

Whole-Genome Sequencing of Invasive Neonatal *Escherichia coli* From Uppsala County, Sweden

Anna Heydecke,^{1,2} Åsa Myrelid,^{3,4} Erik Normann,^{4,5} Karolina Gullsby,^{1,2} Eva Tano,^{6,○} and Susanne Sütterlin^{3,4,○}

¹Center for Research and Development Gävleborg, Uppsala University, Gävle, Sweden; ²Department of Clinical Microbiology, Gävle Hospital, Gävle, Sweden; ³Pediatric Inflammation, Metabolism, and Child Health Research, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden; ⁴Department of Pediatrics, Uppsala University Hospital, Uppsala, Sweden; ⁵Perinatal, Neonatal, and Pediatric Cardiology Research, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden; and ⁶Department of Medical Sciences, Uppsala University, Uppsala, Sweden

Background. This study sought to investigate associations between virulence factors and phylogeny in all neonatal *Escherichia coli* bloodstream infections from patients admitted to the neonatal intensive care unit at Uppsala University Hospital between 2005 and 2020.

Methods. A total of 37 *E. coli* isolates from 32 neonates were whole-genome sequenced and analyzed for virulence factors related to extraintestinal *E. coli*; patient-related data were collected retrospectively from the medical records.

Results. *E. coli* isolates that belong to phylogroup B2 were associated with mortality (odds ratio [OR], 26; $P < .001$), extreme prematurity with delivery before gestational week 28 (OR, 9; $P < .05$), and shock (OR, 9; $P < .05$) compared with isolates of non-B2 group. Female neonates were more often infected with isolates of phylogroup B2 *E. coli* compared with male neonates (OR, 7; $P = .05$). The identification of the genotoxin determinant *clb* coding for colibactin exhibited strong associations with mortality (OR, 67; $P < .005$), gestational age (OR, 18; $P < .005$), and shock (OR, 26; $P < .005$).

Discussion. The study highlighted the correlation between neonatal *E. coli* bacteremia caused by phylogroup B2 and the role of colibactin. Results emphasize difference between male and female neonates in *E. coli* populations in bloodstream infections.

Keywords. *Escherichia coli*; neonatal sepsis; virulence factors; ExPEC; population genetics.

Escherichia coli bloodstream infections present a substantial challenge within neonatal care and are acknowledged to be an important global public health concern [1]. In particular, the incidence of early-onset infections—occurring within the first 72 hours after delivery—have risen among extremely premature infants born before 28 gestational weeks [2]. Identified risk factors for early-onset *E. coli* infection in neonates are intrapartum fever, prolonged rupture of membranes, and chorioamnionitis, and it is associated with prematurity [2, 3].

In contrast, late-onset infections, which manifest after 72 hours postnatally and are caused by invasive *E. coli*, are more often linked to compromised barrier defenses that lead to severe conditions such as necrotizing enterocolitis or birth defects that contribute to infections in the urinary tract [4]. Additionally, an extended stay in the intensive care unit heightens the risk for nosocomial infections due to gram-negative rods such as *E. coli* [5]. The widely adopted prevention strategies for neonatal infections involve risk stratification and antibiotic treatment policies for both mother and child. However, these approaches pose challenges in balancing the risks associated with the spread of antimicrobial resistance and the prevention of infections in a vulnerable patient group [3].

E. coli is renowned for its highly adaptable pathogenic nature, and its pathovars are well-adapted to diverse ecological niches. Extraintestinal pathogenic *E. coli* (ExPEC) are responsible for infections in typically sterile body sites such as the bloodstream and urinary tract, or neonatal meningitis. These strains are disproportionately found in phylogenetic groups B2 and D [6]. A multitude of virulence factors contribute to the pathogenicity of ExPEC isolates. Interestingly, when present in phylogroups A and B1, commonly associated with commensalism, these virulence factors can lead to lethal infection in mice [7].

This study sought to investigate associations between a broad spectrum of virulence factors and phylogeny in all neonatal *E. coli* bloodstream infections from patients admitted to the

Received 01 March 2024; editorial decision 04 June 2024; accepted 11 June 2024; published online 13 June 2024

Presented in part: European Society for Paediatric Infectious Diseases, Copenhagen, Denmark, 2024, oral presentation.

Correspondence: Anna Heydecke, MD, Region Gävleborg; Department of Clinical Microbiology, Uppsala University, 801 87 Gävle, Sweden (anna.heydecke@regiongavleborg.se). Susanne Sütterlin, MD, PhD, International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, 75185 Uppsala, Sweden (susanne.sutterlin@uu.se).

The Journal of Infectious Diseases® 2024;230:e1136–45

© The Author(s) 2024. Published by Oxford University Press on behalf of Infectious Diseases Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

<https://doi.org/10.1093/infdis/jiae309>

neonatal intensive care unit at Uppsala University Hospital between 2005 and 2020, and to examine relationships with neonatal characteristics.

METHODS

Patients

At Uppsala University Hospital, all blood stream and cerebrospinal fluid isolates were stored in -80°C freezers in glycerol stock at the Department of Clinical Microbiology. For the purpose of the study, a search for *E. coli* isolates was conducted for all patients treated at the Neonatal Intensive Care Unit, Uppsala University Hospital, during the calendar years 2005 to 2020, and a total of 32 neonates were identified. For all patients, a review of the clinical records regarding the perinatal and neonatal care periods was performed. Informed consent from the included subjects was not necessary; the study was approved by the Swedish Ethical Review Board (document number 2020-03297).

The review of the neonates' records included collection of information about mortality, gestational week, weight, and sex. Infections were divided into early onset (debut 0–3 days postnatally) and late onset (after 3 days postnatally), which was confirmed by a positive blood culture. Furthermore, information about perinatal infections in the mother and/or the neonate was registered, and included necrotizing enterocolitis, chorioamnionitis, and urinary tract infection. Other laboratory parameters for the investigated *E. coli* bacteremia were included as follows: C-reactive protein (mg/L), bilirubin ($\mu\text{mol/L}$), hemoglobin (g/L), white blood cell count ($\times 10^9/\text{L}$), and platelet count ($\times 10^9/\text{L}$) when blood cultures were taken and peak values for C-reactive protein and bilirubin after blood culture samples. Furthermore, the symptoms when the blood culture was taken were evaluated and categorized as (pre)shock (fluid substitution and/or catecholamines were given), impaired (affected general condition regarding vital signs, respiration, circulations), and no symptoms (good general condition).

Culture-Based Methods

Blood or cerebrospinal fluid specimens were inoculated in BacT/Alert pediatric bottles including antimicrobial neutralization media and promptly placed in a BacT/Alert (Biomérieux) incubator. Cultures with growth were plated on blood agar, hematin agar, cysteine-lactose electrolyte-deficient, and fastidious anaerobic agar plates, and incubated for 2 days at 35°C in 5% CO_2 , air, and anaerobic conditions, respectively. Species identification was performed using standard laboratory procedures and VITEK (before 2013; Biomérieux) and MALDI-TOF (since 2013; Bruker). All isolates were stored in glycerol stock at -80°C .

Whole-Genome Analysis

All *E. coli* isolates were analyzed by whole-genome sequencing. Extraction and library preparation was performed as described

elsewhere [8]. Sequencing was performed on an Illumina HiSeq4000 platform (Oxford Genome Center, Oxford, United Kingdom) or Illumina NovaSeq 6000 SP platform (National Genomics Infrastructure, Stockholm, Sweden), generating 150-bp paired-end reads.

Paired-end reads were assembled using SPAdes assembler (version 3.11.1, <https://github.com/ablab/spades>) using the `-careful` flag and kmer length 21, 33, 55, 77, 99, and 127. Species confirmation was performed using the ribosomal multilocus sequence typing (rMLST) species tool on <http://pubmlst.org/rmlst>. All sequence data are available through the European Nucleotide Archive (www.ebi.ac.uk/ena) under project accession number PRJEB62832.

Composition of an ExPEC Virulence Factor Database and Application With the Study Isolates

The ExPEC database was constructed through an intense literature search for previously described virulence factors that have been associated with ExPEC. ExPEC-associated genes and gene alleles were included in the database when a nucleotide entry in the NCBI GenBank nucleotide collection (www.ncbi.nlm.nih.gov/nucleotide) and/or UniProt (www.uniprot.org) was linked to experimental data showing evidence for significant association with extraintestinal *E. coli* infections. The database included 674 entries corresponding to 133 virulence factors that are assigned to the following functional groups: fimbriae and adhesins ($n = 67$), iron metabolism ($n = 13$), exotoxins ($n = 19$), immunomodulation ($n = 15$), bacteriocins ($n = 29$), and capsule types ($n = 2$) (Table 1, Supplementary Table 1, and database ExPEC.fasta). Gene prediction on draft genomes of the 37 study isolates was performed using BLASTn algorithm with standard settings and virulence factors with a nucleotide coverage $\geq 99\%$ and nucleotide identity $\geq 98\%$ [9].

Calculations and Statistics

Genetic determinants for virulence factors were treated as presence/absence data; mortality, gestational age categorized by delivery before and after 28 gestational weeks, symptoms, necrotizing enterocolitis, chorioamnionitis, urinary tract infection, and maternal infection perinatally as dichotomous variables; and C-reactive protein, hemoglobin, bilirubin, white blood cell count, and platelet count as continuous variables. All data were tabulated and prepared in Microsoft Excel for further statistical computations, which were performed using the statistical software R (version 4.3.0, 2023-04-21; R Foundation for Statistical Computing).

Univariate description of data involved median and range or absolute and relative frequencies where appropriate. Analysis of associations between dichotomous variables was performed using Pearson correlation (ϕ coefficient for binary variables) as implemented in basic R (`"cor"` function, `"corrplot,"` `corrplot` package) for all variables with more than 4 observations.

Table 1. Fimbriae and Adhesins

Functional Group	Virulence Factor
Chaperone-usher pathway fimbriae	
α-Fimbriae	CFA/I (F2 antigen; <i>cfa</i>); CS1 (CFA/II, F3 antigen; <i>coa</i>); CS2 (CFA/II, F3 antigen; <i>cof</i>); CS4 (CFA/IV; <i>csa</i>); CS5 (CFA/IV; <i>csf</i>); CS14 (<i>csu</i>); CS17 (F17; <i>csb</i>); CS19 (<i>csd</i>); PCF071 (<i>cos</i>); ECP (Yag, Mat; <i>ecp</i>)
β-Fimbriae	Yhc (<i>yhc</i>)
γ1-Fimbriae	Auf (<i>auf</i>); F1A (Type 1 fimbriae; <i>fim</i>); F1C (Type 1C fimbriae; <i>foc</i>); S (<i>sfa</i>); Sfm (<i>sfm</i>); Ycb (<i>ycb</i>); Yde (F9; <i>ydb</i>); Lpf (Long polar fimbriae; <i>lpf</i>); Stg (<i>stg</i>); Yra (<i>yra</i>); <i>yqhG</i>
γ2-Fimbriae	CS12 (<i>csw</i>); CS18 (<i>fof</i>); 987P (F6 fimbrial antigen; <i>fas</i>)
γ3-Fimbriae	Afa-Dr (<i>dra</i>); DAF (<i>dafa</i>); AAF/I (<i>agg</i>); AAF/II (AAF-2; <i>aaf</i>); AAF/III (<i>agg3</i>); CS3 (CFA/II); CS6 (<i>css</i>); Dr (<i>dra</i>); F1845 (<i>daa</i>); Hda (<i>hda</i>); AFA-1, AFA-2, AFA-3, AFA-5, AFA-7, AFA-8 (<i>afa</i>)
γ4-Fimbriae	Yeh (<i>yeh</i>); Yad (<i>yad</i>); type 3 Fimbriae (<i>mrk</i>)
κ-Fimbriae	F4 (K88ab fimbrial antigen; <i>fae</i>); F5 (K99 fimbrial antigen; <i>fan</i>); F18ab (F107 fimbrial antigen; <i>fed</i>); Csh (<i>csh</i>); Lda (locus for diffuse adhesion; <i>lda</i>); AF/R1 (<i>afn</i>)
π-Fimbriae	<i>P-pili</i> (<i>pap</i>); F7(1) (<i>fso</i>); Pix (<i>pix</i>); PRF (P-related fimbriae; <i>prf</i>); Sfp (<i>sfp</i>); Ybg (<i>ybg</i>); Yfc (<i>yfc</i>); Ygi (<i>ygi</i>)
Other fimbriae and adhesins	
Type IV-fimbriae (TFP)	BFP (<i>bfp</i>); R64 (<i>pil</i>); Lng (<i>lng</i>); CFA/III (<i>cof</i>); CS31A (<i>cpf</i>); F17b-G (<i>f17G</i>); <i>gafD</i> (<i>gafD</i>)
Nonfimbrial adhesins	Ag8786 (<i>nfaA</i>); NFA-I (<i>nfa</i>)
AIDA-1-type adhesins	AIDA-I (<i>aidA</i>); Ag43 (<i>flu</i>); UpaB (<i>upaB</i>); UpaC (<i>upaC</i>); UpaH (<i>upaH</i>)
Other fimbriae	Curli fibres (<i>csg</i>); FdeC (<i>fdeC</i>); Iha (<i>iha</i>)
Iron metabolism	
Siderophore	Enterobactin (<i>fep</i>); Aerobactin (<i>iuc/iutA</i>); Yersiniabactin (<i>fyuA</i>); Salmochelin (<i>iro</i>)
Iron acquisition	ChuA (<i>chuA</i>); Hma (<i>hma</i>); Fec (<i>fec</i>); Fhu (<i>fhuA</i>); FhuE (<i>fhuE</i>); Sit (<i>sit</i>); Feo (<i>feo</i>); IreA (<i>ireA</i>); SsbL (<i>ssbL</i>)
Exotoxins	
RTX-toxin	Ehx (<i>hyl</i>); Tos (Upx; <i>tos</i>)
SPATE	EspC (<i>espC</i>); EspI (<i>espI</i>); EspP (<i>espP</i>); Pet (<i>pet</i>); Pic (<i>pic</i>); Sat (<i>sat</i>); Hbp (<i>hbp</i>); Tsh (<i>tsh</i>); Vat (<i>vat</i>); EatA (<i>eatA</i>); Eaa (<i>eaa</i>); EpeA (<i>epeA</i>)
Agglutinins	BmaE (<i>bmaE</i>); Hra (<i>hra</i>)
Bacterial genotoxins	Usp (<i>usp</i>); Colibactin (<i>clb</i>); CDT (<i>cdt</i>); CNF (<i>cnf</i>)
Enterotoxins	Set (<i>set1A</i> , <i>set1B</i>)
Immunomodulation	SisA, SisB (<i>sisA</i> , <i>sisB</i>); Eib (<i>eib</i>); TraT (<i>traT</i>); Iss (<i>iss</i> , <i>borD</i>); GalU (<i>galU</i>); TcpC (<i>tcpC</i>); AslA (<i>aslA</i>); Ibe (<i>ibe</i>); Bam (<i>bam</i>); Tol-Pal (<i>tol</i> , <i>pal</i> , <i>fil</i>); Lpp (<i>lpp</i>); Tsp (Prc, <i>prc</i>); OmpA (<i>ompA</i>); OmpT (<i>ompT</i>); OmpC (<i>ompC</i>); AslA (<i>aslA</i>)
Bacteriocins	
Colicins	Colicin 10 (<i>cta</i> , <i>cti</i> , <i>ctf</i>); Colicin 5 (<i>cfa</i> , <i>cfi</i> , <i>ctf</i>); Colicin B (<i>cba</i> , <i>cbi</i>); Colicin D (<i>cda</i> , <i>cdi</i>); Colicin E1 (<i>cea</i> , <i>immE1</i> , <i>lysE1</i>); Colicin E3 (<i>colE3</i> , <i>immE3</i> , <i>lysE3</i>); Colicin E4 (<i>colE4</i> , <i>lysE4</i>); Colicin E5 (<i>colE5</i> , <i>immE5</i> , <i>lysE5</i>); Colicin E6 (<i>colE6</i> , <i>immE6</i> , <i>lysE6</i>); Colicin E7 (<i>colE7</i>); Colicin E8 (<i>colE8</i> , <i>immE8</i> , <i>lysE8</i>); Colicin E9 (<i>colE9</i> , <i>immE9</i> , <i>lysE9</i>); Colicin Ia (<i>cia</i> , <i>cil</i>); Colicin Ib (<i>cib</i>); Colicin K (<i>cka</i> , <i>cki</i> , <i>ckl</i>); Colicin M (<i>cma</i> , <i>cmi</i>); Colicin N (<i>cna</i>); Colicin S4 (<i>csa</i> , <i>csi</i> , <i>csf</i>); Colicin Y (<i>cya</i> , <i>cyi</i> , <i>cyl</i>)
Microcins	Microcin 24 (<i>mtf</i> , <i>mbd</i>); Microcin B17 (<i>mcb</i>); Microcin C7 (<i>mcc</i>); Microcin H47 (<i>mch</i>); Microcin J25 (<i>mcj</i>); Microcin L (<i>mcl</i>); Microcin M (<i>mcm</i>); Microcin PDI (<i>mcp</i>); Microcin S (<i>mcs</i>); Microcin V (<i>cva</i> , <i>cvi</i>)
Capsule types	Group-II-capsules (<i>kps</i>); Group-III-capsules (<i>kps</i>), Region 2 of K1 capsule (<i>neu</i>);

For variable pairs with correlation or ϕ coefficients of > 0.4 or < -0.4 , the correlation was examined for meaningfulness. Subsequently, for meaningful associations, hypotheses regarding clinical parameters were formulated and odds ratios (OR) and statistical significance were calculated by Fisher exact test for small sample sizes.

The phylogenetic relationship of all isolates was analyzed using rMLST allele types in a cluster analysis calculated with MS Tree V2 algorithm as implemented in Enterobase (<https://enterobase.warwick.ac.uk>). The isolates' phylotypes were determined by Clermont typing, core genome MLST (cgMLST), and sequence types (ST) with concordant ST-clonal complexes using the 7-gene Achtman scheme were determined by Enterobase. Associations were evaluated graphically among mortality, gestational age, sex, symptoms, necrotizing enterocolitis, chorioamnionitis, urinary tract infection, and maternal infection perinatally to phylogenetic lineages according to

Clermont, and where appropriate ORs and statistical significance were calculated using Fisher exact test for small sample sizes. Graphical illustrations of metadata were accomplished using GraphTree [10].

RESULTS

Patients

During the 15-year period (2005–2020), in 14 very premature neonates (gestational age 21.8–26.3 weeks), 10 premature neonates (28.3–35.8 weeks), and 8 full-term neonates (36.7–42 weeks) *E. coli* blood stream infections were identified. For all 32 neonates, information on gestational age, weight, and sex were accessible. The neonates were born at a median gestational age of 29.6 weeks (range, 21.8 to 42 weeks) with a medium weight of 1140 g (range, 417 to 4080 g) corresponding to 7 of 32 (22%) infants being small for gestational age (below

–1.5 × standard deviation), 22 of 32 (69%) being appropriate for their gestational age, and 2 of 32 (6%) being large for their gestational age (above 1.5 × standard deviation). More neonates were male (22/32, 69%) than female (10/32, 31%).

For 1 patient, no record was accessible, thus information on mortality, symptoms, and other clinical conditions was available for 20 of 31 neonates (65%) that survived and 11 of 31 (35%) deceased neonates. A summary of the patient data and virulence factors is given in Table 2; no association between gestational age or mortality and chorioamnionitis or urinary tract infection was seen (Table 2).

General Comments on Bacterial Analysis

Bacteria

A total of 37 isolates were cultured from 32 neonates; 29 of 32 (91%) patients had 1 isolate, and in 3 of 32 (9%) neonates more than 1 isolate was detected. One patient had 3 bacteremic episodes and 1 simultaneous infection of cerebrospinal fluid during the second episode. The first bacteremic episode was at day 2 postnatally, the second 10 days later, and the last 6 weeks postnatally. One further patient had 2 bacteremic episodes with 12 days in between and 1 neonate had simultaneous bloodstream and cerebrospinal fluid infections. The first cultured isolate per patient and corresponding parameters were included in the statistical analysis. High-quality reads were obtained for 37 isolates and species verification on sequence data confirmed the species identification as *E. coli* and the purity of the whole-genome extracts.

Phylogenetics

All phylogenetic lineages according to Clermont were represented, and the majority was found to belong to phylogroup B2 (16/32, 50%). The distribution among other lineages was D (7/32, 22%), A (4/32, 13%), F (3/32, 9%), C (1/32, 3%), and E (1/32, 3%). According to 7-gene MLST, 26 isolates could be assigned to ST-clonal complexes: CC ST69 (5/32, 16%), CC ST10 (4/32, 13%), CC ST95 (4/32, 13%), CC ST12 (2/32, 6%), CC ST59 (2/32, 6%), CC ST568 (2/32, 6%), with single isolates for CC ST14, CC ST23, CC ST28, CC ST31, CC ST38, CC ST73, and CC ST131. The remaining 6 isolates corresponded to sequence types ST62, ST141, ST127, ST567, ST526 and ST1583. For patients from whom multiple isolates were cultured, both cgMLST and rMLST analysis consistently revealed identical sequence types. Specifically, 1 patient's isolate found during 3 bacteremic episodes and 1 isolate from cerebrospinal fluid, were all designated as ST69. Likewise, the other isolates originating from a patient who had 2 blood stream infections, could be assigned to ST62. The patient with *E. coli* isolates from both the bloodstream and cerebrospinal fluid exhibited isolates assigned to ST59. Additionally, 2 patients shared identical sequence types according to cgMLST and rMLST analysis. However, in the 7-gene MLST, only 1 isolate was assigned to ST69 and the other to ST31. Notably, these individuals

underwent intensive care simultaneously in the neonatal intensive care unit; therefore the isolates were assumed to be epidemiologically related (Figure 1).

Virulence Factors

Overall, virulence factors from all virulence factor groups were found: 29 of 67 (43%) fimbriae and adhesins, 13 of 13 (100%) iron metabolism, exotoxins 9 of 19 (47%), immunomodulation 15 of 15 (100%), bacteriocins 14 of 29 (48%), and capsule types 2 of 2 (100%). Identical isolates according to cgMLST shared the same virulence factor pattern.

Associations Between Bacterial Analysis and Patient Data

Phylogenetics, Patient Data, and Virulence Factors

Neonates infected with isolates belonging to phylogenetic lineage B2 according to Clermont (15/31, 48%) were more likely to die compared with neonates infected with isolates from other lineages (B2 10/15, 67% vs non-B2 1/16, 16%; OR, 26; $P < .001$). Extremely premature neonates delivered before 28 gestational weeks were more often infected with *E. coli* belonging to lineage B2 compared with neonates delivered after 28 gestational weeks (11/14, 78% vs 5/18, 28%; OR, 9; $P = .01$). Similarly, neonates with isolates belonging to lineage B2 were more likely to present with shock compared with those with impaired or good general condition (B2 6/15, 40% vs non-B2 1/16, 6%; OR, 9; $P = .03$). Also, *E. coli* that belong to phylogroup B2 were more frequently found in female patients compared with male patients (female 8/10, 80% vs male 8/22, 36%; OR, 7; $P < .05$). The mortality of infection with phylogroup B2 isolates was comparable between female and male patients (female 5/7, 71%; male 5/8, 63%). No associations between phylogenetic lineage and necrotizing enterocolitis, chorioamnionitis, urinary tract infection, or maternal infection perinatally were seen. Presence of determinants encoding fimbriae Sfm, Ycb, and Yra were statistically significantly associated with isolates belonging to non-B2 lineages. In contrast, determinants involved in iron metabolism (salmochelin, Sit, Hma, Fec, ChuA), exotoxins (colibactin, CNF, Usp), immunomodulation (TcpC, AslA, and OmpT), and bacteriocins (microcin H47 and M) were statistically significantly associated with phylogroup B2 (Table 3, Figure 1, and Figure 2).

Virulence Factors and Patient Data

Most statistically significant associations between virulence factors and patient data were found for mortality (14/133, 11%) and gestational age (8/133, 6%). Very premature neonates delivered before gestational week 28 had a high mortality; however, prematurity was also associated with 7 virulence factors (*cnf*, *yde*, *clb*, *hma*, *sfm*, *aslA*, and *iro*). The overall function of the 14 virulence factors that were associated with mortality could be assigned to exotoxins (4/14, 29%), immunomodulation (4/14, 29%), bacterial iron metabolism (2/14, 14%), fimbriae (2/14,

Table 2. Summary of the Patient Data and Presence of Virulence Factors Presented by Gestational Age With Delivery Before and After 28 Gestational Weeks, and Mortality

Characteristic	Gestational Age			P	Mortality	
	Delivery Before Gestational wk 28	Delivery in Gestational wk 28 to 42	18		Deceased	Survived
Total	14	18	11 ^a		20 ^a	
Mortality (No. in group/total No. [%])	9/13 (70) ^a	2/18 (11)	<.001***		...	
Gestational weight, g (median [range])	564.5 (417–897)	2412 (538–4080)	<.001***	631 (417–2414)	2289 (530–4080)	.003**
Gestational weight (SD) (median [range])	–1 (–3–3)	0 (–4.5–3)		0 (–3–3)	–0.25 (–4.5–3)	
Sex						
Male (No. in group/total No. [%])	9/14 (64)	13/18 (72)	.71	6/11 (55)	16/20 (80)	.218
Female (No. in group/total No. [%])	5/14 (36)	5/18 (28)		5/11 (45)	4/20 (20)	
Early onset (No. in group/total No. [%])	4/14 (29)	9/18 (50)	.29	5/11 (45)	8/20 (40)	1
Symptoms when blood culture taken						
Shock (No. in group/total No. [%])	6/13 (46) ^a	1/18 (6) ^a	<.05*	6/11 (55)	1/20 (5)	.007**
Impaired (No. in group/total No. [%])	6/13 (46) ^a	12/18 (67) ^a		4/11 (36)	14/20 (70)	
No symptoms (No. in group/total No. [%])	1/13 (8) ^a	5/18 (28) ^a		1/11 (9)	5/20 (25)	
Necrotizing enterocolitis (No. in group/total No. [%])	8/13 (62) ^a	2/18 (11)	.006**	7/11 (64)	3/20 (15)	.013*
Chorioamnionitis (No. in group/total No. [%])	8/12 (67) ^a	5/17 (29) ^a	.14	6/10 (60) ^a	7/19 (37) ^a	.270
Urinary tract infection (No. in group/total No. [%])	0/12 (0) ^a	5/18 (28)	.36	0/10 (0) ^a	5/20 (25)	.6
Maternal infection perinatally (No. in group/total No. [%])	7/12 (58) ^a	7/17 (41) ^a	.46	3/10 (30)	5/18 (28) ^a	1
CRP when blood culture taken, mg/L (median [range])	25 (5–93) ^a	52 (0.73–159) ^a	<.05*	32 (5–117)	47 (0.73–159)	.20
CRP peak after blood culture, mg/L (median [range])	52.5 (5–116) ^a	148 (14–294) ^a	<.005**	62 (5–161) ^a	127 (14–294) ^a	<.05*
Hemoglobin, g/L (median [range])	134.5 (96–160) ^a	144 (84–203) ^a	.23	120 (84–161) ^a	143.5 (88–203) ^a	<.05*
Bilirubin when blood culture taken, µmol/L (median [range])	61.5 (16–113) ^a	109 (10–235) ^a	.093	64 (16–113) ^a	81 (10–235) ^a	.21
Bilirubin peak after blood culture, µmol/L (median [range])	70 (21–134) ^a	181 (17–281) ^a	<.05*	90.5 (21–213) ^a	141 (17–281) ^a	.32
White blood cell count, x10 ⁹ /L (median [range])	7.5 (0.9–51.4) ^a	10.3 (1.2–29.1) ^a	.36	6.25 (0.9–32.8) ^a	10.8 (1.2–51.4) ^a	.59
Platelet count, x10 ⁹ /L (median [range])	132 (11–269) ^a	126 (22–265) ^a	.86	187 (25–269) ^a	115 (11–265) ^a	.27

Abbreviation: CRP, C-reactive protein.

*** *P* < .001, ** *P* < .01, * *P* < .05.

^aNumerator adjusted due to missing data.

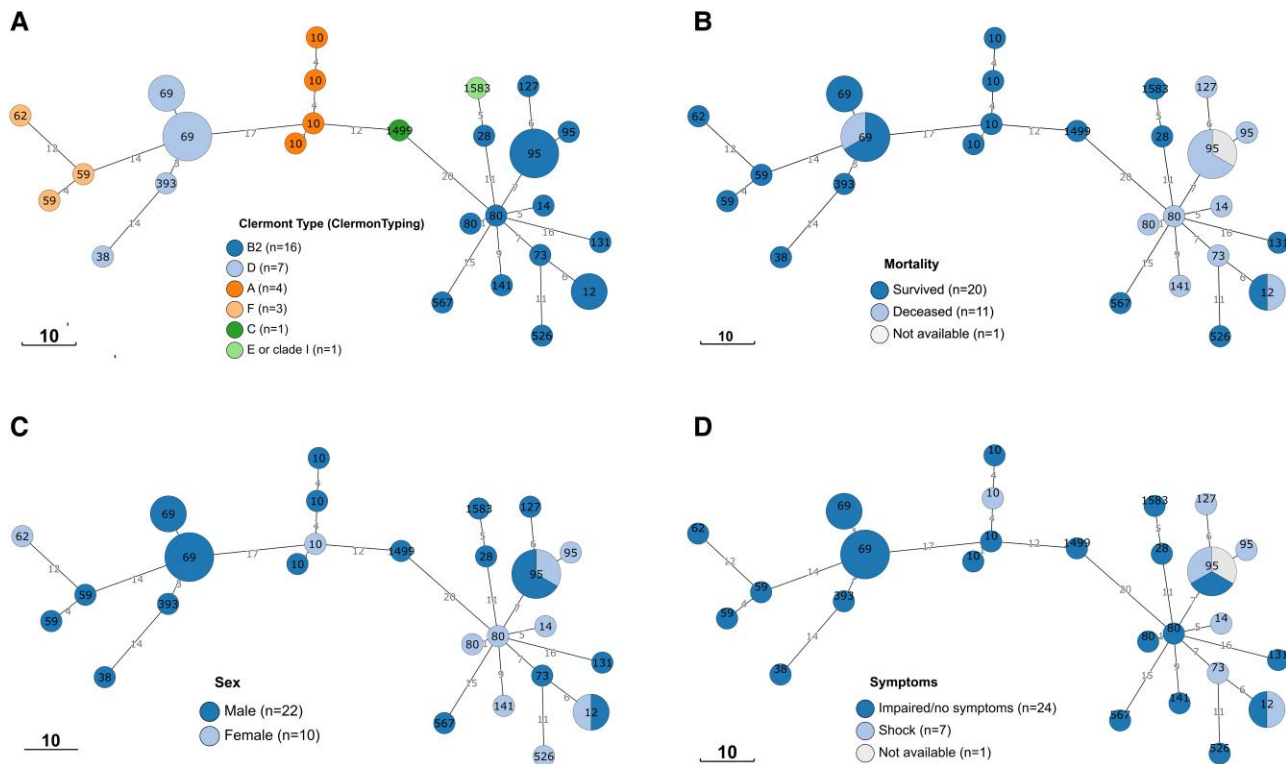


Figure 1. Minimum spanning tree (MST) for 32 *Escherichia coli* isolates based on rMLST with 52 allele positions, pairwise ignoring missing values. Numbers in the circles indicate clonal clusters according to legacy multilocus sequence type (MLST), the circles are colored by Clermont type (A); mortality (B); sex (C); and symptoms (D).

Table 3. Summary of Associations of Phylogroup B2 and Non-B2 and Mortality, Gestational Age, Sex, and Severity of Symptoms

Characteristic	B2, No. (%) (n = 16)	Non-B2, No. (%) (n = 16)	OR	P
Mortality ^a	10/15 (67)	1/16 (16)	26	< .001
Gestational age			9	< .05
Delivery before gestational wk 28	11/14 (78)	3/14 (21)		
Delivery after gestational wk 28	5/18 (28)	13/18 (72)		
Sex			7	.05
Male	8/22 (36)	14/22 (64)		
Female	8/10 (80)	2/10 (20)		
Symptoms ^a			9	< .05
Shock	6/7 (86)	1/7 (14)		
Impaired or no symptoms	9/24 (37)	15/24 (63)		

Odds ratios (OR) and probabilities (P) were calculated using Fisher exact test.

^aDue to missing data adjusted numerator.

14%), and bacteriocins (2/14, 14%). While most virulence factors were positively associated with stillbirth and extremely premature delivery, the γ_1 -fimbriae encoded by *sfm* and the self-recognizing antigen 43, encoded by *ag43* were associated with survival and delivery after gestational age 28, respectively. Of the 8 virulence factors that were associated with gestational age, most common were fimbriae (4/8, 50%), exotoxins (2/8, 25%), immunomodulation (1/8, 12.5%), and bacterial iron metabolism (1/8, 12.5%). Above all, colibactin showed

strong associations with mortality, gestational age, sex, and necrotizing enterocolitis. Patients with *E. coli* isolates carrying genes encoding the exotoxins colibactin, *clb*, were more likely to develop shock symptoms (shock 6/7, 86% vs nonshock 4/24, 17%) and to have necrotizing enterocolitis (6/10, 60% vs 4/21, 19%; OR, 6; $P = .04$). There were no strong associations of virulence factors with necrotizing enterocolitis, chorioamnionitis, urinary tract infection, or maternal infection perinatally (Table 4).

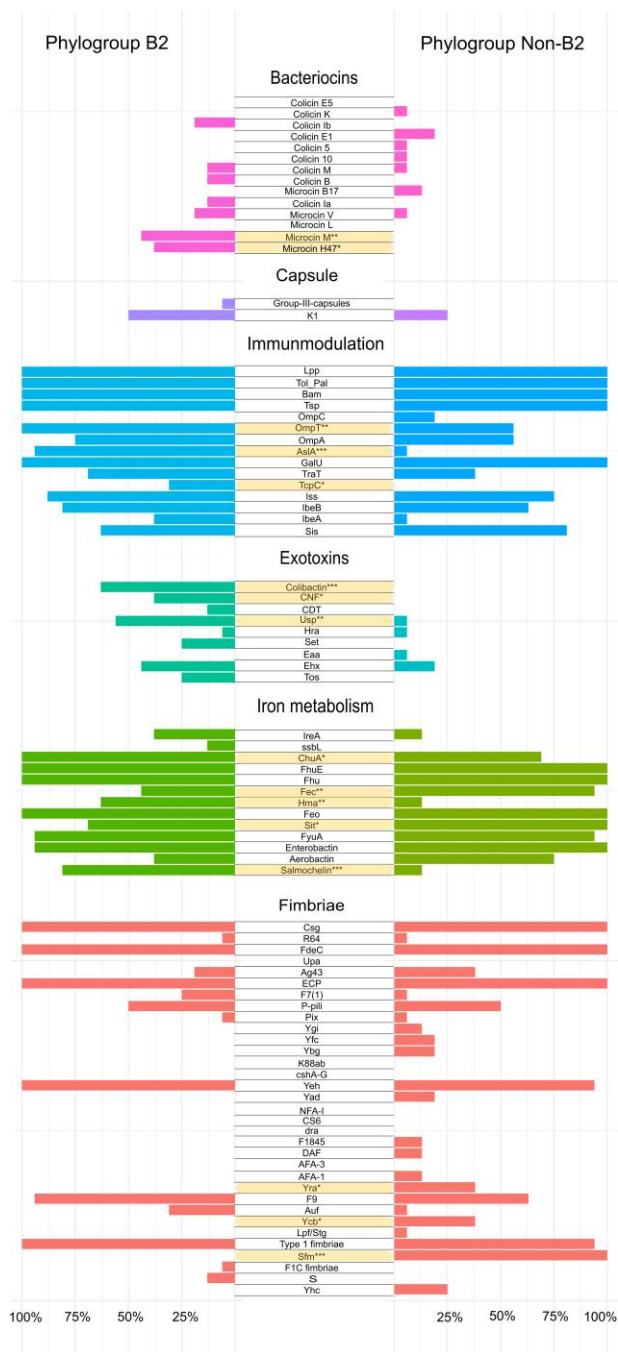


Figure 2. Absence/presence data of virulence factors (middle column) in isolates belonging to phylogroup B2 (left column) and phylogroups A, B1, C, or D (right column). Bar plots illustrate the percent presence of virulence factors. *** $P < .001$, ** $P < .01$, * $P < .05$.

DISCUSSION

The present study examined 37 neonatal *E. coli* isolates, obtained from infants admitted to the neonatal intensive care unit, Uppsala University Hospital across a 15-year period, for ExPEC virulence factors. Notably, *E. coli* isolates categorized under phylogroup B2 were associated with mortality, extreme

prematurity, shock, and female sex. The identification of exotoxin determinants, particularly the colibactin determinant *clb*, exhibited associations with mortality, gestational age, and the occurrence of shock. Additionally, the study introduced a comprehensive database of ExPEC virulence factor determinants, capable of efficiently analyzing whole-genome sequenced *E. coli* isolates.

The *E. coli* population is categorized into at least 8 phylogroups (A, B1, B2, C, D, E, F, and G), with phylogroup B2 notably linked to extraintestinal infections [11]. This association is believed to originate from colonization advantages of the human intestinal tract, enabling the acquisition of virulence traits and antimicrobial resistance determinants [12]. *E. coli* blood stream isolates associated with fetal and neonatal infections are often found to belong to phylogroup B2, which correlate with increased morbidity and mortality rates compared with other phylogroups [13, 14]. Half of the isolates belonged to phylogroup B2, and were associated with higher mortality compared with non-B2 and presented more often with shock symptoms. Phylogroup B2 was also found more often amongst extremely premature neonates, which have a higher mortality irrespective of infection status. However, a noteworthy observation is that while only 36% of the female neonates were extremely premature, 80% of all female infants had bacteremia linked to phylogroup B2, which resulted in higher mortality rates compared with male infants. This is in line with previous observations of heightened virulence potential of the B2 phylogroup and indicates sex-related differences in susceptibility to infections.

Sex-related differences in morbidity and mortality in early life have been documented since the 1930s, but evidence concerning sex-dependent distribution of *E. coli* populations is still limited [15, 16]. After perinatal and respiratory or nervous system conditions, infections are the fourth leading cause of infant mortality and exhibit a notable skew towards higher mortality in male patients [17]. While cultural factors can contribute to mitigating the male disadvantages in infant morbidity and mortality, an increasing body of evidence highlights biological sex differences in immune response to infections [18]. The previously described male vulnerability is reflected in the findings of this study, because bacteremia caused by *E. coli* was more common in male neonates born before or after 28 gestational weeks. Female infants had both a significantly higher likelihood of being infected with *E. coli* isolates associated with phylogroup B2, and had an increased risk of infection with isolates expressing capsule types K1 and K5 and genes encoding colibactin production, an association that might also be explained by phylogenetic linkages [13]. Possible biological explanations have been found in a genome-wide study involving 233 cases with late-onset neonatal sepsis from 6 countries that linked single-nucleotide polymorphisms in interleukin 10 (IL-10), tumor necrosis factor- α (TNF- α), and the NOTCH signaling

Table 4. Summary of Associations of Virulence Factors and Mortality, Sex, Necrotizing Enterocolitis, and Severity of Symptoms

Virulence Factor	Deceased (n = 11)	Survived (n = 20)	OR	P
Mortality ^a				
<i>clb</i> (colibactin, genotoxin)	9 (82)	1 (5)	67	< .005
<i>sfm</i> (Sfm, γ_1 -fimbriae)	1 (9)	15 (75)	0.04	< .005
<i>asfA</i> (arylsulfatase like gene)	10 (91)	5 (25)	26	< .005
<i>yde</i> (F9, γ_1 -fimbriae)	11 (100)	6 (30)	21	< .005
<i>mcm</i> (microcin M, class IIb microcin)	6 (55)	1 (5)	20	< .005
<i>iro</i> (salmochelin, siderophore)	9 (82)	5 (25)	12	< .005
<i>hma</i> (Hma, iron acquisition)	8 (73)	4 (20)	10	< .005
<i>tos</i> (hemolysin TosA, RTX-toxin)	4 (36)	0 (0)	10	.01
<i>set</i> (Set, enterotoxin)	4 (36)	0 (0)	10	.01
<i>cnf</i> (cytotoxic necrotizing factor)	5 (45)	1 (5)	14	.01
<i>mch</i> (microcin H47, class IIb microcin)	5 (45)	1 (5)	14	.01
<i>ompT</i> (OmpT, outer membrane protease VII)	11 (100)	13 (65)	5	.03
<i>tcpC</i> (TcpC, Toll-like receptor inhibitory homolog)	4 (36)	1 (5)	10	.04
<i>ompA</i> (OmpA, outer membrane porin A)	10 (91)	10 (50)	9	.04
Gestational age	Delivered before gestational wk 28 (n = 14)	Delivered in gestational age 28 to 42 (n = 18)		
<i>cnf</i> (cytotoxic necrotizing factor)	6 (42)	0 (0)	18	< .005
<i>yde</i> (F9, γ_1 -fimbriae)	14 (100)	11 (61)	20	< .005
<i>clb</i> (colibactin, genotoxin)	8 (57)	2 (11)	10	< .005
<i>hma</i> (Hma, iron acquisition)	9 (64)	3 (17)	13	< .005
<i>sfm</i> (Sfm, γ_1 -fimbriae)	3 (21)	13 (72)	0.11	.01
<i>asfA</i> (arylsulfatase like gene)	11 (79)	5 (28)	9	.01
<i>iro</i> (salmochelin, siderophore)	10 (71)	5 (28)	6	.03
<i>ag43</i> (self-recognizing antigen 43)	1 (7)	8 (44)	0.21	.04
Symptoms ^a	Shock (n = 7)	Impaired/no symptoms (n = 24)		
<i>clb</i> (colibactin, genotoxin)	6 (86)	4 (17)	26	< .005
<i>hma</i> (Hma, iron acquisition)	6 (86)	6 (25)	16	.007
<i>cnf</i> (cytotoxic necrotizing factor)	4 (57)	2 (8)	13	.01
<i>hyl</i> (hemolysin Ehx, RTX-toxin)	5 (71)	5 (21)	7	.02
<i>ireA</i> (IreA, iron acquisition)	4 (57)	3 (13)	8	.03
<i>sfm</i> (Sfm, γ_1 -fimbriae)	1 (14)	15 (63)	0.11	.04
<i>asfA</i> (arylsulfatase like gene)	6 (86)	9 (38)	9	.04
Sex	Female (n = 10)	Male (n = 22)		
Group II capsules (K1 and K5)	8 (80)	5 (23)	12	.005
<i>clb</i> (colibactin, genotoxin)	6 (60)	4 (20)	6	.04
NEC ^a	NEC (n = 10)	No NEC (n = 21)		
<i>tcpC</i> (TcpC, Toll-like receptor inhibitory homolog)	4 (40)	1 (5)	12	.03
<i>clb</i> (colibactin, genotoxin)	6 (60)	4 (19)	6	.04
Bilirubin ^a	VF present	VF not present	R ²	P
<i>ompA</i> (OmpA, outer membrane porin A)	58 (0–205), n = 19	149 (61–235), n = 8	0.3	< .005
C-reactive protein peak, mg/L ^a	VF present	VF not present	R ²	P
<i>ompT</i> (OmpT, outer membrane protease VII)	71 (5–291), n = 21	171 (31–264), n = 7	0.2	< .005
Hemoglobin, g/L ^a	VF present	VF not present	R ²	P
<i>tcpC</i> (TcpC, Toll-like receptor inhibitory homolog)	110 (84–143), n = 5	142.5 (98–203), n = 22		.007

OR and probabilities (P) were calculated using Fisher exact test.

Abbreviations: NEC, necrotizing enterocolitis; OR, odds ratio; VF, virulence factor.

^aDue to missing data adjusted numerator.

pathway through sex stratified analysis [19]. Moreover, placentas from male fetuses exhibited more chronic inflammation, potentially offering female fetuses greater survival chances

due to cardiovascular stability and lower levels of cytokines [20, 21]. Hormonal influences on the immune system have been documented, with testosterone demonstrating an

essentially immunosuppressive effect and estrogens exhibiting an immune amplifying effect [22].

Blyton et al investigated the distribution of fecal *E. coli* phylogeny in relation to sex and found no significant influence of sex on the phylogenetic lineage of colonizing *E. coli* [23]. However, the same study did note that female patients, who predominantly harbored *E. coli* B2, carried fewer strains of *E. coli* A and B1, suggesting a potential advantage for *E. coli* B2 colonization in female patients. Consideration of sex-related differences in immune response to infections prompts speculation on whether the higher prevalence of *E. coli* B2 in female neonates found in the present study reflects sex differences in establishing bacteremia or instead indicates differences in symptom manifestations of bacteremia. The latter might introduce a selection bias, potentially resulting in a reduced frequency of blood culturing in female patients.

Unlike non-B2 *E. coli*, group B2 *E. coli* demonstrate significant adaptation to the human intestine. Many virulence factors prevalent in this phylogroup confer a colonization advantage, contributing to prolonged persistence [24]. Colibactin, linked genetically to phylogroup B2 through the pathogenic island, induces DNA damage by alkylating DNA, resulting in interstrand cross-links that promote mutagenesis and cell death [25, 26]. Infants acquire colibactin-carrying *E. coli* isolates during delivery from their mother, aiding in persistent colonization in the neonatal gut [24, 27, 28]. Similarly, factors associated with bacterial iron acquisition enhance both bacterial virulence and colonization properties, predominantly linked to group B2 [29].

In the context of evolution where bacterial populations strive for persistence, the properties enhancing colonization fitness and transmission to new hosts seemingly conflict with virulence traits that impede transmission by killing the host. Le Gall et al have suggested, based on phylogenetic evidence, that virulence in the B2 group might be a fortuitous side effect of successful commensalism [30]. Extraintestinal infections primarily involve an array of virulence factors; however, no single factor deems *E. coli* an obligate pathogen in extraintestinal infections. Notably, several virulence factors correlated with mortality in our study. Intriguingly, our findings indicate a strong association between colibactin-positive *E. coli* bacteremia and severe outcomes, including shock and mortality in both female and male patients. Conversely, the absence of colibactin was associated with survival. This prompts speculation about whether *E. coli* B2 harboring colibactin in the bloodstream might represent an obligate pathogen.

The sample size of this study is a clear limitation; however, we believe that the detailed examination of clinical features yielded some new insights, particularly concerning sex differences, mortality, and symptoms. Several variables in the study are interconnected; for instance, very premature age is linked to mortality from conditions other than infection, warranting a multifactorial analysis to unravel these associations. While the retrospective retrieval of patient data was

incomplete for some individuals, prospective studies on infrequent infections demand an extended time horizon and poses logistical challenges. Sequence analysis is limited to showing the presence of genes and provides no information about the expression and effects of virulence factors. Nonetheless, the study underscores the diverse nature of neonatal *E. coli* septicemia: comprehensive analysis that integrates whole-genome sequencing of bacteria with corresponding extended patient profiles is imperative to unveil associations. Additionally, the study offers a comprehensive repository of extraintestinal virulence factors, potentially valuable for further investigations.

This study highlights the correlation between neonatal *E. coli* bacteremia caused by phylogroup B2, as well as isolates harboring certain virulence factors associated with adverse outcomes such as mortality, shock, and extreme prematurity. Notably, female patients were predominantly affected by group B2 infections, whereas male patients had bacteremia caused by both group B2 and non-B2 *E. coli*. Furthermore, the severity of *E. coli* bacteremia was notably higher in female patients, correlating with increased mortality rates. It is intriguing to note that despite recent recognition of significant sex-based differences in immune response and infectious diseases, these factors are not yet considered in clinical practice.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (<http://jid.oxfordjournals.org/>). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Author contributions. A. H. and S. S. contributed conceptualization and methodology. A. H., Å. M., E. N., E. T., and S. S. performed investigation, analysis, and project administration. A. H., Å. M., E. N., K. G., E. T., and S. S. wrote the article.

Data availability. All sequence data are publicly available through the European Nucleotide Archive (www.ebi.ac.uk/ena) under accession number PRJEB62832.

Financial support. This work was supported by the Centre for Research and Development, Region Gävleborg, Sweden (grant numbers CFUG-698631 to A. H. and CFUG-978549 to K. G.); the Region Uppsala, Sweden (grant number 810901 to S. S.); and the Swedish Research Council (grant number 2019-05909 to S. S.). Generation of sequence data was supported by Afa Insurance, Sweden (grant number 150411) and the Alf de Ruvo Memorial Foundation, Sweden. The National Genomics Infrastructure in Stockholm (funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation,

and the Swedish Research Council), and Swedish National Infrastructure for Computing/Uppsala Multidisciplinary Center for Advanced Computational Science supported assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

Potential conflicts of interest. All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Qazi SA, Stoll BJ. Neonatal sepsis: a major global public health challenge. *Pediatr Infect Dis J* **2009**; 28(1 Suppl):S1–2.
2. Stoll BJ, Puopolo KM, Hansen NI, et al. Early-onset neonatal sepsis 2015 to 2017, the rise of *Escherichia coli*, and the need for novel prevention strategies. *JAMA Pediatr* **2020**; 174:e200593.
3. Glaser MA, Hughes LM, Jnah A, Newberry D. Neonatal sepsis: a review of pathophysiology and current management strategies. *Adv Neonatal Care* **2021**; 21:49–60.
4. Wynn JL, Wong HR. Pathophysiology of neonatal sepsis. *Fetal Neonatal Physiol* **2017**:1536–52.e10.
5. Johnson J, Akinboyo IC, Schaffzin JK. Infection prevention in the neonatal intensive care unit. *Clin Perinatol* **2021**; 48:413–29.
6. Riley LW. Distinguishing pathovars from nonpathovars: *Escherichia coli*. *Microbiol Spectr* **2020**; 8 doi:10.1128/microbiolspec.ame-0014-2020.
7. Picard B, Garcia JS, Gouriou S, et al. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* **1999**; 67:546–53.
8. Heydecke A, Yin H, Tano E, Sütterlin S. Limitations in predicting reduced susceptibility to third generation cephalosporins in *Escherichia coli* based on whole genome sequence data. *PLoS One* **2023**; 18:e0295233.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* **1990**; 215:403–10.
10. Zhou Z, Alikhan NF, Sergeant MJ, et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res* **2018**; 28:1395–404.
11. Clermont O, Dixit OVA, Vangchhia B, et al. Characterization and rapid identification of phylogroup G in *Escherichia coli*, a lineage with high virulence and antibiotic resistance potential. *Environ Microbiol* **2019**; 21:3107–17.
12. Dale AP, Woodford N. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): disease, carriage and clones. *J Infect* **2015**; 71:615–26.
13. Weissman SJ, Hansen NI, Zaterka-Baxter K, Higgins RD, Stoll BJ. Emergence of antibiotic resistance-associated clones among *Escherichia coli* recovered from newborns with early-onset sepsis and meningitis in the United States, 2008–2009. *J Pediatric Infect Dis Soc* **2016**; 5:269–76.
14. Surgers L, Bleibtreu A, Burdet C, et al. *Escherichia coli* bacteraemia in pregnant women is life-threatening for foetuses. *Clin Microbiol Infect* **2014**; 20:O1035–41.
15. Wells JC. Natural selection and sex differences in morbidity and mortality in early life. *J Theor Biol* **2000**; 202:65–76.
16. Trivers RL, Willard DE. Natural selection of parental ability to vary the sex ratio of offspring. *Science* **1973**; 179:90–2.
17. Read JS, Troendle JF, Klebanoff MA. Infectious disease mortality among infants in the United States, 1983 through 1987. *Am J Public Health* **1997**; 87:192–8.
18. Rosenstock S, Katz J, Mullany LC, et al. Sex differences in neonatal mortality in Sarlahi, Nepal: the role of biology and environment. *J Epidemiol Community Health* **2013**; 67:986–91.
19. Ciesielski TH, Zhang X, Tacconelli A, et al. Late-onset neonatal sepsis: genetic differences by sex and involvement of the NOTCH pathway. *Pediatr Res* **2023**; 93:1085–95.
20. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol* **2016**; 16:626–38.
21. Goldenberg RL, Andrews WW, Faye-Petersen OM, Goepfert AR, Cliver SP, Hauth JC. The Alabama preterm birth study: intrauterine infection and placental histologic findings in preterm births of males and females less than 32 weeks. *Am J Obstet Gynecol* **2006**; 195:1533–7.
22. Fish EN. The X-files in immunity: sex-based differences predispose immune responses. *Nat Rev Immunol* **2008**; 8:737–44.
23. Blyton MDJ, Cornall SJ, Kennedy K, Colligon P, Gordon DM. Sex-dependent competitive dominance of phylogenetic group B2 *Escherichia coli* strains within human hosts. *Environ Microbiol Rep* **2014**; 6:605–10.
24. Nowrouzian FL, Wold AE, Adlerberth I. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J Infect Dis* **2005**; 191:1078–83.
25. Johnson JR, Johnston B, Kuskowski MA, Nougayrede JP, Oswald E. Molecular epidemiology and phylogenetic distribution of the *Escherichia coli* pks genomic island. *J Clin Microbiol* **2008**; 46:3906–11.
26. Wilson MR, Jiang Y, Villalta PW, et al. The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **2019**; 363:eaar7785.
27. Nowrouzian FL, Oswald E. *Escherichia coli* strains with the capacity for long-term persistence in the bowel microbiota carry the potentially genotoxic pks island. *Microb Pathog* **2012**; 53:180–2.
28. Tsunematsu Y, Hosomi K, Kunisawa J, et al. Mother-to-infant transmission of the carinogenic colibactin-producing bacteria. *BMC Microbiol* **2021**; 21:235.
29. Martin P, Marcq I, Magistro G, et al. Interplay between siderophores and colibactin genotoxin biosynthetic pathways in *Escherichia coli*. *PLoS Pathog* **2013**; 9:e1003437.
30. Le Gall T, Clermont O, Gouriou S, et al. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol Biol Evol* **2007**; 24:2373–84.