogenomics of neonatal meningitis causing *Escherichia coli*. (PhD), Pennsylvania State University. Retrieved fromhttps://etda.libraries.psu.edu/paper/21085/

100. Wijetunge, D. S. S., Karunathilake, K. H. E. M., Chaudhari, A., Katani, R., Dudley, E. G., Kapur, V., . . . Kariyawasam, S. (2014). Complete nucleotide sequence of pRS218, a large virulence plasmid, that augments pathogenic potential of meningitis-associated *Escherichia coli* strain RS218. *BMC Microbiol, 14*, 203. doi:10.1186/s12866-014-0203-9

101. Wilson, C. G. M., Kajander, T., & Regan, L. (2005). The crystal structure of NlpI. *FEBS J, 272*(1), 166-179. doi:10.1111/j.1432-1033.2004.04397.x

102. Wyborn, N. R., Clark, A., Roberts, R. E., Jamieson, S. J., Tzokov, S., Bullough, P. A., . . . Green, J. (2004). Properties of haemolysin E (HlyE) from a pathogenic *Escherichia coli* avian isolate and studies of HlyE export. *Microbiol, 150*(5), 1495-1505. doi:10.1099/mic.0.26877-0

103. Xie, Y., Kolisnychenko, V., Paul-Satyaseela, M., Elliott, S., Parthasarathy, G., Yao, Y., . . . Kim, K. S. (2006). identification and characterization of *Escherichia coli* RS218–derived islands in the pathogenesis of *E. coli* meningitis. *J Infect Dis, 194*(3), 358-364. doi:10.1086/505429

104. Yamaguchi, Y., Park, J.-H., & Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Ann Rev Genet, 45*(1), 61-79. doi:doi:10.1146/annurev-genet-110410-132412

105. Yoshida, Y., Murakami, A., Iwai, K., & Tanaka, K. (2007). A Neural-specific F-box protein Fbs1 functions as a chaperone suppressing glycoprotein aggregation. *J Biol Chem, 282*(10), 7137-7144. doi:10.1074/jbc.M611168200

106. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Res, 39*, W347-W352. doi:10.1093/nar/gkr485

107. Zhou, Y., Tao, J., Yu, H., Ni, J., Zeng, L., Teng, Q., . . . Yao, Y. (2012). Hcp family proteins secreted via the type VI secretion system coordinately regulate *Escherichia coli* K1 interaction with human brain microvascular endothelial cells. *Infect*

Immun, 80(3), 1243-1251. doi:10.1128/iai.05994-11

Insights from the Whole Genome Sequence of *Escherichia coli* RS218 in Identifying Potential Virulence Traits Common to *E. coli* Causing Neonatal Meningitis and Other Extra-intestinal Diseases

Supporting information

S1 Table. Characteristics of genomic islands (GIs) identified in the *E. coli* RS218 genome.
Genomic islands of RS218 chromosome were detected in comparison to the *E. coli* K12 genome. Adjacent genomic regions presence in RS218 chromosome but absent in *E. coli* K12 were combined to determine GIs.

GI #	Start	Stop	Size (Kbp)	Proteins encoded by the gene/region	Flanking tRNA
1	150613	160018	9.4	Fimbrial proteins Yad	
2	228804	266534	37.7	Type IV secretion system	tRNA Asp-GGT
3	289950	321781	31.8	Vat, CFA/I fimbrial cluster	tRNA-Thr-CGT
4	328912	333812	4.9	Hypothetical	
5	352727	357854	5.1	Ribose ABC transport	
6	666141	672069	5.9	Putative transporter, alcohol dehydrogenase	
7	833360	867256	33.9	Phage (intact)	
8	905310	913855	8.5	CRISPR-associated proteins	tRNA-Ser-GGA
9	1060208	1123164	63.0	S fimbrial cluster, salmochelin, antigen 43	tRNA-Ser-GGA
10	1219583	1268928	49.3	Phage proteins, SitABCD	
11	1298752	1307372	8.6	Ton B dependent OMP, hemin ABC transporter, trehalase	
12	1381227	1416358	35.1	Phage	
13	1453629	1461121	7.5	RND efflux system	
14	1628585	1643663	15.1	Type I fimbriae, phage related proteins	
15	1658146	1668394	10.2	PTS sytem-chitobiose-specific components	
16	1872257	1910488	38.2	Phage	
17	2090634	2128906	38.3	Yersiniabactin, adhesin and invasin	tRNA-Asn-GTT
18	2133298	2235245	101.9	Phage proteins, yersiniabactin	tRNA-Asn-GTT
19	2259035	2266332	7.3	O antigen modification protein, glycosyltransferase	
20	2322143	2331637	9.5	Putative inner membrane and cytoplasmic proteins	
21	2412032	2422963	10.9	Sugar fermentation and transport	
22	2616461	2658013	41.6	Phage	tRNA-Arg-CCT
23	2919732	2924825	5.1	Putative proteins	
24	3307311	3335791	28.5	Capsular polysaccharide island	tRNA-Phe GAA
25	3121760	3157032	35.3	Putative type VI secretory proteins, sugar fermentation	
26	3336285	3341820	5.5	Prepillin, accessory colonization factor (AcfD) precursor	
27	3354984	3362306	7.3	Conserved TRP repeat protein	
28	3408660	3418262	9.6	Putative iron ABC transporters, ferrochrome-	

				iron receptor
29	3431868	3438964	7.1	Uncharacterized fimbrial-like proteins
30	3635820	3644315	8.5	Ribose ABC transport system
31	3795890	3804697	8.8	Putative fimbrial operon
32	3838120	3845038	6.9	PTS system, mannose-specific components
33	3864570	3870170	5.6	Putative carbohydrate PTS system
34	3891923	3900874	9.0	Hemin transport proteins , ChuA
35	4012262	4017740	5.5	Putative lipoprotein
36	4036410	4045507	9.1	O antigen glycosylation proteins
37	4069493	4076436	6.9	PTS system, fructose-specific components
38	4259240	4275085	15.8	PTS system, glucose-specific IIB components, sialic acid utilization
39	4276986	4282425	5.4	Tricarboxylate transport
40	4325074	4333727	8.7	Sugar kinase, putative transport protein
41	4385586	4395426	9.8	Nucleotidase
42	4492336	4500704	8.4	PTS system, sorbose-specific components
43	4529460	4576064	46.6	Phage
44	4575338	4582368	7.0	Hydroxybutyryl-CoA dehydratase, propionate CoA-transferase
45	4588020	4601367	13.3	Succinyl-CoA ligase, dihydrolipoamide dehydrogenase, malate dehydrogenase
46	4629636	4637226	7.6	Putative hemin protease, formate dehydrogenase
47	4762663	4772351	9.7	Hydroxybutyryl-CoA dehydratase, CoA-transferase
48	4815175	4821558	6.4	Arginine/ornithine metabolism
49	4840815	4963833	123.0	Pap operon, CNF1, RTX toxin transporter, putative hemin receptors, tRNA-Leu-CAA
50	4982442	5003447	21.0	GimA, IbeA
51	5014081	5018158	4.1	Type I restriction-modification system, specificity subunit S

