AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

Highly Specialized Carbohydrate Metabolism Capability in *Bifidobacterium* Strains Associated with Intestinal Barrier Maturation in Early Preterm Infants

Bing Ma,[a,b] Sripriya Sundararajan,[c] Gita Nadimpalli,[d] Michael France,[a,b] Elias McComb,[a] Lindsay Rutt,[a] Jose M. Lemme-Dumit,[c,e] Elise Janofsky,[c] Lisa S. Roskes,[c] Pawel Gajer,[a,b] Li Fu,[a] Hongqiu Yang,[a] Mike Humphrys,[a] Luke J. Tallon,[a] Lisa Sadzewicz,[a] Marcela F. Pasetti,[b,c,e] Jacques Ravel,[a,b] Rose M. Viscardi[c]

[a]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA
[b]Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA
[c]Department of Pediatrics, University of Maryland School of Medicine, Baltimore, Maryland, USA
[d]Department of Epidemiology, University of Maryland School of Medicine, Baltimore, Maryland, USA
[e]Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, Maryland, USA

ABSTRACT "Leaky gut," or high intestinal barrier permeability, is common in preterm newborns. The role of the microbiota in this process remains largely uncharacterized. We employed both short- and long-read sequencing of the 16S rRNA gene and metagenomes to characterize the intestinal microbiome of a longitudinal cohort of 113 preterm infants born between $24^{0/7}$ and $32^{6/7}$ weeks of gestation. Enabled by enhanced taxonomic resolution, we found that a significantly increased abundance of *Bifidobacterium breve* and a diet rich in mother's breastmilk were associated with intestinal barrier maturation during the first week of life. We combined these factors using genome-resolved metagenomics and identified a highly specialized genetic capability of the *Bifidobacterium* strains to assimilate human milk oligosaccharides and host-derived glycoproteins. Our study proposes mechanistic roles of breastmilk feeding and intestinal microbial colonization in postnatal intestinal barrier maturation; these observations are critical toward advancing therapeutics to prevent and treat hyperpermeable gut-associated conditions, including necrotizing enterocolitis (NEC).

IMPORTANCE Despite improvements in neonatal intensive care, necrotizing enterocolitis (NEC) remains a leading cause of morbidity and mortality. "Leaky gut," or intestinal barrier immaturity with elevated intestinal permeability, is the proximate cause of susceptibility to NEC. Early detection and intervention to prevent leaky gut in "at-risk" preterm neonates are critical for decreasing the risk of potentially life-threatening complications like NEC. However, the complex interactions between the developing gut microbial community, nutrition, and intestinal barrier function remain largely uncharacterized. In this study, we reveal the critical role of a sufficient breastmilk feeding volume and the specialized carbohydrate metabolism capability of *Bifidobacterium* in the coordinated postnatal improvement of the intestinal barrier. Determining the clinical and microbial biomarkers that drive the intestinal developmental disparity will inform early detection and novel therapeutic strategies to promote appropriate intestinal barrier maturation and prevent NEC and other adverse health conditions in preterm infants.

KEYWORDS preterm infant, gut microbiome, leaky gut, intestinal barrier maturation, human milk oligosaccharides, *Bifidobacterium*

Early preterm neonates are particularly vulnerable to life-threatening events and routinely require intensive care and medical intervention to survive (1). The physiological immaturity of their gastrointestinal (GI) tract is commonly associated with

Editor Maria Gloria Dominguez Bello, Rutgers, The State University of New Jersey

Address correspondence to Bing Ma, bma@som.umaryland.edu.

The authors declare no conflict of interest.

This article is a direct contribution from Jacques Ravel, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Steven Gill, University of Rochester School of Medicine and Dentistry, and Henrik Roager, Technical University of Denmark.

Received 6 May 2022
Accepted 11 May 2022
Published 13 June 2022

deficiencies in barrier functions that result in a clinical syndrome known as "leaky gut" (2–5). Under leaky gut conditions, the bacteria and bacterial products normally confined to the intestinal lumen are able to translocate into the peripheral circulation through the hyperpermeable epithelial barrier, which could lead to the widespread invasion of the intestinal epithelium and gut lamina propria, mucosal inflammation, epithelial cell damage, intestinal necrosis, systemic infection, and, ultimately, multiorgan failure and death (4, 6, 7). Necrotizing enterocolitis (NEC) is a prominent bacterial translocation-associated GI condition that affects 7 to 10% of preterm neonates or 1 to 5% of all neonatal intensive care unit (NICU) admissions, with a devastating mortality rate as high as 50% (8–12). The early detection of an aberrant leaky gut and early intervention to limit intestinal injury are of paramount importance to reduce the incidence of subsequent complications, including NEC (12, 13).

A functional intestinal barrier combines a physical barrier that encompasses chemical, immunological, and microbiological components (14). We and others have found that the first week of life (day 8 $\pm$ 2 after birth) is a critical window during which the most rapid postnatal intestinal maturation occurs (15–17). More importantly, these previous studies demonstrated that intestinal barrier function, which develops mostly *in utero* in term infants, can be improved postnatally. They also showed that intestinal barrier maturation does not occur at the same rate, with ~40% of preterm neonates (<33 weeks of gestation) failing to develop a functional intestinal barrier within the first 2 weeks of life (15, 16). Determining the factors that drive this developmental disparity will inform early detection and novel therapeutic strategies to promote intestinal barrier maturation.

Efforts to characterize the microbiological factors that are associated with intestinal barrier maturation have thus far yielded unsatisfactory results (18). There are no microbial biomarkers predictive of intestinal development. A major limitation is the use of partial 16S rRNA gene sequences to evaluate the taxonomic composition of the gut microbiota. Short sequences lack the phylogenetic signal necessary to describe the taxonomic composition at the species or even the genus level. Many of the PCR primers used to amplify variable regions of the 16S rRNA gene fail to amplify members of the genus *Bifidobacterium* (19–21). *Bifidobacterium* species are known to be frequent colonizers of the infant gut (22), are considered to play beneficial roles in intestinal development, and influence the maturation of the neonatal gut, potentially through stimulating colonic epithelial proliferation, modulation of host defense responses, and protection against bacterial infections (23, 24). Investigating *Bifidobacterium* and other bacterial groups predictive of early intestinal development and maturation is of pivotal importance.

In this study, we sought to characterize the role of the early assembly of the infant gut microbiota and its metabolism in postnatal intestinal barrier maturation. We build upon the results of previous studies (15, 16), using an expanded cohort (n = 113) of early preterm neonates (24 weeks and 0 day to 32 weeks and 6 days of gestation) from whom stool samples were collected daily up to 21 days after birth. High-resolution approaches were applied to characterize the composition of the developing gut microbiota with a substantially enhanced taxonomic resolution, including *Bifidobacterium* species, which we identified as the microbial biomarker associated with postnatal intestinal barrier maturation within the first week of life. Whole-community metagenomes using both short- and long-read sequences provided a detailed characterization of the genetic content of these *Bifidobacterium* species, which were shown to have distinct genetic features affording complete carbohydrate-foraging capabilities, including human milk oligosaccharides (HMOs) and host-derived glycoprotein. The presence of specific strains of *Bifidobacterium* may inform the early detection of aberrant intestinal permeability (IP). Supplementation with these bifidobacterial strains could be leveraged in novel intervention strategies for the prevention of leaky gut and its devastating sequelae in preterm newborns.

RESULTS

Clinical cohort. We examined a prospective cohort of 113 preterm infants at $24^{0/7}$ to $32^{6/7}$ weeks of gestation, including 37 subjects described in a previous analysis (see

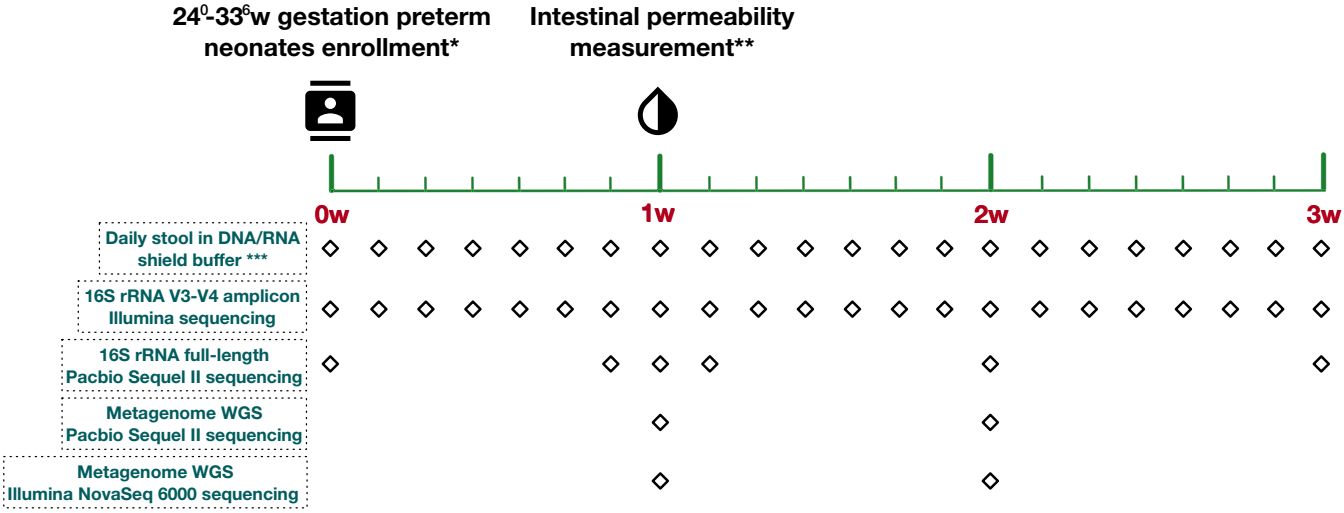

FIG 1 Study design. *, demographic, clinical, and nutritional information was collected for each enrolled preterm neonate. Inclusion criteria include 24^0 to 32^6 weeks and <4 days of age. Exclusion criteria include nonviable or planned withdrawal of care, severe asphyxia, chromosome abnormalities, cyanotic congenital heart disease, intestinal atresia or perforation, abdominal wall defects, significant GI dysfunction, and galactosemia or other forms of galactose intolerance. **, intestinal permeability was measured using the urine non-metabolized sugar probes lactulose and rhamnose at days 7 to 10 after birth. ***, stool specimens were collected daily at every stooling event, stored in storage buffer, and archived at −80°C. WGS, whole-genome sequencing.

Table S1 in the supplemental material). Fecal samples were collected daily until postnatal day 21 or discharge from the neonatal intensive care unit (NICU) (Fig. 1). The mean gestational age (GA) of infants at birth was 29.9 ± 2.3 weeks. A total of 28 infants (24.8%) were <28 weeks GA, and 85 (75.2%) were $28^{0/7}$ to $32^{6/7}$ weeks GA. The mean birth weight was 1,381 g (±415 g); 66 (58.4%) newborns were classified as having a very low birth weight (VLBW) (birth weight of <1,500 g), and 26 (23.0%) were classified as having an extremely low birth weight (ELBW) (<1,000 g).

Intestinal permeability (IP) was determined 7 to 10 days after birth, when rapid intestinal barrier maturation normally takes place (15, 16). IP was calculated as the ratio of two enterally administered sugar probes, lactulose (La) and rhamnose (Rh), markers of intestinal paracellular and transcellular pathways, respectively (25, 26). IP ranged between 0.001 and 0.394, with an average of 0.07 ± 0.007 (mean ± standard error [SE]), and was not significantly different among postnatal days 7 to 10 (Fig. S1A). High IP was defined by an La/Rh ratio of >0.05, as validated and applied previously (16). Of the 113 subjects, 48 (42.5%) were found to have high IP. Infants <28 weeks GA were more likely to have elevated IP ($n = 18$) than infants $28^{0/7}$ to $32^{6/7}$ weeks GA (64.3% versus 35.3% [$P < 0.01$]).

Postmenstrual age and mother's own breastmilk feeding are associated with intestinal permeability in early preterm neonates. Among the collected demographic and maternal variables for each infant, four host factors were observed to be inversely related to IP, including GA, postmenstrual age (PMA) corresponding to chronological age and GA, birth weight, and 1-min APGAR (appearance, pulse, grimace response, activity, and respiration) score (Table 1). These variables are also highly correlated with one another, with high covariates of multicollinearity (variance inflation factor of >10) (Fig. S1). PMA was the most significant factor associated with IP among the four factors ($P = 0.01$; q value = 0.015) based on the Hilbert-Schmidt independence criterion (HSIC) (Table S2). Other host factors such as sex and race were not significantly associated with IP. Maternal factors, including preterm premature rupture of membranes (PPROM), maternal antibiotics, antenatal corticosteroids, preeclampsia, and delivery mode, were not associated with IP. These data indicate that younger infants have significantly higher incidences of high IP, likely attributed to their more immature intestinal development.

However, host factors could only partially explain IP. Longer feeding and a higher intake volume of the mother's own breastmilk (MOM) and a shorter antibiotic treatment

TABLE 1 Study cohort demographics and clinical variables stratified by intestinal permeability category

Variable	Value			*P* value
	Total cohort (*n* = 113)	High IP (*n* = 48)	Low IP (*n* = 65)	
No. (%) of subjects of sex				0.28
Male	61 (54.0)	24 (50.0)	37 (57.0)	
Female	52 (46.0)	24 (50.0)	28 (43.1)	
No. (%) of subjects of race				
White	42 (37.2)	18 (37.5)	24 (37.0)	1.00
African American	63 (55.8)	30 (62.5)	33 (50.8)	0.25
Other	8 (7.1)	0	8 (12.3)	0.02
Mean birth wt (g) ± SD	1,377.8 ± 415.2	1,237.3 ± 378.1	1,496.5 ± 403.0	<0.01
No. (%) of VLBW subjects (<1,500 g)	66 (58.4)	32 (66.7)	34 (52.3)	0.18
Mean gestational age (wks) ± SD	29.8 ± 2.3	29.0 ± 2.3	30.5 ± 2.1	<0.01
No. (%) of early-GA subjects (≤28 wks)	28 (24.8)	18 (37.5)	10 (15.4)	<0.01
Mean postmenstrual age (wks) ± SD	31.1 ± 2.3	30.3 ± 2.3	31.7 ± 2.1	<0.01
No. (%) of early-PMA subjects (<31 wks)	41 (36.3)	23 (47.9)	18 (27.7)	0.03
No. (%) of subjects with caesarean delivery	77 (68.1)	37 (77.1)	40 (61.5)	0.10
No. (%) of mothers with PPROM	36 (31.9)	15 (31.3)	21 (32.3)	1.00
No. (%) of mothers with preeclampsia	25 (22.1)	11 (23.0)	14 (21.5)	1.00
No. (%) of mothers receiving antenatal corticosteroids	106 (94.0)	46 (96.0)	60 (92.3)	0.70
No. (%) of mothers receiving antibiotics	69 (61.1)	30 (62.5)	39 (60.0)	0.85
Mean APGAR score at 1 min ± SD	5.8 ± 2.5	5.3 ±2.8	6.2 ± 2.1	0.04
Mean APGAR score at 5 min ± SD	7.7 ± 1.6	7.5 ± 1.9	7.9 ± 1.6	0.12
No. (%) of subjects receiving antibiotic				
Ampicillin	64 (56.7)	30 (62.5)	34 (52.3)	0.33
Gentamicin	56 (49.6)	25 (52.1)	31 (47.7)	0.70
Vancomycin	8 (7.1)	6 (12.5)	2 (3.1)	0.07
Cefotaxime	9 (8.0)	6 (12.5)	3 (4.6)	0.16
No. (%) of subject who received at least 1 antibiotic vs no antibiotics[a]	68 (60.2)	33 (68.8)	35 (53.9)	0.12
No. (%) of subjects who received antibiotic for[a]:				
≤3 days	83 (73.5)	30 (62.5)	53 (81.5)	0.03
>3 days	30 (26.6)	18 (37.5)	12 (18.5)	
No. (%) of subjects who received MOM for[a]:				
<4 days	26 (23.0)	20 (41.7)	6 (9.2)	<0.01
≥4 days	87 (77.0)	28 (58.3)	59 (90.8)	
Mean feeding duration (no. of days) ± SD[a]				
MOM	4.8 ± 2.3	4 ± 2.7	5.5 ± 1.5	<0.01
Formula	1.3 ± 2.3	2 ± 2.7	0.8 ± 1.6	0.02
Mean feeding intake vol received ± SD[a]				
MOM	200.8 ± 178.8	123.4 ± 154.2	263.0 ± 175.6	<0.01
Formula	61.7 ± 146.7	99.8 ± 194.7	32.8 ± 91.2	0.03

[a]Variable measured during the time period starting from the enrollment day (within 1 to 4 days after birth depending on clinical stability) until the day when IP was measured (day 8 ± 2 after birth).

duration were also significantly associated with low IP (Table 1). Compared to infants with low IP, neonates with high IP had fewer days of MOM feeding (4 days versus 5.5 days [$P < 0.01$]), a lower total MOM volume (123.4 mL/kg of body weight versus 263 mL/kg [$P < 0.01$]), as well as a longer duration (>3 days) of antibiotic use (37.5% versus 18.5% [$P = 0.03$]). We adjusted host factors associated with IP and fit a generalized logistic regression model. Newborns who were fed MOM for ≥4 days during the first week were demonstrated to be 10.3-fold more likely to have low IP than those who were fed MOM for <4 days (adjusted odds ratio [aOR], 10.3 [95% confidence interval {CI}, 3.21 to 33.33]) (Table 2). Additionally, newborns who had longer antibiotic treatment (≥3 days) were 2.6 times more likely to have high IP; however, this association was mitigated when adjusting for confounders like PMA. This result is in line with our previous observations that antibiotic use is significantly more common in the early-GA subjects

TABLE 2 Odds ratios for factors associated with low IP adjusted for postmenstrual age and birth weight[a]

Factor	OR	95% CI for OR	*P* value for OR[d]	Adjusted OR[c]	95% CI for adjusted OR	*P* value for adjusted OR[d]
Duration of antibiotic use[b]						
≤3 days	2.65	1.12, 6.25	0.02	1.56	0.58, 4.16	0.37
>3 days	1.0 (ref)			1.0 (ref)		
Duration of MOM feeding[b]						
≥4 days	7.04	2.5, 19.6	<0.01	10.30	3.21, 33.33	<0.01
<4 days	1.0 (ref)			1.0 (ref)		

[a]Fisher's exact test was used to calculate *P* values for categorical variables. Student's *t* test was used for continuous variables (birth weight, gestational age [GA], postmenstrual age [PMA], and APGAR scores at 1 min and 5 min). Intestinal permeability (IP) was calculated as the ratio of urine lactulose (La) to rhamnose (Rh), and an La/Rh ratio of <0.05 was defined as low IP. MOM, mother's own breastmilk; OR, odds ratio; CI, confidence interval.
[b]Variable measured during the time period starting from the enrollment day (within 1 to 4 days after birth depending on clinical stability) until the day when IP was measured (day 8 ± 2 after birth).
[c]The adjusted OR model includes PMA and birth weight.
[d]*P* value calculated using logistic regression.

(92% for <28 weeks GA versus 32% for >28 weeks GA [$P < 0.001$]) (16). Statistical dependence analyses showed that the cumulative intake volume of MOM prior to the IP measurement was the most significant factor associated with IP ($P < 0.001$; q value < 0.01; HSIC statistics = 1.53 and 1.46), at a significance level even higher than those for host factors, including GA ($P < 0.001$; q value < 0.01; HSIC statistic = 1.12), PMA ($P = 0.01$; q value = 0.015; HSIC statistic = 0.93), and body weight ($P = 0.01$; q value = 0.035; HSIC statistic = 1.12) (Table S2).

Breastmilk intake is associated with improved intestinal barrier integrity. Unfortunately, mothers who deliver preterm often produce less milk than those who deliver at term, and milk administration is often delayed, especially in early preterm infants (27). Formula and/or pasteurized donor human breastmilk (PDHM) is often a necessary dietary supplement. Only 55.7% of neonates in the cohort were exclusively breastfed ($n = 63$); others had their diet complemented with either formula ($n = 31$) or PDHM ($n = 12$) or were fed exclusively formula ($n = 9$) (Fig. 2A). For this reason, we investigated IP in neonates grouped by feeding type. Exclusive formula feeding was significantly associated with high IP, in either the number of days ($P = 0.02$) or the intake volume ($P = 0.03$) (Table 1). However, when formula was used in combination with MOM, even at a minor portion (35.2% ± 31.7% [mean ± SE]), IP was significantly decreased to a level that was no different from that of the cohort fed exclusively MOM (Fig. 2B). Infants whose diet was supplemented with PDHM in addition to MOM had IP similar to that of the group fed exclusively MOM. We further investigated how much MOM is "sufficient" relating to improved IP during the first week after birth. A highly elevated IP was observed in infants who received no MOM (exclusively formula or no feed), and a rapid decrease in IP was inversely correlated with an increased MOM intake volume (Fig. 2C). Discriminatory machine learning schemes suggested that a threshold of around 150 to 180 mL/kg of cumulative intake of MOM by 7 to 10 days of age is associated with low IP. Together, our results indicate that sufficient MOM, used alone or combined with other forms of feeding, significantly impacts IP in early preterm infants. Even more importantly, these results imply that the benefits of breastmilk feeding are beyond nutrition alone but extend to postnatal intestinal barrier maturation.

Increased *Bifidobacterium* species abundance correlates with improved intestinal barrier integrity. We further performed high-resolution characterization of the intestinal microbiota in 517 fecal samples, using both short-read sequencing of the V3-V4 region of the 16S rRNA gene on an Illumina HiSeq 2500 instrument (300 bp paired-end reads) ($n = 472$) and long-read sequencing of the full-length 16S rRNA gene on the Pacific Biosciences (PacBio) Sequel II platform ($n = 192$). For short-read sequencing, we obtained a total of 25,838,078 high-quality, nonchimeric ASVs (amplicon sequence variants) after the assembly of forward and reverse reads and quality assessment, representing 51,165 ± 620 (mean ± SE) ASVs per sample (see Table B at https://doi.org/10.6084/m9.figshare.19723252.v1). On the other hand, long-read sequencing generated using circular consensus

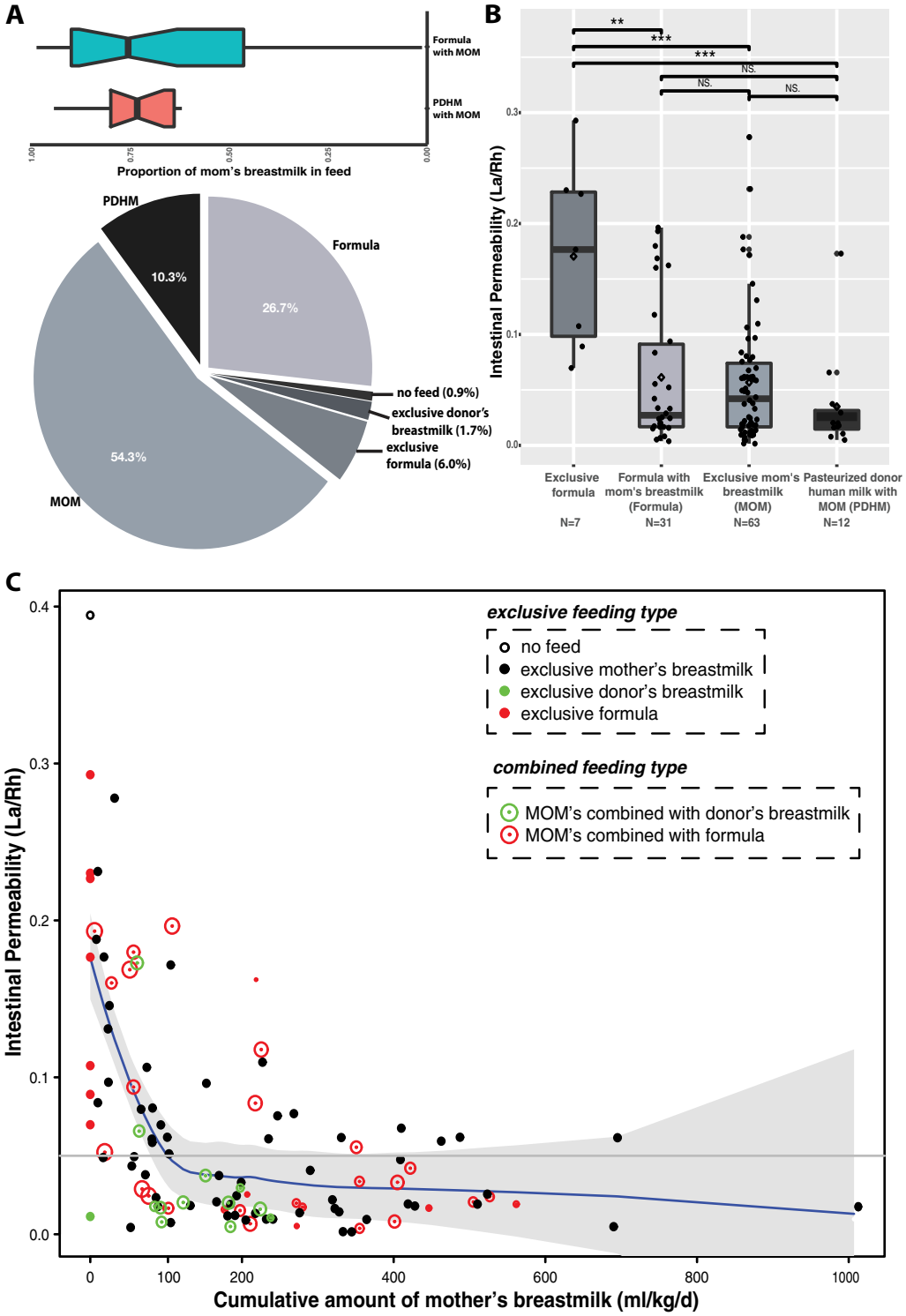

FIG 2 (A) Pie chart of feeding types for the preterm infant population in this study. Abbreviations: MOM, mother's own breastmilk; PHDM, pasteurized human donor's milk. (B) Box plot of IP measurements grouped by feeding types. (C) Correlation between intestinal permeability and the cumulative amount of mother's own breastmilk feeding (milliliters per kilogram) for a total of 113 enrolled preterm infants at $24^{0/7}$ to $32^{6/7}$ weeks of gestation. IP was calculated using the ratio of urine lactulose (La) to rhamnose (Rh), and low or high IP was defined by an La/Rh ratio of >0.05 or ≤ 0.05, respectively. The total amount of mother's own breastmilk feeding was calculated as the sum of the daily amount of milk intake per kilogram of body weight until days 7 to 10, when IP was measured. Initial feeding was calculated based on 10 mL/kg expressed breastmilk between the first and fourth days of life depending on clinical stability. After 3 to 5 days of initial feeds, feedings were advanced by 20 mL/kg/day until 100 mL/kg/day was reached. Plotted are interquartile ranges (IQRs) (boxes), medians (lines in boxes), and means (diamonds). Significance values were calculated using a Wilcoxon rank sum test. * denotes the level of significance. NS, nonsignificant.

sequences (CCSs) yielded 1,271,873 high-quality full-length 16S rRNA sequences or 992.9 ± 16.8 (mean ± SE) nonchimeric ASVs per sample. The full-length 16S rRNA gene sequences (1,462 bp on average) extended the partial V3-V4 region (428 bp on average) 3.2 times and afforded species-level assignment for 87.6% of the long-read ASVs (the remaining ASVs were not assigned due to the lack of a reference), compared to 15.3% for the short-read ones (Fig. S2; see also Table D at https://doi.org/10.6084/m9.figshare.19723252.v1). Using samples sequenced by both methods, taxonomic assignments for long-read ASVs were conveyed to short-read ASVs using perfect sequence matches, thus achieving species-level assignment for 65.3% of the short-read sequences (see Table E at https://doi.org/10.6084/m9.figshare.19723252.v1).

In total, 508 ASVs belonging to 212 species in 15 orders and 6 phyla were identified (see Tables A to C at https://doi.org/10.6084/m9.figshare.19723252.v1). The four most abundant taxa were *Klebsiella pneumoniae*, *Escherichia coli*, *Staphylococcus epidermidis*, and *Enterobacter* spp. These taxa were predominant (>50% relative abundance) and dictated four distinct community types (Fig. S3). These four taxa belong to two classes, *Enterobacteria* (*K. pneumoniae*, *E. coli*, and *Enterobacter* spp.) and *Bacilli* (*S. epidermidis*), and were highly prevalent (present in 86.2 to 94.8% of the samples) in both high- and low-IP subjects (Fig. 3A). They are also known as "first colonizers" of the infant gut (15, 28, 29). Five other taxa, including *Enterococcus faecalis*, *Clostridium perfringens*, *Proteus mirabilis*, *Bifidobacterium breve*, and *Veillonella dispar*, were found to contribute to 17.4% of all sequences and were detected in 47.7 to 86.6% of all samples. These obligate and facultative anaerobes were considered the "succession" microorganisms that succeed the first colonizers (15, 30–32). Together, these nine taxa accounted for 76.0% of all sequences in this data set. The remaining sequences were from a diverse array of obligate and facultative anaerobes (Fig. S3, cluster 5).

A zero-inflated negative binomial random-effects (ZINBRE) model was applied to investigate microbial biomarkers correlated with IP. *B. breve* was the taxon that was most significantly associated with low IP ($P < 0.001$) during the first 7 to 10 days after birth (Fig. 3B, Table S3B, and Fig. S4B). The low-IP group had significantly higher levels of *B. breve*, more *Bifidobacterium* overall, and more MOM. An adaptive spline logistic regression model was used independently to confirm the association of *B. breve* with IP and MOM (Fig. S4C and D). Other phylotypes associated with MOM or PMA are shown in Table S3. The high-IP-associated ASVs of *S. epidermidis*, *E. coli*, and *Parabacteroides distasonis* were associated with early PMA (Table S3A). *Veillonella dispar* was revealed to strongly associate with later PMA ($P < 0.001$) but not with IP. *S. epidermidis* and *E. coli* were also associated with less MOM during the first week (Table S3C). *B. breve* was found in 71.7% of samples containing *Bifidobacterium*, followed by *B. longum* (21.7%). The other *Bifidobacterium* species were either rare or present at very low abundances (<0.1%). Temporal microbiota profiling indicated that *Bifidobacterium* species reached higher abundances (~5 to 20%) after >3 days of MOM (Fig. 3E) (see https://doi.org/10.6084/m9.figshare.19709923.v1). When stratified by major feeding types, *Bifidobacterium* was most abundant in the cohort fed exclusively MOM or MOM supplemented with formula (Fig. S4A). We plotted community diversity against MOM feeding volume as a function of time and observed that low-IP infants had significantly higher microbiota diversity and higher *Bifidobacterium* species diversity when MOM reached >150 mL/kg of cumulative intake within the first week (Fig. 3C and D). It is worth noting that MOM is a critical but not the only contributor to the abundance of *Bifidobacterium*. Fifteen percent of the subjects who received no MOM had >1% *Bifidobacterium*, and 32.5% had a detectable level of *Bifidobacterium* (>0.1%). Overall, this result further supports the importance of achieving the critical threshold of MOM intake and its critical association with low IP.

Population dynamics of *Bifidobacterium* species in early postnatal colonization. Phylogenetic analyses of full-length 16S rRNA gene sequences demonstrated that *B. breve* forms a monophyletic clade, and the four most abundant ASVs were nearly identical, while *B. longum* was more phylogenetically diverse, with four distinct clades (Fig. 4A and B). Clade I was the most abundant and represented *B. longum* subsp.

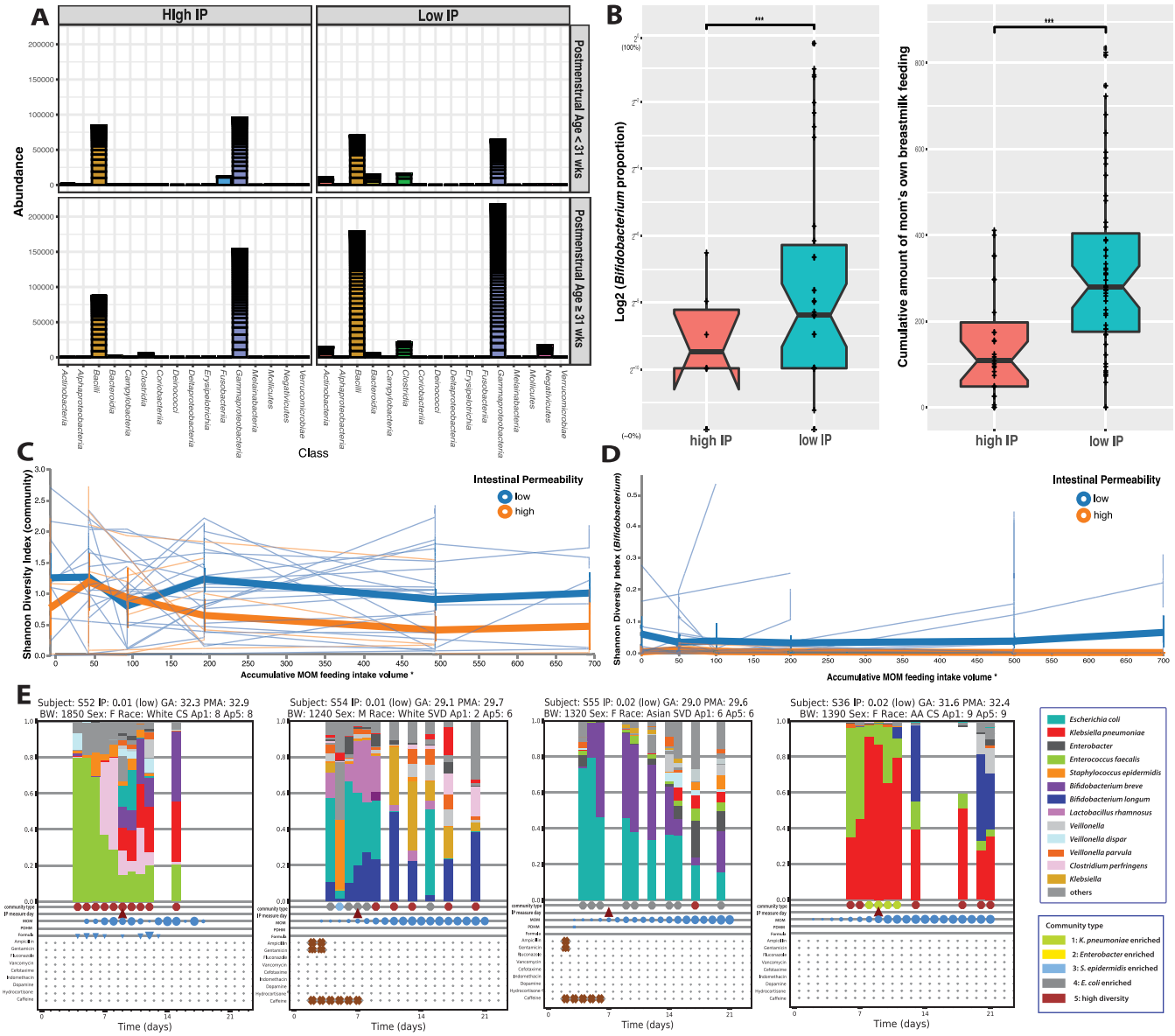

FIG 3 Microbial biomarkers and breastmilk feeding in early preterm subjects with high and low IP. (A) Abundance of bacterial groups stratified by postmenstrual age at study days 7 to 10. The results indicate that the *Actinobacteria* (*Bifidobacterium*) and *Clostridia* (*Clostridiales*) were observed mainly in low-IP subjects but not in high-IP subjects (red). The abundance values of read counts for each ASVs are stacked in order from highest to lowest, separated by a horizontal line. (B) Box plot of *Bifidobacterium* relative abundance and the cumulative amount of mother's breastmilk feeding (milliliters per kilogram) during the first 7 to 10 postnatal days in subjects with high or low IP. IP was calculated using the ratio of urine lactulose (La) to rhamnose (Rh), with low or high IP defined by an La/Rh ratio of >0.05 or ≤0.05, respectively. Plotted are interquartile ranges (IQRs) (boxes), medians (lines in boxes), and means (diamonds). Significance values were calculated using a Wilcoxon rank sum test. * denotes the level of significance. NS, nonsignificant. (C and D) Volatility plots demonstrating the fluctuation of microbial community diversity (characterized by the Shannon diversity index) (C) and *Bifidobacterium* diversity over mother's own breastmilk (MOM) feeding volumes in high- or low-IP groups (D). The plot was generated in QIIME (October 2019 version) (106). Nonoverlapping vertical error bars at each measuring point were considered significantly different. (E) Temporal characterization of the intestinal microbiota of early preterm infants with profile changes over the first 21 days after birth. The taxonomic profile was generated using 16S rRNA gene sequencing. Community type is shown in heatmap clusters in Fig. S3 in the supplemental material. The dates when IP was measured, MOM, pasteurized human donor's milk (PHDM), formula feeding day, and antibiotic administration are shown in the plots. Each circle is sized proportionally to the feeding volume. Abbreviations: BW, body weight; F, female; M, male; CS, cesarean section; SVD, spontaneous vaginal delivery; AA, African American; Ap1, Apgar score 1 minute category; Ap5, Apgar score 5 mintues category; GA, gestational age; PMA, postmenstrual age; BW, birthweight.

longum, while *B. longum* in the other three clades, II to IV, was present at low abundances. ASVs assigned to *Bifidobacterium* showed high sequence diversity (Fig. 4A) as well as inter- and intrasubject variability (Fig. 4C), in that multiple ASVs can be detected in the same subject and a single ASV can be detected in multiple subjects at multiple time points. For instance, 35 *B. longum* ASVs of four different clades were observed in one subject. Furthermore, some ASVs (i.e., unclassified *Bifidobacterium* spp.) were

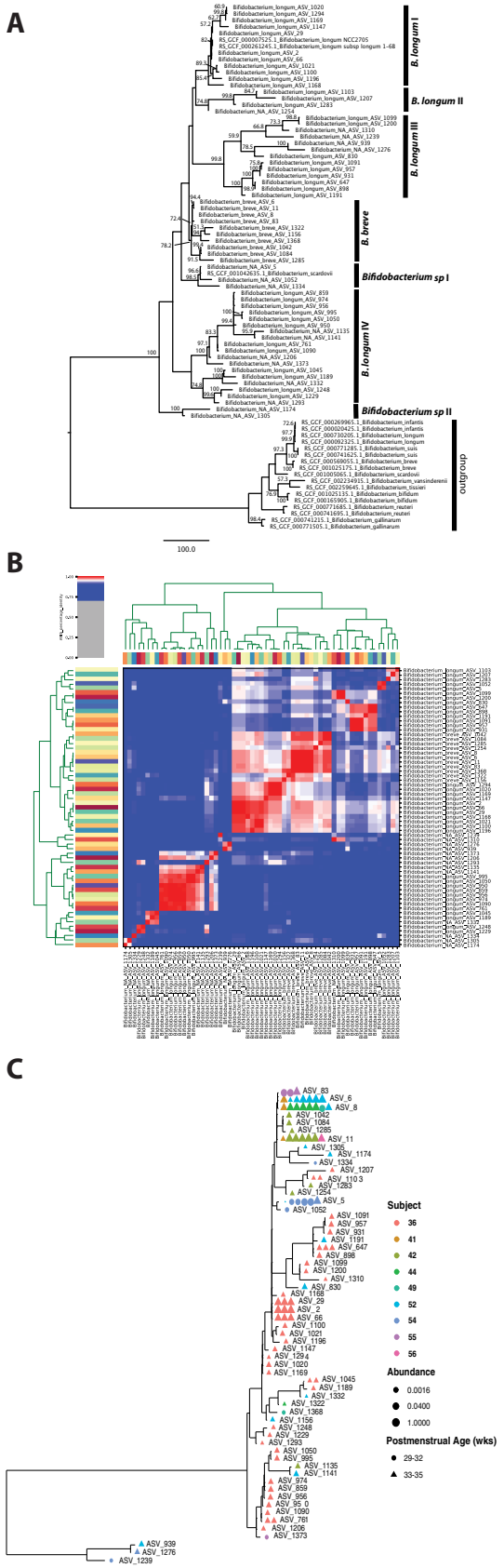

FIG 4 (A) Phylogenetic tree constructed using 81 unique, full-length 16S rRNA gene ASV sequences of *Bifidobacterium*. (B) ANI clustering of full-length 16S rRNA gene sequences. (C) Phylogenetic tree of *Bifidobacterium* ASVs in the stool microbiota of the cohort. All full-length 16S rRNA genes assigned to *Bifidobacterium* were used in the analyses. Color denotes individual subjects.

observed only in infants with an early PMA (<33 weeks), while others did not vary in abundance across PMA (i.e., *B. breve*), supporting high subspecies-level diversity and population dynamics in the preterm infant gut community.

To characterize the genome content of *Bifidobacterium* species, we performed whole metagenomic sequencing of 30 samples with >10% *Bifidobacterium* species using an Illumina NovaSeq 6000 platform (see Table A at https://doi.org/10.6084/m9.figshare.19723255.v1) and generated 26 *B. breve* and 4 *B. longum* nearly complete metagenome-assembled genomes (MAGs) (see Table B at https://doi.org/10.6084/m9.figshare.19723255.v1). We further performed metagenomic sequencing of two samples using the Pacific Biosciences Sequel II platform, which afforded one closed and one nearly complete genome of *B. breve* strains. The closed genome was 2.34 M in size (Fig. S6; see also Table C at https://doi.org/10.6084/m9.figshare.19723255.v1), similar to the median *B. breve* genome size of 2.33 M (million bp) in the NCBI database. For pangenome analysis, we supplemented the 26 *B. breve* in-house MAGs with 107 published genomes (see Table A at https://doi.org/10.6084/m9.figshare.19709917.v2) and the 4 *B. longum* MAGs with 310 published genomes (see Table B at https://doi.org/10.6084/m9.figshare.19709917.v2) to identify homologous gene clusters (HGCs) (see Tables C and D at https://doi.org/10.6084/m9.figshare.19709917.v2). Among the total of 4,922 *B. breve* HGCs, 54.2% were considered dispensable (present in <10% of the genomes), 29.4% were core (present in >95% of the genomes), and the rest were accessory (see Table E at https://doi.org/10.6084/m9.figshare.19709917.v2). The pangenome of *B. longum* (7,265 HGCs) was roughly twice the size of that of *B. breve* (3,363 HCGs), although the core genomes of the two species were similar (1,511 versus 1,448 HCGs). The large pangenome size of *B. longum* may reflect its broader host range, which includes both infant and adult intestines, than that of *B. breve* or *B. infantis*, which were observed exclusively in the infant gut (33). In particular, the genes involved in the fructose 6-phosphate phosphoketolase-dependent glycolytic pathway for ATP-efficient carbohydrate catabolism, or the "bifid shunt," are conserved in both species (https://doi.org/10.6084/m9.figshare.19907113). Furthermore, *B. longum*'s dispensable genome, which comprised 46.3% of its pangenome (2,666 HGCs), was smaller than that of *B. breve* (54.2%; 3,363 HCGs) in both size and proportion, indicating high genome plasticity in *B. breve*.

We identified 46 genes specific to *B. breve* strains colonizing infants with low IP (see Table F at https://doi.org/10.6084/m9.figshare.19709917.v2). While a large number of these genes have unknown functions, others encoded functions such as glycosyl transferases, glycosyl hydrolases, cell surface adhesion and transport, polysaccharide biosynthesis, quorum sensing, and phage integration. Furthermore, a number of functions were significantly enriched in these genomes compared to the publicly available species genomes (adjusted q value < 0.05) (see Table F to I at https://doi.org/10.6084/m9.figshare.19709917.v2), such as cation transmembrane transporter activity; glucuronate isomerase; methyladenine glycosylase; glycosyl hydrolase families 59, 2, 85, and 30; and bacterial rhamnosidases A and B. Of note, *B. breve* HGC profiles appear to be highly similar within subjects, indicating that *B. breve* genomes detected at different time points in the same infants shared greater similarity than did those from different subjects (https://doi.org/10.6084/m9.figshare.19907113) (see Table J at https://doi.org/10.6084/m9.figshare.19709917.v2). Together, compared to *B. longum*, *B. breve* strains colonizing infants with low IP have high genome plasticity and are enriched in genetic features of carbohydrate metabolism and transport that underlie the strong niche-adaptive capabilities of the species.

Specialized human milk oligosaccharide assimilation capabilities of *Bifidobacterium* strains in early preterm infants. As both *Bifidobacterium* species abundance and MOM were associated with postnatal intestinal barrier maturation, we next investigated whether these two factors were linked through the ability of *Bifidobacterium* species to utilize the oligosaccharides present in breastmilk. Previously characterized major HMO utilizers like *Bacteroides* species and *Lactobacillus* (34, 35) were largely absent from our cohort (see https://doi.org/10.6084/m9.figshare.19723252.v1), indicating that *Bifidobacterium* species likely provide the genetic capabilities to metabolize HMOs. We thus examined the set of

genes encoding extracellular hydrolases, sugar transporters, and intracellular hydrolases (Table S4), which comprise the machinery necessary to take up and metabolize HMO substrates to feed central fermentative metabolism (36–38).

Intracellular HMO utilization functions were found to be encoded exclusively by both *B. breve* and *B. longum*. We examined eight essential extracellular enzymes and their homologs (for details, see Materials and Methods) known to be required for the extracellular breakdown of HMOs into smaller molecules that are then transported intracellularly. Interestingly, none of these extracellular enzymes were found in this cohort. We investigated five essential bacterial ABC transporters and homologs involved in the import of various oligosaccharides, known to have a high specificity for HMOs conferred by substrate binding protein (SBP) domains (39). Both *B. breve* and *B. longum* contained *gltA* (Table S4A), a gene considered crucial for the import of lacto-*N*-tetraose (LNT). LNT comprises the core HMO structure that is catabolized via lacto-*N*-biose (LNB) intermediates (40). Furthermore, a family 1 solute binding protein (F1SBP) gene cluster, Blon_2177, was found in both *B. breve* and *B. longum* (Table S4B). This cluster was found to be critical for the import of nonfucosylated type 1 oligosaccharides (41). None of the *B. longum* strains but the majority of the *B. breve* strains of this cohort (92.4%) harbor the LNnT (lacto-*N*-neotetraose) transporter that is encoded by *nahS*. These findings indicate that both *B. breve* and *B. longum* could transport LNB and LNT, while *B. breve* can further metabolize LNnT.

We then evaluated the capability of consuming the transported oligosaccharides, and compared to *B. longum*, we revealed expanded metabolic capabilities of *B. breve* strains of this cohort to utilize a variety of HMO molecules, including fucosylated or sialylated forms, in addition to the neutral types of HMOs (i.e., LNB, LNT, and LNnT). Seventeen key glycoside hydrolases (GHs) involved in essential HMO degradation and utilization were investigated (Table S4C). The key intracellular enzymes GH2 (β-1,4-galactosidases) (LacZ2/6), GH112 (galacto-*N*-biose [GNB]/LNB phosphorylase) (*lnpA*), GH20 (β-*N*-acetylglucosaminidase), and GH42 (β-1,3-galactosidase) (*lntA*; bga42A) are highly conserved in both *B. breve* and *B. longum*. These enzymes lack transmembrane domains or signal peptide sequences and are required to degrade HMOs intracellularly (42). While almost all *B. breve* strains contained GH95 α-fucosidase (*afcA*) (homolog of Blon_2335), GH33 α-sialidase (homolog of Blon_0646), and GH20 β-*N*-acetylglucosaminidase (*nahA*) (homolog of Blon_0459) (Table S4C), only a small portion of *B. longum* strains (~10%) contained these enzymes. Furthermore, *B. breve* strains present in these preterm infants carry the gene encoding GH29 α-fucosidases more often (53.8% versus 12.7%) than *B. breve* strains isolated from other sources obtained from GenBank. The presence of GH29 α-fucosidase genes underlines the ability to consume fucosylated oligosaccharides such as 2′-fucosyllactose (2′-FL) and larger fucosylated HMOs such as lacto-*N*-fucopentaose (38, 42). The GH29-containing *B. breve* strains in our cohort also encode GH95. In fact, GH29 and GH95 α-fucosidases are highly complementary since they target specific substrates of α-1,3/4- and α-1,2-fucosyl linkages, respectively (42), and the activation of both enzymes enables the degradation and utilization of a larger variety of HMOs. Moreover, a prominent gene cluster termed FHMO (fucosylated human milk oligosaccharide) that contains both GH29 and GH95 α-fucosidase-encoding genes was observed in some *B. breve* strains but was largely absent from *B. longum* strains (Table S4D). This cluster was reported to enable *B. breve* strains to preferentially consume fucosylated HMOs over neutral HMOs during early bacterial growth (42). In particular, the putative fucosyllactose SBP (BLNG_1257) present in this cluster confers glycan binding specificity and is present consistently in *B. breve* strains of this cohort but rarely in other *B. breve* strains in GenBank. Overall, our results revealed an expanded, specialized HMO assimilation capability of *B. breve* strains, conferring a competitive growth advantage in the gut of this preterm infant cohort when fed breastmilk.

Host-derived glycoprotein utilization is limited to *B. breve* in early preterm infants. Besides HMOs, host-derived glycoproteins such as mucin and proteoglycan (mucus or milk) are critical carbon sources for bacteria in the infant intestinal microenvironment.

Human glycoproteins are often heavily sulfated and could not be metabolized without bacterial glycosidases (43, 44). We investigated two sulfatase-encoding gene clusters essential for sulfatase metabolism, *ats1* and *ats2* (45, 46), and they each encode glycosulfatases and the accompanying anaerobic sulfatase-maturing enzymes (anSMEs) with an associated transport system and transcriptional regulator (46). The primary mucin degradation capabilities of this cohort are shown to be limited to *B. breve* strains (Table S4F), as the two clusters are present in 100% of *B. breve* strains in our cohort and ~70% of all *B. breve* genomes available. *B. longum* strains rarely harbor *ats1*, and no strains carry *ats2*.

In addition to sulfated residues, more than half of human colonic mucin oligosaccharides also contain sialic acid residues (47). The release of sialic acid is an initial step in the sequential degradation of mucins and sialylated HMO substrates (46, 48). Hence, we investigated the two gene clusters essential for the uptake and metabolism of sialic acid, the *nagA2-nagB3* cluster (Bbr_1247-Bbr_1248) and the *nan-nag* cluster (Bbr_0160-Bbr_0172) (49–51). These two gene clusters are highly conserved in *B. breve*, while they are present in only 14% of *B. longum* genomes (https://doi.org/10.6084/m9.figshare.19709917). Our results demonstrate that the capability of foraging sulfated and/or sialylated host-derived glycoproteins is attributed to *B. breve* strains in this cohort. This metabolic versatility of *B. breve* may greatly improve its fitness and facilitate its mucosal adherence, hence facilitating colonization under nutrient- or energy-limited conditions in the preterm infant gut environment.

DISCUSSION

Early preterm neonates are a vulnerable and challenging population that often requires intensive medical care. As a result of their premature birth, these neonates often have an aberrantly permeable intestinal barrier that fails to limit bacterial translocation. Our group has previously reported positive associations between persistently elevated intestinal permeability and delayed feeding, prolonged antibiotic exposure, and altered development of the intestinal microbiota as well as a lack of a progressively increased abundance of *Clostridiales* (15, 16). These *Clostridiales* became abundant mostly at the end of the second week after birth; this is after the extensive barrier maturation that occurs during the first week. In this study, we determined the minimal intake of maternal breastmilk necessary to significantly decrease IP and identified specific *Bifidobacterium* species and strains as biomarkers associated with low-IP development in preterm infants in the first week of life.

We posited that the benefits of breastmilk extend beyond nutrition and include improved gut barrier function and that the two factors associated with reduced IP, MOM feeding and *Bifidobacterium* strains, are, at least in part, linked by the ability of *Bifidobacterium* to metabolize human milk oligosaccharides (illustrated in Fig. 5). To investigate this link, we evaluated the carbohydrate-metabolizing capabilities of *Bifidobacterium* strains and uncovered a complement of genes dedicated to utilizing a wide variety of HMO molecules as well as host-derived glycoproteins. These genetic features were enriched in preterm infant gut-associated *Bifidobacterium* strains compared to those isolated from other sources like dairies or the adult gut. Our results are concordant with those of previous studies showing that the establishment of a bifidobacterium-dominant community was facilitated by specific gene clusters supporting HMO metabolism, which are absent in many adult-associated bifidobacterial strains (52–55). Functional characterization of the contribution of *B. breve* metabolizing MOM to low IP would be critical for its translational significance. Future studies modeling both the transcriptional activities of bifidobacterial biomarkers and host responses in a longitudinal design are warranted to address the cause-effect relationships of MOM and *Bifidobacterium* for intestinal barrier maturation. Furthermore, the production of short-chain fatty acids via carbohydrate consumption by bifidobacteria, particularly acetate and butyrate, was demonstrated to correlate with their anti-inflammatory properties and promoted the defense functions of the epithelium (56–58). Together, the

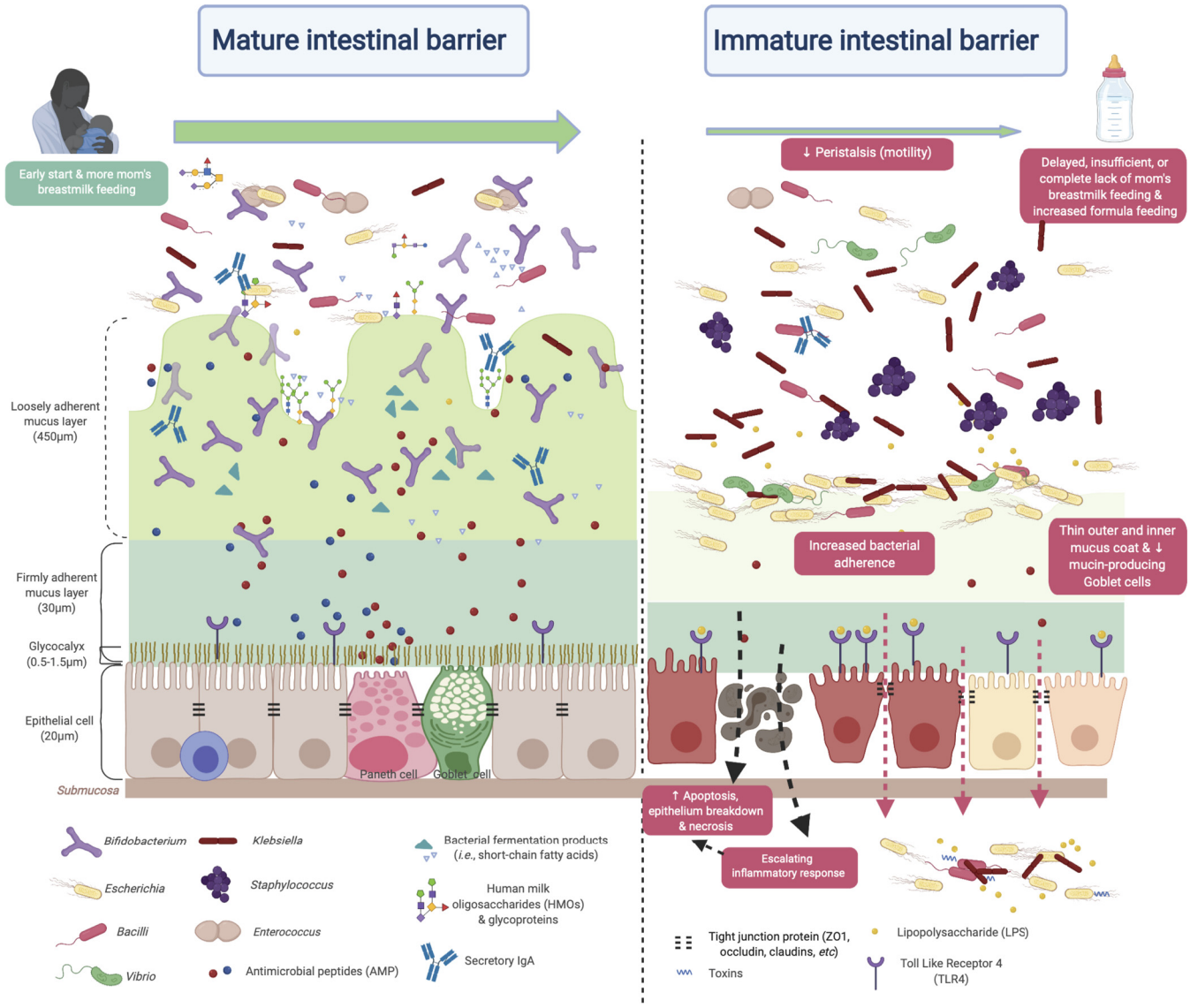

FIG 5 Illustration of mature and immature intestinal barriers in neonates. Peristalsis (reduced intestinal motility), maldigestion of nutrient sources, and a compromised gut barrier may render the mucosa susceptible to invasion by opportunistic pathogens in the gut environment. The resulting imbalance between epithelial cell injury and repair leads to a vicious cycle of maldigestion, bacterial invasion, immune activation, and uncontrolled inflammation. The illustration is not drawn to scale. (Created with BioRender.com.)

results of our study support the notion that intestinal barrier function can develop postnatally, and this process could be induced through supplementation with breastmilk substrates as well as *Bifidobacterium* strains that consume them. These elements are promising therapeutic targets to reduce NEC and other life-threatening conditions associated with intestinal hyperpermeability.

B. breve is a known dominant *Bifidobacterium* species in both preterm and term infant gut microbiota (59) and was also observed in breastmilk and vaginal microbiota (60, 61). In humans, *B. breve* appears to be found exclusively in these environments and is largely absent in the adult gut. The factors contributing to *B. breve* persistence in infants are not well understood. Most studies were performed using the type strain *B. breve* ATCC 15700 (JCM1192), which has a limited ability to consume HMOs (62, 63). As demonstrated by us and others, strains of *B. breve* vary greatly in their abilities to metabolize HMOs (55). The *B. breve* strains in our cohort displayed extensive enzymatic capabilities designed to efficiently utilize a broad range of dietary and host-derived carbohydrates, thus maximizing their colonization of the infant intestinal environment.

In particular, we demonstrated that LNnT utilization was limited exclusively to strains of *B. breve*. Growth on LNnT was shown *in vitro* to enable *B. infantis* to outcompete other species such as *Bacteroides* species (64). LNnT can be fermented by specific strains of *Bifidobacterium* found only in the infant gut (65). Digestion of neutral HMOs (i.e., LNT and LNnT) was actually shown to induce a significant shift in the ratio of secreted acetate to lactate compared to the catabolism of the simpler carbohydrates that they contain (66). Furthermore, the GH29 α-fucosidase, an uncommon enzyme correlated with the ability to grow on fucosylated HMOs (38), was enriched in only *B. breve* strains in this cohort. The presence of key gene sets expands *B. breve* metabolic capabilities (i.e., FHMO, GH29, and GH95) and is reminiscent of those found in *B. infantis* ATCC 15697, the model strain that can also consume a broad repertoire of HMOs (41, 67). Previous clinical trials administering *B. breve* strains in early preterm infants yielded contradicting results, which may be related to the selection of different strains. For example, Kitajima and coauthors reported that a *B. breve* BBG strain could colonize the immature bowel effectively, with significantly fewer abnormal abdominal signs and greater weight gain in VLBW infants (68). However, the clinical trial of the type strain BBG-001 in very-preterm infants observed no evidence of a benefit in terms of preventing NEC and late-onset sepsis (LOS) (69). These data highlight the importance of strain characterization in prophylactic supplementation with live biotherapeutics. Further characterization of these key genes will be necessary to understand the range of oligosaccharides that *B. breve* strains can transport and consume. Collections of *B. breve* strains isolated from both preterm infants with rapidly decreasing IP and healthy term infants should be established to achieve this important goal.

The specialized HMO and glycoprotein utilization capabilities of *B. breve*, particularly the degradation of sulfated and sialic residues, further confer a competitive capability that improves *B. breve* fitness and facilitates its adherence to and colonization of the gut mucosa (70). The release of sialic acid is an initial step in the sequential degradation of mucins and sialylated HMO substrates (46, 48), and the abilities to utilize the heavily sulfated mucin glycoprotein and sialic residues were found to be highly correlated (46, 49). Sialic acid concentrations are highest in the colostrum in preterm infants but decrease by almost 80% after 3 months (71). Furthermore, breastmilk from mothers who delivered preterm was reported to be a rich source of oligosaccharide-bound sialic acids, with 20% more sialic acid residues than in breastmilk from term mothers and 25% more than that found in formula (72). A recent *in vivo* study showed that sialylated HMOs are on the causal pathway of a microbiota-dependent infant growth outcome and hence were considered the most growth-discriminatory HMO structures (73). Interestingly, and supporting its importance in infant health, only strains of *B. infantis* and *B. breve* isolated from the infant gut have been reported to be capable of utilizing sialic acid and sialylated lacto-*N*-tetraose as sole carbon sources (54, 74, 75). A few *B. breve* strains were actually reported to preferentially consume sialylated HMOs, in particular sialyl-LNT b (LSTb) and sialyl-lacto-*N*-hexaose (S-LNH), over neutral HMOs (38, 49). Given that bacteria with pathogenic potential are capable of utilizing sialic acid, *B. breve* strains could rapidly sequester sialic acid away from these pathogens and offer nutritional immunity, i.e., sequestering nutrients to limit infection, thus contributing to a healthy intestinal environment (76). It would be highly insightful to further characterize maternal HMO variations in MOM and the composition of specific formulas, in addition to the information on HMO assimilation capabilities of bifidobacterial strains, for a comprehensive understanding of the essential factors contributing to postnatal intestinal maturation.

HMO utilization by *Bifidobacterium* species in this cohort appears to be exclusively an intracellular process, which would be unlikely to allow cross-feeding of intermediates with other gut bacterial species. Extracellular digestion of HMOs would afford fucose and sialic acid monomers to be cross-fed to other bacteria, some of which have pathogenic properties (77). *Bacteroides* spp., which are largely absent in this cohort, are known to employ an exclusively extracellular process in HMO utilization (64). The "internalize,

then degrade" approach for HMO consumption is a critical *Bifidobacterium* property that affords protection against infection for the infants. Interestingly, the preference for intracellular digestion of HMOs is not conserved across all infant gut *Bifidobacterium* species or strains. A recent study revealed that *Bifidobacterium* strains in the gut microbiome of breastfed infants in Malawi and Venezuela similarly employed an intracellular HMO digestion strategy, while *Bifidobacterium* strains in a cohort of U.S. infants fed formula and breastmilk preferentially employed extracellular HMO digestion strategies (36). This difference may relate to galactooligosaccharide (GOS) transporter genes present in strains that internalize HMOs to metabolize them, especially the GNB/LNB-BP (*gltA*) gene (36, 78), although the mechanisms remain unclear.

Our study highlights the strong potential for the prophylactic administration of specific *B. breve* strains early in life along with specific HMOs to enhance the intestinal barrier in preterm neonates. We previously defined a "window of opportunity" of day 8 ± 2 after birth for intervention prior to the onset of leaky gut-associated conditions such as NEC (15, 16). Our study proposed the role of breastmilk feeding in promoting the growth of beneficial *Bifidobacterium* species and strains that could consume breastmilk HMOs during that critical window period of time. In the absence of these prophylactic *Bifidobacterium* strains, the benefit of breastmilk feeding is expected to be dramatically reduced. Counting on the vertical transmission of these *Bifidobacterium* strains from the mother's gut or vaginal microbiota or breastmilk is not reliable and could leave many infants unprotected (79, 80). It is thus critical to gain further mechanistic insight into bifidobacterium-rich microbiota formation in the infant gut by prophylactic supplementation with live biotherapeutics that possess the ability to effectively utilize them. Such an understanding will inform the design of clinical interventions with supplementation with HMOs and *Bifidobacterium* as live biotherapeutic prophylaxis to enhance intestinal barrier integrity early in life and ultimately reduce the risk of NEC.

MATERIALS AND METHODS

Study cohort and feeding protocol. The study protocol was approved by the institutional review boards of the University of Maryland, Baltimore, and Mercy Medical Center. Written informed parental consent was obtained. Eligibility criteria were described previously (16). One hundred thirteen eligible preterm infants at $24^{0/7}$ to $32^{6/7}$ weeks of gestation were enrolled within 4 days after birth from combining cohorts enrolled from June 2013 to October 2014 and from October 2018 to November 2019. Prior to the study procedures, a complete physical examination, including vital signs, weight, height, and head circumference, was performed. Demographic, obstetric, clinical, medication exposure, feeding practice, and adverse event data were collected from the medical record.

Enteral feeds by the orogastric or nasogastric route were initiated between the first and fourth day of life depending on clinical stability. After initial feeds of 10 mL/kg expressed breastmilk or 20 kcal/oz preterm formula daily for 3 to 5 days, feedings were advanced by 20 mL/kg/day until 100 mL/kg/day was reached. Subsequently, caloric density was advanced to 24 kcal/oz prior to increasing the feeding volume by 20 mL/kg/day to 150 mL/kg/day. The total volume of each source of feeds was calculated as the sum of the daily amount of milk intake per kilogram of the administered expressed mother's breastmilk, donor milk, or preterm formula from the initial feed day until postnatal days 7 to 10, when intestinal permeability (IP) was measured. Feedings were held or discontinued for signs of feeding intolerance such as abdominal distension, gastric residuals, or hematochezia or for clinical deterioration. Pooled pasteurized human donor breastmilk (PHDB) was purchased from Prolacta Biosciences (Duarte, CA, USA). PHDB was collected from mothers of term infants who have breastfed for at least 6 months (81).

***In vivo* intestinal permeability measurement.** In our previous pilot studies that employed a small cohort of neonates ($n = 37$) (15, 16) with IP measured at study days 1, 8 ± 2, and 15 ± 2, it was shown that IP is high within 4 days of birth in all preterm infants, with a rapid maturation of the intestinal barrier over the first week of life. A persistently high IP and/or a late increase in IP indicates the physiological immaturity of intestinal tract barrier function. Hence, the first 7 to 10 days in preterm infants are a critical observation period for monitoring IP. Eligible preterm infants received 1 mL/kg of the nonmetabolized sugar probes on postnatal days 7 to 10, which included lactulose (La; Cumberland Pharmaceuticals, Nashville, TN), which is a marker of intestinal paracellular transport, and rhamnose (Rh; Saccharides, Inc., Calgary, Alberta, Canada), which is a marker of intestinal transcellular transport. One milliliter of a solution containing 8.6 g La plus 140 mg Rh/100 mL was administered enterally by nipple or by gavage via a clinically indicated orogastric tube (82). A minimum of 2 mL of urine was collected over a 4-h period following the administration of the La/Rh dose as previously described (16). La and Rh concentrations were measured by high-pressure liquid chromatography (HPLC) at the University of Calgary (Calgary, Canada). High or low intestinal permeability was defined by an La/Rh ratio of >0.05 or ≤ 0.05, respectively, as validated and applied previously (16). Postmenstrual age at sugar probe dosing was calculated as the gestational age at birth plus the postnatal age on the dosing day (83).

Fecal specimen collection and nucleic acid extraction. Fecal samples (~1 g) collected daily from enrollment until postnatal day 21 or NICU discharge were stored immediately in 1 mL of DNA/RNA Shield (Zymo Research, Irving, CA, USA). Stool specimens were collected from within the stool mass from the diaper as much as feasible to avoid frequent air exposure. The stool sitting time was 0 to 3 h, and the sample was collected during diaper changes every 3 h. Urine and fecal samples were archived at −80°C until processing.

Genomic DNA was extracted from homogenized fecal samples using the MagAttract PowerMicrobiome DNA/RNA kit (Qiagen) implemented on a Hamilton Star robotic platform and after a bead-beating step on a TissueLyser II instrument (Qiagen) in 96-deep-well plates at the Microbiome Service Laboratory (MSL) at the University of Maryland, Baltimore (Baltimore, MD, USA). DNA purification from lysates was done on a QIAsymphony automated platform.

Short-read sequencing of 16S rRNA gene amplicons and whole-community metagenomes. PCR amplification of the 16S rRNA gene V3-V4 hypervariable region was performed using dual-barcoded universal primers 318F and 806R as previously described (84). In brief, amplicon pools were prepared for sequencing with AMPure XT beads (Beckman Coulter Genomics, Danvers, MA), and the size and quantity of the amplicon library were assessed on the LabChip GX system (PerkinElmer, Waltham, MA) and with a library quantification kit for Illumina (Kapa Biosciences, Woburn, MA), respectively. The PhiX control library (v3) (Illumina, San Diego, CA) was combined with the amplicon library. High-throughput sequencing of the amplicons was performed on an Illumina MiSeq platform using the 300-bp paired-end protocol. Sequence libraries were prepared from the extracted DNA using the Nextera DNA Flex kit (Illumina, San Diego, CA) according to the manufacturer's specifications. Libraries were then pooled in equimolar proportions and sequenced on a single Illumina NovaSeq 6000 S2 flow cell providing an average of 6.5 million pairs of 150-bp reads per library at the Genomic Resource Center at the University of Maryland School of Medicine.

Long-read sequencing of the full-length 16S rRNA gene and whole-community metagenomes on the Pacific Biosciences Sequel II platform. Amplification of the full-length 16S rRNA gene was performed using dual-barcode, two-step PCR on diluted (1:10) genomic DNA. The first round of PCR amplification of the 16S rRNA full-length gene was performed using universal primers 27F (AGRGTTYGATYMTGGCTCAG) and 1492R (RGYTACCTTGTTACGACTT) according to Pacific Biosciences (Menlo Park, CA, USA) specifications for 20 cycles. The cycling conditions for the first-step PCR were 95°C for 30 s, 57°C for 30 s, and 72°C for 60 s. The PCR mixture was then diluted in water (1:5) and amplified with Pacific Biosciences universal forward/reverse 96-plate primers for an additional 20 cycles according to Pacific Biosciences specifications. Cycling conditions are described in the manufacturer's protocol (85). DNA quantification was carried out using the Quant-iT PicoGreen double-stranded DNA assay (Invitrogen) and visualized on a 2% agarose E-gel. The amplicon libraries were normalized, cleaned, and concentrated using AmPure XP SPRI beads (Beckman Coulter, Brea, CA, USA) at 0.6× the reaction volume.

Library pools were prepared with SMRTBell template prep kit 1.0 with barcoded adaptors. Libraries were then size selected on a BluePippen system (Sage Science, Beverly, MA) with a cutoff of 5 kb. Sequencing was performed on the Sequel II platform (PacBio, Menlo Park, CA) with loading at 60 pM. Multiplexed samples were sequenced on PacBio Sequel II cells using S/P3-C1/5.0-8M sequencing chemistry. Demultiplexing was done with lima (version 1.9.0) using default parameters, except for a minimum barcode score of 26 and a minimum length of 50 bp; both tools are part of the SMRTLink 6.0.1 software package with updated CCS version 3.4.1 (Pacific Biosciences, 2019). Raw reads were assembled via Canu v1.8 and the -pacbio-raw protocol (86). The resulting contigs were taxonomically annotated using BLASTN v2.8.1 (87) and the nonredundant nucleotide database (updated on 3 May 2019) to pool all contigs identified under the same species name to form metagenomic bins. Binned contigs were circularized and rotated using Simple-circularise (88) and were retained if the circularized contigs were in the range of the full genome size according to published closed genomes of that species based on the GenBank genome database. Metagenome bins were further confirmed using GTDB-Tk v1.1.0 (89). Genomes were annotated using PROKKA v1.13 (90).

Epidemiological analyses. Covariates identified based on previous literature and biological plausibility were collected at the time of enrollment of the participants and evaluated. Categorical data were compared using Fisher's exact test, and continuous data were compared using Student's *t* test. Multicollinearity between covariates was assessed using the variance inflation factor (VIF) and tolerance, where covariates with a VIF of >10 were considered collinear. Covariates with a *P* value of <0.05 in the bivariate analysis were considered confounding factors and were adjusted in the multivariable analysis as random factors. Generalized logistic regression was used to determine the association between IP category and continuous variables, including the duration of antibiotics and the duration of MOM feeding. Analyses were conducted using SAS version 9.4 software (SAS Institute, Cary, NC), and the code used for this statistical analysis was deposited at https://github.com/igsbma/IP_microbiome/tree/main/statistical_analyses.

Bioinformatics analysis of the intestinal microbiota. For 16S rRNA V3-V4 gene amplicon analysis, raw data were demultiplexed, and barcode, adaptor, and primer sequences were trimmed using tagcleaner v0.16 (91). Quality assessment and sequencing error correction were performed using the DADA2 v1.14 software package (92) and the following parameters: forward reads were truncated at position 240 and reverse reads were truncated at position 210 based on the sequencing quality plot, and no ambiguous bases and a maximum of 2 expected errors per read were allowed (93). The quality-trimmed reads were used to infer amplicon sequence variants (ASVs) and their relative abundances in each sample after removing chimeras. The SILVA database (94), release 132, was used to assign taxonomy. The following criteria were applied for an ASV: (i) the ASV was at least 400 bp in length for long-read sequencing, (ii) it was

observed in at least two samples, (iii) there were at least 5 counts in at least one sample, and (iv) it was not assigned to taxonomic groups of mitochondria or chloroplasts.

For full-length 16S rRNA gene analyses, CCS reads were generated using the ccs application with a minPredictedAccuracy of 0.99, and the rest of the parameters were default, including a minimum of 3 subread passes. Demultiplexing was done with lima (version 1.9.0) with a minimum barcode score of 26 and a minimum length of 50 bp; both tools are part of the SMRTLink 6.0.1 software package with updated CCS version 3.4.1 (Pacific Biosciences, 2019). The microbiota analyses were modified from a previously reported bioinformatics pipeline that incorporates the DADA2 protocol (95). The quality-trimmed reads were used to infer ribosomal sequence variants and their relative abundance in each sample after removing chimeras. Taxonomy was assigned to each ASV generated by DADA2 using both the SILVA (release 132) database and the Genome Taxonomy Database (GTDB) (96) and the RDP naive Bayesian classifier as implemented in the DADA2 R package (97, 98). In a few cases when conflicting taxonomic assignments appeared, the NCBI RefSeq 16S rRNA database combined with the RDP database (99, 100) and the Human Intestinal 16S rRNA database (HITdb v1) (101) was used to resolve the conflict. Pacific Biosciences long-read sequencing complements short-read sequencing for its high accuracy and extended length. To boost taxonomy assignments for short-read sequencing, we performed a BLASTN search of the short-read ASVs to the long-read ASVs and assigned a taxonomic name to the short reads if there was 100% identity and a unanimous assignment if there were multiple hits for long-read sequences.

A heatmap was constructed from the relative abundances of the 50 most abundant intestinal bacterial taxa in samples collected from the 113 preterm infants enrolled in the study. The ASVs were normalized using the total sum to calculate their relative abundances. Ward linkage clustering was used to cluster samples based on their Jensen-Shannon distance calculated using the vegan package in R (102). The number of clusters was validated using gap statistics implemented in the cluster package in R (103) by calculating the goodness-of-clustering measure. The raxml package (v8.0.0) (104) was used to construct the phylogeny, and the Phyloseq R package (v1.38.0) (105) was used to display the phylogeny and the bar plot. A volatility plot was used to demonstrate the fluctuation of microbial community diversity (characterized by the Shannon diversity index) over the MOM feeding volume in the high- or low-IP groups. The plot was generated in QIIME (October 2019 version) (106) (option-longitudinal plot-feature-volatility).

Statistical analysis of the intestinal microbial community. The Hilbert-Schmidt independence criterion (HSIC) R package dHSIC (107) was used to examine the independence between any variables and IP. Longitudinal modeling was performed using zero-inflated negative binomial random-effects (ZINBRE) models. These models account for the possibility of the existence of more than the expected zeros (from the negative binomial distribution) as well for correlations between samples from the same subject. Although IP was categorized into high and low groups, it is inherently continuous, and hence, we modeled IP as a continuous value in our analyses. Subject was included as a random factor. Read count data of phylotypes detected in at least 15% of the samples were modeled using ZINBRE models. The same principle was applied to MOM and PMA. The model was fitted using the JAGS R package (108), and 10,000 iterations with the same number of burn-in iterations were used. The convergence of the model was assessed using Gelman and Rubin's potential scale reduction factor (109) and visual inspection of each coefficient's Markov chains. The means of the posterior distributions of the estimated coefficients and their corresponding 95% credible intervals were calculated using the model's Markov chains. The credible intervals without overlap are considered significant. *P* values were computed by assuming the normality of the posterior distributions of the corresponding coefficients. An adaptive spline logistic regression model implemented in the spmrf R package (110) was used independently to confirm the association of *B. breve* with IP and MOM. This model is a locally adaptive nonparametric fitting method that operates within a Bayesian framework, which uses shrinkage-prior Markov random fields to induce sparsity and provides a combination of local adaptation and global control (110). The Bayesian goodness-of-fit *P* value implemented in the R package rstan (111) was used to assess the significance of the association. The R code implementation of the model has been deposited at https://github.com/igsbma/IP_microbiome/tree/main/statistical_analyses. Discriminatory machine learning scheme computations were implemented in weka (112, 113), including J48 decision tree, REPTree, decision stump, and logistic model trees. The functional enrichment test was performed for each functional group (based on Clusters of Orthologous Groups [COG] and Pfam annotations) and each of the homologous gene clusters (HGCs) generated in genome comparison analyses. Frequency tables of each function or HGC in each category (i.e., metagenome-assembled genomes [MAGs] from this study versus the GenBank genomes) were generated, which were used to fit a generalized linear model with the logit linkage function to compute an enrichment score and *P* value for each unit (114). False detection rate correction for *P* values was used to account for multiple tests using the R package qvalue (115).

Intestinal microbiome analyses. Metagenomic sequence data were preprocessed using the following steps: (i) human sequence reads and rRNA large-subunit (LSU)/small-subunit (SSU) reads were removed using BMTagger v3.101 (116) using a standard human genome reference (GRCh37.p5) (117); (ii) rRNA sequence reads were removed *in silico* by aligning all reads using Bowtie v1 (118) to the SILVA PARC ribosomal-subunit sequence database (94), and sequence read pairs were removed even if only one of the reads matched the human genome reference or rRNA; (iii) the Illumina adaptor was trimmed using Trimmomatic (119); (iv) sequence reads with an average quality greater than Q_{15} over a sliding window of 4 bp were trimmed before the window, assessed for length, and removed if they were $<$75% of the original length; and (v) no ambiguous base pairs were allowed. The taxonomic composition of the microbiomes was established using MetaPhlAn version 2 (120). The MAG pipeline includes de Bruijn genome assembly using SPAdes v.3.10.1 (121), and the bins were defined through distance clustering based on coverage and tetranucleotide signature using MetaBat v2 (122) and refined using GTDB-Tk (89).

Genomes were annotated using PROKKA v1.13 (90), annotated through evidence from the nomenclature of the Consortium for Function Glycomics, eggNOG (v4.5) (123), KEGG (18 March 2013 release) (124), Pfam (v30.0) (125), and CAZy (2014 release) (126, 127). Similarity searches were performed for comparison with previously annotated enzymes or transporter proteins based on the accession numbers (36–38) using BLASTP and confirmed with COG, Pfam, and CAZy annotation evidence to ensure the integrity of the results. The 8 essential extracellular enzymes that are known to be required for the extracellular cleavage of HMOs before importing selected products of degradation were investigated (36–38), including 1,2-α-L-fucosidase (AfcA), 1,3/4-α-L-fucosidase (AfcB), 2,3/6-α-sialidase (SiaBb2), lacto-*N*-biosidase (LnbB and LnbX), the chaperone for LnbX (LnbY), β-1,4-galactosidase (BbgIII), and β-*N*-acetylglucosaminidase (BbhI). Five essential bacterial ABC transporters and homologs involved in the import of oligosaccharides were examined, which are known to show an exquisite specificity conferred by substrate binding protein (SBPs) for different HMO molecules (39), including the GNB/LNB (galacto-*N*-biose/lacto-*N*-biose I) transporter SBP (GltA), the FL transporter SBPs (FL1-BP and FL2-BP), and the LNnT transporter SBP (NahS). In addition to similarity searches on *Bifidobacterium* genomes and MAGs, we also confirmed the results by searching the metagenomic community gene content to verify that the target genes are not from species other than *Bifidobacterium*.

Metapangenomes were prepared using the MAGs constructed in this study and publicly available genomes under the species names *B. breve* (taxID 1685) and *B. longum* (taxID 216816) (https://doi.org/10.6084/m9.figshare.19709917). The metapangenome was constructed using anvi'o version 6.2 (128) according to the pangenome workflow (114). HGCs were identified in this set of genomes based on all-versus-all sequence similarity. Briefly, this workflow uses BLASTP to compute the average nucleotide identity (ANI) between all pairs of genes, uses the Markov cluster algorithm (MCL) (129) to generate homologous gene clusters, and aligns amino acid sequences using MUSCLE (130) for each gene cluster. Each gene was assigned as core or accessory according to the hierarchical clustering of the gene clusters. Sourmash version 3.3 (131) was used to compute ANIs across genomes. To count a gene as being present in the sample, it had to be of at least 50 reads mapping to at least one *Bifidobacterium* species genome, and the total abundance had to be at least 0.1% after normalizing over the total number of reads. For long-read data sequenced on the Pacific Biosciences Sequel II platform, quality control (QC) and assembly were performed using Canu-1.8 (86). The assemblies were assigned species names through BLAST to the RefSeq data set and confirmed with GTDB-Tk v1.1.0 (89). Genomes of the assemblies assigned to *B. breve* were aligned to reference *B. breve* genome JCM1192 using MAUVE aligner (132, 133).

Code availability. The R code for processing these sequences and the SAS code used in this statistical analysis have been deposited at https://github.com/igsbma/IP_microbiome/tree/main/statistical_analyses. Detailed information on sequences and annotation of the pangenome can be retrieved at https://github.com/igsbma/IP_microbiome/tree/main/pangenome.

Data availability. All metagenomic, metataxonomic, and genomic data were deposited in the NCBI database under BioProject accession number PRJNA774819 for open assessment. Illumina 16S rRNA V3-V4 gene amplicon and Pacific Biosciences full-length 16S rRNA gene data were deposited in the Sequence Read Archive under accession numbers SRX12805867 to SRX12806634. Data deposition includes samples of positive and negative controls in each plate. Metagenomic data using the Pacific Biosciences platform were deposited in the Sequence Read Archive under accession numbers SRR16598000 and SRR16598001. Metagenomic data using the Illumina platform were deposited in BioProject under accession numbers SRX12798907 to SRX12798933. The assembled genomes of *B. breve* were deposited in GenBank under accession numbers JAJGBR000000000 and JAJGBS000000000.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.1 MB.
FIG S2, PDF file, 0.2 MB.
FIG S3, PDF file, 0.1 MB.
FIG S4, PDF file, 0.1 MB.
FIG S5, PDF file, 0.4 MB.
FIG S6, PDF file, 1.1 MB.
TABLE S1, XLSX file, 0.02 MB.
TABLE S2, DOCX file, 0.02 MB.
TABLE S3, XLSX file, 0.02 MB.
TABLE S4, XLSX file, 0.03 MB.

ACKNOWLEDGMENTS

This study was supported in part by a Gerber Foundation 2018 award (project identifier 6361), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health under award number R21DK123674, and an Institute for Clinical and Translational Research (ICTR) at the University of Maryland Accelerated Translational Incubator Pilot (ATIP) award.

We thank Jonathan Meddings and Kim Le at the University of Calgary, Calgary, Alberta, Canada, for the HPLC analysis of serum and urine samples; Ivette Santana-Cruz (GRC at IGS) for constructive discussion on long-read data processing; and Shilpa Narina and Sarah Arbaugh for great assistance in clinical specimen collection.

B.M., J.R., S.S., and R.M.V. designed the research; S.S., E.J., E.M., B.M., G.N., H.Y., L.S.R., and R.M.V. conducted the clinical study; B.M., H.Y., L.F., and M.H. conducted the short-read sequencing; B.M., E.M., M.H., L.S., and L.J.T. conducted the long-read sequencing; B.M., M.F., and P.G. performed the statistical analyses; B.M. and G.N. performed the epidemiological analyses; and B.M., S.S., J.M.L.-D., G.N., M.F., M.F.P., J.R., and R.M.V. wrote the paper.

REFERENCES

1. Hunter CJ, Upperman JS, Ford HR, Camerini V. 2008. Understanding the susceptibility of the premature infant to necrotizing enterocolitis (NEC). Pediatr Res 63:117–123. https://doi.org/10.1203/PDR.0b013e31815ed64c.
2. Fitzgibbons SC, Ching Y, Yu D, Carpenter J, Kenny M, Weldon C, Lillehei C, Valim C, Horbar JD, Jaksic T. 2009. Mortality of necrotizing enterocolitis expressed by birth weight categories. J Pediatr Surg 44:1072–1075. https://doi.org/10.1016/j.jpedsurg.2009.02.013.
3. Fox TP, Godavitarne C. 2012. What really causes necrotising enterocolitis ? ISRN Gastroenterol 2012:628317. https://doi.org/10.5402/2012/628317.
4. Anand RJ, Leaphart CL, Mollen KP, Hackam DJ. 2007. The role of the intestinal barrier in the pathogenesis of necrotizing enterocolitis. Shock 27:124–133. https://doi.org/10.1097/01.shk.0000239774.02904.65.
5. Bergmann KR, Liu SX, Tian R, Kushnir A, Turner JR, Li HL, Chou PM, Weber CR, De Plaen IG. 2013. Bifidobacteria stabilize claudins at tight junctions and prevent intestinal barrier dysfunction in mouse necrotizing enterocolitis. Am J Pathol 182:1595–1606. https://doi.org/10.1016/j.ajpath.2013.01.013.
6. Fasano A. 2008. Physiological, pathological, and therapeutic implications of zonulin-mediated intestinal barrier modulation: living life on the edge of the wall. Am J Pathol 173:1243–1252. https://doi.org/10.2353/ajpath.2008.080192.
7. Nanthakumar N, Meng D, Goldstein AM, Zhu W, Lu L, Uauy R, Llanos A, Claud EC, Walker WA. 2011. The mechanism of excessive intestinal inflammation in necrotizing enterocolitis: an immature innate immune response. PLoS One 6:e17776. https://doi.org/10.1371/journal.pone.0017776.
8. Claud EC, Walker WA. 2008. Bacterial colonization, probiotics, and necrotizing enterocolitis. J Clin Gastroenterol 42(Suppl 2):S46–S52. https://doi.org/10.1097/MCG.0b013e31815a57a8.
9. Berman L, Moss RL. 2011. Necrotizing enterocolitis: an update. Semin Fetal Neonatal Med 16:145–150. https://doi.org/10.1016/j.siny.2011.02.002.
10. Guner YS, Friedlich P, Wee CP, Dorey F, Camerini V, Upperman JS. 2009. State-based analysis of necrotizing enterocolitis outcomes. J Surg Res 157:21–29. https://doi.org/10.1016/j.jss.2008.11.008.
11. Ganapathy V, Hay JW, Kim JH. 2012. Costs of necrotizing enterocolitis and cost-effectiveness of exclusively human milk-based products in feeding extremely premature infants. Breastfeed Med 7:29–37. https://doi.org/10.1089/bfm.2011.0002.
12. Neu J, Walker WA. 2011. Necrotizing enterocolitis. N Engl J Med 364:255–264. https://doi.org/10.1056/NEJMra1005408.
13. Claud EC. 2009. Neonatal necrotizing enterocolitis—inflammation and intestinal immaturity. Antiinflamm Antiallergy Agents Med Chem 8:248–259. https://doi.org/10.2174/187152309789152020.
14. Neish AS. 2009. Microbes in gastrointestinal health and disease. Gastroenterology 136:65–80. https://doi.org/10.1053/j.gastro.2008.10.080.
15. Ma B, McComb E, Gajer P, Yang H, Humphrys M, Okogbule-Wonodi AC, Fasano A, Ravel J, Viscardi RM. 2018. Microbial biomarkers of intestinal barrier maturation in preterm infants. Front Microbiol 9:2755. https://doi.org/10.3389/fmicb.2018.02755.
16. Saleem B, Okogbule-Wonodi AC, Fasano A, Magder LS, Ravel J, Kapoor S, Viscardi RM. 2017. Intestinal barrier maturation in very low birthweight infants: relationship to feeding and antibiotic exposure. J Pediatr 183:31–36.e1. https://doi.org/10.1016/j.jpeds.2017.01.013.
17. Weaver LT, Laker MF, Nelson R. 1984. Intestinal permeability in the newborn. Arch Dis Child 59:236–241. https://doi.org/10.1136/adc.59.3.236.
18. Halpern MD, Denning PW. 2015. The role of intestinal epithelial barrier function in the development of NEC. Tissue Barriers 3:e1000707. https://doi.org/10.1080/21688370.2014.1000707.
19. Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 2015. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. Microbiome 3:26. https://doi.org/10.1186/s40168-015-0087-4.
20. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Appl Environ Microbiol 74:2461–2470. https://doi.org/10.1128/AEM.02272-07.
21. Sim K, Cox MJ, Wopereis H, Martin R, Knol J, Li MS, Cookson WO, Moffatt MF, Kroll JS. 2012. Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing. PLoS One 7:e32543. https://doi.org/10.1371/journal.pone.0032543.
22. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, Ross MC, Lloyd RE, Doddapaneni H, Metcalf GA, Muzny D, Gibbs RA, Vatanen T, Huttenhower C, Xavier RJ, Rewers M, Hagopian W, Toppari J, Ziegler A-G, She J-X, Akolkar B, Lernmark A, Hyoty H, Vehik K, Krischer JP, Petrosino JF. 2018. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature 562:583–588. https://doi.org/10.1038/s41586-018-0617-x.
23. O'Connell Motherway M, Houston A, O'Callaghan G, Reunanen J, O'Brien F, O'Driscoll T, Casey PG, de Vos WM, van Sinderen D, Shanahan F. 2019. A bifidobacterial pilus-associated protein promotes colonic epithelial proliferation. Mol Microbiol 111:287–301. https://doi.org/10.1111/mmi.14155.
24. Sonnenburg JL, Chen CT, Gordon JI. 2006. Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. PLoS Biol 4:e413. https://doi.org/10.1371/journal.pbio.0040413.
25. van Wijck K, Bessems BAFM, van Eijk HMH, Buurman WA, Dejong CHC, Lenaerts K. 2012. Polyethylene glycol versus dual sugar assay for gastrointestinal permeability analysis: is it time to choose? Clin Exp Gastroenterol 5:139–150. https://doi.org/10.2147/CEG.S31799.
26. van Wijck K, Verlinden TJ, van Eijk HM, Dekker J, Buurman WA, Dejong CH, Lenaerts K. 2013. Novel multi-sugar assay for site-specific gastrointestinal permeability analysis: a randomized controlled crossover trial. Clin Nutr 32:245–251. https://doi.org/10.1016/j.clnu.2012.06.014.
27. Asztalos EV. 2018. Supporting mothers of very preterm infants and breast milk production: a review of the role of galactogogues. Nutrients 10:600. https://doi.org/10.3390/nu10050600.
28. Bittinger K, Zhao C, Li Y, Ford E, Friedman ES, Ni J, Kulkarni CV, Cai J, Tian Y, Liu Q, Patterson AD, Sarkar D, Chan SHJ, Maranas C, Saha-Shah A, Lund P, Garcia BA, Mattei LM, Gerber JS, Elovitz MA, Kelly A, DeRusso P, Kim D, Hofstaedter CE, Goulian M, Li H, Bushman FD, Zemel BS, Wu GD. 2020. Bacterial colonization reprograms the neonatal gut metabolome. Nat Microbiol 5:838–847. https://doi.org/10.1038/s41564-020-0694-0.
29. Del Chierico F, Vernocchi P, Petrucca A, Paci P, Fuentes S, Pratico G, Capuani G, Masotti A, Reddel S, Russo A, Vallone C, Salvatori G, Buffone E, Signore F, Rigon G, Dotta A, Miccheli A, de Vos WM, Dallapiccola B, Putignani L. 2015. Phylogenetic and metabolic tracking of gut microbiota during perinatal development. PLoS One 10:e0137347. https://doi.org/10.1371/journal.pone.0137347.
30. La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM, Stevens HJ, Bennett WE, Jr, Shaikh N, Linneman LA, Hoffmann JA, Hamvas A, Deych E, Shands BA, Shannon WD, Tarr PI. 2014. Patterned progression of bacterial populations in the premature infant gut. Proc Natl Acad Sci U S A 111:12522–12527. https://doi.org/10.1073/pnas.1409497111.
31. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, Khan MT, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee YS, Kotowska D, Colding C, Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L, Kristiansen K, Dahlgren J, Wang J, Jun W. 2015. Dynamics and

stabilization of the human gut microbiome during the first year of life. Cell Host Microbe 17:690–703. https://doi.org/10.1016/j.chom.2015.04.004.
32. Robertson RC, Manges AR, Finlay BB, Prendergast AJ. 2019. The human microbiome and child growth—first 1000 days and beyond. Trends Microbiol 27:131–147. https://doi.org/10.1016/j.tim.2018.09.008.
33. Odamaki T, Bottacini F, Kato K, Mitsuyama E, Yoshida K, Horigome A, Xiao JZ, van Sinderen D. 2018. Genomic diversity and distribution of Bifidobacterium longum subsp. longum across the human lifespan. Sci Rep 8:85. https://doi.org/10.1038/s41598-017-18391-x.
34. Zuniga M, Monedero V, Yebra MJ. 2018. Utilization of host-derived glycans by intestinal Lactobacillus and Bifidobacterium species. Front Microbiol 9:1917. https://doi.org/10.3389/fmicb.2018.01917.
35. Martens EC, Chiang HC, Gordon JI. 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe 4:447–457. https://doi.org/10.1016/j.chom.2008.09.007.
36. Sakanaka M, Gotoh A, Yoshida K, Odamaki T, Koguchi H, Xiao JZ, Kitaoka M, Katayama T. 2019. Varied pathways of infant gut-associated Bifidobacterium to assimilate human milk oligosaccharides: prevalence of the gene set and its correlation with bifidobacteria-rich microbiota formation. Nutrients 12:71. https://doi.org/10.3390/nu12010071.
37. Odamaki T, Horigome A, Sugahara H, Hashikura N, Minami J, Xiao JZ, Abe F. 2015. Comparative genomics revealed genetic diversity and species/strain-level differences in carbohydrate metabolism of three probiotic bifidobacterial species. Int J Genomics 2015:567809. https://doi.org/10.1155/2015/567809.
38. Ruiz-Moyano S, Totten SM, Garrido DA, Smilowitz JT, German JB, Lebrilla CB, Mills DA. 2013. Variation in consumption of human milk oligosaccharides by infant gut-associated strains of Bifidobacterium breve. Appl Environ Microbiol 79:6040–6049. https://doi.org/10.1128/AEM.01843-13.
39. Tam R, Saier MH, Jr. 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol Rev 57:320–346. https://doi.org/10.1128/mr.57.2.320-346.1993.
40. Suzuki R, Wada J, Katayama T, Fushinobu S, Wakagi T, Shoun H, Sugimoto H, Tanaka A, Kumagai H, Ashida H, Kitaoka M, Yamamoto K. 2008. Structural and thermodynamic analyses of solute-binding protein from Bifidobacterium longum specific for core 1 disaccharide and lacto-N-biose I. J Biol Chem 283:13165–13173. https://doi.org/10.1074/jbc.M709777200.
41. Garrido D, Kim JH, German JB, Raybould HE, Mills DA. 2011. Oligosaccharide binding proteins from Bifidobacterium longum subsp. infantis reveal a preference for host glycans. PLoS One 6:e17315. https://doi.org/10.1371/journal.pone.0017315.
42. Garrido D, Ruiz-Moyano S, Kirmiz N, Davis JC, Totten SM, Lemay DG, Ugalde JA, German JB, Lebrilla CB, Mills DA. 2016. A novel gene cluster allows preferential utilization of fucosylated milk oligosaccharides in Bifidobacterium longum subsp. longum SC596. Sci Rep 6:35045. https://doi.org/10.1038/srep35045.
43. Robertson AM, Wright DP. 1997. Bacterial glycosulphatases and sulphomucin degradation. Can J Gastroenterol 11:361–366. https://doi.org/10.1155/1997/642360.
44. Filipe MI. 1979. Mucins in the human gastrointestinal epithelium: a review. Invest Cell Pathol 2:195–216.
45. Berteau O, Guillot A, Benjdia A, Rabot S. 2006. A new type of bacterial sulfatase reveals a novel maturation pathway in prokaryotes. J Biol Chem 281:22464–22470. https://doi.org/10.1074/jbc.M602504200.
46. Egan M, Jiang H, O'Connell Motherway M, Oscarson S, van Sinderen D. 2016. Glycosulfatase-encoding gene cluster in Bifidobacterium breve UCC2003. Appl Environ Microbiol 82:6611–6623. https://doi.org/10.1128/AEM.02022-16.
47. Corfield AP, Wagner SA, Safe A, Mountford RA, Clamp JR, Kamerling JP, Vliegenthart JF, Schauer R. 1993. Sialic acids in human gastric aspirates: detection of 9-O-lactyl- and 9-O-acetyl-N-acetylneuraminic acids and a decrease in total sialic acid concentration with age. Clin Sci 84:573–579. https://doi.org/10.1042/cs0840573.
48. Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta JP, Michalski JC. 2003. Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. J Biol Chem 278:46337–46348. https://doi.org/10.1074/jbc.M302529200.
49. Egan M, O'Connell Motherway M, Ventura M, van Sinderen D. 2014. Metabolism of sialic acid by Bifidobacterium breve UCC2003. Appl Environ Microbiol 80:4414–4426. https://doi.org/10.1128/AEM.01114-14.
50. Podolsky DK. 1985. Oligosaccharide structures of human colonic mucin. J Biol Chem 260:8262–8271. https://doi.org/10.1016/S0021-9258(17)39465-6.
51. Kunz C, Rudloff S, Baier W, Klein N, Strobel S. 2000. Oligosaccharides in human milk: structural, functional, and metabolic aspects. Annu Rev Nutr 20:699–722. https://doi.org/10.1146/annurev.nutr.20.1.699.
52. Turroni F, Milani C, Duranti S, Mahony J, van Sinderen D, Ventura M. 2018. Glycan utilization and cross-feeding activities by bifidobacteria. Trends Microbiol 26:339–350. https://doi.org/10.1016/j.tim.2017.10.001.
53. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, Armanini F, Truong DT, Manara S, Zolfo M, Beghini F, Bertorelli R, De Sanctis V, Bariletti I, Canto R, Clementi R, Cologna M, Crifo T, Cusumano G, Gottardi S, Innamorati C, Mase C, Postai D, Savoi D, Duranti S, Lugli GA, Mancabelli L, Turroni F, Ferrario C, Milani C, Mangifesta M, Anzalone R, Viappiani A, Yassour M, Vlamakis H, Xavier R, Collado CM, Koren O, Tateo S, Soffiati M, Pedrotti A, Ventura M, Huttenhower C, Bork P, Segata N. 2018. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. Cell Host Microbe 24:133–145.e5. https://doi.org/10.1016/j.chom.2018.06.005.
54. Sela DA, Chapman J, Adeuya A, Kim JH, Chen F, Whitehead TR, Lapidus A, Rokhsar DS, Lebrilla CB, German JB, Price NP, Richardson PM, Mills DA. 2008. The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. Proc Natl Acad Sci U S A 105:18964–18969. https://doi.org/10.1073/pnas.0809584105.
55. James K, O'Connell Motherway M, Bottacini F, van Sinderen D. 2016. Bifidobacterium breve UCC2003 metabolises the human milk oligosaccharides lacto-N-tetraose and lacto-N-neo-tetraose through overlapping, yet distinct pathways. Sci Rep 6:38560. https://doi.org/10.1038/srep38560.
56. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, Tobe T, Clarke JM, Topping DL, Suzuki T, Taylor TD, Itoh K, Kikuchi J, Morita H, Hattori M, Ohno H. 2011. Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature 469:543–547. https://doi.org/10.1038/nature09646.
57. Alcon-Giner C, Dalby MJ, Caim S, Ketskemety J, Shaw A, Sim K, Lawson MAE, Kiu R, Leclaire C, Chalklen L, Kujawska M, Mitra S, Fardus-Reid F, Belteki G, McColl K, Swann JR, Kroll JS, Clarke P, Hall LJ. 2020. Microbiota supplementation with Bifidobacterium and Lactobacillus modifies the preterm infant gut microbiota and metabolome: an observational study. Cell Rep Med 1:100077. https://doi.org/10.1016/j.xcrm.2020.100077.
58. Lawson MAE, O'Neill IJ, Kujawska M, Gowrinadh Javvadi S, Wijeyesekera A, Flegg Z, Chalklen L, Hall LJ. 2020. Breast milk-derived human milk oligosaccharides promote Bifidobacterium interactions within a single ecosystem. ISME J 14:635–648. https://doi.org/10.1038/s41396-019-0553-2.
59. Turroni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J, Murray D, Fuligni F, Gueimonde M, Margolles A, De Bellis G, O'Toole PW, van Sinderen D, Marchesi JR, Ventura M. 2012. Diversity of bifidobacteria within the infant gut microbiota. PLoS One 7:e36957. https://doi.org/10.1371/journal.pone.0036957.
60. Soto A, Martin V, Jimenez E, Mader I, Rodriguez JM, Fernandez L. 2014. Lactobacilli and bifidobacteria in human breast milk: influence of antibiotherapy and other host and clinical factors. J Pediatr Gastroenterol Nutr 59:78–88. https://doi.org/10.1097/MPG.0000000000000347.
61. Ma B, France MT, Crabtree J, Holm JB, Humphrys MS, Brotman RM, Ravel J. 2020. A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. Nat Commun 11:940. https://doi.org/10.1038/s41467-020-14677-3.
62. LoCascio RG, Ninonuevo MR, Freeman SL, Sela DA, Grimm R, Lebrilla CB, Mills DA, German JB. 2007. Glycoprofiling of bifidobacterial consumption of human milk oligosaccharides demonstrates strain specific, preferential consumption of small chain glycans secreted in early human lactation. J Agric Food Chem 55:8914–8919. https://doi.org/10.1021/jf0710480.
63. Strum JS, Kim J, Wu S, De Leoz ML, Peacock K, Grimm R, German JB, Mills DA, Lebrilla CB. 2012. Identification and accurate quantitation of biological oligosaccharide mixtures. Anal Chem 84:7793–7801. https://doi.org/10.1021/ac301128s.
64. Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, Desai P, Lebrilla CB, Weimer BC, Mills DA, German JB, Sonnenburg JL. 2011. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. Cell Host Microbe 10:507–514. https://doi.org/10.1016/j.chom.2011.10.007.
65. Miwa M, Horimoto T, Kiyohara M, Katayama T, Kitaoka M, Ashida H, Yamamoto K. 2010. Cooperation of beta-galactosidase and beta-N-acetylhexosaminidase from bifidobacteria in assimilation of human milk oligosaccharides with type 2 structure. Glycobiology 20:1402–1409. https://doi.org/10.1093/glycob/cwq101.

66. Ozcan E, Sela DA. 2018. Inefficient metabolism of the human milk oligosaccharides lacto-N-tetraose and lacto-N-neotetraose shifts Bifidobacterium longum subsp. infantis physiology. Front Nutr 5:46. https://doi.org/10.3389/fnut.2018.00046.
67. Garrido D, Dallas DC, Mills DA. 2013. Consumption of human milk glycoconjugates by infant-associated bifidobacteria: mechanisms and implications. Microbiology (Reading) 159:649–664. https://doi.org/10.1099/mic.0.064113-0.
68. Kitajima H, Sumida Y, Tanaka R, Yuki N, Takayama H, Fujimura M. 1997. Early administration of Bifidobacterium breve to preterm infants: randomised controlled trial. Arch Dis Child Fetal Neonatal Ed 76:F101–F107. https://doi.org/10.1136/fn.76.2.f101.
69. Costeloe K, Hardy P, Juszczak E, Wilks M, Millar MR, Probiotics in Preterm Infants Study Collaborative Group. 2016. Bifidobacterium breve BBG-001 in very preterm infants: a randomised controlled phase 3 trial. Lancet 387:649–660. https://doi.org/10.1016/S0140-6736(15)01027-2.
70. Tailford LE, Crost EH, Kavanaugh D, Juge N. 2015. Mucin glycan foraging in the human gut microbiome. Front Genet 6:81. https://doi.org/10.3389/fgene.2015.00081.
71. Wang B, Brand-Miller J, McVeagh P, Petocz P. 2001. Concentration and distribution of sialic acid in human milk and infant formulas. Am J Clin Nutr 74:510–515. https://doi.org/10.1093/ajcn/74.4.510.
72. De Leoz ML, Gaerlan SC, Strum JS, Dimapasoc LM, Mirmiran M, Tancredi DJ, Smilowitz JT, Kalanetra KM, Mills DA, German JB, Lebrilla CB, Underwood MA. 2012. Lacto-N-tetraose, fucosylation, and secretor status are highly variable in human milk oligosaccharides from women delivering preterm. J Proteome Res 11:4662–4672. https://doi.org/10.1021/pr3004979.
73. Charbonneau MR, O'Donnell D, Blanton LV, Totten SM, Davis JC, Barratt MJ, Cheng J, Guruge J, Talcott M, Bain JR, Muehlbauer MJ, Ilkayeva O, Wu C, Struckmeyer T, Barile D, Mangani C, Jorgensen J, Fan YM, Maleta K, Dewey KG, Ashorn P, Newgard CB, Lebrilla C, Mills DA, Gordon JI. 2016. Sialylated milk oligosaccharides promote microbiota-dependent growth in models of infant undernutrition. Cell 164:859–871. https://doi.org/10.1016/j.cell.2016.01.024.
74. Ward RE, Ninonuevo M, Mills DA, Lebrilla CB, German JB. 2007. In vitro fermentability of human milk oligosaccharides by several strains of bifidobacteria. Mol Nutr Food Res 51:1398–1405. https://doi.org/10.1002/mnfr.200700150.
75. Sela DA, Li Y, Lerno L, Wu S, Marcobal AM, German JB, Chen X, Lebrilla CB, Mills DA. 2011. An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides. J Biol Chem 286:11909–11918. https://doi.org/10.1074/jbc.M110.193359.
76. Hood MI, Skaar EP. 2012. Nutritional immunity: transition metals at the pathogen-host interface. Nat Rev Microbiol 10:525–537. https://doi.org/10.1038/nrmicro2836.
77. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, Naidu N, Choudhury B, Weimer BC, Monack DM, Sonnenburg JL. 2013. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. Nature 502:96–99. https://doi.org/10.1038/nature12503.
78. Sotoya H, Shigehisa A, Hara T, Matsumoto H, Hatano H, Matsuki T. 2017. Identification of genes involved in galactooligosaccharide utilization in Bifidobacterium breve strain YIT 4014(T). Microbiology (Reading) 163:1420–1428. https://doi.org/10.1099/mic.0.000517.
79. Cortes-Macias E, Selma-Royo M, Garcia-Mantrana I, Calatayud M, Gonzalez S, Martinez-Costa C, Collado MC. 2021. Maternal diet shapes the breast milk microbiota composition and diversity: impact of mode of delivery and antibiotic exposure. J Nutr 151:330–340. https://doi.org/10.1093/jn/nxaa310.
80. Wang S, Ryan CA, Boyaval P, Dempsey EM, Ross RP, Stanton C. 2020. Maternal vertical transmission affecting early-life microbiota development. Trends Microbiol 28:28–45. https://doi.org/10.1016/j.tim.2019.07.010.
81. Kim J, Unger S. 2010. Human milk banking. Paediatr Child Health 15:595–602. https://doi.org/10.1093/pch/15.9.595.
82. Rouwet EV, Heineman E, Buurman WA, ter Riet G, Ramsay G, Blanco CE. 2002. Intestinal permeability and carrier-mediated monosaccharide absorption in preterm neonates during the early postnatal period. Pediatr Res 51:64–70. https://doi.org/10.1203/00006450-200201000-00012.
83. Grier A, Qiu X, Bandyopadhyay S, Holden-Wiltse J, Kessler HA, Gill AL, Hamilton B, Huyck H, Misra S, Mariani TJ, Ryan RM, Scholer L, Scheible KM, Lee YH, Caserta MT, Pryhuber GS, Gill SR. 2017. Impact of prematurity and nutrition on the developing gut microbiome and preterm infant growth. Microbiome 5:158. https://doi.org/10.1186/s40168-017-0377-0.
84. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2:6. https://doi.org/10.1186/2049-2618-2-6.
85. Pacific Biosciences. 2019. Procedure & checklist—amplification of full-length 16S gene with barcoded primers for multiplexed SMRTbell library preparation and sequencing. Pacific Biosciences, Menlo Park, CA. https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Full-Length-16S-Amplification-SMRTbell-Library-Preparation-an.
86. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
87. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
88. Kitson E. 2018. Simple-Circularise. https://github.com/Kzra/Simple-Circularise. Accessed 19 September 2019.
89. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
90. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
91. Schmieder R, Lim YW, Rohwer F, Edwards R. 2010. TagCleaner: identification and removal of tag sequences from genomic and metagenomic datasets. BMC Bioinformatics 11:341. https://doi.org/10.1186/1471-2105-11-341.
92. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.
93. Edgar RC, Flyvbjerg H. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31:3476–3482. https://doi.org/10.1093/bioinformatics/btv401.
94. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and Web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.
95. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, McGill SK, Dougherty MK. 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Res 47:e103. https://doi.org/10.1093/nar/gkz569.
96. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229.
97. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res 42:D643–D648. https://doi.org/10.1093/nar/gkt1209.
98. 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. https://doi.org/10.1128/AEM.00062-07.
99. Anonymous. 2019. NCBI 16S RefSeq nucleotide sequence records. NCBI, Bethesda, MD.
100. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37:D141–D145. https://doi.org/10.1093/nar/gkn879.
101. Ritari J, Salojarvi J, Lahti L, de Vos WM. 2015. Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. BMC Genomics 16:1056. https://doi.org/10.1186/s12864-015-2265-y.
102. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2011. vegan: community ecology package. R package version 20-2.
103. Mächler M, Rousseeuw P, Struyf A, Hubert M, Hornik K, Studer M, Roudier P, Gonzalez J. 2017. Cluster: "Finding groups in data": Cluster analysis extended Rousseeuw et al. https://svn.r-project.org/R-packages/trunk/cluster/.
104. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
105. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. https://doi.org/10.1371/journal.pone.0061217.
106. Bokulich NA, Dillon MR, Zhang Y, Rideout JR, Bolyen E, Li H, Albert PS, Caporaso JG. 2018. q2-longitudinal: longitudinal and paired-sample

analyses of microbiome data. mSystems 3:e00219-18. https://doi.org/10.1128/mSystems.00219-18.
107. Gretton A, Herbrich R, Smola A, Bousquet O, Scholkopf B. 2005. Kernel methods for measuring independence. J Mach Learn Res 6:2075–2129.
108. Plummer M. 2011. rjags: Bayesian graphical models using MCMC. http://CRAN.R-project.org/package=rjags.
109. R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
110. Faulkner JR, Minin V. 2018. Locally adaptive smoothing with Markov random fields and shrinkage priors. Bayesian Anal 13:225–252. https://doi.org/10.1214/17-BA1050.
111. Stan Development Team. 2018. RStan: the R interface to Stan. R package version 2.17.3.
112. Frank E, Hall M, Trigg L, Holmes G, Witten IH. 2004. Data mining in bioinformatics using Weka. Bioinformatics 20:2479–2481. https://doi.org/10.1093/bioinformatics/bth261.
113. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. 2009. The WEKA data mining software: an update. SIGKDD Explor 11:10–18. https://doi.org/10.1145/1656274.1656278.
114. Muret EA. 2020. An anvi'o workflow for microbial pangenomics. http://merenlab.org/2016/11/08/pangenomics-v2/.
115. Storey J, Bass A, Dabney A, Robinson D. 2020. qvalue: Q-value estimation for false discovery rate control. R package version 2.22.0. http://github.com/jdstorey/qvalue.
116. Rotmistrovsky K, Agarwala R. 2011. BMTagger: best match tagger for removing human reads from metagenomics datasets. NCBI/NLM, National Institutes of Health, Bethesda, MD.
117. Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, Chen HC, Agarwala R, McLaren WM, Ritchie GR, Albracht D, Kremitzki M, Rock S, Kotkiewicz H, Kremitzki C, Wollam A, Trani L, Fulton L, Fulton R, Matthews L, Whitehead S, Chow W, Torrance J, Dunn M, Harden G, Threadgold G, Wood J, Collins J, Heath P, Griffiths G, Pelan S, Grafham D, Eichler EE, Weinstock G, Mardis ER, Wilson RK, Howe K, Flicek P, Hubbard T. 2011. Modernizing reference genome assemblies. PLoS Biol 9:e1001091. https://doi.org/10.1371/journal.pbio.1001091.
118. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25.
119. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
120. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 9:811–814. https://doi.org/10.1038/nmeth.2066.
121. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
122. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.
123. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293. https://doi.org/10.1093/nar/gkv1248.
124. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109–D114. https://doi.org/10.1093/nar/gkr988.
125. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344.
126. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The Carbohydrate-Active Enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.
127. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233–D238. https://doi.org/10.1093/nar/gkn663.
128. Delmont TO, Eren AM. 2018. Linking pangenomes and metagenomes: the Prochlorococcus metapangenome. PeerJ 6:e4320. https://doi.org/10.7717/peerj.4320.
129. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30:1575–1584. https://doi.org/10.1093/nar/30.7.1575.
130. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
131. Pierce NT, Irber L, Reiter T, Brooks P, Brown CT. 2019. Large-scale sequence comparisons with sourmash. F1000Res 8:1006. https://doi.org/10.12688/f1000research.19675.1.
132. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147.
133. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics 25:2071–2073. https://doi.org/10.1093/bioinformatics/btp356.