The purpose of this study was to evaluate the composition and richness of bacterial communities associated with low-birthweight (LBW) infants in relation to host body site, individual, and age. Bacterial 16S rRNA genes from saliva samples, skin swabs, and stool samples collected on postnatal days 8, 10, 12, 15, 18, and 21 from six LBW (five premature) infants were amplified, pyrosequenced, and analyzed within a comparative framework that included analogous data from normal-birthweight (NBW) infants and healthy adults. We found that body site was the primary determinant of bacterial community composition in the LBW infants. However, site specificity depended on postnatal age: saliva and stool compositions diverged over time but were not significantly different until the babies were 15 days old. This divergence was primarily driven by progressive temporal turnover in the distal gut, which proceeded at a rate similar to that of age-matched NBW infants. Neonatal skin was the most adult-like in microbiota composition, while saliva and stool remained the least so. Compositional variation among infants was marked and depended on body site and age. Only the smallest, most premature infant received antibiotics during the study period; this heralded a coexpansion of *Pseudomonas aeruginosa* and a novel *Mycoplasma* sp. in the oral cavity of this vaginally delivered, intubated patient. We conclude that concurrent molecular surveillance of multiple body sites in LBW neonates reveals a delayed compositional differentiation of the oral cavity and distal gut microbiota and, in the case of one infant, an abundant, uncultivated oral *Mycoplasma* sp., recently detected in human vaginal samples.

**ABSTRACT** The purpose of this study was to evaluate the composition and richness of bacterial communities associated with low-birthweight (LBW) infants in relation to host body site, individual, and age. Bacterial 16S rRNA genes from saliva samples, skin swabs, and stool samples collected on postnatal days 8, 10, 12, 15, 18, and 21 from six LBW (five premature) infants were amplified, pyrosequenced, and analyzed within a comparative framework that included analogous data from normal-birthweight (NBW) infants and healthy adults. We found that body site was the primary determinant of bacterial community composition in the LBW infants. However, site specificity depended on postnatal age: saliva and stool compositions diverged over time but were not significantly different until the babies were 15 days old. This divergence was primarily driven by progressive temporal turnover in the distal gut, which proceeded at a rate similar to that of age-matched NBW infants. Neonatal skin was the most adult-like in microbiota composition, while saliva and stool remained the least so. Compositional variation among infants was marked and depended on body site and age. Only the smallest, most premature infant received antibiotics during the study period; this heralded a coexpansion of *Pseudomonas aeruginosa* and a novel *Mycoplasma* sp. in the oral cavity of this vaginally delivered, intubated patient. We conclude that concurrent molecular surveillance of multiple body sites in LBW neonates reveals a delayed compositional differentiation of the oral cavity and distal gut microbiota and, in the case of one infant, an abundant, uncultivated oral *Mycoplasma* sp., recently detected in human vaginal samples.

**IMPORTANCE** Complications of premature birth are the most common cause of neonatal mortality. Colonization by the indigenous microbiota, which begins at delivery, may predispose some high-risk newborns to invasive infection or necrotizing enterocolitis (NEC), and protect others, yet neonatal microbiome dynamics are poorly understood. Here, we present the first cultivation-independent time series tracking microbiota assembly across multiple body sites in a synchronous cohort of hospitalized low-birthweight (LBW) neonates. We take advantage of archived samples and publically available sequence data and compare our LBW infant findings to those from normal-birthweight (NBW) infants and healthy adults. Our results suggest potential windows of opportunity for the dispersal of microbes within and between hosts and support recent findings of substantial baseline spatiotemporal variation in microbiota composition among high-risk newborns.
lism and immune function (18, 19). In the neonatal intensive care unit (NICU), however, the promotion of potentially beneficial host-microbe interactions must be carefully balanced against the control of pathogen spread among a highly vulnerable patient population (20, 21). This is distinctively challenging with regard to the prevention and treatment of NEC, a disease in which the interrelated roles of antibiotic exposure, enteral feedings, and changes in the intestinal microbiota are imprecisely defined (10, 11). Recent studies of the fecal microbiota of premature infants using cultivation-independent approaches have revealed a low level of diversity, high interindividual variability, and a capacity for abrupt temporal shifts in species- and strain-level composition (22–32). However, most of these studies have been limited to a relatively small number of samples and to a single body site, the distal gut.

In the present study, we simultaneously tracked the distal gut, oral cavity, and skin surface microbiota of six hospitalized LBW infants, including 2 sets of twins, over the 2nd and 3rd weeks of life. Our analysis focused on factors underpinning compositional variation during this critical time span. For the distal gut microbiota, we also made comparisons to age-matched normal-birthweight (NBW) infants using archived samples from a prior study (5); and for all sites, we made comparisons to adults using publicly available sequence data (1, 2). Although the infants sampled here were unaffected by sepsis or NEC, their age range represents an important window of vulnerability for both of these conditions.

RESULTS

LBW infant cohort characteristics. Five of the six infants (all but baby 6) were premature; these five had completed $31\frac{1}{7}$ weeks of gestation at the time of delivery. Among the premature infants, three were born weighing $<1.5$ kg, placing them in the category of “very LBW” (VLBW) and at highest risk for complications of preterm birth. These three infants were born at Comer Children’s Hospital, whereas the others were born at outside hospitals and then transferred to Comer’s NICU prior to enrollment. The cohort included two sets of premature twins, both delivered via Cesarean section. All infants received antibiotics in the first week of life (Table 1). None of their mothers received antepartum antibiotics.

Baby 3, the smallest, most premature infant in the study, was intubated and mechanically ventilated throughout the sampling period, whereas the others either had no history of endotracheal intubation or were extubated before the sampling period.

| Baby | Sex | Delivery mode | Birth wt (kg) | Gestational age at delivery | Birth location | Postdelivery antibiotics and length of treatment | Medical conditions | Complication(s) during pregnancy and/or delivery |
|------|-----|---------------|--------------|-----------------------------|---------------|-----------------------------------------------|-------------------|-----------------------------------------------|
| 1    | F   | Cs            | 1.82         | $31\frac{1}{2}$           | Not at a UC hospital | Ap + Gm 48 h | Premature, respiratory distress, hyperbilirubinemia | Loss of fetal heart tones |
| 2    | M   | Cs            | 1.74         | $31\frac{1}{2}$           | Not at a UC hospital | Ap + Gm 48 h | Premature, respiratory distress, hyperbilirubinemia | Loss of fetal heart tones |
| 3    | M   | V             | 0.75         | $24\frac{1}{2}$           | UC hospital     | Ap + Gm 7 days | Premature, respiratory distress, hyperbilirubinemia | PPROM |
| 4    | M   | Cs            | 1.38         | $30\frac{1}{2}$           | UC hospital     | Ap + Gm 48 h | Premature, hyperbilirubinemia, GE reflux, AOP | Preeclampsia, preterm labor |
| 5    | M   | Cs            | 1.05         | $30\frac{1}{2}$           | UC hospital     | Ap + Gm 48 h | Premature, respiratory distress, hyperbilirubinemia | Preeclampsia, preterm labor |
| 6    | M   | V             | 1.72         | $38\frac{1}{2}$           | Not at a UC hospital | Ap + Gm + Cx 48-72 h | Premature, hyperbilirubinemia, GE reflux, AOP | IUGR, chromosome 4p deletion syndrome |

- Babies 1 and 2 are dizygotic (DZ) twins; babies 4 and 5 are monozygotic (MZ) twins (monochorionic diamniotic).
- F, female; M, male.
- Cs, Cesarean section; V, vaginal.
- $31\frac{1}{2}$, 31 weeks and 1 day.
- UC, University of Chicago.
- Ap, ampicillin; Cx, cefotaxime; Gm, gentamicin.
- GE, gastroesophageal; AOP, anemia of prematurity; IUGR, intrauterine growth restriction.
- PPROM, preterm premature rupture of membranes.
- A third antibiotic may have been given (non-UC chart unclear/unavailable).
- The most likely treatment duration (non-UC chart unclear/unavailable).
- Including various syndrome-associated medical problems.
intubation (babies 4 and 5) or had been extubated by the time sampling commenced (babies 1, 2, and 6). Baby 3 was also treated with antibiotics on days 13 to 19 for a suspected case of sepsis (Table 2), but all cultures (blood, urine, and cerebrospinal fluid [CSF] samples) were negative; no respiratory tract samples were cultured. Baby 3 was the only subject to receive antibiotics during the sampling period. Finally, in some cases, modifications to the infants’ feeding regimens and/or hospital locations were made during the sampling period (Table 2). Most feedings were delivered via nasogastric or orogastric tube.

Baby 3 received antibiotics for a suspected case of NEC around day of life (DOL) 40, but his clinical signs resolved quickly without further intervention. To our knowledge, none of the other infants went on to have invasive infections or NEC after DOL 21.

**Overview of bacterial taxonomic representation.** Of the 108 samples collected for the study, 106 yielded sufficient quantities of 16S rRNA gene sequences to warrant subsequent analysis (range, 219 to 1,914 sequences/sample; median, 1,066 sequences/sample). Due to low sequencing yielding, two samples were dropped.

Overall, nine bacterial phyla were represented (Fig. 1). On average, the most abundant were the **Firmicutes** (71.6%), **Proteobacteria** (21.4%), **Bacteroidetes** (5.4%), **Tenericutes** (1.0%), and **Actinobacteria** (0.5%). Rare phyla (those with average abundances of ≤0.01%) included the **Cyanobacteria**, **Deinococcus-Thermus**, **Chloroflexi**, and **Fusobacteria**. In total, 119 bacterial genera were detected, the most abundant of which are displayed in Fig. 1. Dominant genera were as follows: from the phylum **Firmicutes**, **Staphylococcus**, **Streptococcus**, **Enterococcus**, and **Gemella**; from the class **Gamma-proteobacteria**, **Klebsiella/Enterobacter** (genera indistinguishable using the available gene fragment), **Haemophilus**, **Citrobacter**, **Proteus**, and **Pseudomonas**; and from the phylum **Bacteroidetes**, the genus **Bacteroides**.

The identities of the abundant taxa found here are generally consistent with those observed in prior studies of LBW infants (22,
including studies of premature infants recruited from the same NICU as that which served as the setting for the current study (23, 27, 29).

**Microbiota composition is primarily shaped by body site.** Patterns of bacterial community-wide compositional variation were evaluated using the unweighted UniFrac metric. Pairs of samples containing similar (i.e., closely related) lineages have relatively small UniFrac distances, whereas those containing divergent (i.e., distantly related) lineages have relatively large ones (34). The unweighted UniFrac metric is incidence based (i.e., presence/absence of taxa).
Neonatal personalization of microbiota composition. Compositional variation existed among the LBW infants (PERMANOVA main test, \( P < 0.001 \)), but the effect of “individual” (PERMANOVA \( \eta^2 = 0.13 \)) was smaller than the effect of “body site” \( (\eta^2 = 0.21) \) (Fig. 2). It was also not the case that every baby harbored a highly personalized microbiota: in pairwise a posteriori tests, the microbiota of babies 1 and 2 (the dizygotic [DZ] twins) were compositionally similar to each other and to the microbiota of baby 5 \( (P \text{ values of } >0.05) \). By day 21, the genus-level profiles for the fecal bacterial communities of co-twins were remarkably similar (Fig. 1c); as follows, overall interindividual variability for the distal gut decreased modestly as the cohort grew older (see Fig. S3 in the supplemental material). Throughout their hospitalization, co-twins were generally colocated; however, specific aspects of their care may have varied. For example, on DOL 21, babies 1 and 2 (the DZ twins) received different diets (Table 2).

The relative abundance of three genera differed significantly across the three body sites (ANOVA adjusted for Bonferroni’s correction, \( P < 0.001 \)). Among those with an average abundance of \( >1.0\% \), Klebsiella/Enterobacter (genera indistinguishable using the available gene fragment), Enterococcus, and Citrobacter were particularly abundant in stool, as was Staphylococcus (largely Staphylococcus epidermidis) on skin, and Streptococcus in saliva (Fig. 1). Controlling for sequencing effort, the number of operational taxonomic units (OTUs) on skin was significantly higher than the number in saliva or stool (see Fig. S2 in the supplemental material).

Notably, Staphylococcus and Streptococcus, which are characteristically found on skin and in saliva, respectively, were surprisingly abundant at other sites in the LBW infants (Fig. 1), and the level of body site-driven compositional differentiation in the LBW infants (as shown in Fig. 2) seemed lower than that reported for healthy adults (1–3). Indeed, when we compared these groups directly (see Fig. S1 in the supplemental material), we found that the effect of “body site” was smaller in LBW infants (PERMANOVA \( \eta^2 = 0.21 \)) than in healthy adults \( (\eta^2 = 0.34) \). This direct comparison also revealed that, among the three sites examined, LBW infant skin was the most adult-like in terms of microbiota composition (Fig. 3).

The relative abundance of three genera differed significantly among the six infants (ANOVA adjusted for Bonferroni’s correction, \( P < 0.001 \)). Bacteroides \( (B. \text{ caccae}) \) was particularly abundant in baby 6’s stool samples (at all ages), as was Proteus \( (P. \text{ mirabilis}) \) in baby 3’s saliva and stool samples (early ages), and Haemophilus \( (H. \text{ parainfluenzae}) \) in baby 4’s saliva and stool samples (early ages; also present at low abundance in monzygotic [MZ] co-twin) (Fig. 1). Baby 6 was the only term infant in the study; he was also delivered vaginally. Numerous studies link vaginal delivery to early colonization by Bacteroides (35, 36).

A high degree of interindividual variation in fecal microbiota composition has been observed in preterm (23, 24, 28) and term (5, 7) infants. Our data suggest that this pattern extends to the neonatal skin and oral microbiota. The ultimate cause of interindividual variation may be difficult to ascertain—e.g., despite receiving remarkably similar medical treatment (Tables 1 and 2),...
clear differences existed between the microbiomes of the MZ co-twins (Fig. 1).

Delayed compositional divergence of gut and oral communities. As a categorical predictor, infant age was not associated with differences in bacterial community composition (PERMANOVA main test, $P = 0.935$), and the relative abundance of only one genus (Staphylococcus) changed consistently as the cohort grew older (modest decline in stool and saliva; linear correlation, $r = -0.485$ and $-0.387$, respectively; adjusted for Bonferroni’s correction, $P = 0.011$ and 0.064, respectively). Thus, microbiota composition was more stable over time (here, DOL 8, 10, 12, 15, 18, and 21) than across body sites and host individuals.

Next, we addressed whether the degree of body site-associated compositional differentiation depended on the age of the infant. We found that at all ages, microbiota composition on skin was significantly different from that in saliva and stool; however, we also found that the microbiota compositions of saliva and stool were not significantly different from each other until the babies were at least 15 days old (Table 3). Indeed, saliva and stool compositions grew progressively more distinct as the infants grew older (Table 3). Furthermore, on average within infants, the compositional difference between saliva and stool samples increased significantly with infant age (linear regression, $R^2 = 0.7075$, $P = 0.0359$). We asked whether this divergence was driven by compositional turnover in the distal gut, oral cavity, or both. Pairwise a posteriori tests mainly implicated the distal gut, where the amount of variation explained by time was positively correlated with the size of the time step (see Table S1 in the supplemental material)—a pattern that was not as apparent in saliva or on skin (Table S2). These results were well supported by correlation tests, which further emphasize that the temporal pattern of neonatal microbiome assembly depends on the observed body site (Table 4).

### Table 3 Effect of body site on the composition of LBW infant-associated bacterial communities over time

| Test         | Age | Body sites (age) | $F/t$   | $P$ value | No. of permutations |
|--------------|-----|-----------------|---------|-----------|---------------------|
| Main test    | All |                 | 2.9245  | 0.001     | 996                 |
| Pairwise tests | DOL 8 | Saliva, skin   | 1.9426  | 0.002     | 402                 |
|              |     | Saliva, stool   | 0.90264 | 0.696     | 407                 |
|              |     | Skin, stool     | 2.0107  | 0.005     | 405                 |
|              | DOL 10 | Saliva, skin  | 2.1699  | 0.001     | 402                 |
|              |     | Saliva, stool   | 0.96292 | 0.544     | 401                 |
|              |     | Skin, stool     | 2.0022  | 0.001     | 407                 |
|              | DOL 12 | Saliva, skin  | 1.4472  | 0.013     | 414                 |
|              |     | Saliva, stool   | 1.0734  | 0.314     | 405                 |
|              |     | Skin, stool     | 1.9793  | 0.004     | 407                 |
|              | DOL 15 | Saliva, skin  | 1.8748  | 0.001     | 409                 |
|              |     | Saliva, stool   | 1.416   | 0.025     | 411                 |
|              |     | Skin, stool     | 1.9982  | 0.004     | 416                 |
|              | DOL 18 | Saliva, skin  | 1.5016  | 0.014     | 399                 |
|              |     | Saliva, stool   | 1.7443  | 0.010     | 416                 |
|              |     | Skin, stool     | 2.1177  | 0.006     | 411                 |
|              | DOL 21 | Saliva, skin  | 1.5612  | 0.004     | 395                 |
|              |     | Saliva, stool   | 1.6077  | 0.006     | 407                 |
|              |     | Skin, stool     | 2.2291  | 0.001     | 402                 |

a Results of main and pairwise a posteriori using unweighted UniFrac-based permutational multivariate ANOVA and the $t$ statistic. DOL, day of life.

b The model is “body site” nested within levels of “age.”

c The $P$ values in bold type highlight the divergence of saliva and stool sample values over time.

### Table 4 Correlation between the compositional dissimilarity of LBW infant-associated bacterial communities and elapsed time

| Parameter | Body site | Saliva | Skin | Stool |
|-----------|-----------|--------|------|-------|
| No. of pairs | 85        | 85     | 90   |
| Spearman’s $r$ | 0.2919    | 0.0351 |
| 95% confidence interval | -0.1856 to 0.2524 | 0.2301 to 0.5831 |
| $P$ value (one-tailed) | 0.0034     | 0.7500 | <0.0001 |

Spearman’s rank order correlation measuring the dependence of community distance (unweighted UniFrac metric) on temporal distance (number of days) within subjects for each body site.
Relationship between neonatal stool microbiota composition and time. (a) Average (95% CI) within-subject, unweighted UniFrac distance plotted against the age difference (lag) in days for LBW ($R^2 = 0.5$) and NBW ($R^2 = 0.5$) infants. Lines indicate best fit linear regressions. The NBW infants were not sampled on day 18; thus, for this analysis, the corresponding age was excluded from the LBW data set. (b) For LBW infants at various ages, average (95% CI) unweighted UniFrac distance compared to a healthy reference cohort (the 21-day-old NBW infants).

FIG 4  Relationship between neonatal stool microbiota composition and time. (a) Average (95% CI) within-subject, unweighted UniFrac distance plotted against the age difference (lag) in days for LBW ($R^2 = 0.5$) and NBW ($R^2 = 0.5$) infants. Lines indicate best fit linear regressions. The NBW infants were not sampled on day 18; thus, for this analysis, the corresponding age was excluded from the LBW data set. (b) For LBW infants at various ages, average (95% CI) unweighted UniFrac distance compared to a healthy reference cohort (the 21-day-old NBW infants).

We also found that there was no significant difference between the cohorts with respect to the rate of compositional turnover (Fig. 4a, $P = 0.7911$). On average, the stool microbiotas of LBW infants were slightly enriched in the observed number of OTUs (controlling for sequencing effort; $P = 0.01$) and significantly enriched in OTUs assigned to Enterobacteriaceae, Enterococcus, and Staphylococcus (adjusted for Bonferroni’s correction, $P$ values of $<0.001$). Escherichia was abundant in the NBW infants and virtually absent from the LBW infants ($P < 0.001$). Despite these differences, over time, the community-wide composition of LBW infant stool grew more similar to that of 21-day-old NBW infant stool (i.e., to that of a healthy reference group; Fig. 4b). These results suggest that while gestational age at delivery, delivery mode, or other factors may affect gut microbiota makeup, its rate of development may depend more on intrinsic community-level factors, e.g., the amount of time the site has been available to colonists, microbe-microbe interactions, microbe-host interactions (that are independent of host gestational age), or increasing hypoxia/anaerobiosis.

Dynamics of particular taxa in LBW infants, including an uncultivated Mycoplasma. Several noteworthy taxa were briefly abundant in LBW infant stool samples (Fig. 1c). On day 18, Clostridium perfringens represented ~40% of sequences from baby 2, but it was below the detection level on all other days. On day 15, Dysgonomonas capnocytophagoides comprised ~8% of sequences from baby 3; this fastidious organism (and opportunistic pathogen) has not, to our knowledge, been reported in pediatric clinical samples. Finally, on day 15, a Peptoniphilus sp. represented ~7% of sequences from baby 4, having been detected previously in his day 12 skin swab (1%; Fig. 1b)—a possible bellwether for the taxon’s appearance in the distal gut.

However, the most striking example emerged from the oral data set and involved taxa from baby 3’s saliva samples: specifically, the genera Mycoplasma (several species) and Pseudomonas (P. aeruginosa), which became dominant on days 15, 18, and 21 (Fig. 1a). Indeed, the sequences comprising one, highly abundant Mycoplasma-related OTU appeared to be phylogenetically novel. This finding prompted an in-depth analysis of these and related sequences belonging to the phylum Tenericutes.

Among the OTUs detected in the LBW infants, three were assigned to the phylum Tenericutes; together, they contained 788 sequences. Representatives of the first and second OTUs were >99% identical to Mycoplasma hominis and Ureaplasma parvum, respectively. However, the representative of the third OTU, which contained 771 sequences, was only 88% identical to the closest named species in GenBank (e.g., Mycoplasma iowae, Mycoplasma microti, and Mycoplasma muris). This novel OTU was virtually exclusive to baby 3, the only extremely LBW (ELBW) infant in the study (ELBW is defined as <1.0 kg). Its expansion in baby 3’s oral cavity, which peaked on DOL 18 at 47.2% of sequences, coincided with antibiotic treatment for suspected (but ultimately unconfirmed) sepsis (Fig. 5 and Table 2).

Phylogenetic analysis suggests that the novel OTU belongs to a single, well-supported clade comprising uncultivated lineages from cow rumen, which are among its closest relatives at 94.3 to 94.8% sequence identity, and termite gut (see Fig. S4 in the supplemental material). Interestingly, a recently deposited GenBank (uncultured Mycoplasma sp. clone Mnola; accession no. JX508800) is 99% identical to our infant-derived OTU (Fig. S4); this clone was isolated from a vaginal swab from a Trichomonas vaginalis-infected patient (37). Finally, we amplified and cloned near-full-length 16S rRNA gene sequences from baby 3’s DOL 18 saliva (see Materials and Methods). This yielded sequences belonging to the novel OTU that confirmed the phylogenetic place-
Among the three sites examined, LBW infant skin was the most adult-like in terms of microbiota composition (Fig. 3); this may result from infant skin being more selective for, and/or more heavily exposed to, the skin microbiota of adult caretakers in the NICU compared to other body sites (33), although we did not quantify the amount of time each infant spent in direct contact with mothers or other caregivers. (In the mouth and gut, the main difference between neonates and adults seems to be a relative lack of strict anaerobes [38].) While developmental changes over the first year of life have been reported for the infant skin microbiome (8), they were not apparent within the relatively short, neonatal time frame of the current study (Table 4).

Finally, delivery mode has been noted to exert a strong influence on the composition of the newborn microbiota (4); while this effect was conceivably manifest in our study (e.g., *Ureaplasma* in baby 3; *Bacteroides* in baby 6 [Fig. 1] [36, 39]), its pervasiveness and persistence will require examination in larger cohorts of high-risk infants.

We found that microbiota composition was relatively stable over time within LBW neonates. This small effect size for time, compared to those for body site and host individual, is also consistent with patterns observed in healthy adults (1–3, 40). Nonetheless, our comparative approach uncovered subtle yet important temporal changes that occurred over the 8- to 21-day age range: in particular, a gradual (i.e., delayed) compositional divergence of the oral and fecal microbiota (Table 3), largely driven by progressive temporal turnover in the distal gut (Table 4), the latter of which proceeded at a rate indistinguishable from that of age-matched NBW infants (Fig. 4a). Long recognized as a key process taking place in early infancy (38, 41–43), our study draws into focus the initiation phase of gut microbiome development, capturing, possibly, the time span over which the site begins to receive and select for gut-specific microbes, which may then grow to outnumber or outcompete transient or generalist immigrants from the oral cavity (or other sources shared by the two sites). However, given our small cohort of six infants for which there were a number of uncontrolled variables (e.g., gestational age at delivery, multiple gestation, medical treatment, delivery mode), we caution that our data are likely limited in terms of their generalizability and capacity to detect subtle effects. The biogeographic patterns we report warrant follow-up in larger, well-controlled, prospective cohort studies.

We also detected a novel, uncultivated lineage of *Mycoplasma* at high abundance in the oral cavity of ELBW baby 3. *Mycoplasma* and *Ureaplasma* spp. colonize the human respiratory and urogenital tracts, and some play roles as perinatal pathogens (39). *M. hominis* and *Ureaplasma* spp. can cause chorioamnionitis (a risk factor for preterm premature rupture of membranes [PPROM]) and pass from mother to newborn, and the latter organisms have been associated with preterm labor and low birthweight (39, 44). In neonates, they cause respiratory, blood, and central nervous system (CNS) infections (39). Lacking cell walls, these organisms are innately resistant to beta-lactam (e.g., ampicillin, cefotaxime) and glycopeptide (e.g., vancomycin) antibiotics (45). Although not innately resistant, their susceptibility to aminoglycosides (e.g., gentamicin) is variable (46).

Baby 3 was delivered vaginally after PPROM at ~24.5 weeks of completed gestation and was treated intravenously with ampicillin and gentamicin for the first 7 days of life. Thus, carriage of *Mycoplasma*- and *Ureaplasma*-related OTUs at low abundance at the start of the study, on DOL 8, may have been due to vertical transmission at delivery, followed by resistance to the initial course of antibiotics, although alternative scenarios are possible (e.g., later exposure in the NICU). Baby 3 was again treated with antibiotics (vancomycin, gentamicin, cefotaxime) from DOL 13 to 19 (Table 2), and this coincided with a marked increase in the proportional abundance of *Pseudomonas aeruginosa* (Fig. 1a) and OTU 15, a member of a novel, uncultivated clade belonging to the *Mycoplastomataceae*, in baby 3’s oral samples (Fig. 5; see Fig. S4 in the supplemental material). Intriguingly, a recent study found
high abundances of this uncultivated Mycoplasma in the vaginal microbiota of Trichomonas vaginalis-infected women (detected in 19/30 T. vaginalis-infected and 1/29 uninfected individuals) (37), again raising the possibility that this organism too was transferred from mother to infant at delivery. Further investigation into the diversity, distribution, and clinical significance of this novel, uncultivated Mycoplasma in human hosts is warranted, particularly in pregnant women and premature infants.

Although the LBW infants in this study were relatively free of major medical problems, we found that their microbiomes were dominated at times by bacterial taxa that have been associated with neonatal infections and NEC, e.g., Staphylococcus, C. perfringens, P. aeruginosa, and others (28, 32, 47, 48). Yet, despite the abundance of taxa with pathogenic potential, it appears that certain normal processes were under way, including the development of body site-specific bacterial communities and progressive compositional turnover in the distal gut, as observed in healthy hosts (2, 38). Our analysis was cohort based; however, it might be useful to know whether individual infants vary in the precise timing of body site-associated compositional differentiation, and if so, whether such variation depends on gestational age at delivery or particular NICU management protocols. Unfortunately, our cohort was not well suited to this analysis because of its small size, but also because gestational age at delivery was confounded with delivery location and the amount of time spent in the NICU (Tables 1 and 2). This underscores a need for larger and distinct cohorts but also highlights a challenge: the smallest, most premature infants will almost always require the most intensive medical support, thus entangling factors such as gut and immune immaturity with, for example, the number of invasive procedures or days on antibiotics. Nevertheless, monitoring of oral and other potential source communities in the NICU might be particularly warranted during the time the gut microbiome remains “undifferentiated” and, possibly, more open to invasion.

**MATERIALS AND METHODS**

**Patients and sample collection.** Six low-birthweight (LBW) infants were recruited from a level III NICU at the University of Chicago Comer Children’s Hospital. The infants were born within 1 week of each other in the summer of 2010. The cause of the low birthweight was preterm delivery in five of the infants (a singleton and two pairs of twins) and fetal growth restriction in the sixth. Birth weights ranged from 0.75 to 1.82 kg (see Table 1 for clinical details; <2.5 kg is considered low birthweight). Stool and saliva samples and skin swabs were obtained from each infant on postnatal days 8, 10, 12, 15, 18, and 21. The age range of 8 to 21 days was selected because it may represent a critical window for the colonization of the infant, and although it did not occur in the present cohort, for the onset of NEC. Stool sampling involved manual perineal stimulation with a lubricated cotton swab, which induced prompt defecation. Oral and skin samples were collected by gently swabbing the dorsum of the tongue and the anterior upper chest wall, respectively. For the oral samples, we simply call the collected materials “saliva,” because it is likely that multiple sites were contacted during the gentle swabbing. Samples were collected using sterile nylon or cotton swabs, placed in 3 ml of universal transport medium (UTM; EMD Millipore, Billerica, MA), and promptly frozen at −80°C. A total of 108 samples were collected for the study. Data pertaining to the care and location of the infants during the sampling period are presented in Table 2. All infants remained hospitalized throughout the study. The Institutional Review Board of the University of Chicago approved the study protocol, and the infants’ parents provided written informed consent.

**DNA extraction, PCR amplification, and pyrosequencing.** Genomic DNA was isolated from each sample (1.5 ml UTM) using a QiAmp DNA stool minikit (Qiagen, Valencia, CA) with modifications, including bead beating (49). A fragment of the 16S rRNA gene spanning the V3-V5 hypervariable regions was amplified. The forward primer (5’ CTG ATC GCC TCC TCT GCG OCA TCA GNN NNN NNN NN G CGT ACT CCT ACG GGA GGC AGC A 3’) contained the 454 Life Sciences primer A sequence, a unique 12-nucleotide (nt) error-correcting Golay barcode used to label each amplicon (designated by the N’s) (50), the broad-range bacterial primer 338F (F stands for forward), and a two-base linker located between the bar code and the rRNA primer (GC). The reverse primer (5’ CTG TTC CCG TTG CCA GGC CGC TCA GAA CCG TCA ATT CCT TGT AGT TT 3’) contained the 454 Life Sciences primer B sequence, a two-base linker (AA), and the broad-range bacterial primer 906R (R stands for reverse). Amplifications were carried out in triplicate 25-μl reactions using 0.4 μM forward and reverse primers, 3-μl template DNA, and 1× HotMasterMix (5 PRIME, Gaithersburg, MD). Bovine serum albumin (BSA) was added at a final concentration of 0.1 μg/μl to reaction mixtures containing fecal DNA. Thermal cycling was carried out at 94°C for 2 min, followed by 35 cycles, with 1 cycle consisting of 94°C for 45 s, 50°C for 30 s, and 72°C for 90 s, with a final extension step of 10 min at 72°C. Replicate reactions were pooled and then purified using an Ultra-Clean-htp 96-well PCR clean-up kit according to the manufacturer’s instructions (MO BIO, Carlsbad, CA).

DNA concentrations were determined using a high-sensitivity Quant-IT double-stranded DNA (dsDNA) kit according to the manufacturer’s instructions (Invitrogen, Carlsbad, CA). Purified amplicons were combined in equimolar ratios into a single tube, ethanol precipitated, and resuspended in 100 μl of nuclease-free water. The pooled DNA was gel purified and recovered using a QiAquick gel extraction kit (Qiagen). Uni-directional amplicon sequencing was performed by the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois, Urbana-Champaign using a 454 Life Sciences genome sequencer FLX instrument, titanium (Ti) series reagents, primer A, and 6 regions of a 16-region gasket (Roche, Branford, CT). Sequencing generated 186,428 raw reads.

**Sequence analysis.** Raw reads were filtered using the QIIME software package (51). Reads were removed from the analysis if they were <200 or >600 nt in length, contained an ambiguous base, had a mean quality score of <25 across the entire read, contained a homopolymer run >6 nt in length, did not contain the forward primer sequence, or contained an uncorrectable barcode. Remaining reads were truncated at the first base of the first 50-nt sliding window with a mean quality score of <25 (if found), and retained unless <200 nt in length after truncation. Filtered reads were assigned to samples by examining the 12-nt barcode. A total of 119,191 filtered reads were associated with samples at this step (mean read length, 535 nt).

Error correction, chimera detection (using UCHIME), and clustering of filtered reads into de novo operational taxonomic units (OTUs) at 97% sequence identity were performed in USEARCH using outuipipe-like scripts enabled in QIIME (http://www.drive5.com/usearch/manual/otu_clustering.html) (52, 53). A representative sequence was chosen from each OTU by selecting the “first” sequence (i.e., the UCLUST cluster seed). Representative sequences were aligned against the Greengenes core set (54) using PyNAST (55) with a minimum alignment length of 150 nt and a minimum identity of 80%. Fifteen OTU representative sequences failed to align; BLASTn searches against GenBank’s nr/nt database revealed 13 human OTUs, 1 Candida albicans OTU (representing 276 reads from baby 6, day 8 stool), and 1 poor-quality OTU, all of which were excluded from further analysis. Taxonomic assignments were made using the Ribosomal Database Project (RDP) classifier version 2.2 with a minimum support threshold of 80% and the RDP taxonomic nomenclature database (56). For the most abundant OTU’s study-wide (here, those with >0.05% average abundance across all samples), RDP assignments were manually confirmed and, when possible, annotated with species-level information.
using BLASTn searches against the nr/nt database. A table of sequence counts per classified OTU × sample was generated in which the criteria for an OTU’s inclusion were that it contained at least 2 sequences and was assigned at least to the genus level. The final OTU table consisted of 321 OTUs containing a total of 105,462 sequences.

Sequences representing OTUs that did not make it into the final table were removed from the alignment. Hypervariable (i.e., uninformative) positions were then excluded using the PH Lane mask (57). A phylogeny was inferred using FastTree version 2.1.3 (58) with the Jukes-Cantor plus CAT model. The final OTU table and phylogeny served as inputs to subse-
quent analyses, including rarefaction, $\alpha$ and $\beta$ diversity calculations, unweighted UniFrac-based principal coordinate analysis (PCoA), and phyllum- and genus-level taxonomic summaries implemented in QIIME. Unweighted UniFrac-based permutational multivariate analysis of variance (PERMANOVA) was performed in PRIMER-E version 6 (59). Other statistical tests were performed in QIIME or Prism (GraphPad Software, Inc.).

**Sequence analysis focused on a novel, uncultivated oral Mycoplasma.** Phylogenetic relationships among sequences belonging to OTUs assigned to the phylum Tenericutes (3 OTUs) were investigated in detail. This analysis was prompted by the identification of an OTU assigned to the genus *Mycoplasma* containing 771 reads (99% of which were from baby 3’s saliva) and exhibiting low sequence identity (~88%) to the most closely related cultivated strains represented in GenBank (http://www.ncbi.nlm.nih.gov/genbank/). Sequences were aligned against the Green-genes core set using the NAST algorithm (60) (http://greengenes.lbl.gov) and imported into ARB (version 08.08.27) (61). In ARB, the alignment was manually improved using secondary structure information and align-
ment to nearest neighbors in the context of an expanded, in-house data-
base founded upon the Greengenes alignment. Phylogenetic relationships among the 3 OTUs found in the present study, their closest relatives (uncultivated mycoplasmas), and selected representatives of cultivated *Tenericutes* were inferred using bootstrapped maximum likelihood infer-
ence methods in RAxML (version 7.2.8) (62). In order to confirm and in-
ferricutes were inferred using bootstrapped maximum likelihood infer-
ence methods in RAxML (version 7.2.8) (62). In order to confirm and in-
ner of the phylogenetic placemat of the novel *Mycoplasma*-related OTU, a small number of near-full-length 16S rRNA gene se-
quences were recovered from baby 3’s day 18 saliva sample via amplifica-
tion (with primers 8F/1391R), cloning, and Sanger sequencing using methods described elsewhere (63). Fifteen high-quality sequences were assembled (4 uncultivated *Mycoplasma* sequences and 11 *Pseudomonas aeruginosa* sequences). The near-full-length *Mycoplasma* sequences were analyzed using NAST, ARB, and RAxML as described above.

**Comparison to microbiota of NBW infants via pyrosequencing of archived stool DNA.** Archived stool DNA samples from healthy, age-matched (i.e., time series spanning 8 to 21 days in age), normal birth-
weight (NBW) (>2.5 kg) infants enrolled in a prior study (5) were ampli-
fied, sequenced, and analyzed using the pyrosequencing and bioinformat-
ics approaches described herein. The archived DNA had been isolated using the QIAamp stool DNA minikit (Qiagen) and stored at −80°C. The Stanford University Administrative Panel on Human Subjects in Medical Research approved this work, and the infants’ parents provided written informed consent.

**Comparison to microbiota of healthy adults using publicly avail-
able sequence data.** Sequence data from the LBW and NBW infants were compared to publicly available sequence data from the corresponding body sites of healthy adults. Adult data were selected from two published studies that used pyrosequencing approaches similar to those used here (1, 2). From the first study (1), we selected samples from 7 adults (3 female), “days 1 and 2” (of 4 sampling dates), including dorsal tongue swabs, skin swabs (forehead and right forearm), and stool samples (36 samples in total). These 16S rRNA gene sequences were V2 region FLX reads originating from the distal primer (338F). From the second study (2), we selected samples from 6 adults (a subset chosen at random but matched for gender to the LBW infants), “visit 2” (of 2 sampling visits), including saliva samples, skin swabs (right retroauricular crease), and stool samples (18 samples in total). These were V3-V5 region Ti reads originating from the distal primer (926R). By comparison, the infant-
derived sequences generated for the present study were V3-V5 region Ti reads originating from the proximal primer (338F). Thus, given the dif-
ferences in sequence length and sequenced region among the data sets, the pooled sequences were trimmed to a length of not more than 300 nt and OTUs were picked against a set of reference sequences. This was accom-
plished in QIIME using uclust_ref-based OTU picking against the Green-
genes gg_97_otus_4feb2011.fasta reference set at an identity threshold of 95% (relaxed from 97% to allow for greater recruitment), with reverse strand matching enabled and no new clusters allowed. A total of 3,158 reference OTUs were detected; these encompassed 96% of the 475,080 total sequences. Rarefied and un rarefied OTU tables, along with a refer-
ce tree (gg_97_otus_4feb2011.1re), were used to calculate unweighted UniFrac distance matrices, which served as inputs for PCoA in QIIME.

The DNA extraction method varied among the studies compared: studies of adults used a MO BIO kit, while studies of infants used a Qiagen kit. To investigate potential kit-associated bias, we pyrosequenced 16S rRNA genes amplified from archived adult stool DNA that had been isolated using a Qiagen kit (from the NBW infants’ fathers [5]; mothers were excluded due to possible pregnancy-associated shifts in microbiota com-
position [64]). These new adult sequences were filtered as described herein and trimmed to a length of not more than 300 nt, pooled with the other sequences, and analyzed as described in the preceding paragraph. Because the Qiagen-extracted adult stool samples (5) clustered with the MO BIO-extracted ones (1, 2) (see Fig. S1 in the supplemental material), we concluded that DNA extraction kit did not grossly bias the results of the unweighted UniFrac-based PCoA.

**Nucleotide sequence accession numbers.** The sequence data gener-
ated for this study were deposited in the QIIME database (study identifi-
cation numbers 2042 and 2046).

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at http://mbio.asm.org /lookup/suppl/doi:10.1128/mBio.00782-13/-/DCSupplemental.

Figure S1, PDF file, 0.4 MB.
Figure S2, PDF file, 0.2 MB.
Figure S3, PDF file, 0.2 MB.
Figure S4, PDF file, 0.4 MB.
Table S1, DOCX file, 0.1 MB.
Table S2, DOCX file, 0.1 MB.

**ACKNOWLEDGMENTS**

We thank the subjects and their families for their participation and the NICU staff for their support. We also thank members of the Morowitz and Relman laboratories, and in particular Valeriy Poroyko (University of Chicago) for technical assistance and Diana Proctor (Stanford University) for critical review of the manuscript.

This work was supported in part by NIH grant 1R01AI092531-01 (M.J.M. and D.A.R.; Jill Banfield, principal investigator [PI]), a Walter V.
and Idun Berry postdoctoral fellowship (E.K.C.), NIH Pioneer award DP1OD009964 (D.A.R.), March of Dimes Foundation research grant 5-FY10-103 (M.J.M.), the March of Dimes Prematurity Research Center at Stanford University School of Medicine (D.A.R.), and by the Thomas C. and Joan M. Merigan Endowment at Stanford University (D.A.R.).

**REFERENCES**

1. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JJ, Knight R. 2009. Bacterial community variation in human body habitats across space and time. Science 326:1694–1697.
2. Microbiome, Human Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214.
3. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JJ, Knight R. 2011. Moving pictures of the human microbiome. Genome Biol. 12: R50. doi:10.1186/gb-2011-12-3-r50.
4. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo...
González-Rivera R, Culverhouse RC, Hamvas A, Tarr PI, Warner BB.

Neu J, Walker WA.

Morowitz MJ, Poroyko V, Caplan M, Alverdy J, Liu DC.

Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, Glickman JN, Cho I, Yamanishi S, Cox L, Methé BA, Zavadil J, Li K, Gao Z, Mahana.

Carlisle EM, Morowitz MJ.

Manzoni P, De Luca D, Stronati M, Jacqz-Aigrain E, Ruffinazzi G, Chang JY, Shin SM, Chun J, Lee JH, Seo JK.

Mshvildadze M, Neu J, Shuster J, Theriaque D, Li N, Mai V.

Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc. Natl. Acad. Sci. U. S. A. 108:1128–1133.

Madan JC, Salari RC, Saxena D, Davidson I, O’Toole GA, Moore JH, Sogin ML, Foster JA, Edwards WH, Palumbo P, Hibberd PL. 2012. Gut microbiota colonization in premature neonates predicts neonatal sepsis. Arch. Dis. Child. (Fetal Neonatal Ed.) 97:F546–F462.

Morowitz MJ, Poroyko V, Culverhouse RC, Hamvas A, Tarr PI, Warner BB. 2010. Fecal microbiota in premature infants prior to necrotizing enterocolitis. PLoS One 6:e20647. doi:10.1371/journal.pone.0020647.

Morowitz MJ, Denef VJ, Costello EK, Thomas BC, Poroyko V, Relman DA, Banfield JF. 2011. Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. Proc. Natl. Acad. Sci. U. S. A. 108:1128–1133.

November/December 2013 Volume 4 Issue 6 e00782-13

Gut, Oral, and Skin Microbiome Dynamics in LBW Neonates
48. Stewart CJ, Marrs EC, Magorrian S, Nelson A, Lanyon C, Perry JD, Embleton ND, Cummings SP, Berrington JE. 2012. The preterm gut microbiota: changes associated with necrotizing enterocolitis and infection. Acta Paediatr. 101:1121–1127.

49. Zoetendal EG, Heilig HG, Klaassens ES, Booijink CC, Kleerebezem M, Smidt H, De Vos WM. 2006. Isolation of DNA from bacterial samples of the human gastrointestinal tract. Nat. Protoc. 1:870–873.

50. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat. Methods 5:233–237.

51. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pettig AG, Goodrich JK, Gordon JJ, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrunn M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7:335–336.

52. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

53. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27: 2194–2200.

54. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72:5069–5072.

55. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26:266–267.

56. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73:5261–5267.

57. Lane DJ. 1991. 16S/23S rRNA sequencing, p 115–148. In Stackebrandt E, Goodfellow M (ed), Nucleic acid techniques in bacterial systematics. John Wiley & Sons, Chichester, United Kingdom.

58. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. doi:10.1371/journal.pone.0009490.

59. Clarke KR, Gorley RN. 2006. PRIMER v6: user manual/tutorial. PRIMER-E, Plymouth, United Kingdom.

60. DeSantis TZ, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM, Phan R, Andersen GL. 2006. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res. 34:W394–W399.

61. Ludwig W, Strunk O, Westram R, Richter I, Meier H, Kumar Y, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Breitske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüssmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Völlig A, Lenke M, Ludwig T, Bode A, Schleifer KH. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32:1363–1371.

62. Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57:758–771.

63. Bik EM, Long CD, Armitage GC, Loomer P, Emerson J, Mongodin EF, Nelson KE, Gill SR, Fraser-Liggett CM, Relman DA. 2010. Bacterial diversity in the oral cavity of 10 healthy individuals. ISME J. 4:962–974.

64. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Bäckhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R, Bäckhed F, Isolauri E, Salminen S, Ley RE. 2012. Host remodeling of the gut microbiome and metabolic changes during pregnancy. Cell 150:470–480.