Microbiome

RESEARCH Open Access

Metagenomic signatures of extraintestinal bacterial infection in the febrile term infant gut microbiome

Anna L. DeVeaux[1], Carla Hall-Moore[2], Nurmohammad Shaikh[2], Meghan Wallace[3], Carey-Ann D. Burnham[2,3,4], David Schnadower[5,6], Nathan Kuppermann[7], Prashant Mahajan[8], Octavio Ramilo[9], Phillip I. Tarr[2,4], Gautam Dantas[1,2,3,4,10] and Drew J. Schwartz[2,4,11,12*]

Abstract

Background Extraintestinal bacterial infections (EBIs), e.g., urinary tract infection, bacteremia, and meningitis, occur in approximately 10% of febrile infants younger than 60 days. Although many EBI-causing species commonly reside in the infant gut, proof that the digestive system is a pre-infection habitat remains unestablished.

Results We studied a cohort of febrile term infants < 60 days old who presented to one of thirteen US emergency departments in the Pediatric Emergency Care Applied Research Network from 2016 to 2019. Forty EBI cases and 74 febrile controls matched for age, sex, and race without documented EBIs were selected for analysis. Shotgun sequencing was performed of the gut microbiome and of strains cultured from the gut and extraintestinal site(s) of EBI cases, including blood, urine, and/or cerebrospinal fluid. Using a combination of EBI isolate genomics and fecal metagenomics, we detected an intestinal strain presumptively isogenic to the EBI pathogen (> 99.999% average nucleotide identity) in 63% of infants with EBIs. Although there was no difference in gut microbiome diversity between cases and controls, we observed significantly increased *Escherichia coli* relative abundance in the gut microbiome of infants with EBIs caused by *E. coli*. Infants with *E. coli* infections who were colonized by the putatively isogenic pathogen strain had significantly higher *E. coli* phylogroup B2 abundance in their gut, and their microbiome was more likely to contain virulence factor loci associated with adherence, exotoxin production, and nutritional/metabolic function.

Conclusions The intestine plausibly serves as a reservoir for EBI pathogens in a subset of febrile term infants, prompting consideration of new opportunities for surveillance and EBI prevention among colonized, pre-symptomatic infants.

Background

Infants younger than 60 days are at considerable risk for extraintestinal bacterial infections (EBIs), which frequently occur in the urinary tract, bloodstream, and meninges [1]. Although early-onset infections (< 7 days of life) have decreased following the introduction of intrapartum antibiotic prophylaxis for *Streptococcus agalactiae* (Group B Streptococcus) [2], the incidence of late-onset infections (≥ 7 days) remains unchanged. EBIs continue to affect an estimated 10% of febrile infants [3–7]. To date, little is known about how to prevent late-onset infections in full-term infants, motivating a better understanding of the risk factors present in this population.

*Correspondence:
Drew J. Schwartz
schwartzd@wustl.edu
Full list of author information is available at the end of the article

EBIs can have substantial consequences in young infants, yet these infections remain challenging to diagnose. Fever and/or non-specific symptoms are often the only indicators of infection, and diagnosis requires a positive bacterial culture from a normally sterile site, which may take 24–48 h to identify, though rapid diagnostic tests have reduced this interval for bloodstream infections [8–10]. Because of the risk of clinical deterioration, infants are often admitted to the hospital and receive broad-spectrum antibiotics pending culture results [9]. Although a necessary practice to rapidly treat possible EBI, antibiotics risk perturbing an infant's developing microbiome and selecting for antimicrobial resistance [11–13]. Furthermore, antibiotic administration may precede obtaining sterile site cultures, complicating their interpretation and often leading to extended treatments for "partially treated" presumed EBI [9, 14]. Clinical and laboratory characteristics (e.g., fever, white blood cell count, urinalysis, absolute neutrophil count, and serum procalcitonin) can inform EBI risk [9, 15], but specific biomarkers of EBI are lacking. Thus, improved detection of EBIs is needed to commence targeted, timely antibiotic therapy and reduce the burden of EBIs in vulnerable infants.

Bacterial species that commonly cause EBIs in infancy, including Enterobacterales (e.g., *Escherichia coli* and *Klebsiella pneumoniae*) and *Enterococcus* spp. [3, 5–7], often reside in the developing gut microbiome [16–18]. In adults and children, the gut microbiome can source urinary tract infections (UTIs) via the fecal-perineal-urethral route [19, 20] (ascending infection). Bacteremia is believed to arise during infancy via bacterial translocation from the gut or skin, especially in preterm infants with immature mucosal barriers [21], or from the urinary system [22], while meningitis may arise from bacteremia [23]. Thus, the gut microbiome may serve as a direct or indirect habitat for pathogens that cause EBIs in infants, and gut colonization may be an important precursor of community-acquired EBIs in otherwise healthy infants. However, the extent of gut colonization by the EBI-causing strain, along with microbial community correlates of EBI risk, has not been well characterized in febrile term infants < 60 days, who are in the most vulnerable age group [24]. Identifying both gut bacterial community- and strain-level attributes of EBIs using high genomic resolution, including alterations in the developing microbiome, may improve our understanding of EBI risk during early life.

Here, we sought to investigate the prevalence of gut pathogen colonization among infants with EBIs and define the correlates of invasion using isolate genomics and fecal metagenomics. Our data demonstrate that gut carriage of the EBI-causing strain is common in febrile term infants < 60 days old and can be associated with differences in virulence factor profiles, supporting a role for the intestine as a reservoir of EBI-causing organisms in term infants.

Methods

Participant selection

This study included infants with EBI and matched controls without EBI to investigate (1) whether the EBI-causing strain was present in the gut of cases based on genomic analysis of cultured isolates from cases and (2) whether EBI-causing species were enriched in the gut of cases compared to controls. Samples were obtained from a completed, prospective observational study of febrile infants presenting to the emergency department (ED) of 22 centers across the US in the Pediatric Emergency Care Applied Research Network (PECARN; Fig. S1). Details of the network have been published [15, 25, 26]. Thirteen centers provided samples for this study. Fever was defined as a rectal temperature ≥ 38 °C in the ED, in a prior healthcare setting, or at home as reported by the family in the preceding 24 h. Infants were excluded if they appeared critically ill, had received antibiotics in the preceding 48 h, were born preterm, or had major comorbidities as previously described [27] (e.g., serious congenital abnormalities, inborn errors of metabolism), indwelling devices, or soft-tissue infections. Patients were not excluded for otitis media. Clinical care was at the discretion of the treating clinician. Bacteremia and meningitis were defined as infants whose culture of blood and/or cerebrospinal fluid (CSF), respectively, yielded a single species of a known pathogen, while UTI was defined as growth of a pathogen above a certain threshold as previously described [25, 28]. Febrile infants without document EBIs were selected for use as controls; these infants were culture negative for known pathogens. Each EBI case was matched to up to 2 febrile controls (median: 2) by age (± 7 days of life, using the closest match available), sex (same sex when available ± 7 days of life), and race (data file S1). Infants whose culture grew a suspected contaminant (i.e., coagulase-negative Staphylococci or *Streptococcus* spp. not including *S. agalactiae* or *S. pneumoniae*) were excluded as cases but eligible as controls. All participants provided an enteric specimen (stool and/or rectal swab) in the ED, which was subsequently stored at − 80 °C without further processing until analysis in St. Louis, MO, USA.

Selective culturing

Extraintestinal bacterial isolates from EBI cases and enteric specimens from all cases and controls were sent to Washington University School of Medicine (St. Louis, MO, USA). Enteric specimens from EBI cases were

selectively cultured for the extraintestinal pathogen species (Fig. S2). Less than 5% of the enteric specimen was removed without thawing by scraping frozen transport media or stool followed by resuscitation in brain heart infusion (BHI) broth for 1 h and incubation in selective broth until turbid and then plated on selective and differential agar. Samples were grown under the following conditions according to the identity of the organism previously isolated from extraintestinal culture (and as reported by the sending institution): *Enterococcus faecalis*, Enterococcosel broth with colistin (10 mg/L), amphotericin B (2 mg/L), and nalidixic acid (15 mg/L) followed by Enterococcosel agar (Becton Dickinson & Company); *Klebsiella pneumoniae*, MacConkey broth with vancomycin (20 mg/L) followed by *Klebsiella* ChromoSelect agar (Millipore 90925); *Escherichia coli* or *Enterobacter cloacae*, MacConkey broth with vancomycin (20 mg/L) followed by MacConkey agar with vancomycin; *Proteus mirabilis*, MacConkey broth with vancomycin followed by MacConkey agar with a 10 µg colistin disc; and *S. agalactiae*, tryptic soy broth (TSB) with added NaCl (6.5% w/v), aztreonam (8 mg/L) and amphotericin B (2 mg/L) followed by tryptic soy agar (TSA) and blood agar plates (BAP). Species identity was confirmed with matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) and one colony matching the target species per intestinal sample was profiled by multi-locus sequence typing (MLST) using PCR amplification and sequencing of housekeeping genes, which were matched to known sequence types (STs) using the PubMLST.org database (data file S1) [29]. Potential matches were defined as extraintestinal and intestinal isolates of the same species and ST, except for *Proteus mirabilis* from one infant, for which neither isolate had a known ST at the time of screening; these isolates were later in silico typed (see below section). Isolates matching the extraintestinal species were retained and stored in glycerol stocks from pure culture at − 80 °C until sequencing. Antibiotic susceptibility testing was performed on Enterobacterales and *Enterococcus* spp. as previously described [30].

Isolate sequencing

Single isolates were recovered with the following media for subsequent whole-genome sequencing (WGS): *Enterococcosel* agar plates followed by BHI broth for *E. faecalis*, MacConkey agar followed by TSB for *Enterobacteriaceae* spp., and TSA followed by TSB for *S. agalactiae*. Genomic DNA was extracted from 5 mL overnight cultures using the QIAamp BiOstic Bacteremia DNA kit (QIAGEN). Short-read sequencing libraries were prepared using a Nextera DNA Library Prep kit (Illumina) with modifications [31] and sequenced at 2 × 150 bp on either the Illumina NextSeq or NovaSeq platforms to a target depth of 50 × coverage based on the expected genome size for each species. Reads were demultiplexed by index pair and quality-trimmed with Trimmomatic [32] (v0.39; SLIDINGWINDOW: 4:20, LEADING: 10, TRAILING: 10, MINLEN:60) followed by decontamination of human genomic reads using DeconSeq (v4.3, -dbs hsref38) and read repair with BBMap (v38.63). Oversequenced samples (> 150X coverage) were subsampled to 150 × coverage using SeqTK (v1.3). SPAdes (v3.15.3) was used to assemble genomes with defined parameters (-k 21,33,55,77,87,97,107,117,127 –careful) and short (< 1000 bp) contigs were subsequently removed. Genome assemblies were assessed via Quast (v5.2.0) and CheckM (v1.2.1) and species were identified using MASH (v2.3). Assemblies were confirmed to have < 500 contigs after filtering, > 90% completeness, and < 5% contamination.

Isolate genome analysis

Sequence types were identified in silico with mlst [33] (v2.22.1) (https://github.com/tseemann/mlst), specifying –scheme ecoli_achtman_4 for *E. coli* genomes. *Proteus mirabilis* isolates were typed by uploading their genome assemblies to PubMLST.org [29]. Whole-genome single-nucleotide polymorphism (SNP) quantification was performed in an all-versus-all manner for the isolates within each species using snippy [34] (v4.6.0). The more complete genome (based on CheckM output) for each pairwise alignment was selected as the reference genome, with the following tiebreaker criteria: (1) higher N50, (2) lower number of contigs, and (3) higher N90 (for two pairs, all four of these metrics were the same, so a reference genome was randomly chosen). All genomes were annotated using bakta [35] (v1.5.1) and antimicrobial resistance genes (ARGs) were profiled with AMRFinder (v3.10.42).

Metagenomic sequencing

One enteric specimen from each infant was selected for sequencing, prioritizing stool over swab samples when available; however, a pilot sequencing run indicated that community composition was similar for paired stool and rectal swab samples from a subset of infants ($p = 0.001$, PROTEST; Fig. S3). Frozen samples were chipped with a target quantity of 100 mg of stool (30–50 mg when quantity was limited) and 250 mg of transport media for rectal swabs. Metagenomic DNA was extracted using the DNeasy PowerSoil Pro kit (QIAGEN) with a modified homogenization step of 2 min at 2500 rpm on a bead beater followed by 5 min on ice and an additional 2 min at 2500 rpm. Libraries were prepared as for WGS and sequenced to a target depth of 5 M reads. Short reads were processed as for WGS.

Microbiome profiling

Preliminary analysis was carried out on a subset of samples using MetaPhlAn2 to determine the appropriate sequencing depth. Samples sequenced to > 10 M reads were subsampled to depths of 10 M, 5 M, 3 M, and 2 M reads using seqtk, and 2 M reads was used as a cutoff based on an insignificant difference of species α-diversity (richness and Shannon diversity) between 2 M reads and the original read set. MetaPhlAn4 [36] (v4.0.2) was used to profile the community taxonomic composition (mpa_vJan21_CHOCOPhlAnSGB_202103 database). Community diversity metrics were profiled with phyloseq (v1.41.1). MaAsLin2 [37] (v1.8.0) was used to investigate taxonomic associations with infection status using the formula Taxa ~ Group + (1 | SampleType) with default parameters. Groups included all cases or stratified by the causative species (of those with at least three cases) compared to all controls. To determine if intestinal *S. agalactiae* was enriched in *S. agalactiae* UTI, –min_prevalence was lowered to 1% [30]. Respiration mode was initially profiled at the species level using bacdive (https://bacdive.dsmz.de/); unassigned species or those with conflicting bacdive classifications (e.g., both "anaerobe" and "aerobe") were subjected to further profiling at the species or a higher taxonomic level when applicable (e.g., all *Lachnospiraceae* are classified as anaerobes) using another source [38] as has previously been done [39]. Species without experimental evidence of respiration mode and those ≤ 1% abundance were filtered out. Obligate aerobes and facultative organisms were grouped as "aerobes," and obligate anaerobes and aerotolerant organisms were classified as "anaerobes." shortBRED [40] (v.0.9.4) was used to profile virulence factors (VFs) in metagenomes. A marker set was constructed from the VFDB 2022 [41] core protein sequences and complete UniRef90 [42] database (both downloaded in May 2024) using shortbred_identify.py (–clustid 0.95). VF-encoding loci were manually assigned and filtered to include those encoded by *Enterobacteriaceae* based on their species attribution. VF abundance was profiled using shortbred_quantify.py with default settings. MaAsLin2 [37] was used to identify VF associations with EBI status using the formula Pathways ~ Group + (1 | SampleType) with normalization set to "NONE."

Strain-level analysis

BWA (v0.7.17) was used to index isolate genomes and align metagenomic reads in a pairwise all-versus-all manner. Alignments were provided to inStrain [43] (v1.6.3) as a .sam file along with the genome.fasta files, and inStrain profile was used with default parameters to determine strain presence and microdiversity metrics. If multiple extraintestinal isolates were available for an EBI case and were isogenic (> 99.999% ANI), the alignment to the higher quality assembly, as assessed above for SNP analyses, was used. For ease of visualization, population ANI (popANI), single nucleotide substitution (SNS) rate, and breadth were converted to log scale with -log10(1-metric); values of 1 were transformed to infinite values, so they were changed to a value of 7. Strain presence was defined as breadth ≥ 0.5 and popANI ≥ 99.999% [43]. Normalized coverage was calculated as (coverage/metagenomic read depth) * 10^6. StrainGST from the StrainGE toolkit [44] was used to metagenomically profile *E. coli* strains. All 2,835 complete *E. coli* genomes on NCBI RefSeq on August 8th, 2023, were downloaded using ncbi-genome-download [45] (v0.3.1) and chromosomes were retained. Reference genomes were *k*-merized with the default $k = 23$ and subsequently clustered into 946 groups using a minimum Jaccard similarity of 0.90, with *k*-mer profiles from one representative genome from each group selected to form the database. Metagenomic read sets were *k*-merized and closely matching strains from the database were identified. Phylogroups were assigned to each reference and isolate genome with ClermonTyping [46] and STs were assigned to references using the mlst software as above. The predicted relative abundance (rapct) was summed for each phylogroup per sample; non-zero values were converted to 1 to determine prevalence. Matching *E. coli* EBI cases were defined as those with the same strain in their gut and cultured from extraintestinal site(s) based on isolate genomics and/or inStrain. Reference strains were annotated as described for isolate genomes, and a core gene alignment was produced via panaroo [47] (v1.2.10) using –core_threshold 0.98. An approximately-maximum-likelihood phylogenetic tree was generated with FastTree [48] (v2.1.10) and visualized using ggtreeExtra [49] (v1.4.2). FastANI [50] was used to assess relatedness between EBI-causing *E. coli* strains colonizing the gut and reference genomes called by strainGE that had discordant in silico MLST calls. The reference was presumed to represent the EBI-causing strain when it shared ≥ 99.5% pairwise whole genome ANI (average of the two measurements in which either genome was used as the query), a reported threshold for STs [51].

Statistical analysis

We performed statistical analyses in R Studio (R 4.1.0, RStudio 2024.04.1 + 748). We tested numeric distributions for normality using the Shapiro–Wilk test. Non-parametric distributions were compared using a two-tailed Wilcoxon rank-sum test. A PERMANOVA test (adonis2, vegan v2.6.4) was used to compare Bray–Curtis dissimilarity measures between cases and controls. Fisher's exact tests were used to compare the

distribution of categorical data between groups. The Benjamini-Hochberg (BH) correction was used to adjust p values for multiple hypotheses when applicable. Spearman correlations were used to determine the linear correlation between two variables (R) for one group. Linear regression using the lm() function in R was used to compare the relationship between continuous variables when multiple groups were present. $p < 0.05$ (BH-corrected when applicable) was considered statistically significant. For MaAsLin2 analyses, the default $q < 0.25$ was considered statistically significant except for the virulence factor analysis, in which $q < 0.05$ was used to reduce the number of observations for ease of visualization.

Results

The gut contains putatively identical bacteria to those isolated from extraintestinal sites in febrile infants < 60 days of life

Forty infants < 60 days of life with EBIs (cases) from whom EBI isolates were available were selected, with a median age of 31 days of life (Fig. S1, Table S1). UTIs without concomitant bacteremia or meningitis were the most common infection and were identified in 70% of infants with culture-proven EBI (28/40; Table 1). The remaining 12 cases had concomitant bacteremia and UTI ($n = 7$), bacteremia and meningitis ($n = 1$), bacteremia alone ($n = 3$), or meningitis alone ($n = 1$). Most EBIs were caused by *E. coli* (26/40), followed by *E. faecalis* (6/40), *S. agalactiae* (3/40), *K. pneumoniae* (2/40), *E. cloacae* (1/40), *P. mirabilis* (1/40), or *E. coli* and *S. agalactiae* (1/40), similar to previous reports of bacterial pathogens isolated from normally sterile sites in this population [3, 5–7] (Table 1).

To investigate if the EBI-causing strain was present in the gut at the time of infection, we first screened intestinal samples by selectively culturing for the EBI-causing species (Fig. S2). Across all infants, intestinal samples were collected a median of 1.07 h before extraintestinal samples (Table S1, data file S1). We found at least one swab/fecal isolate to be concordant with the EBI-causing strain in 60% (24/40) of cases using MLST (Fig. 1A, Table 1). Further, 83% (5/6) of infants with UTIs caused by *E. faecalis* carried the presumptively pathogenic strain in their gut, consistent with an intestinal source of these infections. Of the infants with infections caused by *E. coli*, which is commonly found in the early-life gut microbiome [52], 56% (15/27) had a strain presumptively matching the pathogen cultured from their intestinal sample. The remaining infants for whom we observed gut pathogen colonization based on MLST had EBIs caused by *P. mirabilis* (1/1), *K. pneumoniae* (1/2), or *S. agalactiae* (2/4). These findings are consistent with the hypothesis that the gut is a reservoir for pathogenic bacteria in young infants [30] based on shared phylogeny of strains.

To strengthen confidence in the association between intestinal and extraintestinal isolates as inferred by species and MLST, we performed whole-genome sequencing of all isolates and assessed pairwise genomic distances between isolates of the same species. To set a benchmark for genomic relatedness, we included intestinal isolates not matching their cognate extraintestinal strain(s) by MLST and compared strains of the same species between infants. We found that a maximum of 12 whole genome

Table 1 Distribution of EBI type by causative organism. Number of infants with EBIs based on extraintestinal site affected and causative species. Gut colonization is defined as detection via isolate genomics (culturing) and/or metagenomics. [a]one case was detected via metagenomics only

Urine	Blood	CSF	n	Gut colonization (%)	Culture-positive enteric samples	Sample detected in enteric metagenome
E. coli	–	–	15	10 (**67**)	10	10
E. coli *S. agalactiae*	–	–	1	0 (**0**)	0	0
E. faecalis	–	–	6	5 (**83**)	5	2
E. cloacae	–	–	1	0 (**0**)	0	0
K. pneumoniae	–	–	1	0 (**0**)	0	0
P. mirabilis	–	–	1	1 (**100**)	1	1
S. agalactiae	–	–	3	2 (**67**)	2	1
E. coli	*E. coli*	–	7	4[a] (**57**)	3	3
–	*E. coli*	–	3	2 (**67**)	2	0
–	*E. coli*	*E. coli*	1	0 (**0**)	0	0
–	–	*K. pneumoniae*	1	1 (**100**)	1	1
Total			40	25 (**63**)	24	18

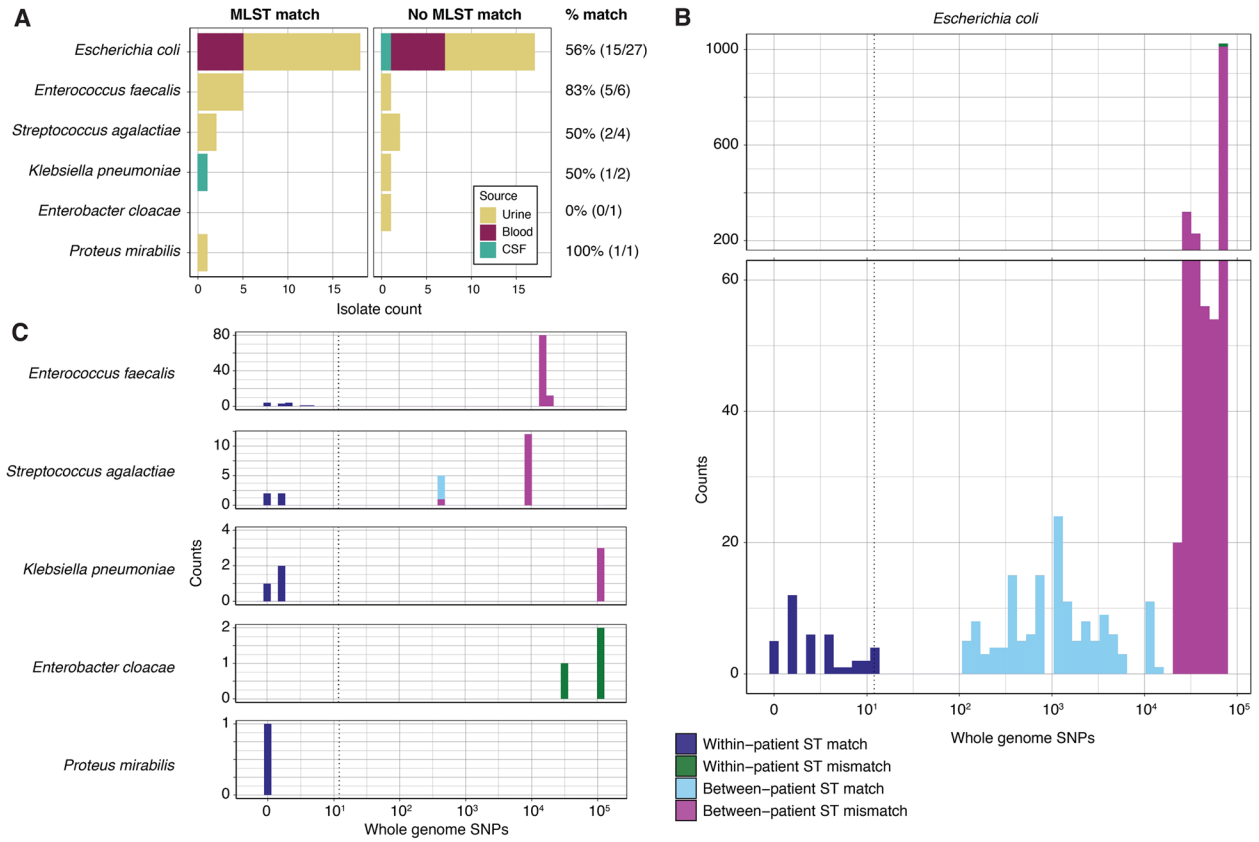

Fig. 1 High genomic relatedness between cognate extraintestinal and intestinal isolates. **A** Bar plot depicting the number of extraintestinal isolates that had a strain of the same species and ST cultured from the gut (MLST match) or not (no MLST match) by species and isolate source. **B**, **C** Histograms of whole-genome SNP counts computed through pairwise comparisons within (**B**) *Escherichia coli* or (**C**) other pathogen species. Colors indicate whether the SNPs represent comparisons of samples from the same or different infants and whether the STs matched. The dotted line at 12 SNPs indicates the upper limit observed for within-patient ST matches. CSF cerebrospinal fluid, ST sequence type, SNP single-nucleotide polymorphism

SNPs separated within-infant ST matches (median 1 SNP; Fig. 1B, C, data file S2). This threshold corresponds to an average nucleotide identity (ANI) >99.999%, indicating high genomic relatedness, i.e., isogenicity [43]. In comparison, the lower limit of SNPs observed between isolates from different infants was 114 for two *E. coli* ST404 isolates. Based on a stringent 12 SNP threshold, we confirmed that isolates of the same ST cultured from the gut and site of EBI were nearly identical by isolate genomics. However, given that only one colony was screened by MLST per intestinal sample, these findings may underestimate the true rate of gut pathogen colonization in young infants with EBI.

We next characterized the antimicrobial resistance profiles of strains isolated from sterile sites in our EBI cases. Specifically, we were interested in determining susceptibility against agents commonly used for empiric management in febrile term infants (i.e., ampicillin plus either gentamicin or a third-generation cephalosporin, such as ceftriaxone or ceftazidime [9]). 61% (19/31) of extraintestinal Gram-negative strains were resistant to ampicillin, while 13% (4/31) were resistant to gentamicin and 6% (2/31) to at least one third-generation cephalosporin (data file S2). Inferring susceptibility to ampicillin for *S. agalactiae* [53], across all EBI-causing strains, 90% (37/41) were susceptible to ampicillin and/or gentamicin, while 98% (40/41) were susceptible to ampicillin and/or a third-generation cephalosporin, mirroring results from similar cohorts [54, 55]. Through genome-wide annotation of antimicrobial resistance genes (ARGs), we also identified extended-spectrum β-lactamase (ESBL)-encoding *E. coli* strains colonizing the gut and causing UTIs in two infants: an ST69 strain encoding $bla_{\text{CTX-M-14}}$, which had intermediate resistance to ceftriaxone, and an ST131 strain encoding $bla_{\text{CTX-M-15}}$, consistent with resistance to ceftazidime, ceftriaxone, and cefepime (data file S2). While both strains were susceptible to gentamicin, these concordances indicate gut colonization and EBI

caused by ESBL-producing *E. coli* that displayed resistance to third-generation cephalosporins commonly used in this population.

Gut metagenomic detection of high abundance EBI-causing strains

We next determined if the EBI-causing strains were prominent among the bacterial gut community in these colonized infants. To address this, we performed metagenomic sequencing of intestinal samples from EBI cases and age- and sex-matched febrile infants without EBI ($n=74$; Fig. S1, Table S1). As we have done previously [30], we performed metagenomic read mapping using inStrain [43] to resolve strains within the gut microbial community (data file S3). We identified > 99.999% popANI as being specific for within-infant alignments, which indicated that the same EBI-causing strain was detected in the gut metagenome of the cognate infant (Fig. 2A). This exceptionally stringent threshold identified the pathogenic strain in the gut of 75% (18/24) of infants with EBIs for whom the isogenic strain had already been cultured from the gut (Table 1). Additionally, we identified one infant with concomitant *E. coli* UTI and bacteremia as having the EBI-causing strain present in the gut metagenome despite no *E. coli* being previously isolated from the intestine. We repeated selective culturing of this sample but were unable to recover *E. coli*, suggesting that although the organism eluded recovery, its genomic DNA was sufficient to infer presence within the intestinal microbial community.

We observed that for infants with the EBI-causing strain detected in their gut metagenome, the fraction of the pathogen genome covered by metagenomic reads (breadth) exceeded that of infants whose EBI-causing strain was cultured from their gut but not detected metagenomically (median = 1.000 versus 0.696, $p=1.547e\text{-}4$, Wilcoxon rank-sum test). Thus, we reasoned that lower pathogen abundance in these communities may explain why these strains eluded metagenomic detection. Indeed, on average, the EBI-causing species was present at a lower relative abundance in infants whose pathogen strain was detected in the gut solely by culturing compared to those detected metagenomically (median 11.1% versus 0.209%, $p=1.547e\text{-}4$, Wilcoxon rank-sum test; Fig. 2B). In these seven infants, the EBI-causing species was present at < 1% relative abundance; in two of these cases, the EBI-causing species was present below the limit of metagenomic detection. While higher metagenomic sequencing depth may increase its sensitivity in these cases, four of these seven samples were already sequenced at > 10 million reads, suggesting that routine metagenomic detection is less practical for low abundance (< 1%) strains.

In sum, although isolating the target organism via culture was more sensitive than gut metagenomic sequencing for identifying the pathogen strain in the gut, in 72% (18/25) of cases, metagenomic DNA was sufficient

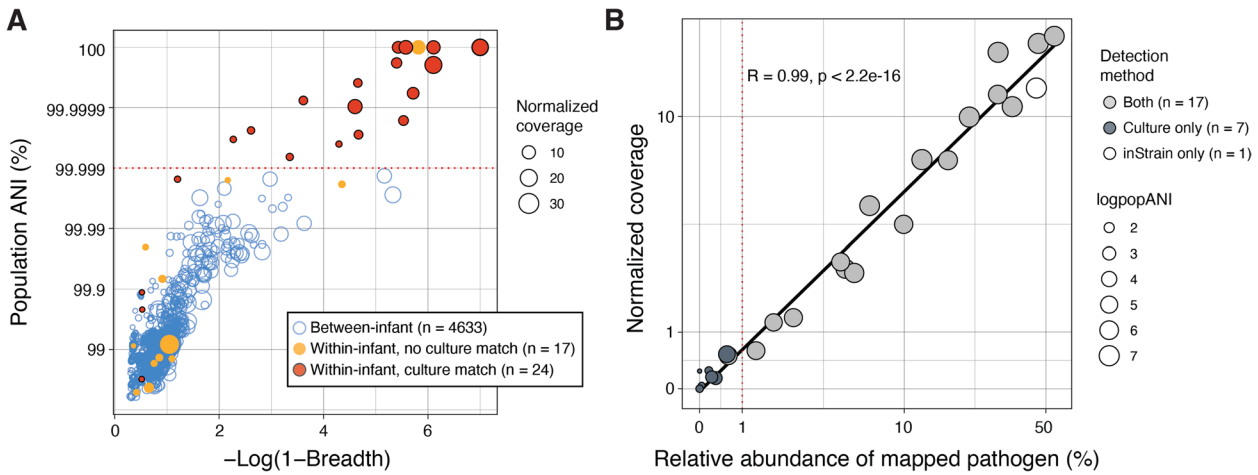

Fig. 2 Metagenomic read mapping delineates gut pathogen colonization at high abundance in EBI cases. **A** Genome breadth of coverage (filtered by ≥ 0.5 for visualization) and popANI of pairwise metagenomic read mapping to EBI pathogens. Each dot represents an individual alignment, and colors correspond to whether the metagenome mapped to a genome from the same infant and whether that pathogen had been cultured from the gut ("culture match"). Dot size is scaled to read depth-normalized coverage. Dashed line indicates the minimum popANI (99.999%) that was determined to be specific to within-infant alignments. **B** Relative abundance of the EBI-causing species compared to normalized coverage for within-infant mappings of metagenomic reads to pathogen genomes for strains determined to be present in the gut via culturing and/or inStrain (as indicated by color). Dashed line indicates the minimum relative abundance (1%) of the pathogen species that enabled metagenomic detection of the EBI-causing strain. Dot size is scaled to popANI. Pearson correlation coefficient (R) and *p* value are shown

to detect the presence of the EBI-causing organism. By combining both traditional microbial isolation and metagenomic sequencing approaches, we identified the EBI-causing strain in the gut of 63% (25/40) of cases in our cohort (Table 1). Metagenomic read mapping also suggested that, when present, the EBI-causing organism is often at high (≥1%) relative abundance, implicating microbiome community composition as a potential signature of EBI in febrile term infants, which we have previously observed with preterm infants [30].

Enrichment of pathogen species observed in the gut of *E. coli* EBI cases

To determine if a microbiome risk profile of EBI exists at the time of presentation, we compared the gut microbial community structure of EBI cases to those of controls. We observed that infants with EBI had on average higher serum procalcitonin ($p=0.005$) and white blood cell counts ($p=0.032$) than controls, consistent with prior reports [56, 57] and confirming systemic differences (Table S1). However, α-diversity of the gut microbiome, measured by richness and Shannon diversity at the species level, did not differ significantly between cases and

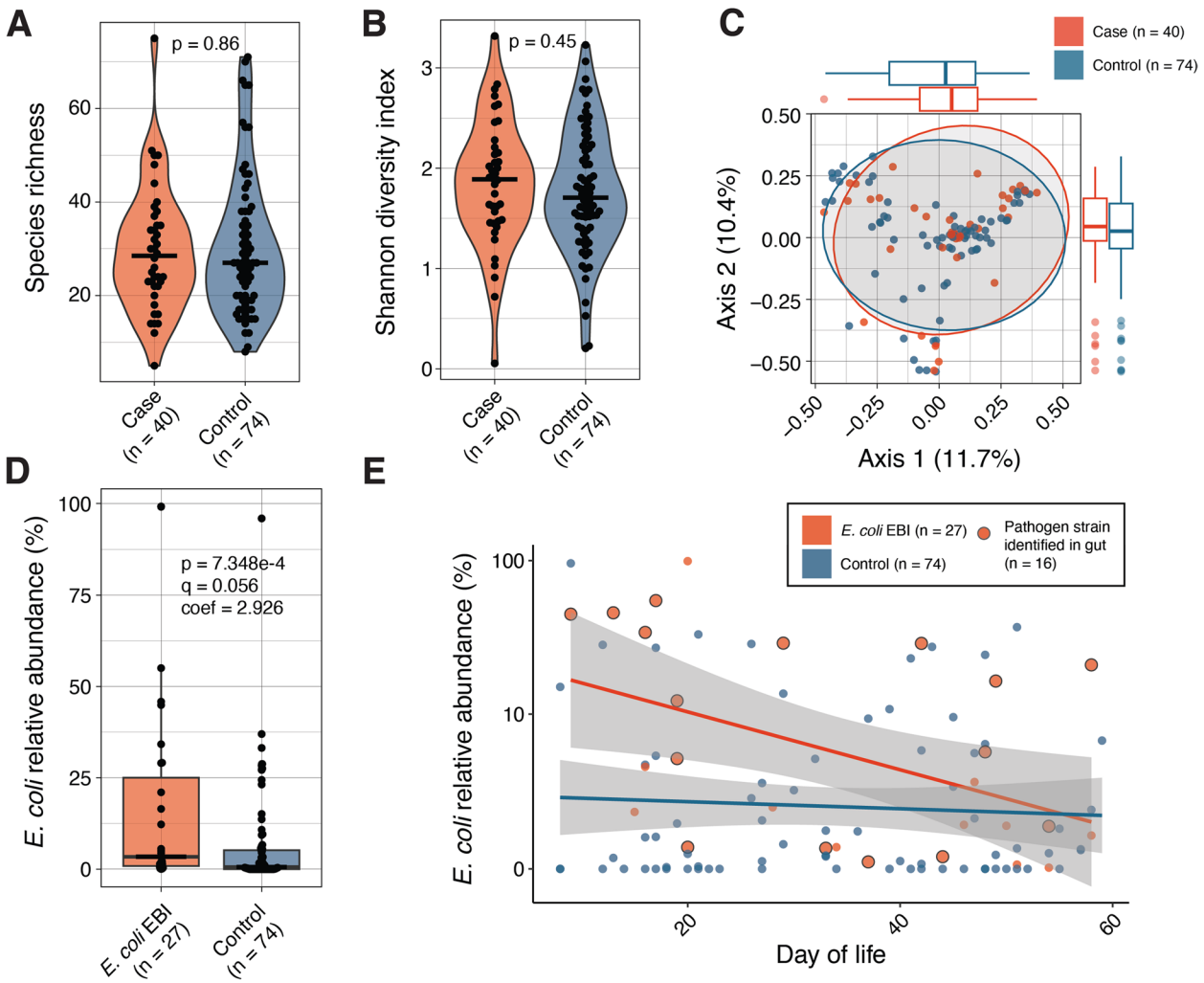

Fig. 3 Microbiome compositional differences are age- and infection-dependent for infants with *E. coli* EBI. α-diversity as measured by **A** species richness and **B** Shannon diversity from the gut metagenomes of all EBI cases compared to control infants. *P* values determined through a Wilcoxon rank-sum test. **C** Principal coordinate analysis (PCoA) plot of Bray–Curtis dissimilarity between samples computed at the species level. PCoA axes have the percent variation explained shown parenthetically. Marginal x- and y-axis boxplots correspond to PCoA axis 1 and 2 values, respectively. **D** Association between *E. coli* relative abundance and *E. coli* EBI status compared to control infants as determined through generalized linear mixed effect models. The BH-adjusted *p* value (q) and coefficient of effect (coef) are shown. **E** *E. coli* relative abundance plotted by day of life for infants with *E. coli* EBI compared to controls with linear regression lines plotted for each group. The shaded region represents the 95% confidence interval

controls (Fig. 3A, B). Similarly, we did not observe differences in β-diversity as measured by Bray–Curtis dissimilarity (PERMANOVA, $p=0.360$; Fig. 3C).

There were no overall taxonomic associations between EBI cases and controls using generalized linear mixed effect models, prompting us to stratify cases based on the species causing infection (for groups with ≥ 3 infants). Infants with EBIs caused by *E. coli* had greater relative abundance of *E. coli* in their intestine compared to control infants ($q=0.056$, coefficient = 2.926; Fig. 3D, Table S2, data file S3). Likewise, infants with *S. agalactiae* UTI had significantly higher abundance of the EBI-causing organism detected in their gut metagenome ($q=4.449e-4$, coef = 1.608; data file S3). However, this association was driven by the presence of *S. agalactiae* in only one of four *S. agalactiae* UTI cases (Table S2). Interestingly, we did not observe a significant enrichment of *E. faecalis* in the gut microbiome of infants with *E. faecalis* UTIs, despite the pathogenic strain having previously been cultured from the gut of 5/6 cases (Table 1). We also found that *E. coli* and *P. mirabilis* were more prevalent in the gut metagenome of infants with EBI caused by those organisms than controls ($p=2e-4$ and $p=0.04$, respectively; Table S2). Additionally, a relative abundance threshold of ≥ 1% *E. coli* was 74% sensitive and 57% specific for *E. coli* EBIs, highlighting its potential utility during clinical risk assessment. In sum, pathogen presence and enrichment in the gut was evident among certain EBI-causing organisms, suggesting that it may be related to infection risk in young infants.

Given that facultative bacteria, including *E. coli*, decline in relative abundance in the gut in the months after birth [16, 17], we postulated that there may be an age-related component to its enrichment in the gut of *E. coli* EBI cases. We compared the relationship of *E. coli* relative abundance with day of life for *E. coli* EBI cases and controls using linear regression. Indeed, there was a significant difference in this association between groups ($p=0.029$), with *E. coli* EBI cases exhibiting a greater age-related decrease in *E. coli* relative abundance than controls ($\beta = -0.673$ vs. $\beta = -0.155$, Fig. 3E). We predicted that control infants may have greater α-diversity than infants with *E. coli* EBIs, which may protect against a bloom in the EBI-causing organism, especially in the youngest infants, when the difference in *E. coli* relative abundance is most apparent. There was a greater age-related increase in Shannon diversity ($p=0.020$), but not species richness ($p=0.123$) for infants with *E. coli* EBIs relative to control infants (Fig. S4A, B), suggesting that microbiome complexity may contribute to susceptibility to *E. coli* infections in this age group. Because we did not find differences in specific taxa other than *E. coli* associated with *E. coli* EBI, we assessed if controls had increased obligate anaerobic taxa, which could contribute to the exclusion of facultative bacteria such as *E. coli*. As expected, across both *E. coli* EBI cases and controls, aerobes (including strictly aerobic and facultatively anaerobic organisms) were depleted over time ($R = -0.280$, $p=0.005$), while the abundance of obligate anaerobes increased ($R=0.260$, $p=0.009$; Fig. S4C, data file S4). However, there were no significant differences in these rates between *E. coli* EBI cases and controls ($p=0.981$ for aerobes, $p=0.533$ for anaerobes). Hence, while a bloom in *E. coli* abundance in the gut was associated with *E. coli* EBI in an age-related manner, this could not be explained by a depletion in anaerobic taxa, leading us to consider whether there were strain-specific features associated with EBI susceptibility.

Phylogroup B2 is overrepresented in infants carrying EBI-causing *E. coli*

Although *E. coli* is a common cause of EBI, most gut-resident *E. coli* do not cause EBI. Based on this reasoning, we hypothesized that strains found in the gut of *E. coli* EBI cases may differ genetically from commensal *E. coli* found in controls. To test this hypothesis, we used the Strain Genome Search Tool (StrainGST) [44] to identify close representatives to a database of publicly available *E. coli* reference genomes and estimate their relative abundance based on metagenomic reads. Fifty-nine distinct *E. coli* genomes from phylogroups A, B1, B2, C, D, and F represented *E. coli* strains present across the 49 controls and 26 *E. coli* EBI cases whose samples had detectable *E. coli* by StrainGST (Fig. 4A, data file S4). Phylogroup B2 was most prevalent at 59% (44/75), but there were no statistically significant differences in phylogroup prevalence between groups (Table S3). When comparing relative abundance, however, phylogroup B2 was significantly more abundant in the stools of *E. coli* EBI cases who had a strain matching the EBI-causing isolate detected within their gut ("matches") compared to those who did not ("non-matches") or control infants (BH-corrected $p=0.042$, Fig. 4B). We did not observe significant differences in abundance for other phylogroups (Fig. S5).

We next asked if the EBI-causing strain was over-represented among *E. coli* in infants colonized by multiple strains [52]. This disproportion could provide evidence that abundance plays a key role in seeding infection; alternatively, EBI-causing strains may be more adept colonizers in the gut, or more capable of systemic invasion or retrograde movement through the urethra. Multi-strain colonization (defined as ≥ 2 STs predicted to be present) was less common than single-strain gut colonization, occurring in 33% (4/16) of *E. coli* EBI matches, 50% (5/10) of non-matches, and 18% (9/49) of controls colonized by *E. coli*; these rates did not significantly differ between

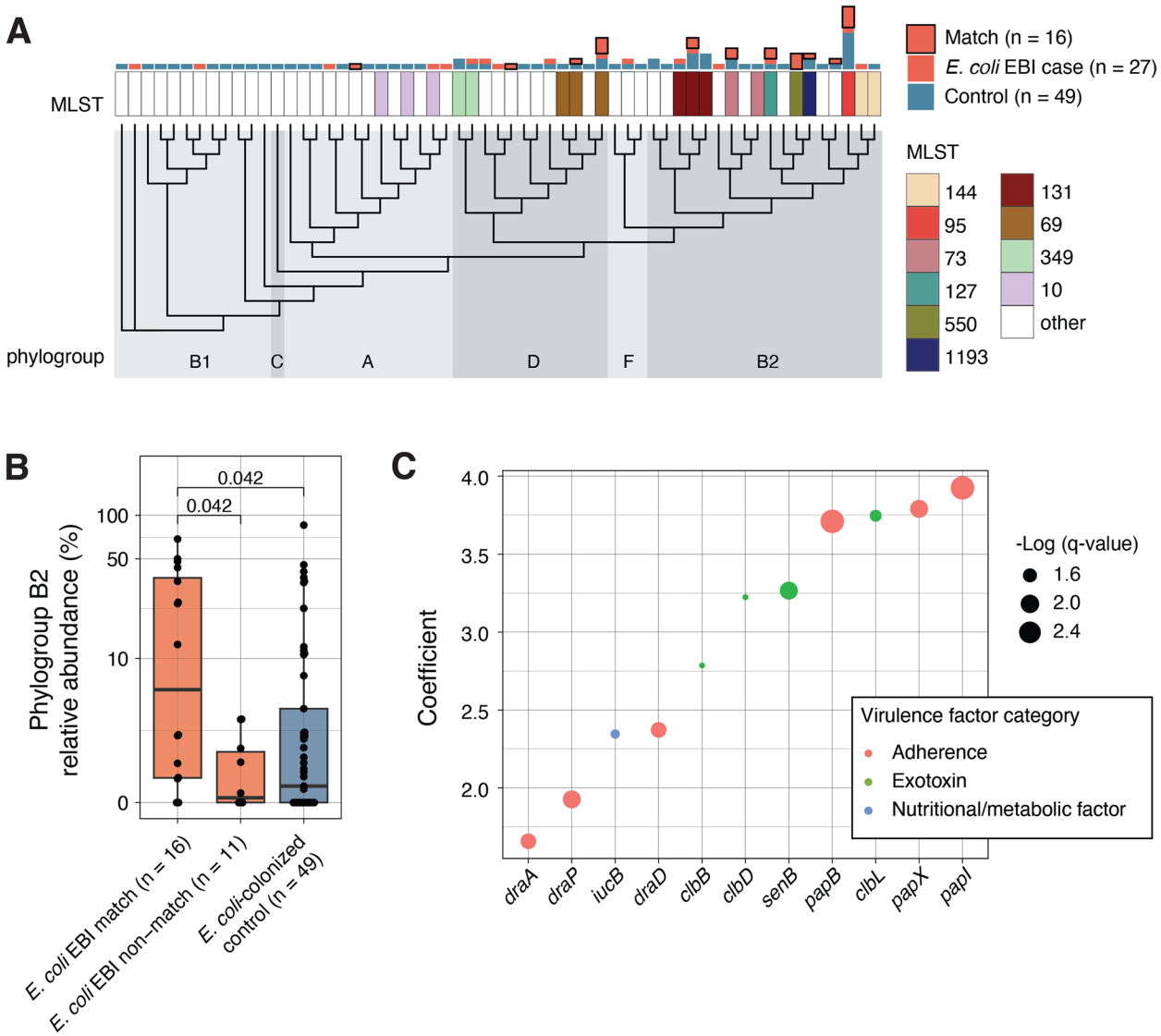

Fig. 4 Strain-level signatures of *E. coli* EBI matches in the gut. **A** Cladogram of all *E. coli* reference genomes for which a related strain was identified in at least one *E. coli* EBI case or *E. coli*-colonized control. Color strip denotes STs identified in at least three infants (the remainder are grouped as "other"). The bar chart indicates the frequency at which each reference genome was identified for E. coli EBI cases with a matching strain in their gut (match), those without a matching strain identified (non-match), or *E. coli*-colonized control infants. **B** Relative abundance of *E. coli* phylogroup B2 between groups as in A. **C** Virulence factor (VF) genes encoded by *Enterobacteriaceae* associated with *E. coli* EBI matches compared to *E. coli*-colonized controls as determined through generalized linear mixed effect models. Points are sized by their BH-adjusted *p* value (q) and colored by their VF functional category

groups ($p > 0.050$, Fisher's exact test). Four *E. coli* UTI matches were co-colonized by multiple strains, including the EBI-causing strain, and in three of these infants, a reference genome presumed to be representing the EBI-causing strain had the greatest relative abundance. Due to the low incidence of co-colonization events, however, we were underpowered to draw conclusions about their relationship to EBI risk. Interestingly, in 15/16 *E. coli* EBI match cases, the presumptively pathogenic strain was estimated to be at high abundance (≥1% of all bacteria present), suggesting that increased abundance may facilitate infection. In sum, our metagenomic strain profiling results demonstrate that both the *E. coli* abundance and lineage are associated with concomitant gut colonization and invasive infection, identifying strain-level microbiome signatures of *E. coli* EBI in young infants.

Virulence factor signature in infants with gut colonization of EBI-causing *E. coli*

We predicted that the stools of *E. coli* EBI matches would contain a higher density of *Enterobacteriaceae* virulence factors (VFs) than *E. coli*-colonized controls, which may indicate strategies by which these pathogenic *E. coli* strains survive in multiple habitats. To test this hypothesis, we profiled the gut metagenomic VF composition of *E. coli* EBI matches and *E. coli*-colonized controls using a core set of experimentally verified VF genes encoded by *Enterobacteriaceae* [41]. The richness and abundance of these VF genes was not different between *E. coli* EBI matches and *E. coli*-colonized controls ($p = 0.066$ and $p = 0.053$, Wilcoxon rank-sum test; Fig. S6A, B). When comparing the abundance of specific genes, 11 VFs of the 225 queried were positively associated with *E. coli* EBI matches ($q < 0.050$, Fig. 4C). These included genes implicated in adherence (*papB*, *papI*, *papX*, *draA*, *draD*, *draP*), exotoxins (*clbB*, *clbD*, *clbL*, *senB*), and nutritional/metabolic factors (*iucB*), each of which have been associated with virulence in extraintestinal pathogenic *E. coli* [58–61]. To ensure that these associations were not merely due to differences in *E. coli* abundance, we compared gene prevalence between groups and confirmed that all 11 VF genes were more frequently found in the metagenome of *E. coli* EBI matches compared to *E. coli*-colonized controls (BH-corrected $p \leq 0.012$ for all enriched VFs, Fisher's exact test; Table S4). These findings point to a potential role for certain VFs in enabling *E. coli* strains to both inhabit the gut and infect normally sterile sites in young infants, which may heighten EBI risk in infants carrying those VFs in their metagenome.

Discussion

We and others have shown that gut pathogen colonization precedes bacteremia in hospitalized preterm (< 37 weeks gestational age) infants [30, 62, 63], and that the relative abundance of the bacteremia-causing strain in the gut microbiome of these infants increases immediately before infection onset [30]. However, hospitalized preterm infants are more commonly exposed to antimicrobials and have less developed gut physiology than term infants, which increases their risk of bacterial translocation into sterile sites [21, 64]. We extend these data to show that gut pathogen colonization is common among febrile term infants with EBIs who are otherwise healthy, supporting the concept that the intestine is a pre-dissemination habitat for these pathogens, which has previously been reported at lower molecular resolution [65]. Our observation that the EBI-causing organism was often at sufficient abundance (≥ 1%) to be detected metagenomically suggests that these strains may be resident, rather than transient, colonizers of the infant gut.

The human digestive system contains a microbiome that is highly dynamic in the months after birth. Aerobic and facultatively anaerobic organisms first seed this organ, followed by a gradual increase in diversity and strictly anaerobic organisms, such as bifidobacteria, as the community matures [16, 17]. However, perturbations to the infant gut microbiome may be associated with community-acquired EBIs, as others recently reported lower *Bifidobacterium* and higher *Bacteroides* abundance in otherwise healthy infants with late-onset sepsis compared to afebrile controls [66]. We observed that infants with *E. coli* EBIs had higher prevalence and relative abundance of *E. coli* in their gut compared to febrile control infants, consistent with previous reports of increased bacterial 16S rRNA reads mapping to either *Escherichia* or *E. coli* in older pediatric populations with UTI caused by *E. coli* [67, 68]. Our identification of *E. coli* relative abundance ≥ 1% as a risk factor for *E. coli* EBI should be validated in additional cohorts, as *E. coli* is the most common organism causing EBI in this population [9], and this finding may serve as a useful biomarker as microbiome profiling enters the clinical sphere [69]. Unlike a recent study that demonstrated that the gut microbiome of older infants (3–11 months of life) with *E. coli* UTI had reduced α-diversity compared to healthy controls [67], our cases did not have lower diversity compared to age-matched febrile infants without EBI. This difference may reflect the relatively low complexity of the gut microbiome in the first 2 months of life or differences in the selection of comparison infants. Specifically, the febrile state or the presence of another pathogen that caused the fever might temporarily diminish diversity. Our results support the concept that a relatively high abundance of pathogenic *E. coli* and/or VF carriage may facilitate dissemination to extraintestinal sites. Nonetheless, we do not find that microbiome diversity was associated with protection from EBI in these infants, at least when compared to controls who were studied during a febrile illness not attributable to EBI.

Metagenomic strain profiling demonstrated strain-specific attributes of *E. coli* inhabiting both the gut and extraintestinal site(s). Our findings support a role for increased relative abundance of *E. coli*, specifically the virulence-associated phylogroup B2 [70], with the onset of *E. coli* EBI in young infants. Phylogroup B2 is the most persistent *E. coli* lineage in the normal infant microbiome [71]. While these strains commonly have a wide VF repertoire [71], we found that certain VF genes were more frequently present in infants with *E. coli* EBIs who were colonized with the EBI-causing strain than *E. coli*-colonized controls. P (*pap*) and Dr (*dra*) fimbriae promote *E. coli* adhesion in extraintestinal sites [60], which may facilitate their persistence in the urinary tract. Similarly, the

presence of machinery encoding the siderophore aerobactin (*iuc*) suggests that these strains may have adapted to survive in varied environmental conditions, such as those encountered in nutrient-limited extraintestinal sites. Interestingly, a set of genes involved in the synthesis of colibactin (*clb*) were more commonly found in infants colonized by the EBI-causing *E. coli* strain. Colibactin is a genotoxin that has been shown to contribute to both *E. coli* gut colonization and systemic infection in neonatal mice [59]. These functional associations potentially explain the difference in capacity to cause EBIs and suggest the likely sequence of dissemination (gut to bladder to bloodstream vs. gut to bloodstream). Future studies will need to determine if these VFs are involved in the seeding of *E. coli* EBI from the intestine and whether these may serve as therapeutic targets.

We note several important limitations of our findings. This study was cross-sectional in design and could not ascribe a causative role for the microbiome in seeding infection. Although a less likely explanation, our data cannot rule out the possibility that gut pathogen presence in stool represents seeding from an extraintestinal source, including the infected blood or bladder. Despite this possibility, longitudinal studies using strain tracking have identified the gut as a reservoir for EBI in other populations, including bloodstream infections in preterm infants [30, 62, 63] and hematopoietic cell transplantation recipients [72, 73] and UTIs in adults [74–76]. Future longitudinal cohort studies will be essential to confirm whether the gut is also a reservoir for organisms causing EBI in term infants. As mentioned above, our control group consisted of febrile infants without EBIs who may have instead had viral infections, which have been associated with a lower risk of bacterial co-infections [27]. Additionally, we cannot rule out the possibility that false negative blood cultures occurred in our control group due to infection with organisms that evaded culturing, which is a limitation of current diagnostics. Despite this possibility, our results present strong evidence for gut pathogen colonization in a subset of infants who also had EBI from those organisms, which we predict may extend to organisms not represented in our study. While we excluded infants with very recent (preceding 48 h) antimicrobial exposure, we did not account for exposures that are known to impact microbiome development in early life, including earlier antimicrobials, other medications, diet, and birth mode [77]. However, given the association between *E. coli* abundance and EBI, we propose that microbiome-targeting interventions, including breastfeeding [78], may provide a protective role by supporting the growth of non-pathogenic organisms. Our work also raises the possibility of screening for carriage of specific organisms (species and sub-species genotypes) with high risk of invasion in infants in the first 60 days of life, separate from the microbiome. The model for such an approach is prenatal screening for *S. agalactiae* carriage. However, demonstrating the efficacy of such targeted interventions against single pathogens would require considerable resources and very large cohorts.

Conclusions

Although many EBI-causing bacterial species colonize the healthy infant gut microbiome, their association with EBIs remains understudied. In this cohort of full-term infants younger than 60 days old presenting with fevers, we demonstrated that 63% of febrile infants with EBIs had the EBI-causing strain present in their gut at the time of presentation. We also identify age- and organism-dependent risk factors of EBI present in the microbiome, including several VFs associated with gut carriage of EBI-causing *E. coli*. In summary, we offer evidence that the intestine is a reservoir for EBI-causing pathogens in young full-term infants. Our work can inform future clinical studies and animal experiments to test the role of microbiome- and possibly pathogen-directed interventions in lowering the risk of EBI in young infants.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-025-02079-w.

Supplementary Material 1: Data file S1.xlsx file containing metadata

Supplementary Material 2: Data file S2.xlsx file containing Snippy, AMRFinder, and antimicrobial susceptibility testing results

Supplementary Material 3: Data file S3.xlsx file containing inStrain and MetaPhlAn profiling results

Supplementary Material 4: Data file S4.xlsx file containing respiration mode, strainGST, and shortBRED results

Supplementary Material 5: Supplementary figures and tables. Fig. 1. Study scheme. Diagram illustrating participant groupings, sample collection, and downstream processing. MALDI-TOF, matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry. Table 1. Cohort characteristics. P-values determined by Fisher's exact test for categorical variables and Wilcoxon rank-sum test for continuous variables. Fig. 2. Culturing workflow. Diagram illustrating workflow for selective culturing of intestinal (stool and/or rectal swab) isolates for the bacterial species causing extraintestinal infection for a given infant. Phase 1: growth in liquid media, Phase 2: growth on solid media, Phase 3: MALDI-TOF, Phase 4: MLST. MALDI-TOF, matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry; MLST, multi-locus sequence typing. Fig. 3. Paired stool and swab samples are highly correlated. Principal coordinate analysis (PCoA) plot of Bray–Curtis dissimilarity measures computed from MetaPhlAn2 species relative abundance values. Percent variation explained by the first two PCoA axes are indicated parenthetically. Stool and swab samples from eight participants were sequenced on a pilot run to determine similarity and black lines indicate samples from the same individual. The non-randomness between sample type configurations was tested using PROTEST and the symmetric Procrustes correlation statistic is noted along with the significance value. Table 1. Cohort characteristics. *P*-values determined by Fisher's exact test for categorical variables and Wilcoxon rank-sum test for continuous variables. Table 2. Prevalence and abundance of EBI-causing organisms. Distribution of the relative abundance (median [interquartile range (IQR)]) and prevalence (n [%]) of EBI-causing organisms in the gut

metagenome. Infants are stratified by EBI cases caused by the indicated organism or not and controls. Relative abundance IQR included for groups with ≥ 3 infants. Prevalence is defined by detection in the metagenome (relative abundance > 0%). Significance assessed between EBI cases caused by the indicated species and controls (cases not caused by the indicated species are reported for reference). Significance was determined by generalized linear mixed effect models for relative abundance (for groups with ≥ 3 infants) and Fisher's exact test for prevalence; the q-value and p-value are reported, respectively. n.s., not significant; n.d., not determined. Fig. 4. α-diversity and community maturation in *E. coli* EBI cases. Species richness (A) or Shannon diversity (B) in *E. coli* EBI cases compared to control infants plotted by day of life. Outlined circles represent *E. coli* EBI cases for whom the EBI-causing strain was detected in the gut. B) Relative abundance of the total aerobe and anaerobe populations by day of life in *E. coli* EBI cases and controls. The Spearman correlation among all included data points is shown. Fig. 6. *Enterobacteriaceae* virulence factor richness and burden are not significantly different in *E. coli* EBI matches. A) Number of unique virulence factor genes and B) virulence factor gene abundance encoded by *Enterobacteriaceae* in the metagenome of *E. coli* EBI matches compared to *E. coli*-colonized controls. *P*-values determined through Wilcoxon rank-sum tests. VF, virulence factor; RPKM, reads per kilobase of reference sequence per million reads. Table 4. *Enterobacteriaceae*-encoded virulence factors found at increased prevalence in the metagenome of *E. coli* EBI matches ($n = 16$) compared to that of *E. coli*-colonized controls ($n = 49$). BH-corrected *P*-values calculated through Fisher's exact test.

Acknowledgements
The authors thank members of the Schwartz lab, including Kate Wardenburg, Galen Wong, and Lingxia Zhao, for assisting in sample processing, and members of the Dantas lab, including Skye Fishbein and Bejan Mahmud, for their valuable insight during the analysis and writing of this manuscript. We would also like to thank the Edison Family Center for Genome Sciences and Systems Biology staff, Eric Martin, Brian Koebbe, MariaLynn Crosby, and Jessica Hoisington-López for their expertise and support in sequencing/data analysis, and Jacky Theodore, Bonnie Dee, James Anderson, and Keith Page for administrative support. We thank the Genome Technology Access Center at the McDonnell Genome Institute at Washington University School of Medicine for help with genomic analysis. The authors are grateful to John VanBuren and Charlie Casper of the University of Utah Data Coordinating Center for retrieving clinical metadata as part of PECARN. The authors would like to acknowledge the following PECARN investigators and institutions involved in the sample collection for this study: Shireen M. Atabaki (Division of Emergency Medicine, Department of Pediatrics, Children's National Medical Center, The George Washington School of Medicine and Health Sciences, Washington, DC), Elizabeth C. Powell (Division of Emergency Medicine, Department of Pediatrics, Ann & Robert H. Lurie Children's Hospital, Northwestern University Feinberg School of Medicine, Chicago, IL), Melissa Vitale (Division of Emergency Medicine, Department of Pediatrics, Children's Hospital of Pittsburgh, University of Pittsburgh School of Medicine, Pittsburgh, PA), Lorin Browne (Departments of Pediatrics and Emergency Medicine, Children's Hospital of Wisconsin, Medical College of Wisconsin, Milwaukee, WI), Lise E. Nigrovic (Division of Emergency Medicine, Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA), Andrea T. Cruz (Sections of Emergency Medicine and Infectious Diseases, Department of Pediatrics, Texas Children's Hospital, Baylor College of Medicine, Houston, TX), Mary Saunders (Children's Hospital of Colorado, University of Colorado School of Medicine, Aurora, CO), Peter S. Dayan (Division of Emergency Medicine, Department of Pediatrics, Columbia University College of Physicians and Surgeons, New York, NY), Daniel M. Cohen (Division of Emergency Medicine, Department of Pediatrics, Nationwide Children's Hospital, Columbus, OH), Elizabeth R. Alpern (Division of Emergency Medicine, Department of Pediatrics, Children's Hospital of Philadelphia, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA), Phillip I. Tarr (see author affiliations), Nathan Kuppermann (see author affiliations), Prashant Mahajan (see author affiliations), and David Schnadower (see author affiliations).

Authors' contributions
D.J.S., D.S., N.K., P.M., O.R., P.I.T., and G.D. conceived and designed the study. D.S., N.K., P.M., O.R., and P.I.T. managed the cohort as part of PECARN and provided clinical insights. C.H.-M. and N.S. performed selective culturing of enteric specimens and processed clinical samples. M.W. and C.-A.D.B. performed organism identification and phenotypic profiling of clinical isolates. A.L.D. performed sequencing of clinical isolates and enteric specimens and analyzed the data. A.L.D. wrote the manuscript with feedback from all authors. D.J.S. and G.D. supervised the project. D.J.S. and G.D. acquired funding for the project.

Funding
PECARN is supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS), in the Maternal and Child Health Bureau (MCHB), under the Emergency Medical Services for Children (EMSC) program through the following cooperative agreements: EMSC Data Center (EDC)-University of Utah (UJ5MC30824), GLACiER-Nationwide Children's Hospital (U03MC28844), HOMERUN-Cincinnati Children's Hospital Medical Center (U03MC22684), PEMNEWS-Columbia University Medical Center (U03MC00007), PRIME-University of California Davis Medical Center (U03MC00001), CHaMP node-State University of New York at Buffalo (U03MC33154), STELAR—Seattle Children's Hospital (U03MC33156), and SPARC node—Emory University School of Medicine (U03MC49671). This research was supported by NIH grants K08 AI159384 (D.J.S.), R01 AI155893 (G.D.), the Washington University Digestive Diseases Research Core Center (P30 DK052574, Biobank Core, to co-director P.I.T.), and a Thrasher Research Fund Early Career Award (D.J.S). D.J.S. was additionally supported by Doris Duke Charitable Foundation Physician Scientist Fellowship grant 2021081 and a Pediatric Infectious Diseases Society St. Jude Children's Hospital Fellowship in Basic and Translational Research. A.L.D. is supported by the NIH under award number T32GM139774 (PI: H. True-Krob). The Genome Technology Access Center at the McDonnell Genome Institute is partially supported by NCI Cancer Center Support Grant P30 CA91842 to the Siteman Cancer Center from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This work was performed by members of the PECARN Registry working group. This information or content and conclusions are those of the authors and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS, PECARN or the U.S. Government.

Data availability
The datasets supporting the conclusions of this article are included within the article (and its additional files). Genome assemblies, shotgun genomic reads, and shotgun metagenomic reads have been deposited to NCBI GenBank and SRA under BioProject ID PRJNA1148144. Scripts used for analysis are available at https://doi.org/10.5281/zenodo.15014489.

Declarations

Ethics approval and consent to participate
This study was approved by the institutional review board at all sites. Secondary analyses conducted here were approved under Washington University HRPO IRB#202106080.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO, USA. [2]Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA. [3]Department of Pathology and Immunology, Division of Laboratory and Genomic Medicine, Washington University School of Medicine, St. Louis, MO, USA. [4]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, USA. [5]Division of Emergency Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [6]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. [7]Departments of Pediatrics and Emergency Medicine, The George Washington School of Medicine and Health Sciences, and Children'S National Hospital, Washington, DC, USA. [8]Departments of Emergency Medicine and Pediatrics, University of Michigan, Ann Arbor, MI, USA. [9]Department of Infectious

Diseases, St. Jude Children'S Research Hospital, Memphis, TN, USA. [10]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA. [11]Department of Obstetrics and Gynecology, Washington University School of Medicine, St. Louis, MO, USA. [12]Center for Women's Infectious Disease Research, Washington University School of Medicine, St. Louis, MO, USA.

Received: 25 August 2024 Accepted: 4 March 2025
Published online: 24 March 2025

References

1. DePorre AG, Aronson PL, McCulloh RJ. Facing the ongoing challenge of the febrile young infant. Crit Care. 2017;21:68.
2. Simonsen KA, Anderson-Berry AL, Delair SF, Davies HD. Early-onset neonatal sepsis. Clin Microbiol Rev. 2014;27:21–47.
3. Greenhow TL, Hung Y-Y, Herz AM, Losada E, Pantell RH. The changing epidemiology of serious bacterial infections in young infants. Pediatr Infect Dis J. 2014;33:595–9.
4. Burstein B, Anderson G, Yannopoulos A. Prevalence of serious bacterial infections among febrile infants 90 days or younger in a canadian urban pediatric emergency department during the COVID-19 pandemic. Jama Netw Open. 2021;4:e2116919.
5. Watt K, Waddle E, Jhaveri R. Changing epidemiology of serious bacterial infections in febrile infants without localizing signs. PLoS ONE. 2010;5:e12448.
6. Mah V, Vanderkooi OG, Johnson DW. Epidemiology of Serious Bacterial Infections in Infants Less Than 90 Days of Age, Presenting to a Tertiary Care Emergency Department, 2010 to 2016. Pediatr Infect Dis J. 2019;38(7):e161–3. https://doi.org/10.1097/INF.0000000000002241.
7. Byington CL, Rittichier KK, Bassett KE, Castillo H, Glasgow TS, Daly J, et al. Serious bacterial infections in febrile infants younger than 90 days of age: the importance of ampicillin-resistant pathogens. Pediatrics. 2003;111:964–8.
8. Friedman J, Matlow A. Time to identification of positive bacterial cultures in infants under three months of age hospitalized to rule out sepsis. Paediatr Child Healt. 1999;4:331–4.
9. Pantell RH, Roberts KB, Adams WG, Dreyer BP, Kuppermann N, O'Leary ST, et al. Evaluation and management of well-appearing febrile infants 8 to 60 days old. Pediatrics. 2021;148:e2021052228. Available from: https://doi.org/10.1542/peds.2021-052228.
10. Juttukonda LJ, Katz S, Gillon J, Schmitz J, Banerjee R. Impact of a rapid blood culture diagnostic test in a children's hospital depends on gram-positive versus gram-negative organism and day versus night shift. J Clin Microbiol. 2019;58. https://doi.org/10.1128/jcm.01400-19.
11. Gasparrini AJ, Wang B, Sun X, Kennedy EA, Hernandez-Leyva A, Ndao IM, et al. Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. Nat Microbiol. 2019;4:2285–97.
12. Fouhy F, Guinane CM, Hussey S, Wall R, Ryan CA, Dempsey EM, et al. High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. Antimicrob Agents Ch. 2012;56:5811–20.
13. Reyman M, van Houten MA, Watson RL, Chu MLJN, Arp K, de Waal WJ, et al. Effects of early-life antibiotics on the developing infant gut microbiome and resistome: a randomized trial. Nat Commun. 2022;13:893.
14. Stocker M, Klingenberg C, Navér L, Nordberg V, Berardi A, el Helou S, et al. Less is more: antibiotics at the beginning of life. Nat Commun. 2023;14:2423.
15. Kuppermann N, Dayan PS, Levine DA, Vitale M, Tzimenatos L, Tunik MG, et al. A clinical prediction rule to identify febrile infants 60 days and younger at low risk for serious bacterial infections. JAMA Pediatr. 2019;173:342–51.
16. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. Cell Host Microbe. 2015;17:690–703.
17. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci Transl Med. 2016;8:343ra82.
18. Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. Nat Med. 2017;23:314–26.
19. Yamamoto S, Tsukamoto T, Terai A, Kurazono H, Takeda Y, Yoshida O. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by Escherichia coli. J Urol. 1997;157:1127–9.
20. Källenius G, Svenson SB, Hultberg H, Möllby R, Helin I, Cedergren B, et al. Occurrence of P-fimbriated Escherichia coli in urinary tract infections. Lancet. 1981;318:1369–72.
21. Sherman MP. New concepts of microbial translocation in the neonatal intestine: mechanisms and prevention. Clin Perinatol. 2010;37:565–79.
22. Mahjoub-Messai F, Bidet P, Caro V, Diancourt L, Biran V, Aujard Y, et al. Escherichia coli isolates causing bacteremia via gut translocation and urinary tract infection in young infants exhibit different virulence genotypes. J Infect Dis. 2011;203:1844–9.
23. Ku LC, Boggess KA, Cohen-Wolkowiez M. Bacterial meningitis in infants. Clin Perinatol. 2015;42:29–45.
24. Greenhow TL, Hung Y-Y, Pantell RH. Management and outcomes of previously healthy, full-term, febrile infants ages 7 to 90 days. Pediatrics. 2016;138: e20160270.
25. Mahajan P, Kuppermann N, Mejias A, Suarez N, Chaussabel D, Casper TC, et al. Association of rna biosignatures with bacterial infections in febrile infants aged 60 days or younger. JAMA. 2016;316:846. Available from: https://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2016.9207.
26. Davis J, Lehman E. Fever characteristics and risk of serious bacterial infection in febrile infants. J Emerg Medicine. 2019;57:306–13.
27. Mahajan P, Browne LR, Levine DA, Cohen DM, Gattu R, Linakis JG, et al. Risk of bacterial coinfections in febrile infants 60 days old and younger with documented viral infections. J Pediatr. 2018;203:86-91.e2.
28. Levine DA, Platt SL, Dayan PS, Macias CG, Zorc JJ, Krief W, et al. Risk of serious bacterial infection in young febrile infants with respiratory syncytial virus infections. Pediatrics. 2004;113:1728–34.
29. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124.
30. Schwartz DJ, Shalon N, Wardenburg K, DeVeaux A, Wallace MA, Hall-Moore C, et al. Gut pathogen colonization precedes bloodstream infection in the neonatal intensive care unit. Sci Transl Med. 2023;15:15.
31. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS ONE. 2015;10: e0128036.
32. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
33. Seemann T. mlst. Github; Available from: https://github.com/tseemann/mlst
34. Seeman T. snippy: fast bacterial variant calling from NGS reads. 2015. Available from: https://github.com/tseemann/snippy.
35. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom. 2021;7: 000685.
36. Blanco-Míguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, Manghi P, Dubois L, Huang KD, Thomas AM, Nickols WA, Piccinno G, Piperni E, Punčochář M, Valles-Colomer M, Tett A, Giordano F, Davies R, Wolf J, Berry SE, Spector TD, Franzosa EA, Pasolli E, Asnicar F, Huttenhower C, Segata N. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat Biotechnol. 2023;41(11):1633–44. https://doi.org/10.1038/s41587-023-01688-w.
37. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. Plos Comput Biol. 2021;17: e1009442.
38. Whitman WB. Bergey's Manual of Systematics of Archaea and Bacteria. 2024. https://doi.org/10.1002/9781118960608. Accessed in 2024.
39. Singer JR, Blosser EG, Zindl CL, Silberger DJ, Conlan S, Laufer VA, et al. Preventing dysbiosis of the neonatal mouse intestinal microbiome protects against late-onset sepsis. Nat Med. 2019;25:1772–82.
40. Kaminski J, Gibson MK, Franzosa EA, Segata N, Dantas G, Huttenhower C. High-specificity targeted functional profiling in microbial communities with ShortBRED. Plos Comput Biol. 2015;11: e1004557.

41. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res. 2021;50:D912–7.
42. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics. 2007;23:1282–8.
43. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nature Biotechnology. 2021; Available from: http://www.nature.com/articles/s41587-020-00797-0.
44. van Dijk LR, Walker BJ, Straub TJ, Worby CJ, Grote A, Schreiber HL, et al. StrainGE: a toolkit to track and characterize low-abundance strains in complex microbial communities. Genome Biol. 2022;23:74.
45. Blin K. ncbi-genome-download (0.3.3). Zenodo. 2023. https://doi.org/10.5281/zenodo.8192486.
46. Beghain J, Bridier-Nahmias A, Nagard HL, Denamur E, Clermont O. ClermonTyping: an easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. Microb Genom. 2018;4: e000192.
47. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020;21:180.
48. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641–50.
49. Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, et al. ggtreeExtra: compact visualization of richly annotated phylogenetic data. Mol Biol Evol. 2021;38:4039–42.
50. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:5114.
51. Rodriguez-R LM, Conrad RE, Viver T, Feistel DJ, Lindner BG, Venter SN, et al. An ANI gap within bacterial species that advances the definitions of intra-species units. mBio. 2023;15:e02696-23.
52. Nowrouzian F, Hesselmar B, Saalman R, Strannegård I-L, Åberg N, Wold AE, et al. Escherichia coli in infants' intestinal microflora: colonization rate, strain turnover, and virulence gene carriage. Pediatr Res. 2003;54:8–14.
53. Randis TM, Baker JA, Ratner AJ. Group B Streptococcal infections. Pediatr Rev. 2017;38:254–62.
54. Woll C, Neuman MI, Pruitt CM, Wang ME, Shapiro ED, Shah SS, et al. Epidemiology and etiology of invasive bacterial infection in infants ≤60 days old treated in emergency departments. J Pediatr. 2018;200:210-217.e1.
55. Feldman EA, McCulloh RJ, Myers AL, Aronson PL, Neuman MI, Bradford MC, et al. Empiric antibiotic use and susceptibility in infants with bacterial infections: a multicenter retrospective cohort study. Hosp Pediatr. 2017;7:427–35.
56. Cruz AT, Mahajan P, Bonsu BK, Bennett JE, Levine DA, Alpern ER, et al. Accuracy of complete blood cell counts to identify febrile infants 60 days or younger with invasive bacterial infections. JAMA Pediatr. 2017;171:e172927–e172927.
57. Milcent K, Faesch S, Guen CG-L, Dubos F, Poulalhon C, Badier I, et al. Use of procalcitonin assays to predict serious bacterial infection in young febrile infants. JAMA Pediatr. 2015;170:1–8.
58. Johnson JR, Johnston B, Kuskowski MA, Nougayrede J-P, Oswald E. Molecular epidemiology and phylogenetic distribution of the Escherichia coli pks Genomic Island. J Clin Microbiol. 2008;46:3906–11.
59. McCarthy AJ, Martin P, Cloup E, Stabler RA, Oswald E, Taylor PW. The genotoxin colibactin is a determinant of virulence in Escherichia coli K1 experimental neonatal systemic infection. Infect Immun. 2015;83:3704–11.
60. Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, Bugla-Ploskonska G, et al. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic Escherichia coli isolated from different sources: recent reports. Gut Pathog. 2019;11:10.
61. Mao B-H, Chang Y-F, Scaria J, Chang C-C, Chou L-W, Tien N, et al. Identification of Escherichia coli genes associated with urinary tract infections. J Clin Microbiol. 2012;50:449–56.
62. Heston SM, Lim CSE, Ong C, Chua MC, Kelly MS, Yeo KT. Strain-resolved metagenomic analysis of the gut as a reservoir for bloodstream infection pathogens among premature infants in Singapore. Gut Pathog. 2023;15:55.
63. Carl MA, Ndao IM, Springman AC, Manning SD, Johnson JR, Johnston BD, et al. Sepsis from the gut: the enteric habitat of bacteria that cause late-onset neonatal bloodstream infections. Clin Infect Dis. 2014;58:1211–8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3982840/.
64. Gibson MK, Wang B, Ahmadi S, Burnham C-AD, Tarr PI, Warner BB, et al. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. Nat Microbiol. 2016;1:16024. Available from: https://www.nature.com/articles/nmicrobiol201624.
65. Jantunen MA, Saxén H, Lukinmaa S, Ala-Houhala M, Siitonen A. Genomic identity of pyelonephritogenic Escherichia coli isolated from blood, urine and faeces of children with urosepsis. J Med Microbiol. 2001;50:650–2.
66. Graspeuntner S, Lupatsii M, van Zandbergen V, Dammann MT, Pagel J, Nguyen DN, Humberg A, Göpel W, Herting E, Rupp J, Härtel C, Fortmann I. Infants < 90 days of age with late-onset sepsis display disturbances of the microbiome-immunity interplay. Infection. 2024. https://doi.org/10.1007/s15010-024-02396-6.
67. Urakami C, Yamanouchi S, Kimata T, Tsuji S, Akagawa S, Kino J, et al. Abnormal development of microbiota may be a risk factor for febrile urinary tract infection in infancy. Microorganisms. 2023;11: 2574.
68. Paalanne N, Husso A, Salo J, Pieviläinen O, Tejesvi MV, Koivusaari P, et al. Intestinal microbiome as a risk factor for urinary tract infections in children. Eur J Clin Microbiol Infect Dis. 2018;37:1881–91. Available from: https://link.springer.com/10.1007/s10096-018-3322-7.
69. Leggett RM, Alcon-Giner C, Heavens D, Caim S, Brook TC, Kujawska M, et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. Nature Microbiology. 2020;5:430–42.
70. Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic Escherichia coli. Nat Rev Microbiol. 2021;19:37–54.
71. 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.
72. Severyn CJ, Siranosian BA, Kong STJ, Moreno A, Li MM, Chen N, et al. Microbiota dynamics in a randomized trial of gut decontamination during allogeneic hematopoietic cell transplantation. Jci Insight. 2022;7:e154344.
73. Tamburini FB, Andermann TM, Tkachenko E, Senchyna F, Banaei N, Bhatt AS. Precision identification of diverse bloodstream pathogens in the gut microbiome. Nat Med. 2018;24:1809–14.
74. Thänert R, Reske KA, Hink T, Wallace MA, Wang B, Schwartz DJ, et al. Comparative genomics of antibiotic-resistant uropathogens implicates three routes for recurrence of urinary tract infections. MBio. 2019;10:e01977–2019.
75. Worby CJ, Schreiber HL, Straub TJ, van Dijk LR, Bronson RA, Olson BS, et al. Longitudinal multi-omics analyses link gut microbiome dysbiosis with recurrent urinary tract infections in women. Nat Microbiol. 2022;7:630–9.
76. Magruder M, Sholi AN, Gong C, Zhang L, Edusei E, Huang J, et al. Gut uropathogen abundance is a risk factor for development of bacteriuria and urinary tract infection. Nat Commun. 2019;10:5521.
77. Enav H, Bäckhed F, Ley RE. The developing infant gut microbiome: a strain-level view. Cell Host Microbe. 2022;30:627–38.
78. McCormick BJJ, Richard SA, Murray-Kolb LE, Kang G, Lima AAM, Mduma E, et al. Full breastfeeding protection against common enteric bacteria and viruses: results from the MAL-ED cohort study. Am J Clin Nutr. 2021;115:nqab391-.

Publisher's Note