Gut pathogen colonization precedes bloodstream infection in the neonatal intensive care unit

Drew J. Schwartz1,2,3,4,5†, Nitan Shalon1†, Kate Wardenburg1, Anna DeVeaux1†, Meghan A. Wallace6, Carla Hall-Moore3, I. Malick Ndao3, Janice E. Sullivan7, Paula Radmacher7, Marilyn Escobedo8, Carey-Ann D. Burnham3,4,6, Barbara B. Warner3, Phillip I. Tarr3,4, Gautam Dantas1,3,4,6,9†

Bacterial bloodstream infections (BSIs) resulting in late-onset sepsis affect up to half of extremely preterm infants and have substantial morbidity and mortality. Bacterial species associated with BSIs in neonatal intensive care units (NICUs) commonly colonize the preterm infant gut microbiome. Accordingly, we hypothesized that the gut microbiome is a reservoir of BSI-causing pathogenic strains that increase in abundance before BSI onset. We analyzed 550 previously published fecal metagenomes from 115 hospitalized neonates and found bacteria that recent ampicillin, gentamicin, or vancomycin exposure was associated with increased abundance of Enterobacteriaceae and Enterococcaceae in infant guts. We then performed shotgun metagenomic sequencing on 462 longitudinal fecal samples from 19 preterm infants (cases) with BSI and 37 non-BSI controls, along with whole-genome sequencing of the BSI isolates. Infants with BSI caused by Enterobacteriaceae were more likely than infants with BSI caused by other organisms to have had ampicillin, gentamicin, or vancomycin exposure in the 10 days before BSI. Relative to controls, gut microbiomes of cases had increased relative abundance of the BSI-causing species and clustered by Bray-Curtis dissimilarity according to BSI pathogen. We demonstrated that 11 of 19 (58%) of gut microbiomes before BSI, and 15 of 19 (79%) of gut microbiomes at any time, harbored the BSI isolate with fewer than 20 genomic substitutions. Last, BSI strains from the Enterobacteriaceae and Enterococcaceae families were detected in multiple infants, indicating BSI-strain transmission. Our findings support future studies to evaluate BSI risk prediction strategies based on gut microbiome abundance in hospitalized preterm infants.

INTRODUCTION

Preterm birth accounts for 10% of births and 17% of infant mortality in the United States (1, 2). Preterm infants in neonatal intensive care units (NICUs) are particularly vulnerable to infection because of their immature organ systems, breach of cutaneous and mucosal barriers, and considerable antimicrobial exposures (3). The frequency of late-onset sepsis, defined as suspected or confirmed infection after the first 3 days of life (DOL) (3, 4), is inversely proportional to gestational age (GA) at birth, with up to 50% of extremely preterm and very low birth weight (<28 weeks and <1000 g) infants experiencing at least one episode (3, 5). In contrast to early-onset sepsis, the incidence and mortality of late-onset sepsis in preterm infants have not diminished appreciably over the past 30 years (6, 7). It is therefore essential to better understand the pathogen sources and risk factors associated with late-onset sepsis to develop mitigation strategies for this highly consequential complication of preterm birth.

Staphylococcus spp., Enterobacteriaceae (such as Escherichia coli, Klebsiella spp., and Enterobacter spp.), and Enterococcus spp. are the most common causes of bloodstream infection (BSI) in NICUs worldwide (3, 8, 9). Risk factors for late-onset sepsis in general include early GA, low birth weight, presence of an intravenous catheter, formula exposure, total parenteral nutrition, and prolonged duration without oral feeding (10–15). Gram-positive late-onset sepsis (predominantly Staphylococcus spp. and Enterococcus spp.) is associated with duration of intravenous catheterization and total parenteral nutrition (13, 14), whereas chorioamnionitis, vaginal birth, prolonged antimicrobial exposure, and necrotizing enterococcalis are associated with Gram-negative, predominately Enterobacteriaceae BSIs (14, 16, 17). Pathogens that cause BSI are commonly found in the NICU environment and in the hospitalized, preterm gut microbiome (18–21). In a subset of preterm infants, the same species or genus can be identified in stool before BSI (15, 22–24). Therefore, a direct gut-to-bloodstream route of dissemination has been postulated (25). Carl et al. (22) recovered isolates from stool culture that provisionally matched the BSI isolate in 7 of 11 cases of late-onset sepsis. These strains differed by an average of 33.5 nucleotides per 100,000 base pairs [average nucleotide identity (ANI) of 99.7%], a rate that may have overrepresented the nucleotide differences based on sequencing and computational technologies available at the time.

Although these findings are consistent with a gut-to-bloodstream mechanism of BSI pathogenesis, this degree of resolution...
of strain similarity cannot separate strains that diverged within the past 500 years (26). As such, prior methods that do not avail whole-genome sequencing, especially those relying on 16S ribosomal RNA gene sequencing or multilocus sequence typing, may overestimate strain similarity (26–28). Further, bona fide subspecies that concurrently colonize the same individual on different body sites, such as skin, gut, and mouth, can differ by as few as 30 genomic substitutions, and multiple strains can coexist in the same microenvironment (27, 29, 30). Hence, high genomic precision is essential for accurate pathogen tracking. Recent computational and sequencing methods (26, 31) permit more precise strain tracking to determine whole genome–resolved BSI-gut similarity and identify abundance changes in pathogens within the gut. These methodologic advances are critical to identify the immediate pre-BSI habitat of microbes to focus on appropriate interventions for prevention and diagnosis.

Recent exposure to antibiotics, which are frequently administered in the NICU for suspected late-onset sepsis (32), is associated with an increased abundance of certain pathobionts (commensal organisms that can be opportunistically pathogenic) in the guts of preterm infants (20). Specifically, Enterococcus faecalis and Staphylococcus epidermidis increased in abundance in response to selective pressure of meropenem therapy, and ampicillin, gentamicin, and vancomycin more variably affected overall diversity and specific species abundance (20). Some studies demonstrate diminished gut microbial diversity in neonates before late-onset sepsis versus controls; however, these controls often had more advanced GA, less formula receipt, and less antibiotic exposure, variables known to be associated with lower gut microbiome diversity (15, 24, 33, 34). Because gut, skin, and NICU hospital microbiomes harbor similar species (19, 27), and antibiotics enrich for potential pathogens in the stool of infants residing in NICUs, it is essential to analyze strain relatedness with maximal precision and investigate clinical features distinguishing BSI cases from controls in the NICU to identify risk factors and the proximate source of late-onset sepsis pathogens.

Fig. 1. The abundance of potential pathogens increases in stool of NICU infants after antibiotics. Family-level changes of bacteria in the gut microbiome were assessed regarding receipt of ampicillin, gentamicin, or vancomycin in the prior 10 days for previously sequenced infants in the cohort irrespective of BSI status. (A) Enterobacteriaceae and (B) Enterococcaceae were the only families found to be changed in the gut after either or both ampicillin and gentamicin. Recent vancomycin or gentamicin increased Enterococcaceae (C) but decreased Enterobacteriaceae (D) abundance, as assessed by MaAsLin2 with DOL as an additional fixed effect and participant as random effect. All P < 0.05 by generalized LME model. q value is Benjamini-Hochberg–corrected for multiple comparisons.
Here, we performed shotgun metagenomic sequencing on 462 stools from 19 preterm infants (cases) and whole-genome sequencing of their BSI isolates and 37 infants without BSI (controls). Our goal was to identify features capable of distinguishing BSI cases and controls without BSI inhabiting the same environments. We used precise BSI strain tracking within the gut microbiome longitudinally to identify sources of BSI pathogens and strain sharing within the NICU. Our results can inform future risk prediction algorithms based on gut microbiome abundance of potential pathogens for hospitalized, preterm infants to anticipate and intervene on at-risk infants.

RESULTS
Abundance of potential pathogens increases in the gut microbiome with recent antibiotics
A previous prospective study of 977 infants born at 27 weeks average GA was performed from 2009 to 2013 in three Midwestern NICUs to characterize the development of the preterm infant gut microbiome (20, 22, 35–37). Because antibiotic administration is associated with subsequent BSI, late-onset sepsis, and death (17, 38–40), and infants’ guts harbor potential pathogens (15, 20, 22, 33, 37), we first asked whether pathogens increased in abundance in the gut after neonates are given common antibiotics. We hypothesized that recent receipt of ampicillin, gentamicin, and vancomycin, the three antibiotics most commonly used in NICUs (20, 40) and often administered in combination, increases the abundance of frequently antibiotic-resistant BSI pathogens (20). The most common antimicrobials in this cohort immediately after birth were the co-administration of ampicillin and gentamicin, whereas the most common antimicrobials administered for late-onset sepsis ruled-out was the co-administration of vancomycin and gentamicin (20) (data file S1). We analyzed 550 previously sequenced stools in the first 100 DOL from 115 hospitalized preterm neonates to determine whether ampicillin, gentamicin, or vancomycin in the 10 days before fecal sample collection were associated with abundance changes in gut microbiome bacterial families using MaAsLin2 (20, 37, 41). We found that only three bacterial families were significantly associated with these antibiotic combinations (data file S2). Ampicillin or gentamicin exposure in the prior 10 days was associated with increased relative abundance of Enterobacteriaceae over receiving neither (Fig. 1A; \( P = 0.013, q = 0.097 \)), but ampicillin and gentamicin exposure was associated with decreased relative abundance of Enterococcaceae (Fig. 1B; \( P = 0.02, q = 0.11 \)). Exposure to vancomycin or gentamicin was associated with increased relative abundance of Enterococcaceae (Fig. 1C; \( P = 0.013, q = 0.097 \)) and decreased Enterobacteriaceae (Fig. 1D; \( P = 0.02, q = 0.11 \)) relative to receiving neither of these two antibiotics, whereas exposure to vancomycin and gentamicin nonsignificantly increased Enterobacteriaceae relative abundance in the gut microbiome (Fig. 1D). These findings are consistent with at least two nonmutually exclusive hypotheses. Abundance increases of these bacterial families may precede antibiotic delivery secondary to clinical changes prompting antibiotic administration. In addition, these organisms may be resistant to one of these frequently chosen antibiotics (20, 42). Because we found increased relative abundance of potential BSI pathogens in the stools of hospitalized NICU infants agnostic of their BSI status, we designed a case-control cohort study to investigate factors associated with BSI and the gut microbiome in the NICU.

Antibiotic-resistant pathogens cause BSIs in the NICU
Within the overall cohort (36), we identified 19 neonates with BSI who met our enrollment criteria (fig. S1). BSI-causing Gram-positive bacteria were Staphylococcus aureus \(( n = 6 \) ), Streptococcus agalactiae \(( n = 3 \) ), and E. faecalis \(( n = 2 \) ). BSI was also caused by Gram-negative Enterobacteriaceae \([ Serratia marcescens \( n = 3 \), E. coli \( n = 2 \), Klebsiella pneumoniae \( n = 2 \), and Enterobacter cloacae \( n = 1 \) \) (Table 1). To control for clinical associations of late-onset sepsis onset and gut microbiome composition, we paired these neonates with control infants without BSI for sex, GA, birth weight, delivery mode, age at onset of BSI, antibiotic days before BSI, and overall antibiotic exposure (data file S1). There were no statistically significant differences for any of these variables between cases and controls (Mann-Whitney, Fisher’s exact, or chi-square test) (Table 1).

| Table 1. Demographics and clinical details of case-control cohort. NA, not applicable; DOL, day of life; NPO: non per os, no oral nutrition. |
|-----------------|-----------------|-----------------|-----------------|
| Participats     | Cases           | Controls        |
| Number of stools per participant before BSI (25th–75th quartile) | 4 (3–6) | 4 (2–4) |
| Number of stools per participant after BSI (25th–75th quartile) | 6 (3–8) | 3 (3–4) |
| Bacteremia DOL (25th–75th quartile) | 32 | 19 (4–46) |
| Antibiotic days before bacteremia or matched DOL for controls | 16 (7–25) | 21 (9–28) |
| Overall antibiotic score from birth to BSI or matched control DOL (25th–75th quartile) | 53 | 26 (12) |
| Gestational age (25th–75th quartile) | 25 | 25 (25–26) |
| Birthweight, g (25th–75th quartile) | 740 | 810 (620–833) |
| % Female | 68 | 68 |
| % Cesarean | 74 | 74 |
| % First 8 weeks NPO | 26 | 20 |
| % First 8 weeks human milk | 47 | 55 |
| % First 8 weeks donor milk | 6 | 7 |
| % First 8 weeks formula | 21 | 18 |
| Bloodstream infection | 19 | NA |
| Staphylococcus aureus infection | 6 | NA |
| Streptococcus agalactiae | 3 | NA |
| Enterococcus faecalis | 2 | NA |
| Serratia marcescens | 3 | NA |
| Escherichia coli | 2 | NA |
| Klebsiella pneumoniae | 2 | NA |
| Enterobacter cloacae | 1 | NA |
We performed antimicrobial susceptibility testing and whole-genome sequencing on the BSI isolates and used ResFinder to infer antibiotic resistance genes (data file S3). Among *S. aureus*, two of six (33%) were phenotypically methicillin resistant and harbored *mecA*, seven of eight (87.5%) of Enterobacteriaceae were ampicillin nonsusceptible, and three of eight (37.5%) were gentamicin nonsusceptible. We identified β-lactam and aminoglycoside antibiotic resistance genes corresponding to the phenotypic antibiotic susceptibility testing for the eight Enterobacteriaceae. Infants whose BSI was caused by Enterobacteriaceae were significantly more likely than infants with BSI caused by other organisms to have had ampicillin, gentamicin, or vancomycin exposure 10 days prior (75% versus 9%; P = 0.0063, Fisher’s exact test). Given our study’s control criteria, infants with BSI had no difference in number of antibiotic days or antibiotic score (43) before versus after bacteremia relative to controls (Table 1 and fig. S2A). However, neonates with BSI had increased antibiotic score after BSI versus before and versus controls owing to antibiotic treatment targeting the BSI organism [fig. S2B and data file S2; linear mixed-effect (LME) model < 0.05]. We found no difference in median antibiotic resistance gene count in the metagenomes of BSI infants before BSI relative to matched controls using ShortBRED (fig. S2C). Specifically, we identified antibiotic resistance genes [in reads per kilobase per million reads (RPKM)] in the metagenomes of cases before BSI and controls during the same matched time span conferring resistance to beta-lactams, aminoglycosides, and vancomycin in decreasing order of prevalence (fig. S3 and data file S3). We could not determine which specific taxa harbor these antibiotic resistance genes (ARGs) from this metagenomic analysis; however, bacteria need not necessarily be resistant to antibiotics to persist because of subinhibitory antibiotic concentrations in the gut (44). We observed a decrease in ARGs in both groups from 2 weeks before BSI to 2 weeks after but no differences between cases and controls in abundance or composition (fig. S4 and data files S2 and S3). Thus, we found antibiotic-resistant organisms causing BSI but no significant differences in the overall gut resistomes of cases before BSI versus controls.

**High abundance of bacteremia-causing species in case metagenomes**

We hypothesized that BSI-causing species originated in the gut and gained access to the bloodstream or via cutaneous access. Five of 16 (31%) case participants from St. Louis Children’s Hospital (SLCH) had either no intravenous access or only a peripheral intravenous catheter in the 7 days before BSI (data file S3). We determined the relative abundance of the causative species before, during, and after BSI within each case’s gut microbiota. Each of the eight infants with Enterobacteriaceae BSI produced a stool before BSI with >10% relative abundance of the causative species, with 75% producing a stool with >45% relative abundance of the BSI-causing species. After completing the antibiotic course for BSI, both infants with *E. coli* still had greater than 20% relative abundance of *E. coli* in their stool (fig. S5). Similarly, the two infants with *E. faecalis* BSI had several stools with greater than 50% relative abundance of *E. faecalis* before and after BSI (fig. S6). Conversely, 83% (five of six) of infants with BSI caused by *S. aureus* had less than 3% relative abundance of *S. aureus* in all stools before BSI, and only one of six produced a stool dominated by *S. aureus*, with over 90% relative abundance (fig. S7). Three of five (60%) infants whose stools had minimal *S. aureus* before BSI had transient increases in fecal *S. aureus* carriage in the few days after BSI (fig. S7). Only one of three infants with *S. agalactiae* BSI had substantial fecal carriage before BSI (fig. S8). Thus, we observed at least two separate patterns of gut pathogen abundance before bacteremia; BSI species from Enterobacteriaceae and Enterococccaeae families largely had high gut abundance before bacteremia, whereas BSI from *S. aureus* and GBS had low or no gut abundance before bacteremia.

Gut microbiome Shannon diversity was stable 2 weeks before BSI but decreased after BSI during antimicrobial therapy (Fig. 2A and data file S2). We found no difference in Shannon diversity between BSI cases and controls within 5-day windows (bins) before BSI (Fig. 2A). Similarly, there was no difference in the number of species (richness) in the gut microbiome between cases and controls relative to BSI (Fig. 2B). We then compared the mean relative abundance of the BSI-causing species in each case before BSI with the mean relative abundance of the same species in all controls in the same DOL range (Fig. 2C). Despite the low relative abundance before BSI for *S. agalactiae* and *S. aureus*, we found significantly increased relative abundance of the causative species in the 2 weeks preceding BSI for all cases versus controls [Fig. 2C; median 17%, interquartile range (IQR) 0.33:48.76 versus median 0%, IQR 0:1.3; P = 0.001, Wilcoxon]. Therefore, the same BSI-causing species is present in stool before BSI onset in a subset of infants, and the relative abundance is higher than in infants not suffering from BSI. We then determined Bray–Curtis dissimilarity between all samples and, using repeated-measures permutational multivariate analysis of variance (PERMANOVA), found that the strongest nonindividial variance was contributed by the BSI-causing organism family (Fig. 2, D and E; 8.4% variance, P < 0.005). This clustering was not attributable to case versus control identity (Fig. 2, D and E) but rather was secondary to differences in the gut microbiome based on the BSI-causing species. Cumulative human milk exposure, DOL, and cumulative formula exposure were the only other metadata significantly contributing to microbiome conformation (Fig. 2E, fig. S9); however, these were not significantly different between cases and controls (Table 1). These data suggest that an organism-specific gut microbiobial signature might precede BSI.

**Gut microbiome features differentiate cases from controls**

We next wanted to determine whether any gut microbiome features permitted differentiation between BSI cases caused by these organisms versus others in the same NICUs. We determined gut microbiome enterotypes at the genus level on our 449 samples with sufficient sequencing reads. We identified four enterotypes that minimized Laplace score at the genus level (fig. S10). Enterotypes were variably dominated by key taxa, including *Staphylococcus* and *Enterococcus* (m1), *Klebsiella* and *Enterococcus* (m2), *Veillonella* and *Klebsiella* (m3), and *Escherichia* and *Enterococcus* (m4). We found that preinfection gut microbiomes (*n* = 214) clustered into distinct enterotypes depending on their control status or the organism that caused subsequent BSI (data file S2). Thus, gut microbiome structure before BSI correlates with the causative organism family and control status.

We next postulated that signatures in the gut microbiome might differentiate individuals who experienced BSI from specific organisms from others. We determined metagenomic differences before BSI between cases that later developed BSI caused by specific species.
Fig. 2. Greater abundance of BSI species in case stools before infection onset despite no gross alpha diversity differences. (A) Shannon diversity and (B) species richness displayed within 5-day windows (bins) relative to BSI. Values are averaged within 5-day windows for infants who produced more than one stool within the same 5-day span. (C) Mean relative abundance of BSI-causing species from each case in the 14 days before BSI versus the same species in controls. Mean value for each case species plotted versus abundance of same BSI-causing species in controls in the same DOL range. (D) PCoA of Bray-Curtis dissimilarity is shown by BSI-causing family. Number in parenthesis is percentage of variance explained. (E) Repeated-measures PERMANOVA results for variance contributed and Benjamini-Hochberg–corrected q value. Boxplots represent 25th to 75th median quartile, with horizontal black bar at the median. (A and B) LME model with participant as a random effect. (C) Wilcoxon signed-rank test. *P < 0.05. HM, human milk; DPI, days post infection.
versus all others, including previously sequenced samples (20, 37), in the same DOL range. The most significant features in the gut microbiome for individuals with *S. aureus* BSI were increased *S. epidermidis* and *Cutibacterium avidum* abundance (fig. S11, A and B; \( q < 0.005 \), coefficient 0.56 and 0.16; data file S2). Despite not being frequently identified in the gut microbiome (fig. S7), we found a small but significant increase in *S. aureus* in the gut before *S. aureus* BSI (Fig. 3A; \( q < 0.01 \), coefficient 0.10). *S. agalactiae* was the most significant feature before BSI in the gut microbiomes of cases with GBS BSI (Fig. 3B; \( q < 0.0001 \), coefficient 0.2). Veillonella atypica was also increased in the gut microbiome for these same individuals who progressed to GBS BSI (fig. S11C; \( q < 0.05 \), coefficient 0.13). *E. faecalis* was the only significant feature differentiating individuals with *E. faecalis* BSI from all others in the NICU (Fig. 3C; \( q < 0.05 \), coefficient 0.58). *S. marcescens* abundance was significantly increased in individuals with *S. marcescens* BSI (Fig. 3D; \( q < 0.0001 \), coefficient 0.63). Klebsiella oxytoca (fig. S11D) and Klebsiella michiganensis (fig. S11E) were both reduced in the gut before *S. marcescens* BSI (\( q < 0.005 \), coefficient −0.61 and −0.20). Dermaclavus hominis was also slightly increased (fig. S11F; \( q < 0.05 \), coefficient 0.03). Klebsiella spp. abundance was the only significant feature differentiating those with *K. pneumoniae* BSI from all others (Fig. 3E; \( q < 0.05 \), coefficient 0.75). Similarly, *E. coli* abundance was significantly elevated in those with *E. coli* BSI (Fig. 3F; \( P < 0.01 \), coefficient 0.75). Therefore, for every BSI-causing organism, we identified an increase in that species, often the only significant difference, before BSI in affected individuals.

For each of the BSI-causing organisms, we calculated the specificity, sensitivity, and positive and negative predictive values (PPV and NPV, respectively) of BSI on the basis of gut relative abundance thresholds of the causative species. For *S. aureus* BSI, we found that a gut relative abundance of *S. aureus* above 0.1% had 6% sensitivity, 98% specificity, 35% PPV, and 86% NPV (\( P < 0.01 \), Fisher’s exact test; data file S2). A fecal abundance greater than 0.1% had 29% sensitivity, 99% specificity, 52% PPV, and 97% NPV for GBS BSI (\( P < 0.0001 \)). *E. faecalis* abundance of >15% was only 5% sensitive but >99% specific for *E. faecalis* BSI, with 93% PPV and 72% NPV (\( P < 0.0001 \)). *S. marcescens* gut microbiome abundance above 8% was 35% sensitive and >99% specific, with a PPV of 78% and NPV of 99% for *S. marcescens* BSI (\( P < 0.0001 \)). For *K. pneumoniae* BSI, we determined these characteristics at 10, 20, 60, and 90% relative abundance thresholds. The sensitivity ranged from 5 to 13%, and specificity ranged from 98 to >99% with PPV of 42 to 88% and NPV of 62 to 93% (all \( P < 0.0001 \), Fisher’s exact test). An *E. coli* relative abundance above 30% was only 4% sensitive but 99.5% specific for *E. coli* BSI with 64% PPV and 85% NPV (\( P < 0.0005 \)). Thus, we have demonstrated that relative abundance thresholds are highly specific for future BSI caused by those same species.

**Nearly identical BSI strains present in stool before BSI**

Because of the increased gut relative abundance of the causative species in BSI cases versus controls, and because gut metagenomes clustered according to BSI-causing bacterial family, we investigated whether the BSI species became more abundant in the days preceding BSI. The median species richness within metagenomes of bacteria cases was 6 to 10 in the 15 days before BSI and did not change before the day of BSI (Fig. 2B). In the 15 days before BSI, in half of cases (4 of 8, 8 of 15, and 8 of 16 within respective 5-day time spans), the future BSI-causing species was in the most abundant quartile within the metagenome (Fig. 4A). At the time

![Fig. 3. Gut microbiome signature before infection onset differentiates infants with BSI caused by specific organisms versus controls.](https://www.science.org/doi/10.1126/scitranslmed.aadg5562)

Abundance of the BSI-causing species (A) *S. aureus*,(B) *S. agalactiae*, (C) *E. faecalis*, (D) *S. marcescens*, (E) *Klebsiella spp.*, and (F) *E. coli* in the gut microbiome before BSI onset shown for those with BSI from that organism versus from all others. Generalized LME models were used with MaAsLin2 and controlled for DOL and participant with Benjamini-Hochberg-corrected false discovery rate \( q \) value. “Yes” indicates bacteremia from that organism, and “No” indicates absence of bacteremia from the listed organism but could include bacteremia from other organisms. Stools after BSI for participants in the “yes” category were not included in this analysis.
of BSI and for the subsequent 5 days, the causative species was in the most abundant quartile in 12 of 18 (67%) of cases. During and after definitive antimicrobial treatment for the BSI organism in the 2 weeks after BSI, the causative species was still in the most abundant quartile in 5 of 12 (42%) of infants. Given these dynamics, we precisely tracked the whole genome of the causative BSI strain in the stool metagenome relative to infection in inStrain (26). We computed the isolate coverage by mapping metagenomic reads and normalized it by read count to allow for cross-cohort comparisons. We found that in the days preceding BSI, the normalized coverage on a scale of 0 to 1 per participant of the BSI isolate within the metagenome increased to an average maximum of 0.48/1 before decreasing once definitive antibiotic treatment was administered (Fig. 4B; LME, \( P = 0.015 \)). We then determined normalized coverage per metagenomic sample relative to BSI divided by read count for participants who had produced at least one sample with breadth greater than 0.99. In total, 125 of 205 (61%) of metagenomes harbored the cognate BSI isolate at greater than 1× coverage and 0.5 breadth (data file S4). BSI strain coverage was highest in the 5 days before and after BSI, and strain abundance decreased at the end of effective antimicrobial therapy 10 to 15 days after BSI (Fig. 4C; \( P < 0.05 \), Wilcoxon). Thus, despite no trends in overall Shannon diversity and richness in case participants before BSI, the coverage of the causative strain within the gut metagenome peaked in the days directly before BSI onset.

To quantify the genetic relatedness between the BSI isolate and the stool metagenome, we determined the breadth and the number of single-nucleotide substitutions per genome (SNS rate) between each infant’s BSI isolate and their stool metagenome longitudinally using inStrain (26). Figure 4D shows negative log (1 – breadth) and
negative log (1 − SNS rate) with a maximum value of 7, where the breadth was 1 and substitutions of 0 in the whole genome. Thus, a y-axis value of 7 signifies that every base pair with sufficient read coverage in the BSI isolate was detected in that cognate metagenome. Among comparisons with breadth >0.5, 45 of 126 (36%) of tested metagenome-BSI isolate pairs had an SNS rate of 0 to 1 (data file S4). Thus, data points in the upper right of Fig. 4D above log breadth of 5 represent those with breadth greater than 0.99999 with 0 to 1 substitutions between their BSI isolate genome and the stool metagenomes over time. We next determined the minimum number of substitutions between the BSI isolate and stool metagenomes before infection or at any time during our sampling interval (Fig. 4E). We found that 11 of 19 (58%) infants produced at least one stool before BSI with fewer than 16 substitutions (Fig. 4D; negative log substitution rate of 5.2 to 7) and at breadth greater than 0.99. This precision confirms that strains present in blood are identical or nearly identical to those in the gut metagenome (26). In addition, the presence of the same strain seemed to correlate with specific bacterial families. Specifically, all Enterobacteriaceae and Enterococcaceae in the stool before BSI had fewer than 16 substitutions at breadth >0.99, whereas only one case with BSI caused by Staphylococcaceae or Streptococcaceae met this threshold (Fig. 4, D and E, and data file S4). Although S. aureus or S. agalactiae were not present in the gut before BSI, we found that four infants produced a stool after BSI with these strains harboring fewer than 13 substitutions from the cognate BSI isolate (Fig. 4, D and E). Therefore, the gut may be seeded by the bacteremic event. Thus, we identified most BSI strains in the gut of that same individual before or immediately after BSI.

We then determined the gut and BSI strain similarity by BSI-causing bacterial family using ANI. The genomes of the BSI organisms in our cohort were 2 to 6 Mb, with S. agalactiae being the smallest and K. pneumoniae being the largest (data file S3). Therefore, ANI values of >0.99999 signify fewer than 20 to 50 substitutions between the whole-genome isolate and the metagenome. In the 15 days before infection, individuals with BSI caused by Enterobacteriaceae and Enterococcaceae had significantly higher ANI between the BSI isolate and the stool metagenome than Staphylococcaceae (P < 0.005) and Streptococcaceae (P < 0.01) in a LME model (Fig. 4F and data file S2). Despite effective antimicrobial therapy directed at the BSI pathogen, the same strain with fewer than 20 substitutions at breadth of >0.99 and logANI > 5 (0.99999) persisted in the stool for 6 of 19 (32%) individuals 10 days or more after BSI onset (Fig. 4, E and F, and data file S4). The abundance of the causative strain decreases, but antibiotics do not completely eliminate them, and, often, the causative species still maintains high relative abundance (Fig. 4, A to C). These data demonstrate that, for most cases, the same strain causing BSI is present in the gut microbiome, increases in abundance until BSI onset, and is often not completely cleared during systemic antimicrobial therapy.

**BSI strains are shared among infants in the NICU**

Environmental surfaces in NICUs are colonized by the same species that are found in gut microbiomes of neonates hospitalized in those

---

**Fig. 5.** BSI strains are infrequently shared in the NICU, and sharing differs by causative organism. (A) InStrain was used to plot BSI isolate reads to metagenomes from the same individual (red) and to all unrelated individuals in the NICU (black). (B) Negative log breadth/substitution rate is plotted by BSI-causing family for each unique non–self-BSI isolate-metagenome pair. Observations above the red dashed line represent negative log breadth/substitution rate above 5 and contain observations with greater than 0.99999 population ANI. Metagenome-BSI mappings below breadth 0.5 were excluded for (A) and (B). Fisher’s exact test for these observations above negative log breadth/substitution rate of 5 compared with all possible nonrelated BSI-metagenome mappings is shown as a two-by-two table.
from future episodes of culture-positive late-onset sepsis and BSI (47). Early identification of at-risk infants can help inform appropriate antibiotic treatment while limiting unnecessary antibiotic use. Other reports have postulated gut-bloodstream overlap in different populations, including in the NICU, with different degrees of genomic precision (22, 31, 46, 48, 49). Here, we precisely track strains at the nucleotide level combined with controls to identify clinical and microbiome correlates of BSI in preterm infants that can be tested in follow-up cohorts and mechanistic models. We found that a gut microbiome signature existed for those individuals who would progress to BSI caused by specific species. We demonstrate that the identical or nearly identical strain causing BSI can be found in the gut microbiome 1 to 2 weeks before late-onset sepsis. In that same interval, we observed increased relative abundance of the cognate species/strain within the gut microbiome without changes in overall richness or diversity.

These findings have several implications for optimizing clinical management to detect and prevent late-onset sepsis in preterm infants. Given the increased abundance of the causative species in cases versus controls and our modeling demonstrating community structure differences between BSI from specific species, gut microbiome profiling could be used to ascribe late-onset sepsis risk to each infant. This finding complements data that show that longitudinal stool volatile profiling could offer late-onset sepsis predictions before onset (50). The possibility of performing rapid sequencing at the start of late-onset sepsis could also help inform antibiotic stewardship and direct more rapid, effective treatment of a causative organism (51).

Our study has several limitations. First, although a direct gut-to-bloodstream translocation is the most likely scenario for Enterobacteriaceae and *E. faecalis* because of the absence of genomic substitutions between gut metagenome and BSI isolate, we cannot exclude the possibility of short-term cutaneous or tracheal colonization before blood translocation. We note that 5 of 16 case participants from SLCH had no intravenous access or only a peripheral intravenous catheter in the 7 days before BSI. However, on the basis of our identified family-level differences in gut colonization before late-onset sepsis, our data suggest at least two different mechanisms for BSI pathogen origin. This distinction will be especially important in determining the proximate source for coagulase-negative *Staphylococcus* BSI. These organisms are common gut microbiome and skin colonizers, so distinguishing them as cutaneous contamination or BSI in a blood culture requires clinical adjudication (3, 52). For that reason, we excluded them from this present study. Another important caveat to the family-based differences in pre-bacteremia colonization is the frequency of sampling for certain species. Carl et al. (22) identified a provisionally matched isolate from a single stool, which we did not interrogate here, each from three cases with GBS bacteremia. We did not identify the same strain in pre-BSI stools for GBS, possibly because colonization was transient for these individuals with GBS BSI or because our threshold for same strain declaration (popANI > 0.99999) was stricter. In all three individuals with GBS BSI, popANI between BSI isolate and stool was 0.997 to 0.999 before BSI, with 13 to 5141 substitutions between BSI isolate and metagenome. Second, we did not include data on maternal prenatal or postnatal antibiotic exposure, which could affect infant microbiome and resistome composition (53). Last, we did not identify a host mechanism of increased susceptibility to BSI between cases and controls, which
MATERIALS AND METHODS

Study design

A previously published prospective study was performed from 2009 to 2013 in three Midwestern NICUs to characterize the development of the hospitalized, preterm gut microbiome and its role in necrotizing enterocolitis and other diseases (20, 22, 35–37). Neonates weighing ≤1500 g were enrolled if they were admitted to the NICU and expected to live beyond 1 week of life at SLCH (St. Louis, MO, USA), Children’s Hospital at Oklahoma University Medical Center (Oklahoma City, OK, USA), and Norton (formerly Kosair) Children’s Hospital (Louisville, KY, USA) (55). Stools were collected and, after brief storage at 4°C, stored indefinitely at −80°C until use. Clinical metadata were stored in the REDCap electronic database at Washington University School of Medicine.

Samples and metadata used in this study were previously collected through "the Neonatal Microbiome and Necrotizing Enterocolitis" study funded by National Institutes of Health (NIH) grant U54AI083265 and approved under Washington University HRPO #201105492 (35–36). This initial study was reviewed by the University of Louisville Institutional Review Board (IRB) (HSPO #11.0136) and University of Oklahoma Health Sciences Center IRB 2472. Each site’s IRB approved the study, and parents were provided informed consent. Secondary analyses as detailed in the current paper were approved under Washington University School of Medicine HRPO #201205152.

Statistical analysis

Analyses were performed using R version 4.0.3, R Studio version 1.4.1103, and GraphPad Prism version 9.3.1. Metadata variables were analyzed with Fisher’s exact (2×2 table) or chi-square test (greater than 2×2 table) using GraphPad Prism or the R package “stats” version 4.1.1. LME models were performed with the R package “lme4” version 1.1–27.1 with participant ID as the random effect. A post hoc Tukey’s test was performed to identify pairwise differences between groups with the package “emmeans” version 1.7.1.1. MaAsLin2 v1.6.0 was used to determine metadata associations with microbiome features (41). Default parameters were used except that we used arcsine square-root transformation (AST). Participant was always used as a random effect. Significance of principal coordinate analysis (PCoA) of Bray-Curtis dissimilarity was determined using repeated-measures PERMANOVA using the R packages “permute” v0.9-7 and vegan v2.6-2 as previously reported (56). Briefly, the variance and significance of each variable were permuted separately with participant-level blocking due to repeated measures. Two samples for whom feeding data did not exist were excluded from this analysis. Benjamini-Hochberg correction was used to account for multiple comparisons. \( P < 0.05 \) and \( q < 0.25 \) were considered significant, and all tests were two tailed.

Supplementary Materials

This PDF file includes:
Materials and Methods
Figs. S1 to S11
References (57–73)

Other Supplementary Material for this manuscript includes the following:
Data files S1 to S5
MDAR Reproducibility Checklist

View/request a protocol for this paper from Bio-protocol.

REFERENCES AND NOTES


43. R. A. A. Moehring, Elizabeth Dodds, Developing patient safety outcome measures and measurement tools for antibiotic stewardship programs metric guide (2017).


Downloaded from https://www.science.org at Washington University on May 05, 2023
190,588), which might earn royalties if the technology is commercialized. G.D. is a consultant to and a member of the Scientific Advisory Board of Pluron Biosciences, which is developing methods for discovering novel environmental microbes for agricultural and pharmaceutical applications. G.D. has consulted for SNIPR Technologies Ltd. in the last 5 years, but not presently. The authors declare no other competing interests. Data and materials availability: All data associated with this study are present in the paper or the Supplementary Materials. Shotgun sequencing data, BSI isolate assemblies, and clinical metadata are available in the NCBI repository under BioProject PRJNA884103. Custom code is available at https://zenodo.org/badge/latestdoi/519858068.

Submitted 5 January 2023
Accepted 11 April 2023
Published 3 May 2023
10.1126/scitranslmed.adg5562
Gut pathogen colonization precedes bloodstream infection in the neonatal intensive care unit


DOI: 10.1126/scitranslmed.adg5562

**View the article online**
https://www.science.org/doi/10.1126/scitranslmed.adg5562

**Permissions**
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the [Terms of service](https://www.science.org)