Assembly, stability, and dynamics of the infant gut microbiome are linked to bacterial strains and functions in mother's milk

Mattea Allert[1,*], Pamela Ferretti[2,*], Kelsey E. Johnson[1], Timothy Heisel[3], Sara Gonia[3], Dan Knights[4,5], David A. Fields[6], Frank W. Albert[1], Ellen W. Demerath[1], Cheryl A. Gale[3], and Ran Blekhman[2]

[1] Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN, USA
[2] Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA
[3] Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA
[4] Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA
[5] BioTechnology Institute, College of Biological Sciences, University of Minnesota, Minneapolis, MN, USA
[6] Department of Pediatrics, the University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
* these authors contributed equally
Corresponding author: Ran Blekhman

Abstract

The establishment of the gut microbiome in early life is critical for healthy infant development. Although human milk is recommended as the sole source of nutrition for the human infant, little is known about how variation in milk composition, and especially the milk microbiome, shapes the microbial communities in the infant gut. Here, we quantified the similarity between the maternal milk and the infant gut microbiome using 507 metagenomic samples collected from 195 mother-infant pairs at one, three, and six months postpartum. We found that the microbial taxonomic overlap between milk and the infant gut was driven by bifidobacteria, in particular by *B. longum*. Infant stool samples dominated by *B. longum* also showed higher temporal stability compared to samples dominated by other species. We identified two instances of strain sharing between maternal milk and the infant gut, one involving a commensal (*B. longum*) and one a pathobiont (*K. pneumoniae*). In addition, strain sharing between unrelated infants was higher among infants born at the same hospital compared to infants born in different hospitals,

suggesting a potential role of the hospital environment in shaping the infant gut microbiome composition. The infant gut microbiome at one month compared to six months of age was enriched in metabolic pathways associated with *de-novo* molecule biosynthesis, suggesting that early colonisers might be more versatile and metabolically independent compared to later colonizers. Lastly, we found a significant overlap in antimicrobial resistance genes carriage between the mother's milk and their infant's gut microbiome. Taken together, our results suggest that the human milk microbiome has an important role in the assembly, composition, and stability of the infant gut microbiome.

Introduction

The gut microbiome composition and maturation in early life plays an important role in the development of the immune system[1,2], nutrient absorption[3,4] and metabolism regulation[5]. Its assembly is influenced by microbes acquired from the mother[6–9], among other sources[10], starting from the first day of life[11]. While the impact of the maternal gastrointestinal, vaginal, oral and cutaneous microbial species on the infant gut microbiome assembly and development have been extensively investigated[6,11–15], the role of the maternal breast milk microbiome in modulating infant gut microbes remains poorly understood.

Human breast milk represents, ideally, the sole source of nutrition for the infant in the first semester of life[16]. Nevertheless, milk remains heavily understudied, estimated to be less than 0.2% of all human-associated public metagenomic samples, based on data from 2021[17]. Maternal milk provides the infant with nutrients, such as proteins, lipids, carbohydrates (including human milk oligosaccharides), maternal immune cells, antibodies, and live bacteria[18–20]. The milk, and the bacteria that it carries, have therefore the potential to significantly impact infant gut microbiome composition, stability, and functionality. In addition, the presence of antimicrobial resistance genes in the milk microbiome can further influence the infant's health and its gut microbiome composition[21]. The microbiome could also underlie the benefits of exclusive breastfeeding for a variety of chronic conditions including asthma[22,23], childhood obesity[24], type 1 diabetes[25], and allergic disease[26].

Despite its clear importance for infant health, research on the human milk microbiome has lagged behind studies of other human body sites. This is partly due to the nature of the sample itself: laced with human cells and high fat content, and characterized by low microbial biomass[27], milk samples are challenging to process and sequence using standard

approaches[27]. The few available milk microbiome cohorts are either limited by a small sample size or by lack of paired infant samples. Studies have also been hindered by the use of 16S rRNA gene sequencing[28,29], which cannot reliably identify microbial strains[30] and does not provide insights into the functional potential of the bacterial communities.

The primary objective of this study was to investigate the maternal milk microbiome in relation to the infant gut microbiome. This was achieved by collecting breast milk and infant stool samples in the first six months postpartum from a large cohort of healthy mothers and their infants, all of whom exclusively breastfed for at least one month. We used high-throughput shotgun metagenomics to assess species composition and stability over time, and to identify strain-sharing events between mother-infant pairs and between unrelated infants. We then analyzed the microbial functional potential, with particular focus on metabolic pathways and the antimicrobial resistance carriage of both maternal milk and infant stool samples over time.

Results

Bifidobacterium longum drives the compositional overlap between the infant gut and the maternal milk microbiomes

We collected and sequenced 507 microbiome samples from 195 mother-infant pairs. All infants were born at term and were exclusively breastfed at one month of age. Breast milk was collected at one and three months postpartum (n=173), while infant stool samples were collected at one and six months postpartum (n=334, **Fig. 1A**). Infants were predominantly vaginally born (76%), exclusively breastfed at six months (66%), and antibiotics naive (66% had not received antibiotics by six months of age) (**Extended Data 1** and **Suppl. Table 1**)**.** After shotgun sequencing, 34 samples were excluded from downstream analysis due to low sequencing yield. Species-level taxonomic profiling was performed with MetaPhlAn4[31] (**Suppl. Table 2**).

The maternal milk microbiome was dominated by Bifidobacteria, in particular by *Bifidobacterium longum*, *Bifidobacterium breve* and *Bifidobacterium bifidum,* in order of prevalence (**Fig. 1B** and **Extended Data 2**). Other prominent members of the milk microbiome included skin-associated species such as *Staphylococcus epidermidis* and *Cutibacterium*

acnes, oral-associated species such as *Streptococcus salivarius*, and gut-associated species such as *Escherichia coli and Phocaeiola vulgatus* (**Fig. 1B**). This is consistent with previous studies showing detecting skin-associated taxa in milk in addition to Bifidobacteria[32,33].

The infant gut microbiome at one month was dominated by *E. coli*, *B. longum*, *B. breve*, *B. fragilis, B. bifidum*, *K. pneumoniae*, *Klebisella michiganensis*, and *Ruminoccus gnavus*, as well as species typically associated with the oral cavity, such as *V. parvula*, and *V. atypica* (**Fig. 1C**). *Phocaeiola vulgatus* and *E. coli* were among the species found in the infant gut as well as in the maternal milk (**Fig. 1B-C)**. Bifidobacteria was the most prevalent and abundant genus in milk and stool samples at six months (**Extended Data 3A**). At the family taxonomic level, the prevalence of Enterobacteriaceae in the infant gut increased over time, while Bifidobacteriaceae showed the opposite trend (**Extended Data 3B**). Based on the dominant (most abundant) species, the infant stool samples could be divided in four groups, hereafter referred to as predominance groups: the first three dominated by *B. longum*, *B. breve*, and *B. bifidum*, respectively, and a fourth group dominated by species not belonging to the Bifidobacterium genus, mostly *E. coli, B. fragilis*, and *P. vulgatus* (**Fig. 1C**). 26.7% of infant stool samples were dominated by *B. longum*, 14.4% by *B. breve*, 4.5% by *B. bifidum* and 54.2% by non-bifidobacteria species (**Extended Data 4A**).

Maternal milk was characterized by a significantly lower species richness than infant stool samples ($p=7.4x10^{-4}$, paired t-test; **Fig. 1D**). Microbial diversity in infant stool samples increased significantly over time ($p=2.4x10^{-7}$, paired t-test), while maternal milk showed a similar but milder trend (p=0.02, paired t-test; **Fig. 1D**). A principal component analysis showed that milk microbiome composition was partially overlapping with that of infant stool samples, especially those collected at six months of age (**Fig. 1E**). Such overlap was mostly driven by Bifidobacteria, in particular by *B. longum* (**Fig. 1E and Extended data 2**). Although all three Bifidobacterium strains were significantly correlated with the first principal coordinate, *B. longum* was more strongly correlated compared to *B. breve* and *B. bifidum* (Spearman's correlation, $R=-0.73$, $p<2.2x10^{-16}$; $R=-0.4$, $p<2.2x10^{-16}$; and $R=-0.2$, $p<5.3x10^{-6}$; respectively; **Fig. 1E)**. Overall, in both milk and infant stool samples, *B. longum* was the most predominant and abundant species (**Fig. 2A**).

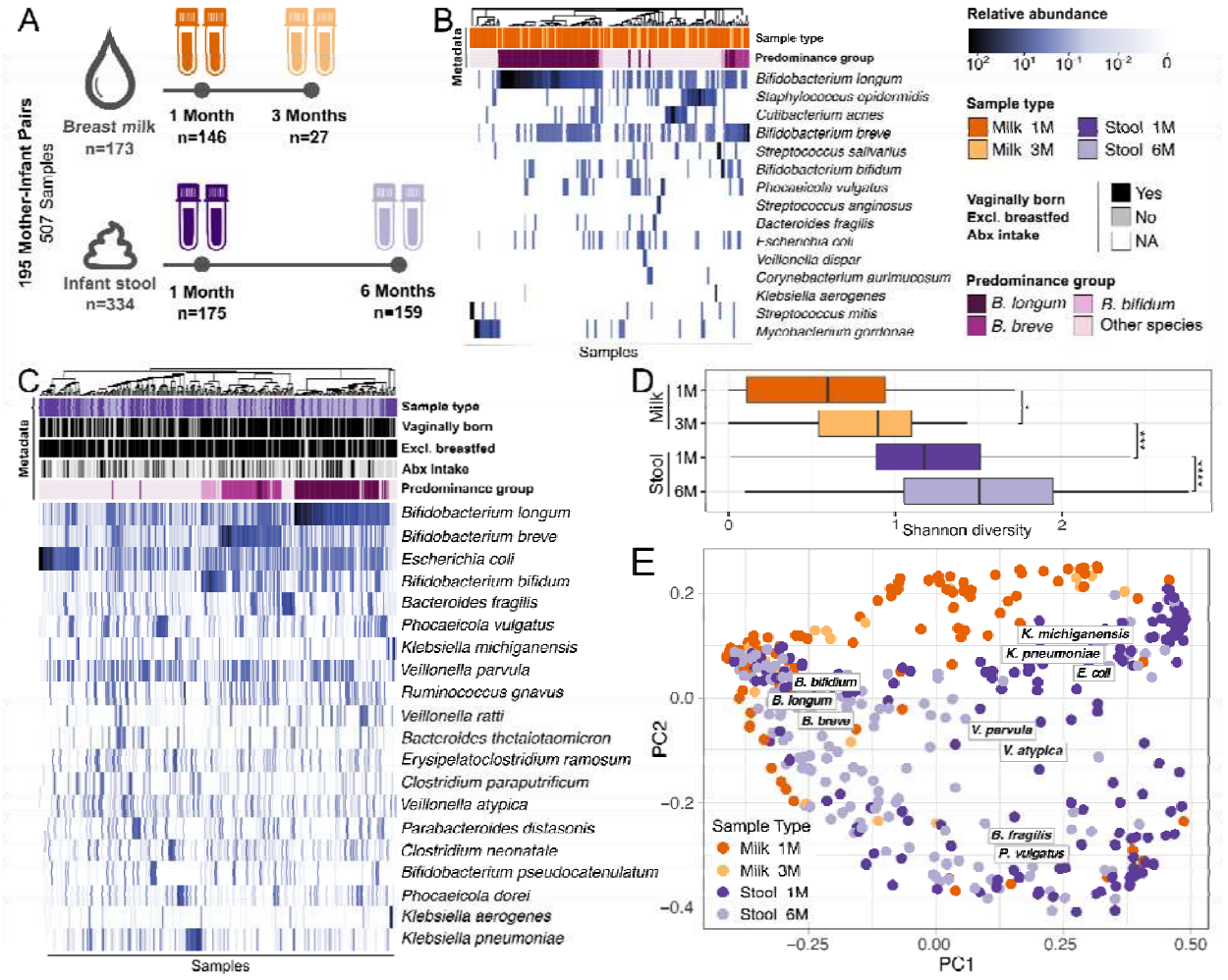


Figure 1. (A) Study design overview for the 507 samples collected from 195 mother-infant pairs. Infant stool samples (n=334) were collected at one and six months of life. Maternal breast milk samples (n=173) were collected one and three months after delivery. Taxonomic composition of (B) the most prevalent and abundant species found in the human breast milk and (C) in the infant gut microbiome samples in relation to sample collection time point, predominance group and other relevant infant metadata. Predominance group identifies the most abundant species in each sample. (D) Shannon diversity distribution for infant stool and maternal milk samples over time. P-values calculated using paired t-test (* p<0.05, ** p<0.01 and *** p<0.001). (E) Ordination plot based on Bray-Curtis distance between samples, colored by body site of origin and sampling time. Boxed species names indicate the species driving the clustering in that area of the PCoA and were obtained using the Weighted Averages Scores for species (see also Extended data 2).

B. longum in the infant gut becomes less prevalent over time and is associated with a more stable microbiome

To investigate the stability of the microbiome in our samples, we first focused on the above described predominance groups. In both breast milk and infant stool, *B. longum* was, among

bifidobacteria, the species with the highest prevalence (49.1% and 26.6%, respectively) and highest relative abundance (15.1% and 12.5%, respectively) (**Fig. 2A-B**). Specifically, 49.1% of milk samples were dominated by *B. longum*, 5.2% by *B. breve,* and 2.3% by *B. bifidum* (**Fig. 2A**). The samples dominated by *B. longum* had the most stable microbiome over time: 76.9% of milk samples dominated by *B. longum* at one month maintained *B. longum* as the predominant species at three months, while the predominant species changed to non-bifidobacteria species in only 23% of samples. In contrast, the samples dominated by *B. breve* were characterized by the lowest stability, as all milk samples dominated by *B. breve* at one month switched to having a non-bifidobacteria as the dominant species at 3 months post-delivery (**Fig. 2B**). In infant stool samples at one month, 41.1% of samples were dominated by *B. longum*, while at three months that percentage decreased to 15.7% (**Fig. 2B**). Conversely, 35% of stool samples at one month were dominated by non-bifidobacteria, and this proportion increased to 67.1% at six months (**Fig. 2B**). 60% of samples dominated by *B. longum* at one month switched to being dominated by non-bifidobacteria species by six months, while 82.4% of samples dominated by non-bifidobacteria species at one month were still dominated by non- bifidobacteria at six months (**Fig. 2B**). Overall, the prevalence of bifidobacteria in the infant gut decreased over time, with a particularly pronounced reduction in infants that had ceased exclusive breastfeeding prior to six months of age (**Fig. 2C**). In contrast, when bifidobacteria were still present at six months, their relative abundance increased compared to the previous time point, irrespective of the breastfeeding practices at six months of age (**Fig. 2D**), and with considerable inter-personal variability (**Extended Data 4B**).

Next, we looked at the longitudinal stability of the microbiome composition at the species-level. We found that *B. longum* and *E. coli* were more abundant at six months compared to one month (BH-adjusted $p=3x10^{-8}$ and $p=1.5x10^{-3}$, respectively, t-test), while *V. dispar* decreased over time (BH-adjusted $p=2.1x10^{-3}$, t-test; **Fig. 2E**). We then investigated if there were species that were differentially abundant between the two timepoints when stratifying the infants for exclusive breastfeeding. Infants that were exclusively breastfed through six months of age showed a significant increase in *B. longum*, *B. breve* and *E. coli* (BH-adjusted $p=4x10^{-8}$, $p= 1.9x10^{-5}$, and $p=1.7x10^{-3}$, respectively, paired t-test; **Fig 2F-G**), and a significant reduction in *K. pneumoniae* and *V. dispar* (BH-adjusted p=0.04 and $p=3.2x10^{-6}$, respectively, paired t-test; **Fig 2F-G**). Infants that were exclusively breastfed at one month but not at six months showed an increase in *R. gnavus* (BH-adjusted $p=1.2x10^{-3}$, t-test) as well as *B. longum*, although at a lesser extent than infants exclusively breastfed through six months (BH-adjusted

p=2.3x10^{-4}, t-test; **Fig 2F-G**). Overall, the infant stool samples that were dominated by *B. longum* at both one and six months of age showed the most stable community composition over time (**Extended data 5**).

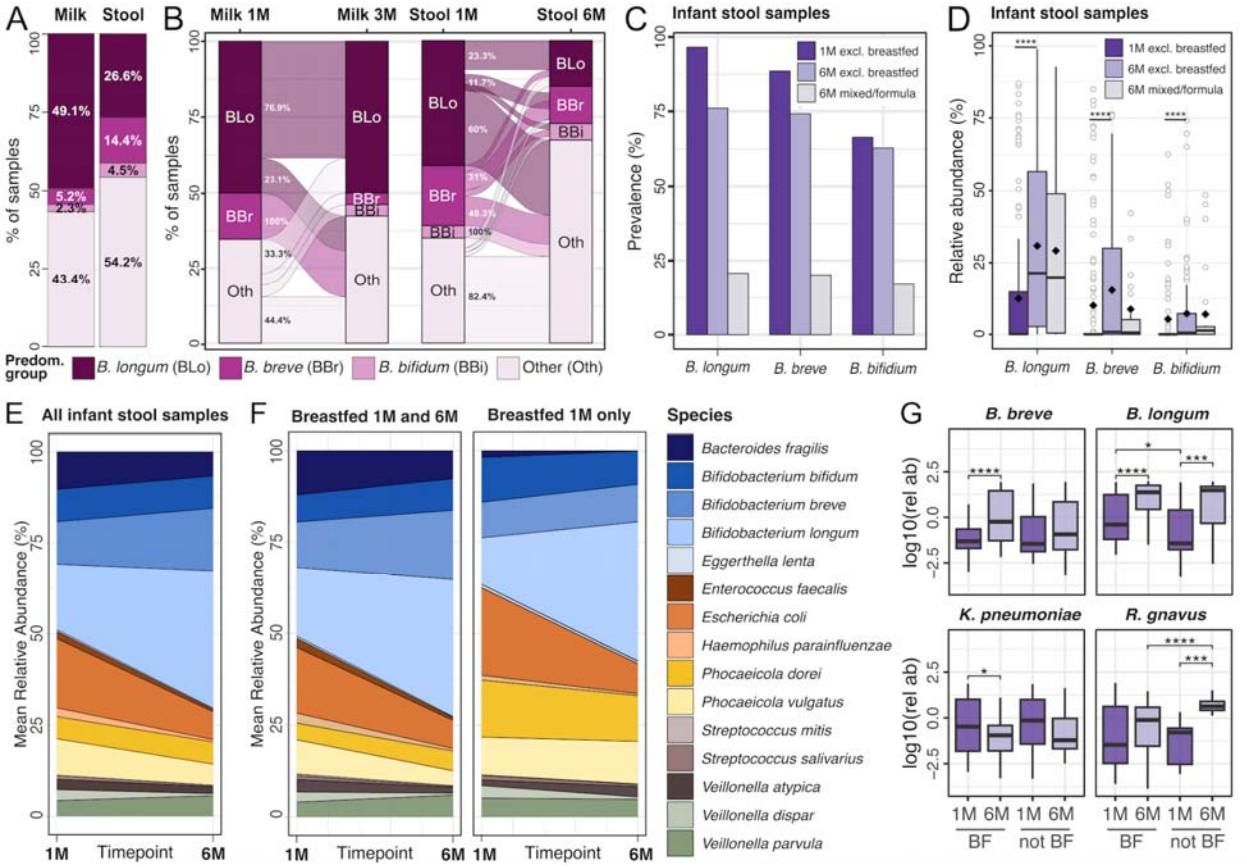


Figure 2. (A) Prevalence of each predominance group in milk and infant stool samples and (B) the transition of samples between predominance groups over time. Each sample is assigned one of four predominance groups indicated by the different colors. (C) Bifidobacteria mean prevalence and (D) distribution of relative abundances in exclusively breastfed infants at one and six months, and non-exclusively breastfed infants at six months of age. Black diamonds indicate the mean relative abundance per group. P-values calculated using Wilcoxon rank sum. Reported p-values are adjusted using Bonferroni method. (E) Species persistence in the infant gut across all samples and (F) stratified by breastfeeding at six months. (G) Relative abundance of some of the differentially abundant species between one and six months when divided by breastfeeding (BF) at six months. P-values calculated with paired t-test and adjusted with BH correction. **** for P ≤ 0.0001, *** for P ≤ 0.001, ** for P ≤ 0.01 and * for P ≤ 0.05.

Strain sharing is more common among unrelated infants born in the same hospitals

As species-level taxonomic profiling is not sufficient to identify potential transmission events, due to the genetic variability between conspecific strains[30], we performed strain-level profiling with StrainPhlAn4[31]. This allows us to reliably identify strains that are found in pairs of samples,

197 such as a mother and her infant or a pair of unrelated infants, defined as strain sharing. We
198 reconstructed a total of 77 strains (**Suppl. Table 3**), of which 75 were found in the infant fecal
199 samples and two in milk samples, due to the lower coverage. Out of the 43 strains present
200 across multiple samples, we identified two instances of strain-sharing between a mother's milk
201 and her infant's stools: the commensal species *B. longum* (in mother-infant pair 1), and the
202 pathobiont *K. pneumoniae* (in mother-infant pair 173) (**Fig. 3A-B**). All strains identified in infant
203 stool samples at one month were also identified at six months (**Suppl. Table 4**). Leveraging the
204 multi-center nature of this cohort, we then compared strain sharing between unrelated infants
205 born in the same hospital with that of infants born in different hospitals. We found that the
206 proportion of unrelated infant pairs sharing at least one strain was significantly higher in infants
207 born at the same hospital, compared to infants born in different hospitals at one month of age
208 ($p=9.6x10^{-8}$, Fisher's Exact Test; **Fig. 3C**). This significant trend persisted also at six months of
209 age (Fisher's exact test $p=9.7x10^{-7}$). When considering temporal overlap (same year of birth), in
210 addition to spatial overlap (same hospital), we found that the proportion of infants born at the
211 same hospital in the same year that shared at least one strain was higher than those born in the
212 same year across different hospitals at one month ($p=2.3x10^{-3}$, Fisher's exact test) but not at six
213 months (p=0.26, Fisher's exact test; **Fig. 3C**).

214

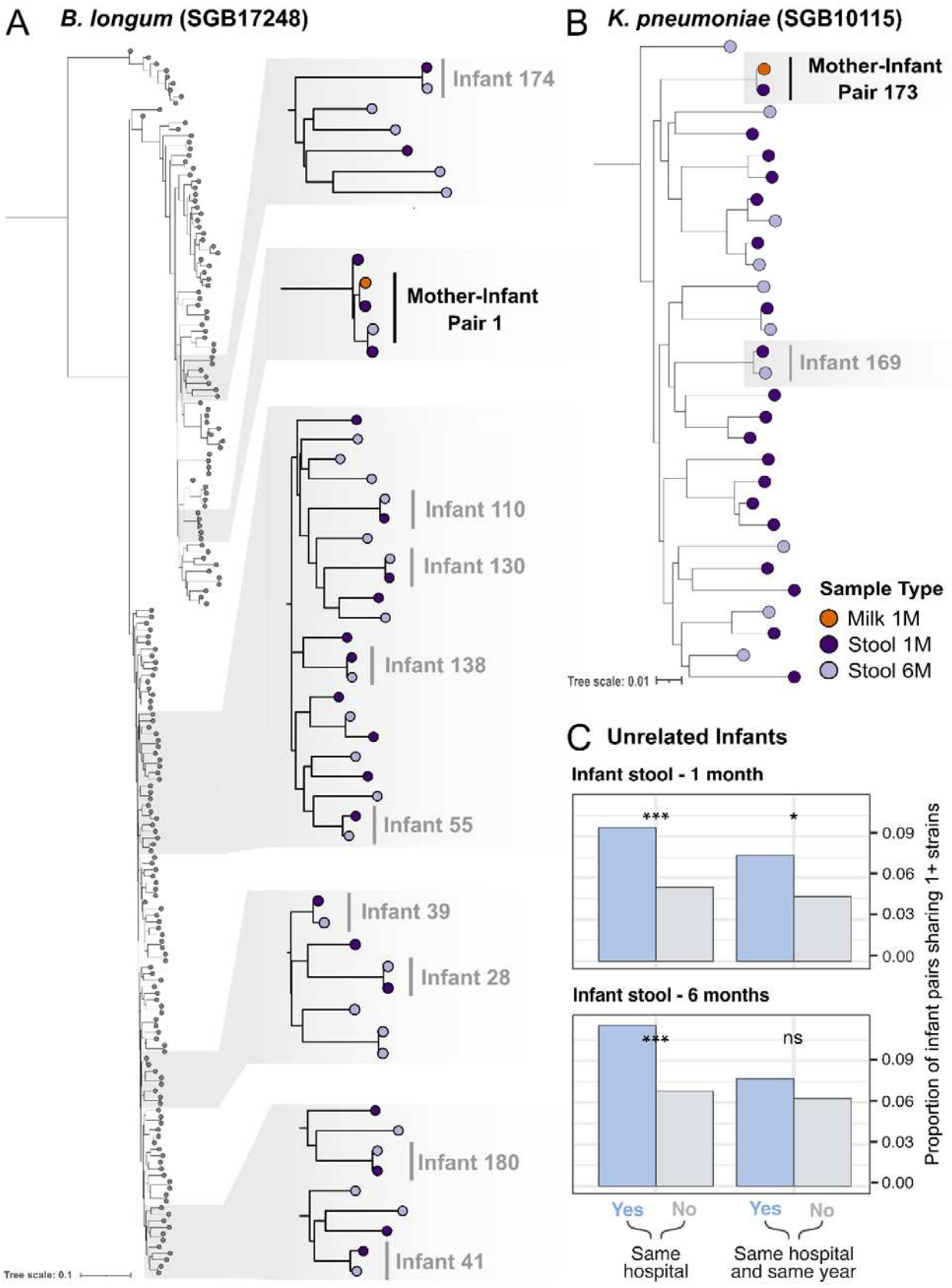


Figure 3. Strain sharing between maternal milk and infant stool samples for (A) the commensal species *B. longum* and (B) the pathobiont *K. pneumoniae*, highlighted in black. Instances of strain persistence within the same infant over time are highlighted with gray. (C) Proportion of unrelated infant pairs at one month (top) and six months (bottom) that share at least one strain (y-axis), considering infant pairs born in the same hospital (left), and same hospital as well as same year (right). Fisher's exact test p-values are reported.

The maternal milk and early infant gut microbiomes are enriched in metabolic pathways for the biosynthesis of essential amino acids

Next, we investigated the functional potential of the maternal milk and infant gut microbiomes using HUMAnN3[34], a method that profiles microbial metabolic pathways from metagenomic data. The most prevalent pathways identified in the infant stool samples were associated with de-novo biosynthesis of molecules, in particular of essential amino acids (such as valine, isoleucine, threonine, and lysine) and ribonucleotides (**Fig. 4A**). Other abundant pathways were associated with cell structure (i.e. peptidoglycan maturation), and energy metabolism. The pathways involved in peptidoglycan maturation were present at higher relative abundances in infant stool samples at one month compared to six months ($p=1.3x10^{-14}$, t-test; **Fig. 4A**), and this increase was particularly prominent for samples dominated by non-bifidobacteria species (**Extended Data 6A**). In milk, the most prevalent and abundant pathways were associated with the biosynthesis of nucleotides and amino acids (such as valine and isoleucine), cell metabolism (sucrose biosynthesis, glycogen degradation, pyruvate fermentation and folate transformations), and cell structure (peptidoglycan maturation) (**Fig. 4B**). In infant stool samples, the pathways associated with essential amino acid biosynthesis were significantly more abundant at one month compared to six months, and this trend was significantly more pronounced for infants that were not exclusively breastfed anymore at six months (Bonferroni-adjusted $p=5.2x10^{-3}$ for infants exclusively breastfed at one and six months, and $p=4.9x10^{-5}$, for infants exclusively breastfed at one month only, t-test; **Fig. 4C**). In particular, milk samples at one month dominated by *B. longum*, but not other Bifidobacteria, were associated with the highest abundance of metabolic pathways associated with the biosynthesis of essential amino acids **(Extended Data 6B**).

We then sought to investigate the relationship between the abundance of the metabolic pathways identified from microbes within each mother's milk with those identified in her infant's stool samples. We found a significant correlation between the metabolic pathways found in breast milk with those found in infant stool samples only for the mother-infant pairs for which we previously identified strain sharing events (BH-adjusted $p=4.3x10^{-23}$ and $p=3.4x10^{-21}$, for mother-

infant pair 1 and 173, respectively, Spearman's correlation; **Fig. 4D and Extended data 7**). When investigating specific pathways, rather than specific mother-infant pairs, we find that no pathway showed a significant correlation in its abundance in milk compared to infant stool samples (**Extended Data 8**).

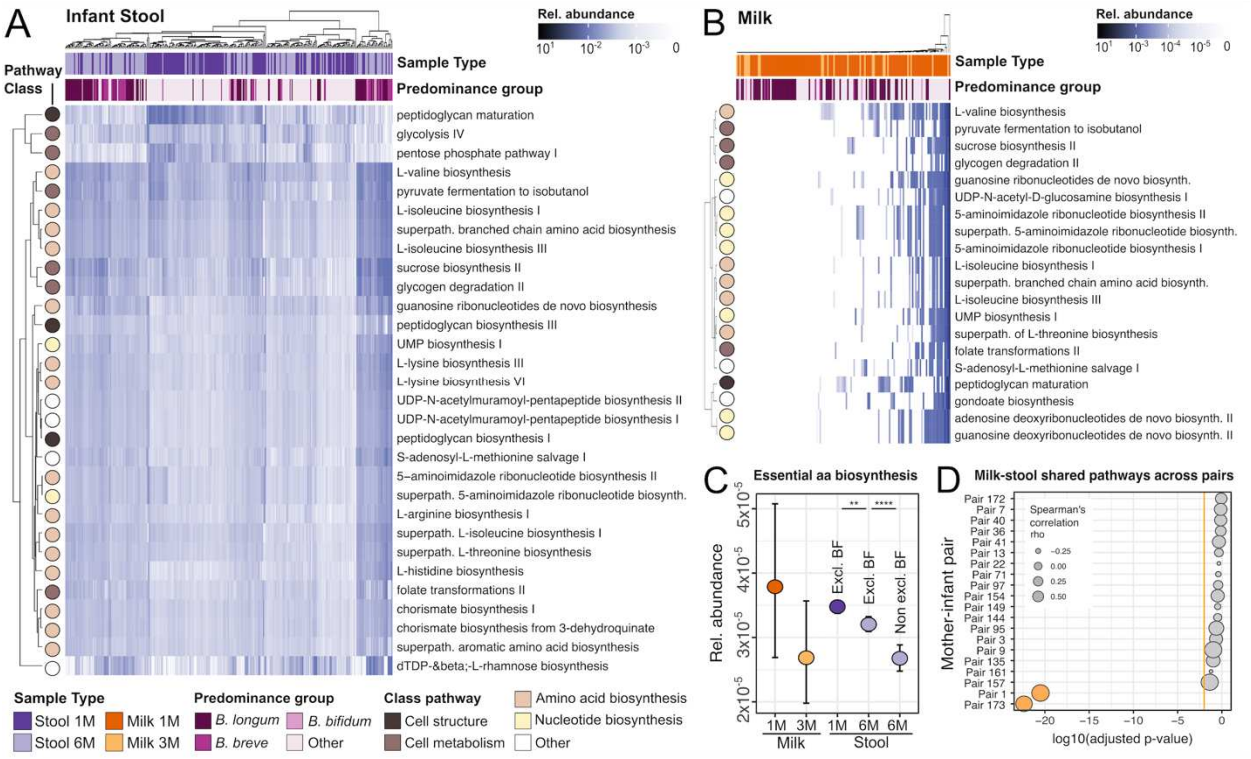


Figure 4. Most abundant pathways identified in (A) breastmilk at one and three months postpartum and (B) infant gut at one and six months of life. (C) Relative abundance of pathways involved in the biosynthesis of essential amino acids across sample type, collection time point and breastfeeding (BF). Error bars represent a 95% confidence interval calculated by bootstrapping (1000 times). P-values calculated using t-test, **** for P ≤ 0.0001 and ** p<0.01. Reported p-values are adjusted using Bonferroni method. (D) P-values of Spearman correlation between the abundance of all metabolic pathways shared between the maternal breast milk and infant stool samples for each mother-infant pair. P-values are corrected for multiple testing using Benjamin Hochberg correction and are shown in log10 scale. Only the top 20 mother-infant pairs are shown. Circle size denotes the correlation coefficient (rho), and the orange line denotes the significance threshold (log10 of p-value=0.01).

The infant gut and the maternal milk microbiomes harbor a diverse landscape of antimicrobial resistance genes

The gut is a reservoir of antimicrobial resistance, yet its composition and longitudinal variability remain poorly characterized in infants[35,36]. Even less is known about the antimicrobial resistance genes (ARGs), collectively known as the resistome[37], present in human breast milk, and their transmission to the infant during lactation[21,38–40]. To investigate the resistome in the maternal

milk and the infant gut, we used DeepARG[41], which leverages deep learning to predict antimicrobial resistance genes from metagenomic data. We then compared ARG classes across sample types and collection timepoints. Maternal milk and infant stool differed in terms of ARGs classes composition and prevalence (**Fig. 5A**, **Extended Data 9A**). In the milk, the most prevalent (25% on average) antimicrobial resistance class identified was against macrolide-lincosamide-streptogramin (MLS). Overall, maternal milk was characterized by a lower diversity of detected ARG classes compared to infant stool samples ($p=1.4x10^{-14}$, t-test; **Extended Data 9B-C**). In both milk and infant stool samples, the diversity of ARGs classes increased over time, although this increase was statistically significant only in the milk ($p=2.4x10^{-3}$ and p=0.61 for milk and stool samples respectively, paired t-test; **Extended Data 9C**).

The infant gut resistome was mostly dominated by resistance to tetracycline, MLS, aminoglycoside, and beta-lactams (**Fig. 5A**). Infants whose stool samples were dominated by bifidobacteria were characterized by a significantly lower carriage rate of ARGs ($p=7.6x10^{-12}$ and $4.2x10^{-2}$ at one and six months respectively, t-test; **Fig. 5B-C**). We found no significant difference in the resistome of infants that were born via C-section compared to those born via vaginal delivery (p=0.34 and p=1 at 1 month and 6 months, respectively, t-test), between those exposed to antibiotics and those that were antibiotics-naive (p=0.3 and p=0.18 at one month and six months, respectively, t-test), nor between those exclusively breastfed at six months of age and those fed with a mixture of breast milk and formula (p=0.13, t-test; **Extended Data 9D-F**). Nevertheless, we found extensive ARGs carriage in infants with no recorded pre-, during-, and postpartum exposure to antibiotics (**Fig. 5B**).

When comparing the overall resistome across all mothers and infants, we found no significant correlation between the ARGs found in milk and the ARGs found in the infant stool samples (R=-0.041, p=0.4, Spearman's test; **Fig. 5D**). However, the infant gut resistome at one month was positively correlated with the resistome at six months of age (R=0.42, $p=8.7x10^{-48}$, Spearman's test; **Fig. 5E**).

Considering sharing of ARGs between milk and stool samples within each mother-infant pair, we found that mother-infant pairs shared a significantly higher number of ARGs than what was expected by chance (p<0.001, by permutation analysis, see Methods; **Fig. 5F-G**). Mother-infant pairs 1 and 173, for which strain sharing events were identified, were the pairs with the highest rate of shared ARGs between the mother's milk and her infant stool samples (**Fig. 5F**). On average, the most commonly shared antimicrobial resistance classes were associated with

resistance to peptides, fluoroquinolone, and MLS, while the most commonly shared antimicrobial resistance genes were MACB (MLS class), ACRD (aminoglycoside class), and TETM (tetracycline class) (**Fig. 5H**).

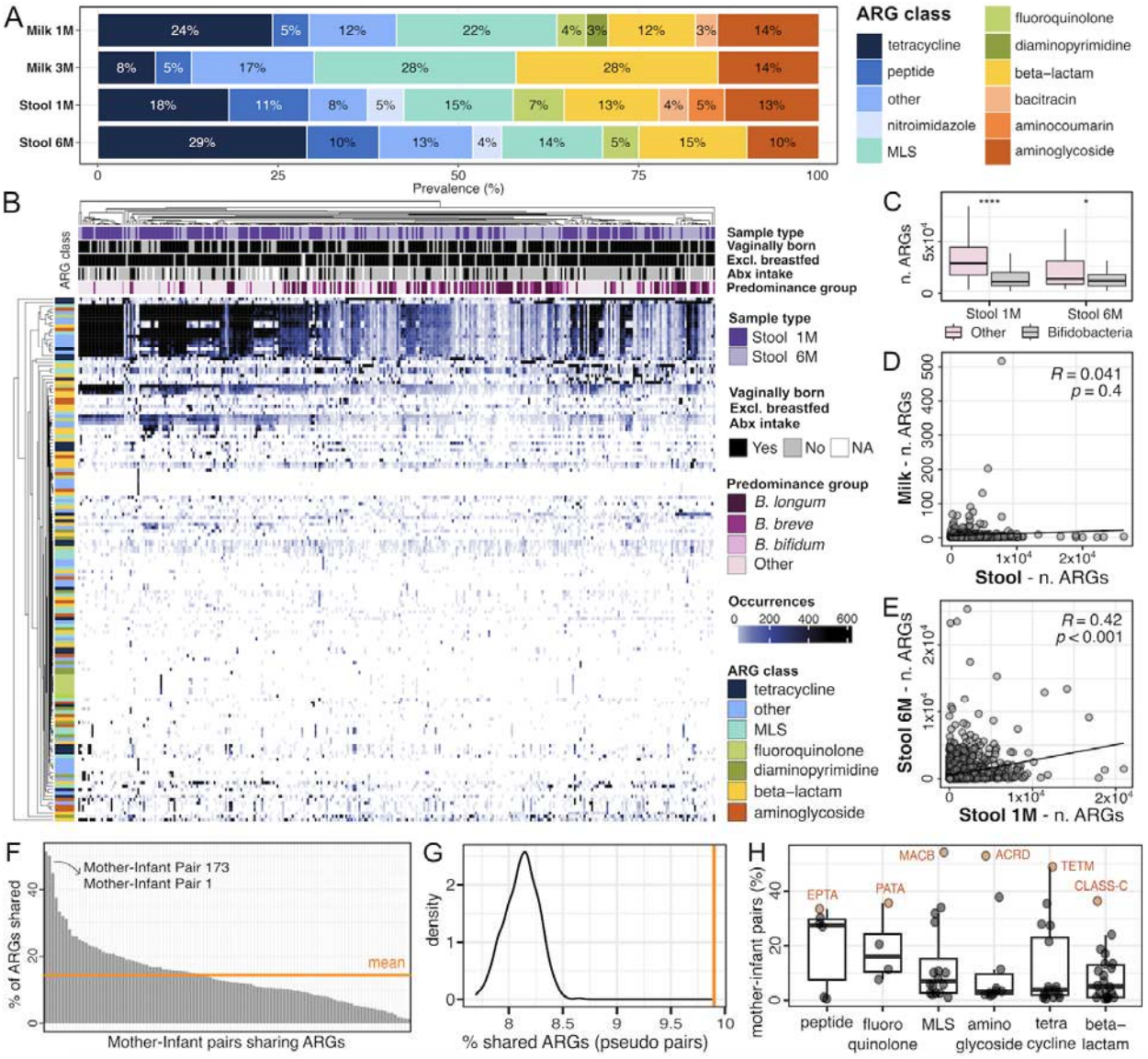


Figure 5. (A) Prevalence of the antimicrobial resistance genes (ARGs) classes predicted in maternal milk and the infant gut microbiome over time, calculated as percentage over the total number of detected ARGs for each sample type. (B) ARGs carriage in infant stool samples, divided by collection time point, infant metadata, predominance group and ARGs class. (C) Number of detected ARGs in infant stool samples dominated by bifidobacteria and non bifidobacteria species at one and six months of life. P-values calculated using t-test, **** for P ≤ 0.0001 and * for P ≤ 0.05. (D) Correlation between ARG carriage in maternal milk and infant stool samples, and (E) between infant stool samples at one month and six months. Each dot is a combination of a mother-infant pair and a predicted ARG class. Spearman's R-values (correlation coefficient) and p-values are reported. (F) Percentage of ARG genes shared between at least one maternal and one infant sample, for each mother-infant pair. Mean value is indicated by the orange line. The mother-infant pairs for which strain sharing events were identified are highlighted with the arrow. (G) Distribution of the percentage of shared ARGs between at least one maternal milk and one infant stool sample on

permuted mother-infant pairs (pseudo-pairs). The mean value obtained from real mother-infant pairs is indicated by the orange line. (H) Percentage of mother-infant pairs sharing antimicrobial resistance genes, divided by ARG class. For each ARG class, percentage value was calculated on the total number of mother-infant pairs in which that ARG class was identified in at least one sample. The most frequently shared antimicrobial resistance genes per class are highlighted in orange.

Discussion

Although breast milk represents a critical source of nutrition for the developing infant, little is known about the milk microbiome, how it changes over time, its metabolic potential, and how it shapes the infant's gut microbiome – all questions necessitating the use of high resolution sequencing techniques, such as metagenomics. Here, we investigated the composition, strain sharing, functional potential, and antimicrobial resistance of the maternal breast milk and the infant gut microbiome in early life in predominantly exclusively breastfeeding mother-infant dyads. In line with previous studies, the milk microbiome yielded a lower number of reads and was characterized by a reduced microbial diversity compared to the infant gut microbiome[11,21,27]. The milk at one month postpartum, considered largely mature[18], was dominated by bifidobacteria, in particular *B. longum*, *B. breve,* and *B. bifidum*. The presence in the milk of typical oral species is likely due to the possible transfer of microbes from the oral cavity of the infant to the breast milk during suckling, a process called retrograde flow[42]. Bifidobacteria also dominated the gut microbiome of the infants, driving overlap in taxonomic composition between the maternal milk and infant gut microbiomes. In particular, infant stool samples largely clustered by which Bifidobacterium species was most abundant. We broadly identified 4 predominance groups, dominated by *B. longum*, *B. breve* and *B. bifidum* respectively, and a fourth group which included the samples dominated by other non bifidobacteria species, most commonly *E. coli*. While *B. longum*, *B. bifidum,* and *B. breve* were often coexisting and present at comparable abundances in milk, these species were largely mutually exclusive in the infant gut. These results suggest a higher level of competition between bifidobacteria species in the infant gut compared to the maternal milk. Bifidobacteria, most commonly found in the human gut[43] and maternal milk[44], are known key foundation taxa of the infant gut microbiome and their abundance is primarily shaped by breast milk intake[45–47]. Indeed, the stools from exclusively breastfed infants had higher prevalence and abundance of bifidobacteria compared to those that ceased exclusive breastfeeding prior to six months of age. Considering change over time, bifidobacteria increased in relative abundance from one to six months of age, in line with what was shown by Heisel et al.[48]. This pattern was more pronounced in infants exclusively breastfed through six months.

We found two cases of strain sharing between the mother's milk and her infant's gut microbiome, specifically for the commensal species *B. longum,* and for the pathobiont *K. pnuemoniae*. Despite *K. pneumoniae* potential association with silent sepsis in infants[49,50] and with mastitis and milk loss in cows[51], the infants in our cohort showed no clinical manifestations nor had any of the mothers reported mastitis/breast inflammation at the time of their study visit and milk collection. While strain sharing between the maternal and the infant gut has been extensively investigated[8,11], only one metagenomic study has so far, to the best of our knowledge, successfully identified strains shared between the maternal breast milk and the infant gut microbiome[52]. Our results provide evidence that strain sharing between the infant gut and the maternal breast milk occurs, even if at a rate considerably lower than what was previously found between the infant and the maternal gut[11]. The fact that we were able to reconstruct only two strains in our milk samples, and yet both were found to be shared between mother's milk and infant gut, indicates that our estimate for strain sharing between breast milk and infant gut is likely an underestimate due to the low sequencing yield of the milk samples. This is a common challenge with this type of sample[27].

We found an enrichment of strain sharing events among unrelated infants born in the same hospital compared to those born in different hospitals, a trend that still persisted at six months of age. Strain sharing between unrelated infants was previously reported in hospitalized premature infants, but this represents, as far as we know, the first evidence of microbial strain sharing between healthy, at-term infants, months after hospital discharge[53,54]. This result suggests that the short-term postpartum hospital stay might play a role in the infant's gut strain acquisition and persistence over time. An alternative hypothesis could be that infants born in the same hospital live in the same geographic area and can acquire microbiomes through other shared environments, such as daycare facilities. However, we could not test this hypothesis as these high-resolution geographical and environmental metadata were not collected.

Resistome analysis showed that both the infant gut and the maternal milk harbor a diverse landscape of antimicrobial resistance gene classes, mostly dominated by resistance genes for tetracycline, aminoglycoside, and macrolide-lincosamide-streptogramin, even in infants that had no recorded exposure to pre-, intra-, and post-partum antibiotics. Tetracycline is not prescribed in pregnancy and its use is not recommended in children younger than 8 years due to potential permanent discoloration of teeth[55]. Still, tetracycline resistance was the most abundant antibiotic resistance class in both milk and infant stool samples at one month. Aminoglycoside (e.g. gentamicin and streptomycin) and macrolides (e.g. azithromycin and erythromycin) are widely

used in the perinatal period[40], therefore resistance is potentially acquired via antibiotic exposure. Milk and infant stool samples at one month were characterized by higher prevalence and diversity in resistance classes, compared to later time points. In addition, the majority (70%) of ARGs found in the infant gut at one month were associated with predominance of non-bifidobacteria species, in particular *E. coli*. The presence of antibiotic resistance in the newborn gut has been reported in previous studies[35,36,56], with indications of resistance transmission from the maternal gut to the infant gut via mobile genetic elements[57]. Our results show a significant overlap between the resistome of infants and that of their mother's milk, and that this overlap was particularly pronounced when evidence of strain sharing was found. This suggests that the infant gut resistome is likely influenced to some degree by acquisition from breastmilk, in addition to other mechanisms, such as mobile genetic elements, vertical strain transmission and environmental exposure, and from other maternal body sites, such as gut[57].

This study has several limitations. Milk sampling was performed via the use of breast pumps, which could potentially impact milk microbiome composition[58]. In addition, the low number of three-month milk samples as compared to one-month milk samples limits our conclusions on milk microbiome composition at three months and strain-sharing with the infant beyond one month postpartum. The low sequencing yield for breast milk samples limited our power for detection of strain sharing between milk and infant gut, suggesting that the sharing patterns are an underestimate. In addition, as we sampled only milk and infant stool samples, we could not confirm that the strain sharing events we identified were indeed cases of microbial transmission from the maternal milk to the infant gut, rather than cases of strain acquisition by both the mother and the infant from external sources not investigated in this study. Finally, infants were born across multiple hospitals, which usually represents a limitation, as it could potentially influence the microbial composition. However, all samples were collected following the same procedures and the samples' taxonomic composition did not cluster based on the sampling location. Furthermore, the multi-center structure of this study enabled us to quantify strain sharing among unrelated individuals across different hospitals.

In this work, we characterized the microbiome composition, function and antimicrobial resistance potential of the breast milk of mothers and the gut microbiome of their infants during the first six months postpartum. We found evidence of strain- and antimicrobial resistance gene-sharing between mother-infant pairs. Taken together our results indicate that the maternal breast milk plays a role in infant gut microbiome and resistome establishment, development and temporal stability. Our results represent an important step towards the strain-level

423 characterization of the maternal milk microbiome in relation to the infant's gut and its better
424 representation in public repositories.

425 Methods

426 Sample collection

427 The participants from this study were enrolled as part of the Mothers and Infants Linked for
428 health (MILk) cohort[59–61]. Recruitment, clinical metadata and sample collection were performed
429 as previously described in [59–61]. All mothers were enrolled prenatally from the University of
430 Minnesota in collaboration with HealthPartners Institute (Minneapolis, MN) and were provided
431 written informed consent. All relevant guidelines and regulations were observed. Inclusion
432 criteria included a healthy, uncomplicated pregnancy and the intention to exclusively breastfeed
433 the infant. All mothers were aged between 21-45 years, were not diabetic and non-smokers,
434 and delivered a full-term infant. No case of breast infection or mastitis was reported during the
435 milk sample collection. All infants were singletons and born at term with a birth weight that was
436 appropriate for their gestational age, and were exclusively breastfed to at least one month of
437 age (**Supplementary Table 1** and **Extended Data 1**). All relevant metadata were collected via
438 the hospital's electronic medical health records and via questionnaires during sample collection
439 (**Supplementary Table 1** and **Extended Data 1**). Breast milk samples were collected at one
440 and three months postpartum. Milk collection was performed as follows: mothers fed their infant
441 from one or both of their breasts, until the infant was satisfied. After two hours, the milk was
442 collected from the right breast using a hospital grade electric breast pump (Medela Symphony;
443 Medela, Inc., Zug, Switzerland), until cessation of production. Each milk sample was gently
444 mixed, and its volume and weight were recorded. Aliquots were stored at -80 °C within 20
445 minutes of collection and kept at that temperature until RNA/DNA extraction.
446 Infant stool samples were collected at one and six months of age. Stool samples collection,
447 storage, and associated metagenomic shotgun DNA extraction were performed as described in
448 previous works[59–62]. Stool samples were either collected from diapers during a study visit or at
449 home by the mother. In case of collection during a study visit, the sample was immediately
450 frozen at -80°C, while in case of home collection the sample was stored in 2 ml cryovials with
451 600 µl RNALater (Ambion/Invitrogen, Carlsbad, CA), and later stored at -80°C upon arrival to
452 the lab at the University of Minnesota.

453 DNA extraction and metagenomic sequencing

454 DNA extraction was performed with PowerSoil kit (QIAGEN, Germantown, MD), eluted with 100
455 µl of the provided elution solution, and stored in microfuge tubes at -80°C. The extracted DNA
456 was used to construct libraries for metagenomic shotgun sequencing using the Illumina Nextera
457 XT 1⁄4 kit (Illumina, San Diego, CA, United States). Metagenomic shotgun sequencing libraries
458 were sequenced on an Illumina NovaSeq system (Illumina, San Diego, CA) using the S4 flow
459 cell with the 2x150 bp paired end V4 chemistry kit by the University of Minnesota Genomics
460 Center.

Quality filtering and removal of human reads

Host DNA was removed using paired-end mapping with Bowtie2[63] version 2.2.4 against a human reference genome hg38. Unmapped paired-end reads were filtered using SAMtools[64] version 1.9 with following parameters “samtools view -bS, samtools view -b -f 12 -F 256, samtools sort -n -m 5G -@ 2, samtools fastq -@ 8 -0 /dev/null -s /dev/null -n”. BEDtools[65] was then used to convert the bam files to fastq files containing the non-human paired-end reads. Then, adapter sequences were removed and the samples were filtered and trimmed using the default parameters of Trimmomatic[66]. Host DNA content in breast milk samples was 74.13 ± 16.37 %. FastQC[67] was used to analyze the quality of the metagenomic reads. Read lengths were 1,224,953 and 1,224,953 bp for the forward and reverse reads, respectively. Mean PHRED of the score was 33. Samples with less that had 500 or less reads that mapped to the Metaphlan4[31] database (mpa_vJan21_CHOCOPhlAnSGB_202103) were discarded. A total of 34 of the 507 samples were excluded from downstream analysis. After pre-processing, milk samples yielded 2.45±1.36 million reads per sample, while stool samples yielded 10.8±3.96 million reads per sample (**Extended Data 10**).

Species- and strain- level taxonomic profiling and strain sharing

Species-level taxonomic profiling was based on marker genes using MetaPhlAn4[31], with the following parameters: “--bt2_ps sensitive” text. MetaPhlAn4 infers taxonomic prevalence and abundance by using unique marker genes for 26,970 species-level genome bins[31]. Merged abundance table was created using the MetaPhlAn4 utils script. Contaminants found in the blank samples as well as additional known contaminant species were removed before downstream analysis. MetaPhlAn4 profiles are available in **Supplementary Table 2**. Heatmaps for species-level taxonomic composition were generated with ComplexHeatmaps[68,69], using euclidean distance for hierarchical clustering. Strain-level profiling was performed using StrainPhlAn4[31], based on single-nucleotide variant calling, with the following parameters: “--mutation_rates --trim_sequences 50 --marker_in_n_samples 50 --secondary_sample_with_n_markers 50 --breadth_thres 50”. Strain_transmission.py output (**Supplementary Table 3**) was used to identify strain sharing events among related and unrelated individuals. Phylogenetic trees were visualized with iTOL[70].

Alpha and beta diversity

Alpha diversity was computed using the MetaPhlAn4 utility script calculate_diversity.R. The ordination plot was computed from the MetaPhlAn4 relative abundances using the “vegan” R package (v2.6-4).

Functional profiling

Functional prediction was performed with HUMAnN3[34]. Biosynthetic potential of essential amino acids were calculated by searching the output pathways for the following keywords: "L-histidine biosynthesis", "L-isoleucine biosynthesis", "L-isoleucine biosynthesis", "L-lysine biosynthesis", "L-methionine biosynthesis", "L-phenylalanine biosynthesis", "L-threonine biosynthesis", and "L-valine biosynthesis". Profiles relative abundance plots show 95% confidence interval using

bootstrapping with 1000 repetitions using the Hmisc R library (v5.0-1). Raw functional profiles are available in **Supplementary Table 5**.

Antimicrobial resistance genes prediction

Antimicrobial resistance genes were predicted from raw metagenomic short reads using DeepARG v1.0.2[41]. DeepARG leverages deep learning models to identify over 30 antimicrobial resistance classes. Predicted resistance is classified as “predicted” or “potential” by the models. To reduce false positives detection, only “predicted” resistance, with a minimum identity threshold of 95% of the target sequence was included in the downstream analysis. ARG predictions classified as “multidrug” or “unclassified” were excluded from downstream analyses. Raw DeepARG profiles are available in **Supplementary Table 6**.
Permutation analysis in Figure 5G was performed as follows: maternal milk sample names were randomly permuted, while data and the infant sample names were preserved, generating pseudo-mother-infant pairs, and the mean number of antimicrobial resistance genes shared between the pseudo-mother-infant pairs was calculated. This was repeated 1000 times, and the distribution of the mean values obtained from pseudo-mother-infant pairs was compared to the mean value obtained from the real mother-infant pairs.

Statistical analysis

Statistical analysis was done in R[71] version 4.2.2 (2022-10-31). All figures if not indicated otherwise were drawn with ggplot[72] version 3.4.1. All analyses have been performed with open source software referenced in the Methods section. In Fig. 3C, Fisher’s exact test was calculated on the occurrences tables defined as follows: number of unrelated infant pairs sharing at least one strain (Yes/No) versus infant pairs born in the same hospital (Yes/No). Fisher’s exact test and occurrence tables were calculated in a similar fashion when considering the same hospital and the same window of birth (same month), and separately for infant pairs at one and six months of age.

Data availability

The raw metagenomic sequences and the associated metadata were deposited and are available on NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1019702. Comprehensive metadata are available in the Supplementary Material. Code is available on git https://github.com/blekhmanlab/milk_infant_microbiome.

Acknowledgements

We thank Katy Duncan, Laurie Foster, Tipper Gallagher, Allyson Kats, and all staff and participants of the MILK study (https://www.milkresearch.umn.edu), and members of the Albert and Blekhman labs for helpful discussions related to this project. This work was supported by

the resources and staff at the Masonic Institute for the Developing Brain at the University of Minnesota (https://www.midb.umn.edu), the University of Oklahoma Health Sciences Center Harold Hamm Diabetes Center, https://www.ouhealth.com/find-a-location/ou-health-harold-hamm-diabetes-center) and the University of Minnesota Genomics Center (https://genomics.umn.edu). We are thankful for the computational resources provided by the Minnesota Supercomputing Institute (https://www.msi.umn.edu/) and by the University of Chicago Research Computing Center (https://rcc.uchicago.edu/).

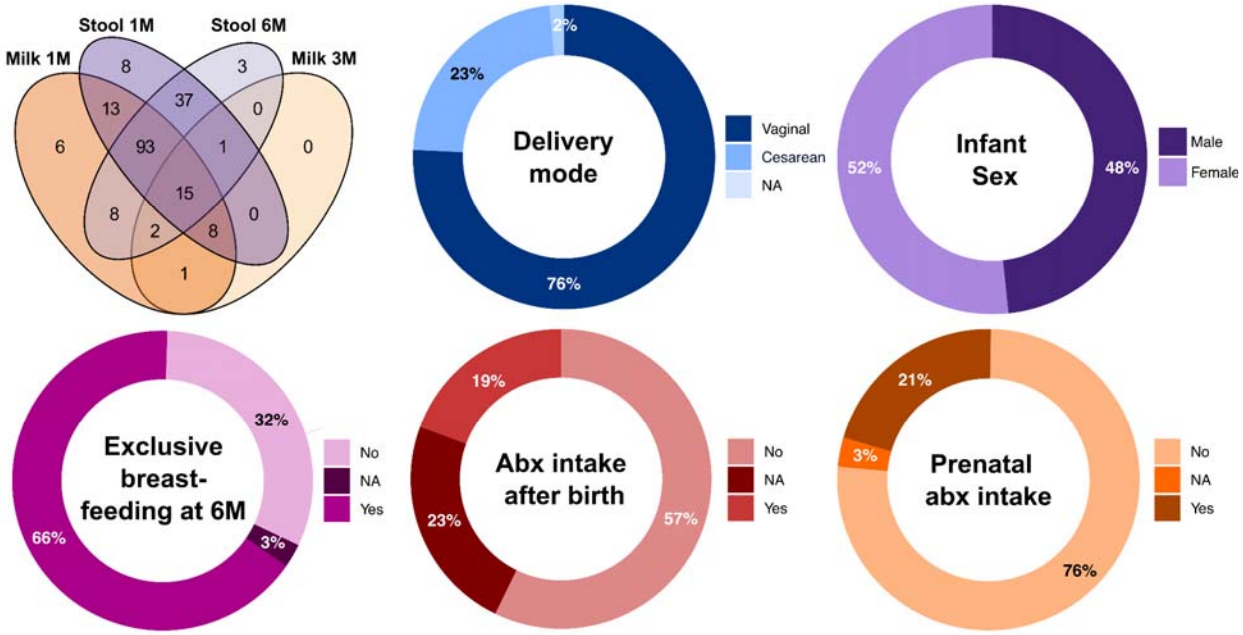
Author contributions

Conceptualization, R.B., C.A.G., E.W.D., and D.K.; Sample and metadata collection, E.W.D, S.G., and C.A.G; Samples processing and data analysis, P.F. and M.A.; Writing – original draft, P.F.; Writing – review & editing, P.F., M.A., K.J., T.H., D.A.F., F.W.A., C.A.G., E.W.D., and R.B.; Supervision, P.F. and R.B.; Funding acquisition, E.W.D., D.A.F., and R.B. All authors read and approved the final manuscript.

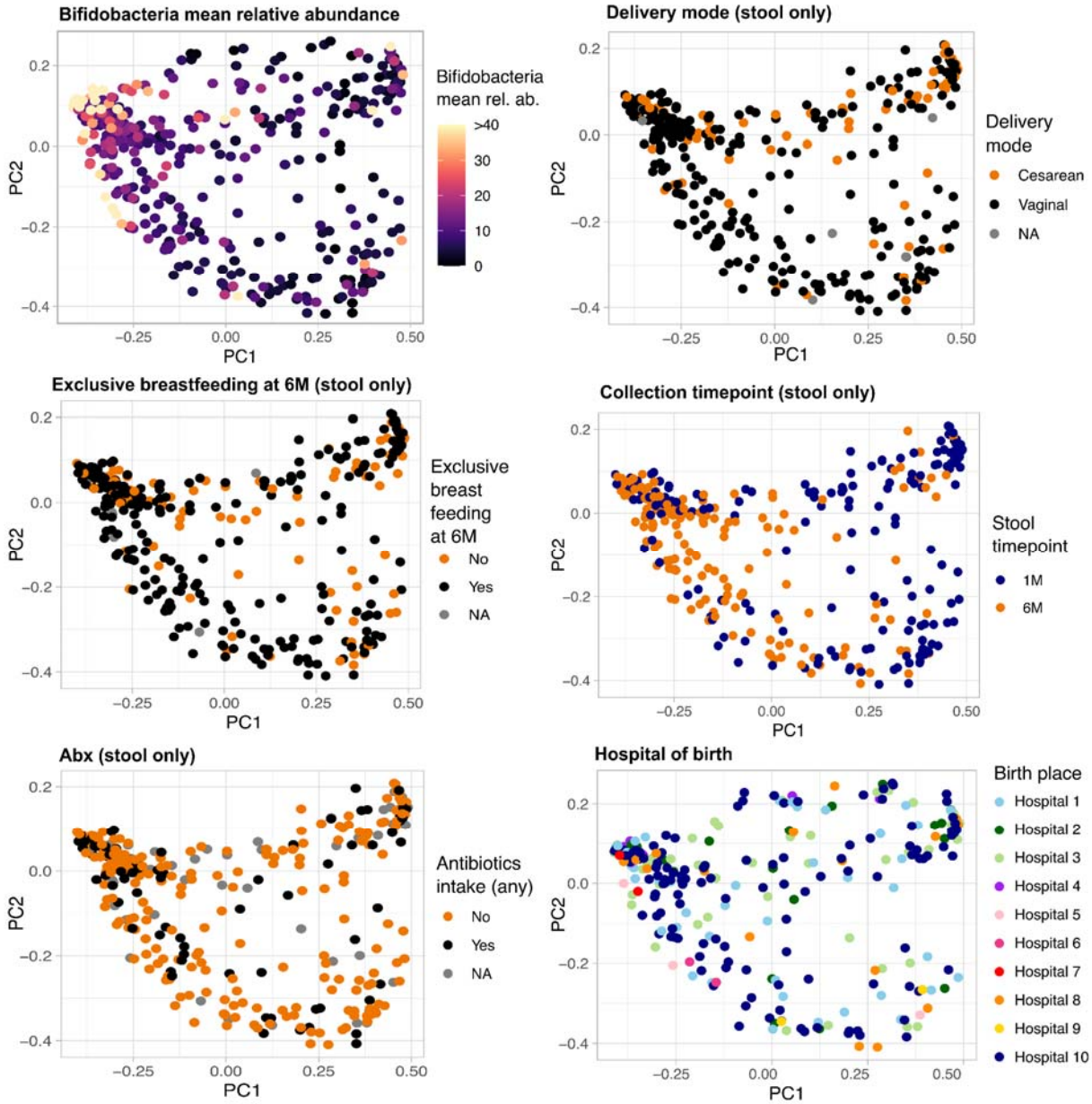
Funding

This study was supported by the University of Minnesota Department of Pediatrics Masonic Cross Departmental Research Grant (FWA, RB, EWD, CAG), University of Minnesota Masonic Children's Hospital Research Fund Award (CAG, EWD, and DK), NIH/NICHD grant R01HD109830 (RB, FWA, EWD, CAG, KEJ), NIH/NICHD grant R21HD099473 (CAG, EWD), NIH/NICHD grant F32HD105364 (KEJ), and a University of Minnesota Office of Academic and Clinical Affairs Faculty Research Development Grant (CAG, EWD, KEJ, and DK). The MILK Study which provided the cohort and milk samples for this study was supported by NIH/NICHD grant R01HD080444 (EWD and DAF).

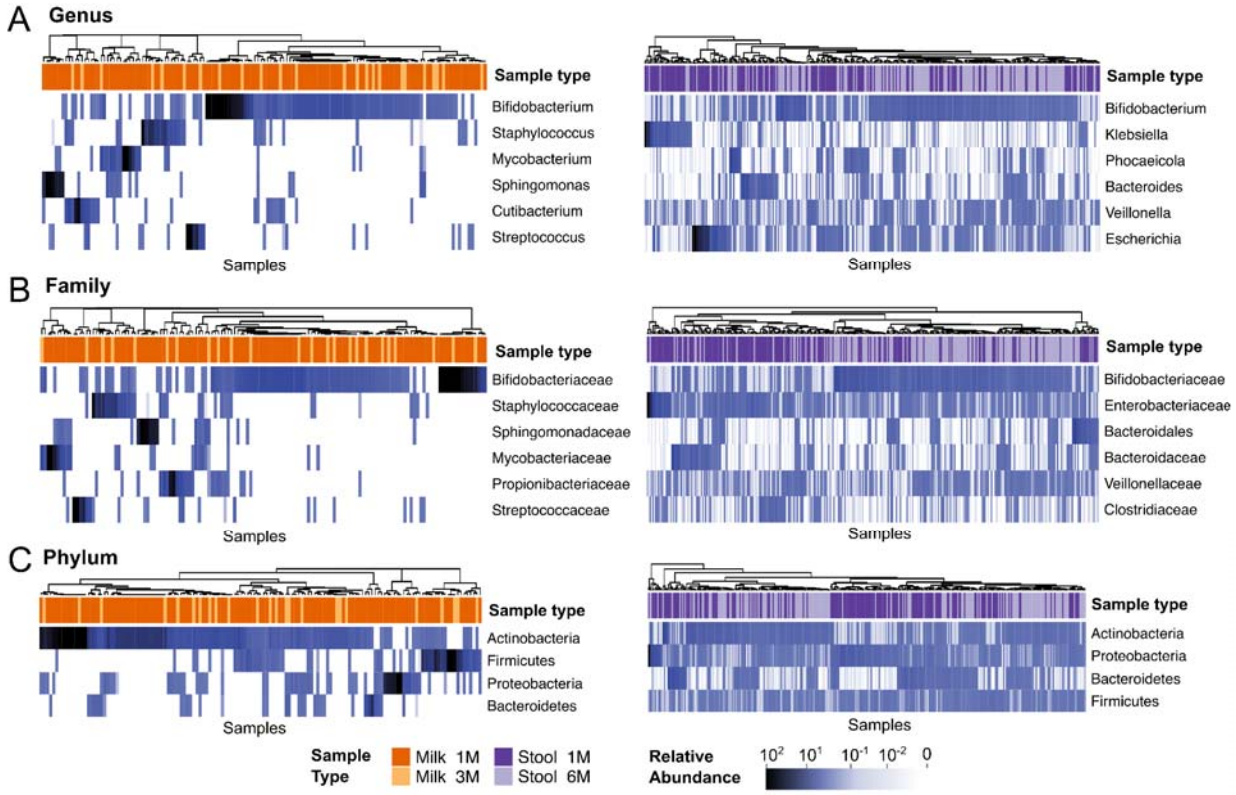
Supplementary Material



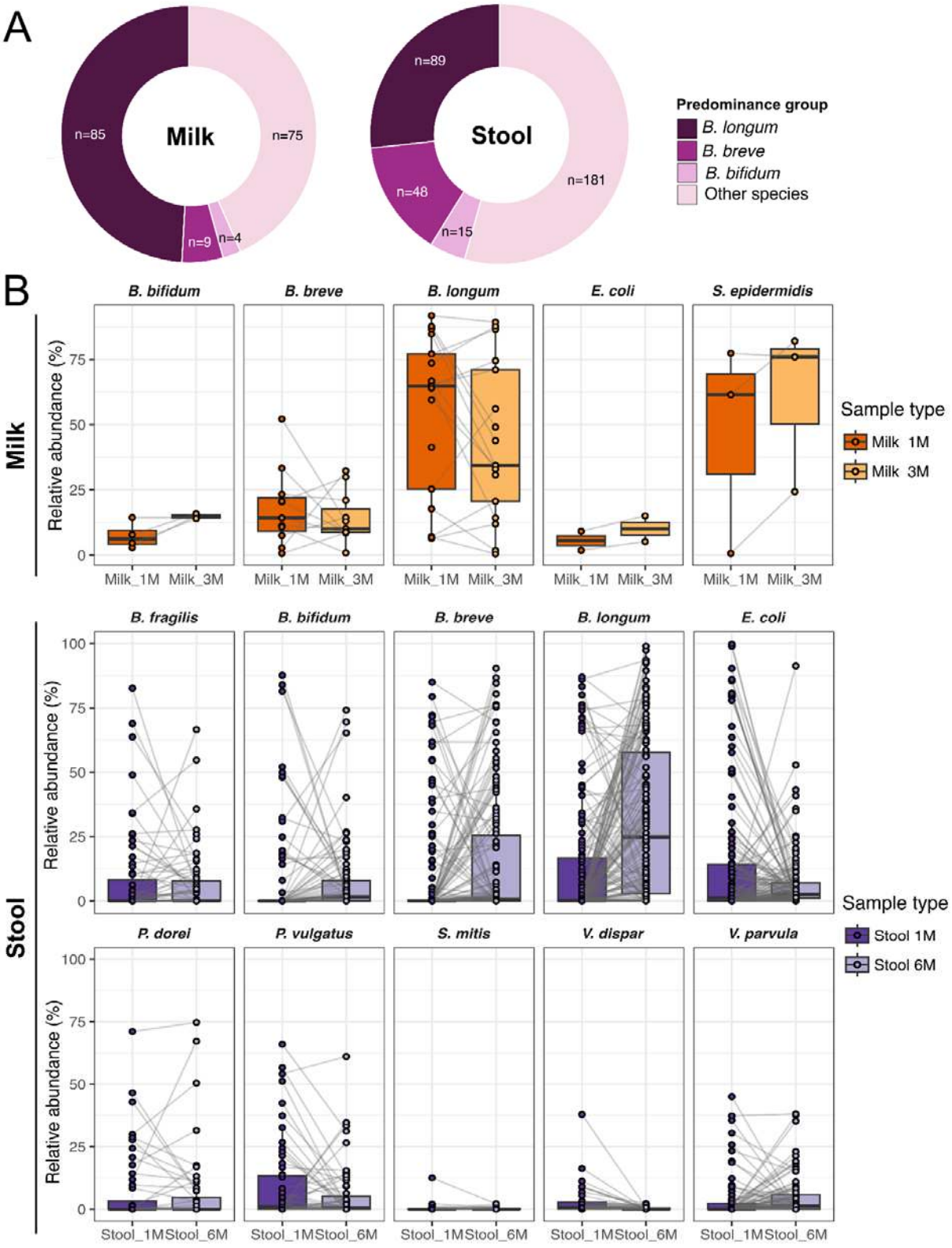
Extended Data 1. Number of mother infant pairs for which samples were collected across body sites and collection timepoints, singularly taken or in combinations. Relevant metadata available for the MILk cohort, including delivery mode, infant sex, exclusive breastfeeding at 6 months of age, and antibiotic (abx) intake pre- and after-partum. All infants were exclusively breastfed at 1 month of age.



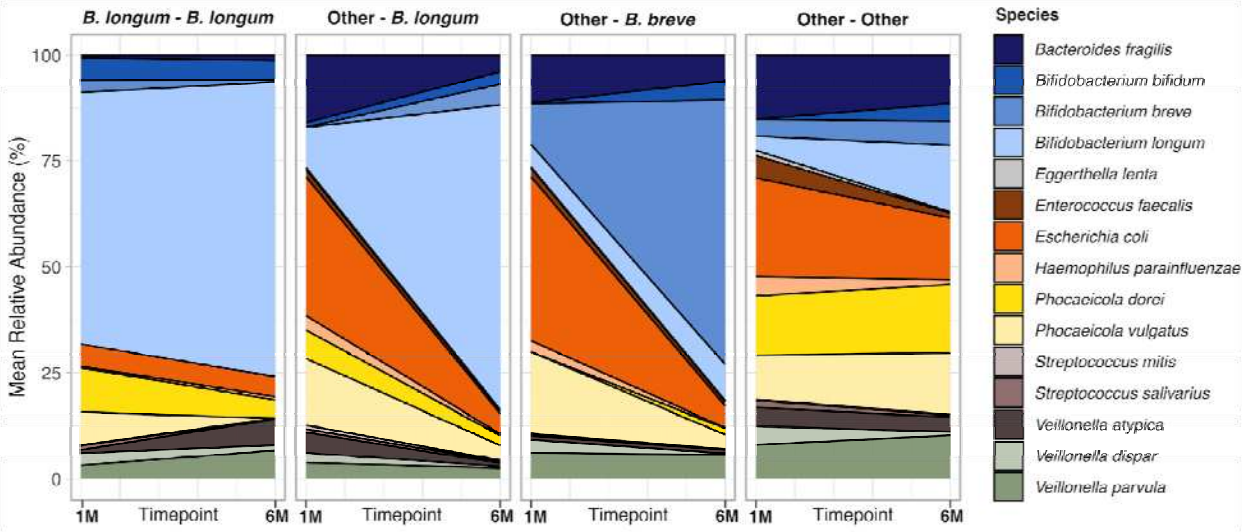
Extended Data 2. Ordination plot coloured by relevant metadata. All body sites and collection time points are included unless otherwise specified. PCoA of hospitals includes only the samples for which the birth hospital or clinic name is known.



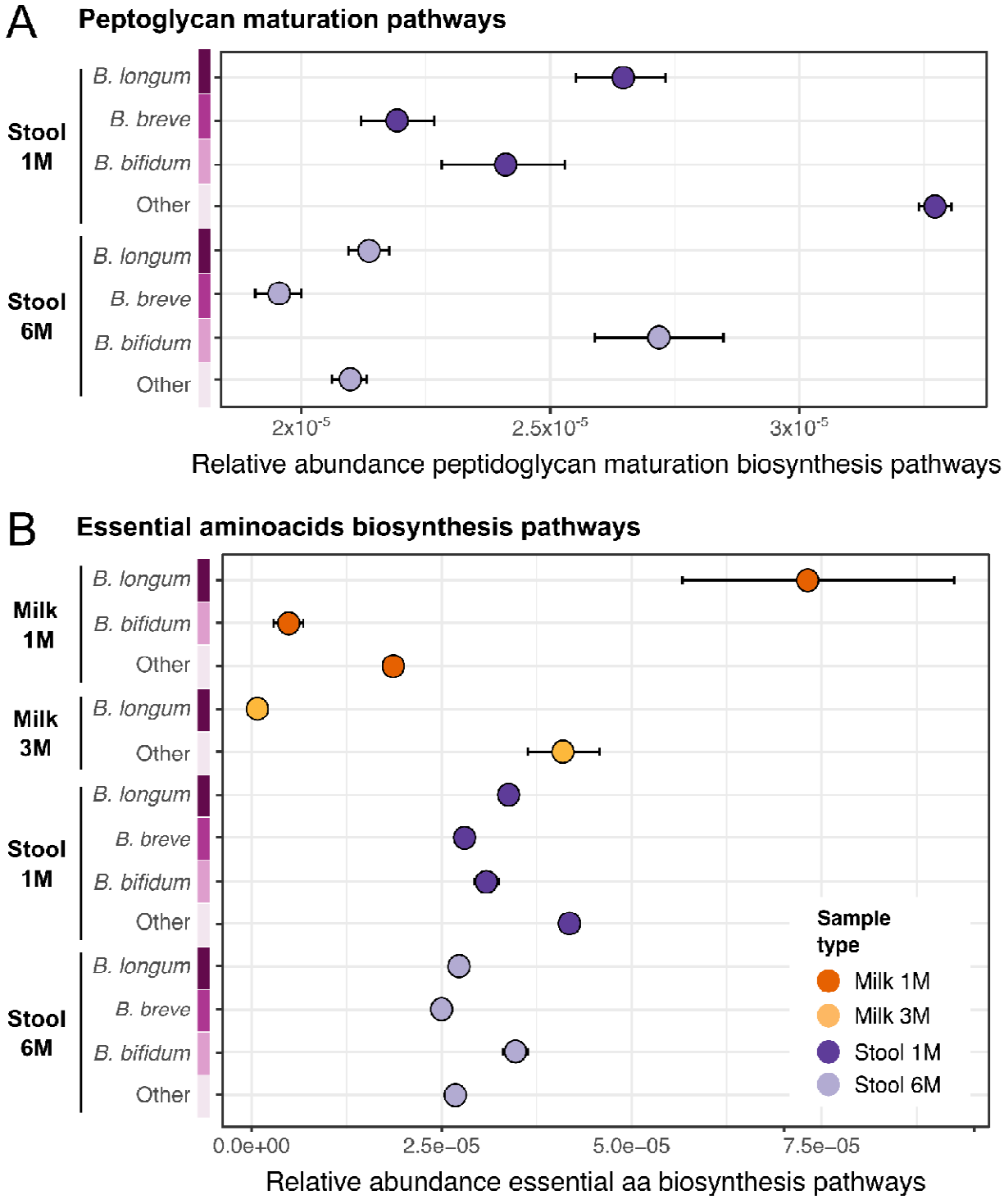
Extended Data 3. Top six genera (A) and families (B) and top four phyla (C) in milk (left) and stool samples (right).



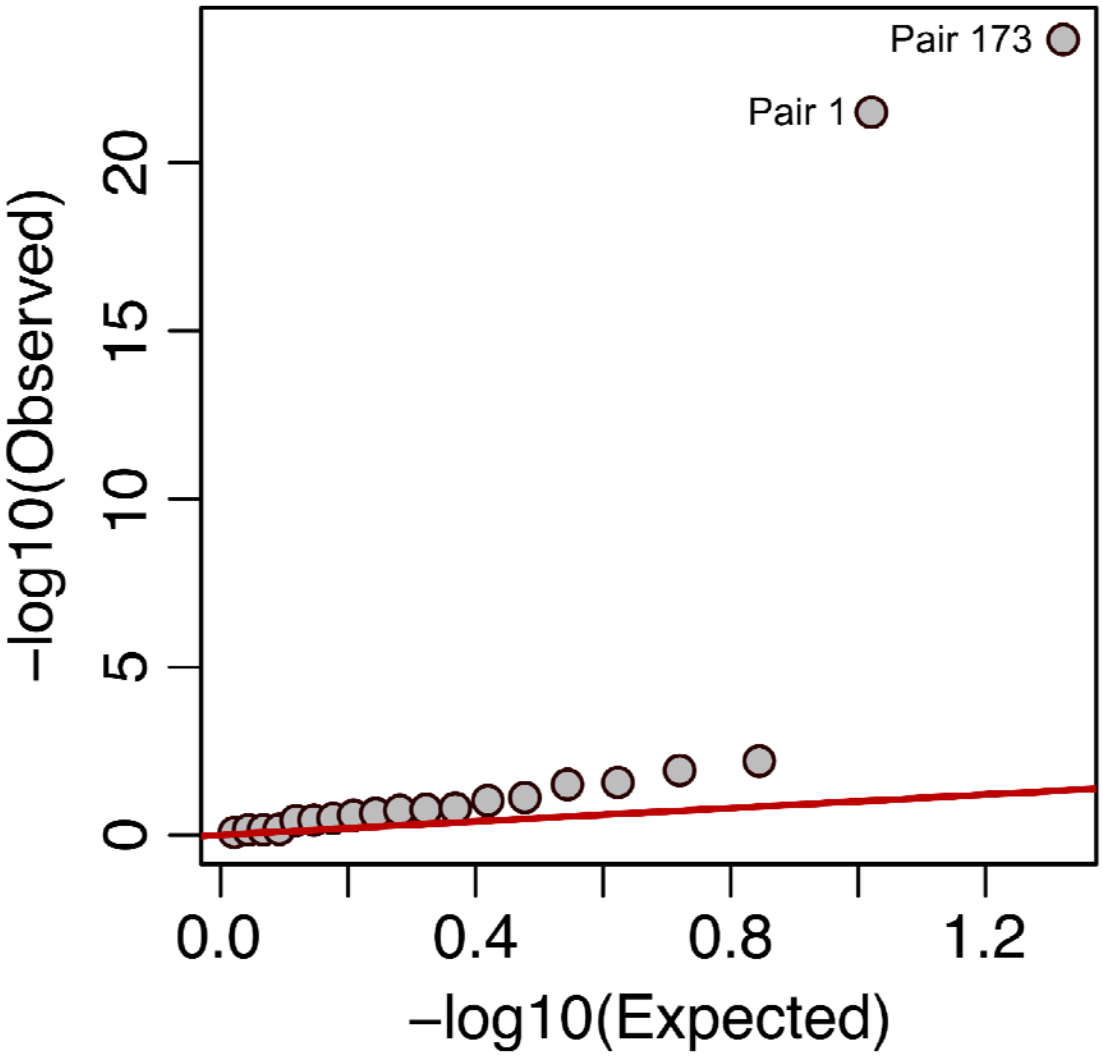
Extended Data 4. (A) Samples distributions across the four predominance groups and their associated metadata. (B) Individual species-specific relative abundance trajectories over time, for milk and infant stool samples.



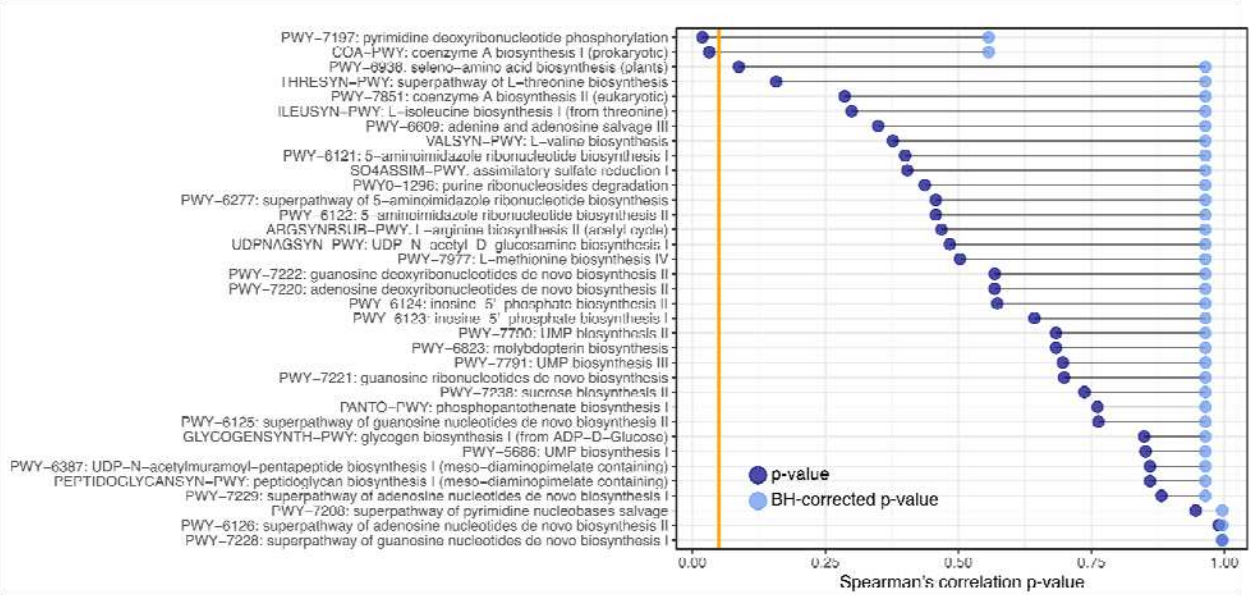
Extended Data 5. Species persistence in the infant gut microbiome over time, stratified by the type of transition between predominance groups from one to six. Only samples with both time points available and transition types with more than ten samples per type were included.



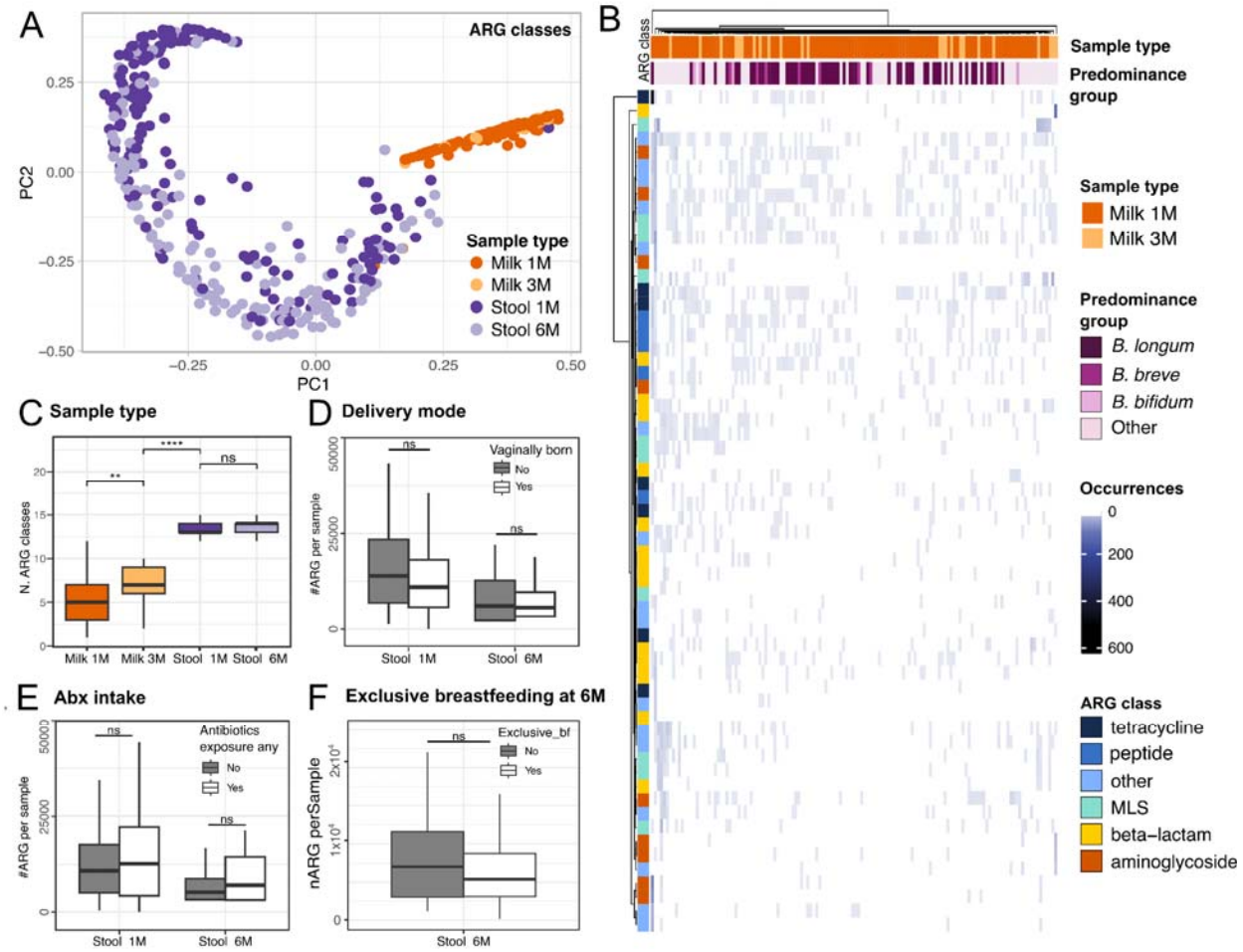
Extended Data 6. Relative abundance of pathways associated with (A) peptidoglycan maturation and (B) essential amino acids biosynthesis across sample types and (stool) predominance groups. CI at 95%, bootstrapping n=1000.



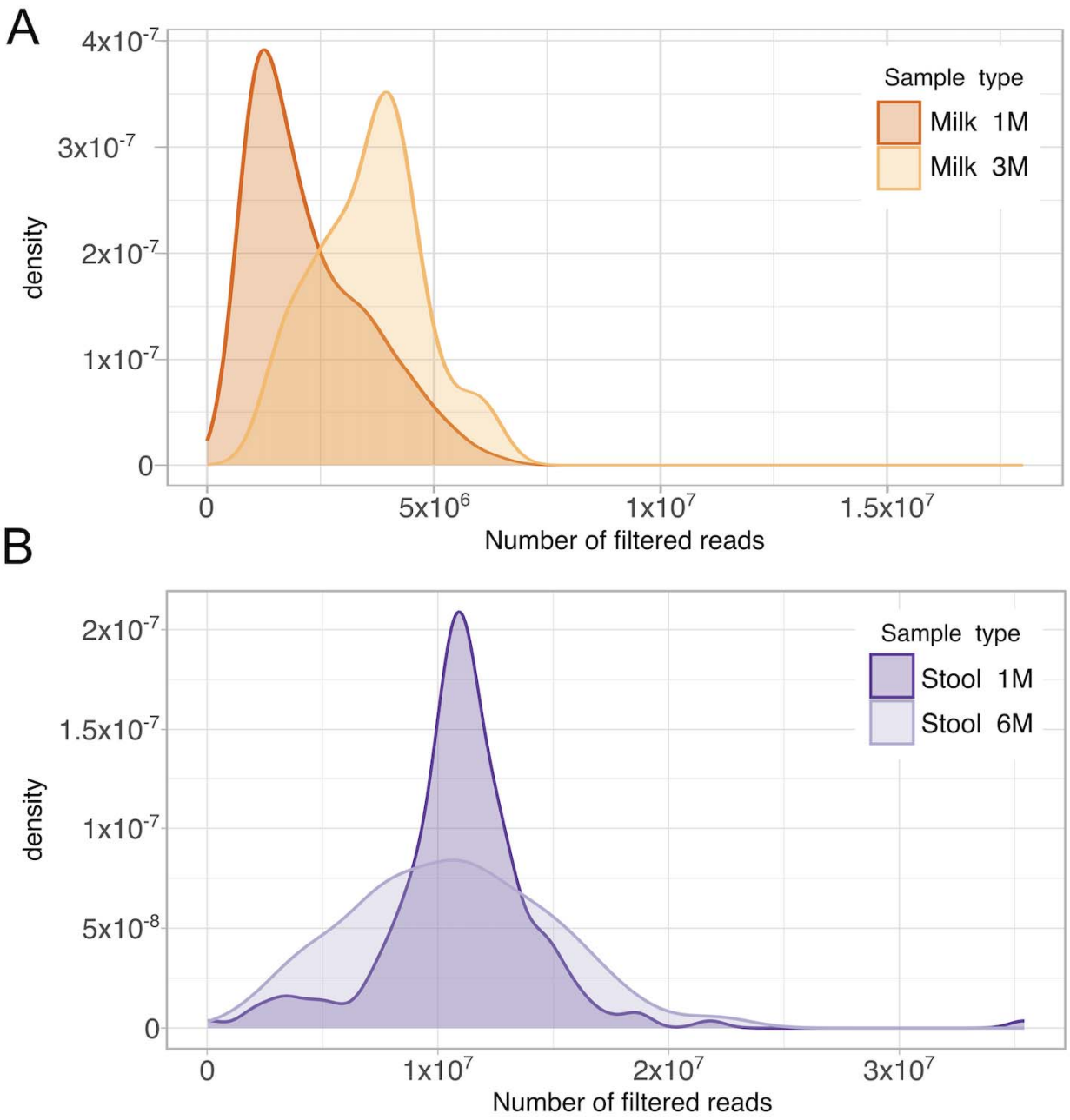
Extended Data 7. Q-Qplot of the Spearman's correlation p-values obtained comparing all metabolic pathways shared between milk and infant stools, for each mother-infant pair. Each dot represents the p-value for one mother-infant pair. Mother-infant pairs 1 and 173 are highlighted. The red line indicates a uniform distribution.



Extended Data 8. Spearman's correlation p-value for all metabolic pathways identified in both maternal milk and infant stool samples considering all mother infant pairs. The orange line identifies the significant thresholds (p=0.05). P-values were corrected for multiple testing using Benjamin Hochberg correction.



Extended Data 9. (A) PCoA of predicted ARGs classes using presence/absence information, as seen by DeepARG. (B) ARGs carriage in milk samples, divided by collection time point, most abundant species (predominance group), number of ARG genes identified and their respective ARGs class. (C) Number of distinct ARG classes identified across body sites and sampling time point. (D-F) Number of ARG genes identified in infant stool samples, divided by delivery mode, history of antibiotic intake and exclusive breastfeeding at 6 months of age, respectively. P-values calculated using t-test, **** for P ≤ 0.0001, ** p<0.01, ns for non significant.



Extended Data 10. Distribution of reads after preprocessing reads divided by sample type and collection time point.

Supplementary Table 1. Samples metadata
Supplementary Table 2. Species-level taxonomic profiles as seen by MetaPhlAn4
Supplementary Table 3. Strain-level taxonomic profiles as seen by StrainPhlAn4
Supplementary Table 4. Strain persistence between stool samples at one and six months
Supplementary Table 5. HUMAnN3 functional profiles
Supplementary Table 6. DeepARG profiles

References

1. Legoux, F. *et al.* Microbial metabolites control the thymic development of mucosal-associated invariant T cells. *Science* **366**, 494–499 (2019).
2. Zegarra-Ruiz, D. F. *et al.* Thymic development of gut-microbiota-specific T cells. *Nature* **594**, 413–417 (2021).
3. LeBlanc, J. G. *et al.* Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr. Opin. Biotechnol.* **24**, 160–168 (2013).
4. Yao, Y. *et al.* The Role of Microbiota in Infant Health: From Early Life to Adulthood. *Front. Immunol.* **12**, 708472 (2021).
5. Wilkins, A. T. & Reimer, R. A. Obesity, Early Life Gut Microbiota, and Antibiotics. *Microorganisms* **9**, (2021).
6. Asnicar, F. *et al.* Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *mSystems* **2**, (2017).
7. Korpela, K. *et al.* Selective maternal seeding and environment shape the human gut microbiome. *Genome Res.* **28**, 561–568 (2018).
8. Podlesny, D. & Fricke, W. F. Strain inheritance and neonatal gut microbiota development: A meta-analysis. *Int. J. Med. Microbiol.* **311**, 151483 (2021).
9. Duranti, S. *et al.* Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome* **5**, 66 (2017).
10. Enav, H., Bäckhed, F. & Ley, R. E. The developing infant gut microbiome: A strain-level view. *Cell Host Microbe* **30**, 627–638 (2022).
11. Ferretti, P. *et al.* Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe* **24**, 133–145.e5 (2018).
12. Song, S. J. *et al.* Naturalization of the microbiota developmental trajectory of Cesarean-born neonates after vaginal seeding. *Med* **2**, 951–964.e5 (2021).

13. Shao, Y. *et al.* Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* **574**, 117–121 (2019).
14. Mortensen, M. S. *et al.* Modeling transfer of vaginal microbiota from mother to infant in early life. *Elife* **10**, e57051 (2021).
15. Mitchell, C. M. *et al.* Delivery Mode Affects Stability of Early Infant Gut Microbiota. *Cell Rep Med* **1**, 100156 (2020).
16. CDC. Recommendations and benefits. *Centers for Disease Control and Prevention* https://www.cdc.gov/nutrition/infantandtoddlernutrition/breastfeeding/recommendations-benefits.html (2023).
17. Ferretti, P. *et al.* C. difficile may be overdiagnosed in adults and is a prevalent commensal in infants. *Elife* **12**, (2023).
18. Ballard, O. & Morrow, A. L. Human milk composition: nutrients and bioactive factors. *Pediatr. Clin. North Am.* **60**, 49–74 (2013).
19. Notarbartolo, V., Giuffrè, M., Montante, C., Corsello, G. & Carta, M. Composition of Human Breast Milk Microbiota and Its Role in Children's Health. *Pediatr Gastroenterol Hepatol Nutr* **25**, 194–210 (2022).
20. Das, L., Virmani, R., Sharma, V., Rawat, D. & Singh, Y. Human Milk Microbiota: Transferring the Antibiotic Resistome to Infants. *Indian J. Microbiol.* **59**, 410–416 (2019).
21. Pärnänen, K. *et al.* Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nat. Commun.* **9**, 3891 (2018).
22. Xue, M. *et al.* Breastfeeding and risk of childhood asthma: a systematic review and meta-analysis. *ERJ Open Res* **7**, (2021).
23. Dogaru, C. M., Nyffenegger, D., Pescatore, A. M., Spycher, B. D. & Kuehni, C. E. Breastfeeding and childhood asthma: systematic review and meta-analysis. *American journal of epidemiology* vol. 179 1153–1167 (2014).
24. Uwaezuoke, S. N., Eneh, C. I. & Ndu, I. K. Relationship Between Exclusive Breastfeeding

669 and Lower Risk of Childhood Obesity: A Narrative Review of Published Evidence. *Clin.*
670 *Med. Insights Pediatr.* **11**, 1179556517690196 (2017).
671 25. Patelarou, E. *et al.* Current evidence on the associations of breastfeeding, infant formula,
672 and cow's milk introduction with type 1 diabetes mellitus: a systematic review. *Nutr. Rev.*
673 **70**, 509–519 (2012).
674 26. Kull, I., Wickman, M., Lilja, G., Nordvall, S. L. & Pershagen, G. Breast feeding and allergic
675 diseases in infants-a prospective birth cohort study. *Arch. Dis. Child.* **87**, 478–481 (2002).
676 27. Ruiz, L., García-Carral, C. & Rodriguez, J. M. Unfolding the Human Milk Microbiome
677 Landscape in the Omics Era. *Front. Microbiol.* **10**, 1378 (2019).
678 28. Bogaert, D. *et al.* Mother-to-infant microbiota transmission and infant microbiota
679 development across multiple body sites. *Cell Host Microbe* **31**, 447–460.e6 (2023).
680 29. Qi, C. *et al.* Lactation-dependent vertical transmission of natural probiotics from the mother
681 to the infant gut through breast milk. *Food Funct.* **13**, 304–315 (2022).
682 30. Van Rossum, T., Ferretti, P., Maistrenko, O. M. & Bork, P. Diversity within species:
683 interpreting strains in microbiomes. *Nat. Rev. Microbiol.* **18**, 491–506 (2020).
684 31. Blanco-Míguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with
685 uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-
686 023-01688-w.
687 32. Singh, P. *et al.* Unveiling the dynamics of the breast milk microbiome: impact of lactation
688 stage and gestational age. *J. Transl. Med.* **21**, 784 (2023).
689 33. Selma-Royo, M., Calvo Lerma, J., Cortés-Macías, E. & Collado, M. C. Human milk
690 microbiome: From actual knowledge to future perspective. *Semin. Perinatol.* **45**, 151450
691 (2021).
692 34. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse
693 microbial communities with bioBakery 3. *Elife* **10**, (2021).
694 35. Bäckhed, F. *et al.* Dynamics and Stabilization of the Human Gut Microbiome during the

First Year of Life. *Cell Host Microbe* **17**, 852 (2015).

36. Lebeaux, R. M. *et al.* The infant gut resistome is associated with E. coli and early-life exposures. *BMC Microbiol.* **21**, 201 (2021).

37. Kim, D.-W. & Cha, C.-J. Antibiotic resistome from the One-Health perspective: understanding and controlling antimicrobial resistance transmission. *Exp. Mol. Med.* **53**, 301–309 (2021).

38. Chen, P.-W., Tseng, S.-Y. & Huang, M.-S. Antibiotic Susceptibility of Commensal Bacteria from Human Milk. *Curr. Microbiol.* **72**, 113–119 (2016).

39. Behari, P., Englund, J., Alcasid, G., Garcia-Houchins, S. & Weber, S. G. Transmission of methicillin-resistant Staphylococcus aureus to preterm infants through breast milk. *Infect. Control Hosp. Epidemiol.* **25**, 778–780 (2004).

40. Samarra, A. *et al.* Maternal-infant antibiotic resistance genes transference: what do we know? *Gut Microbes* **15**, 2194797 (2023).

41. Arango-Argoty, G. *et al.* DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**, 23 (2018).

42. Lopez Leyva, L., Brereton, N. J. B. & Koski, K. G. Emerging frontiers in human milk microbiome research and suggested primers for 16S rRNA gene analysis. *Comput. Struct. Biotechnol. J.* **19**, 121–133 (2021).

43. Turroni, F., Berry, D. & Ventura, M. *Bifidobacteria and Their Role in the Human Gut Microbiota. 2nd Edition*. (Frontiers Media SA, 2020).

44. Soto, A. *et al.* Lactobacilli and bifidobacteria in human breast milk: influence of antibiotherapy and other host and clinical factors. *J. Pediatr. Gastroenterol. Nutr.* **59**, 78–88 (2014).

45. Trosvik, P. & de Muinck, E. J. Ecology of bacteria in the human gastrointestinal tract—identification of keystone and foundation taxa. *Microbiome* **3**, 1–12 (2015).

46. Le Doare, K., Holder, B., Bassett, A. & Pannaraj, P. S. Mother's Milk: A Purposeful

721 Contribution to the Development of the Infant Microbiota and Immunity. *Front. Immunol.* **9**,
722 361 (2018).
723 47. Lawson, M. A. E. *et al.* Breast milk-derived human milk oligosaccharides promote
724 Bifidobacterium interactions within a single ecosystem. *ISME J.* **14**, 635–648 (2020).
725 48. Heisel, T. *et al.* Bacterial, fungal, and interkingdom microbiome features of exclusively
726 breastfeeding dyads are associated with infant age, antibiotic exposure, and birth mode.
727 *Front. Microbiol.* **13**, 1050574 (2022).
728 49. Dorota, P. *et al.* Klebsiella pneumoniae in breast milk: a cause of sepsis in neonate. *Arch*
729 *Med* **9**, (2017).
730 50. Vasilyev, I. Y., Nikolaeva, I. V., Siniagina, M. N., Kharchenko, A. M. & Shaikhieva, G. S.
731 Multidrug-Resistant Hypervirulent Klebsiella pneumoniae Found Persisting Silently in Infant
732 Gut Microbiota. *Int. J. Microbiol.* **2020**, 4054393 (2020).
733 51. Liu, K., Zhang, L., Gu, X. & Qu, W. The Prevalence of Klebsiella spp. Associated With
734 Bovine Mastitis in China and Its Antimicrobial Resistance Rate: A Meta-Analysis. *Front Vet*
735 *Sci* **9**, 757504 (2022).
736 52. Feehily, C. *et al.* Detailed mapping of Bifidobacterium strain transmission from mother to
737 infant via a dual culture-based and metagenomic approach. *Nat. Commun.* **14**, 3015
738 (2023).
739 53. Raveh-Sadka, T. *et al.* Evidence for persistent and shared bacterial strains against a
740 background of largely unique gut colonization in hospitalized premature infants. *ISME J.* **10**,
741 2817–2830 (2016).
742 54. Brooks, B. *et al.* Strain-resolved analysis of hospital rooms and infants reveals overlap
743 between the human and room microbiome. *Nat. Commun.* **8**, 1814 (2017).
744 55. Tetracycline (class) (oral route, parenteral route). https://www.mayoclinic.org/drugs-
745 supplements/tetracycline-class-oral-route-parenteral-route/proper-use/drg-20069585
746 (2023).

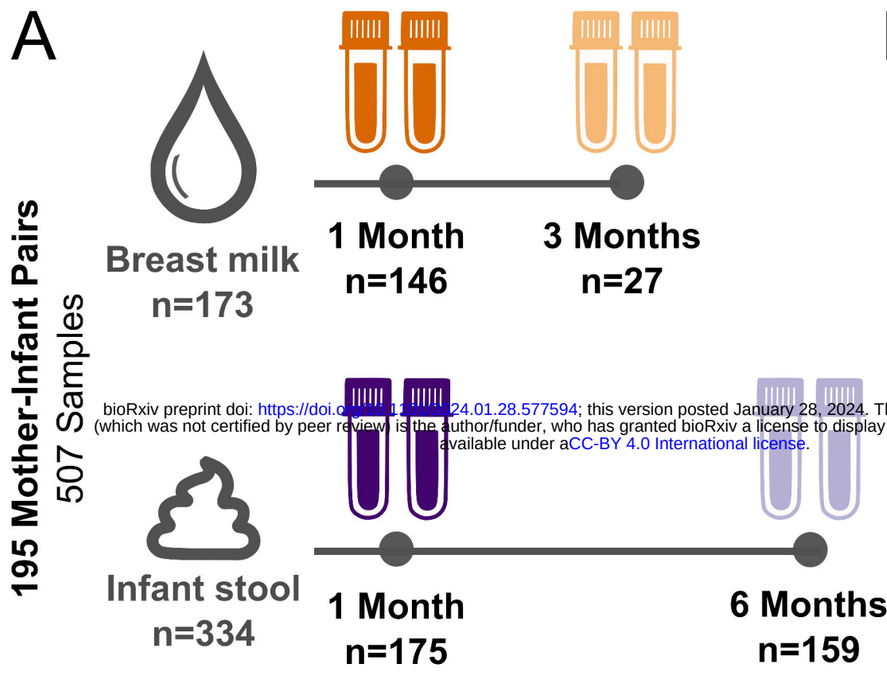
747 56. Leo, S. *et al.* Metagenomics analysis of the neonatal intestinal resistome. *Front Pediatr* **11**,
748 1169651 (2023).
749 57. Vatanen, T. *et al.* Mobile genetic elements from the maternal microbiome shape infant gut
750 microbial assembly and metabolism. *Cell* **185**, 4921–4936.e15 (2022).
751 58. Reyes, S. M. *et al.* Pumping supplies alter the microbiome of pumped human milk: An in-
752 home, randomized, crossover trial. *Am. J. Clin. Nutr.* **114**, 1960–1970 (2021).
753 59. Fields, D. A. *et al.* Associations between human breast milk hormones and adipocytokines
754 and infant growth and body composition in the first 6 months of life. *Pediatr. Obes.* **12**
755 **Suppl 1**, 78–85 (2017).
756 60. Sadr Dadres, G. *et al.* Relationship of Maternal Weight Status Before, During, and After
757 Pregnancy with Breast Milk Hormone Concentrations. *Obesity* **27**, 621–628 (2019).
758 61. Whitaker, K. M. *et al.* Associations of Maternal Weight Status Before, During, and After
759 Pregnancy with Inflammatory Markers in Breast Milk. *Obesity* **26**, 1659–1660 (2018).
760 62. Johnson, K. E. *et al.* Human milk variation is shaped by maternal genetics and impacts the
761 infant gut microbiome. *bioRxiv* (2023) doi:10.1101/2023.01.24.525211.
762 63. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
763 **9**, 357–359 (2012).
764 64. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
765 2079 (2009).
766 65. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
767 features. *Bioinformatics* **26**, 841–842 (2010).
768 66. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
769 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
770 67. Andrews, S. & Others. FastQC: a quality control tool for high throughput sequence data.
771 Preprint at (2010).
772 68. Gu, Z. Complex heatmap visualization. *Imeta* **1**, (2022).

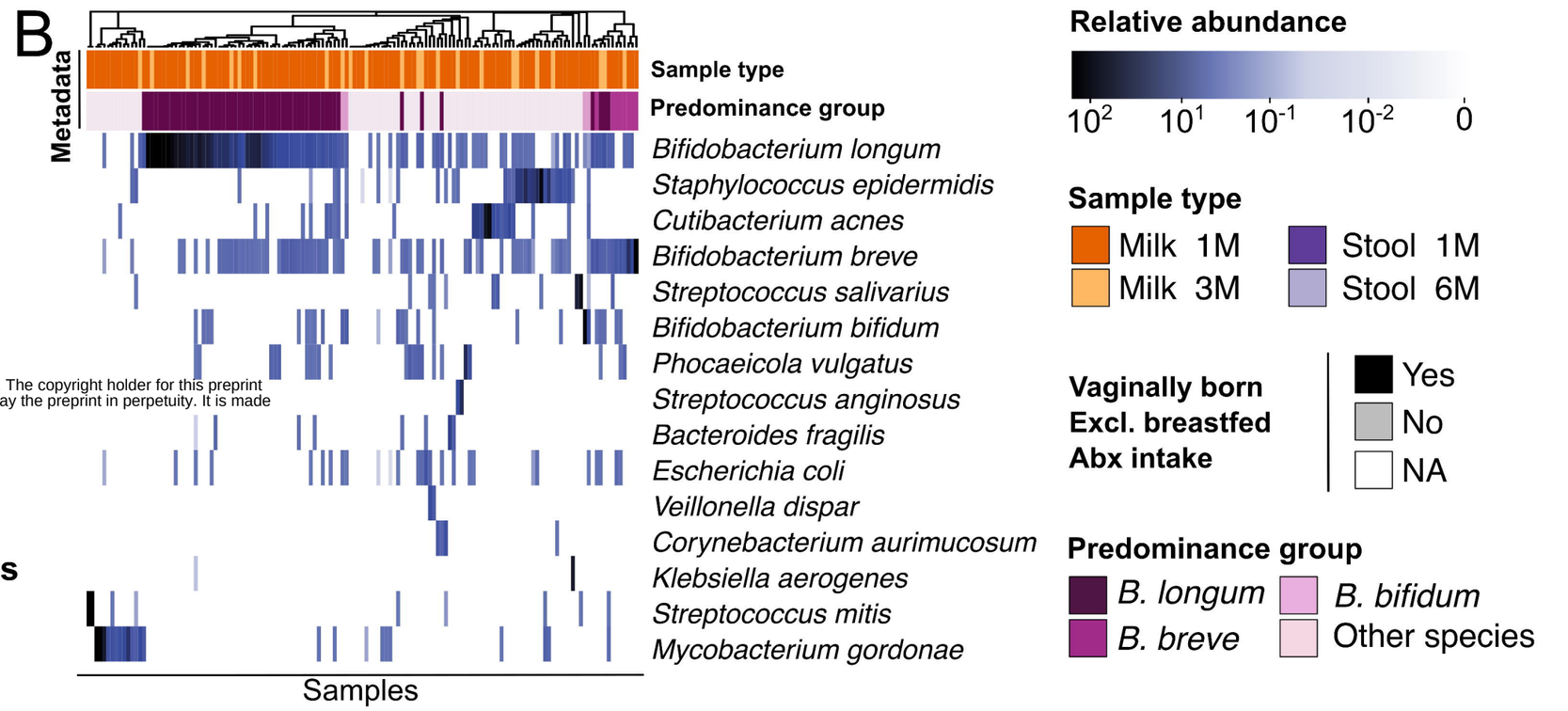
69. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

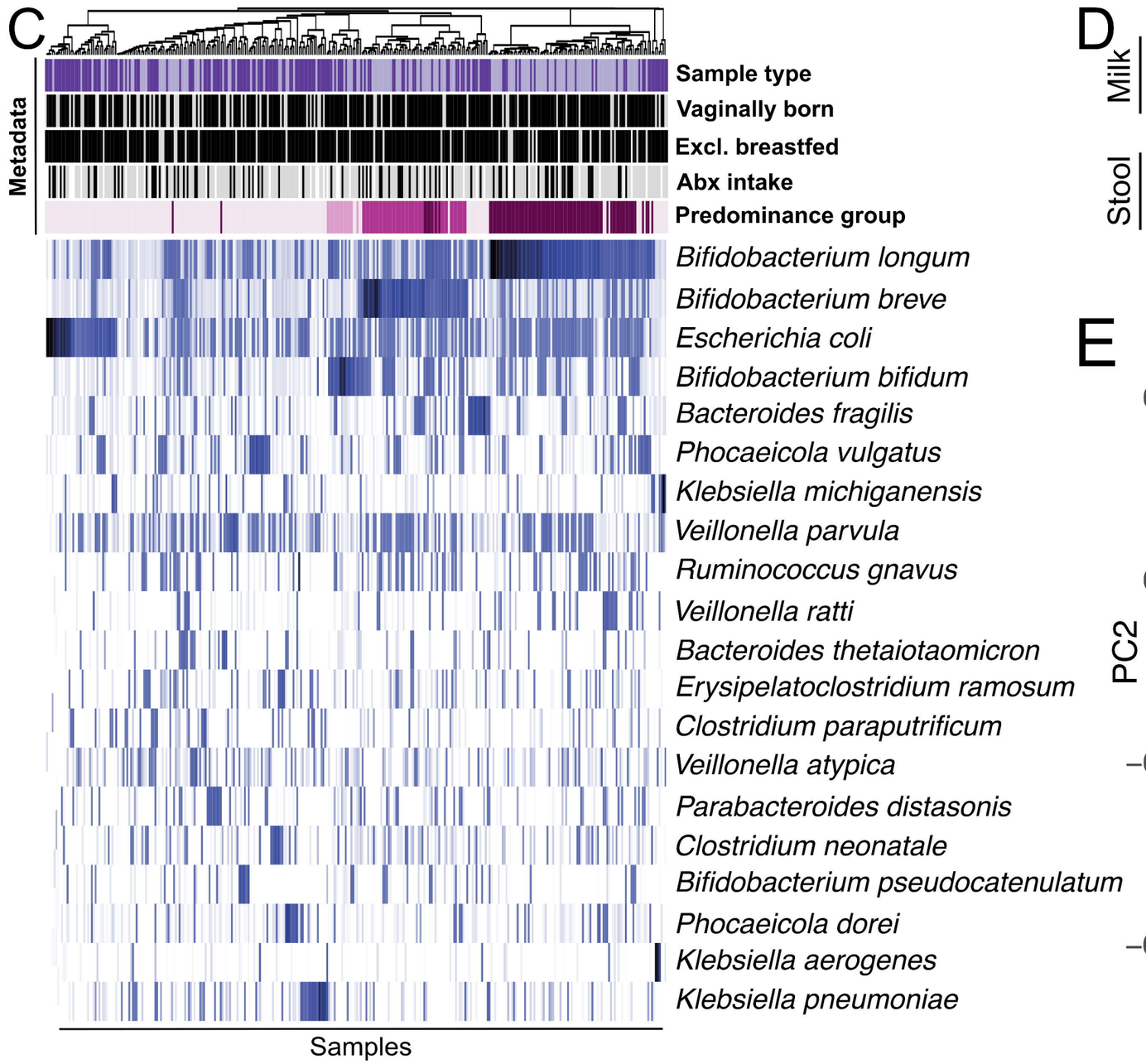
70. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

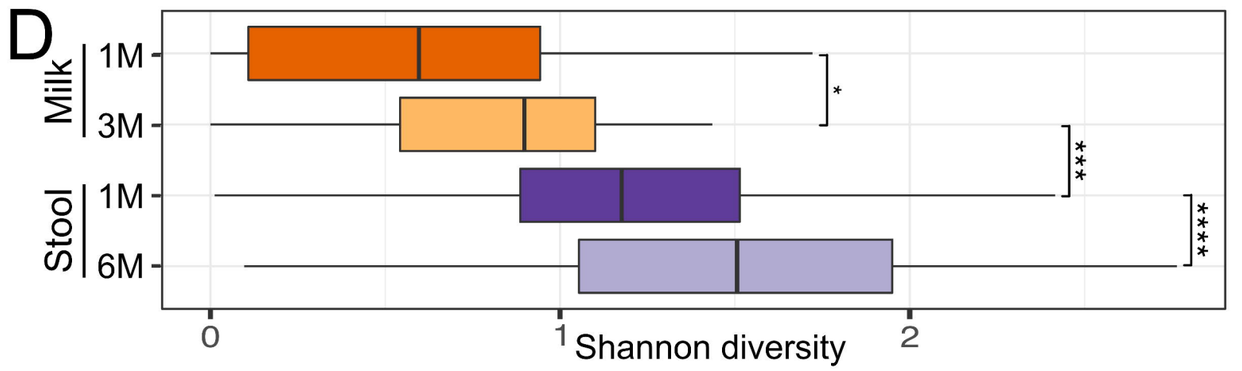
71. Foundation for Statistical Computing, R. R. R: A language and environment for statistical computing. *RA Lang Environ Stat Comput*.

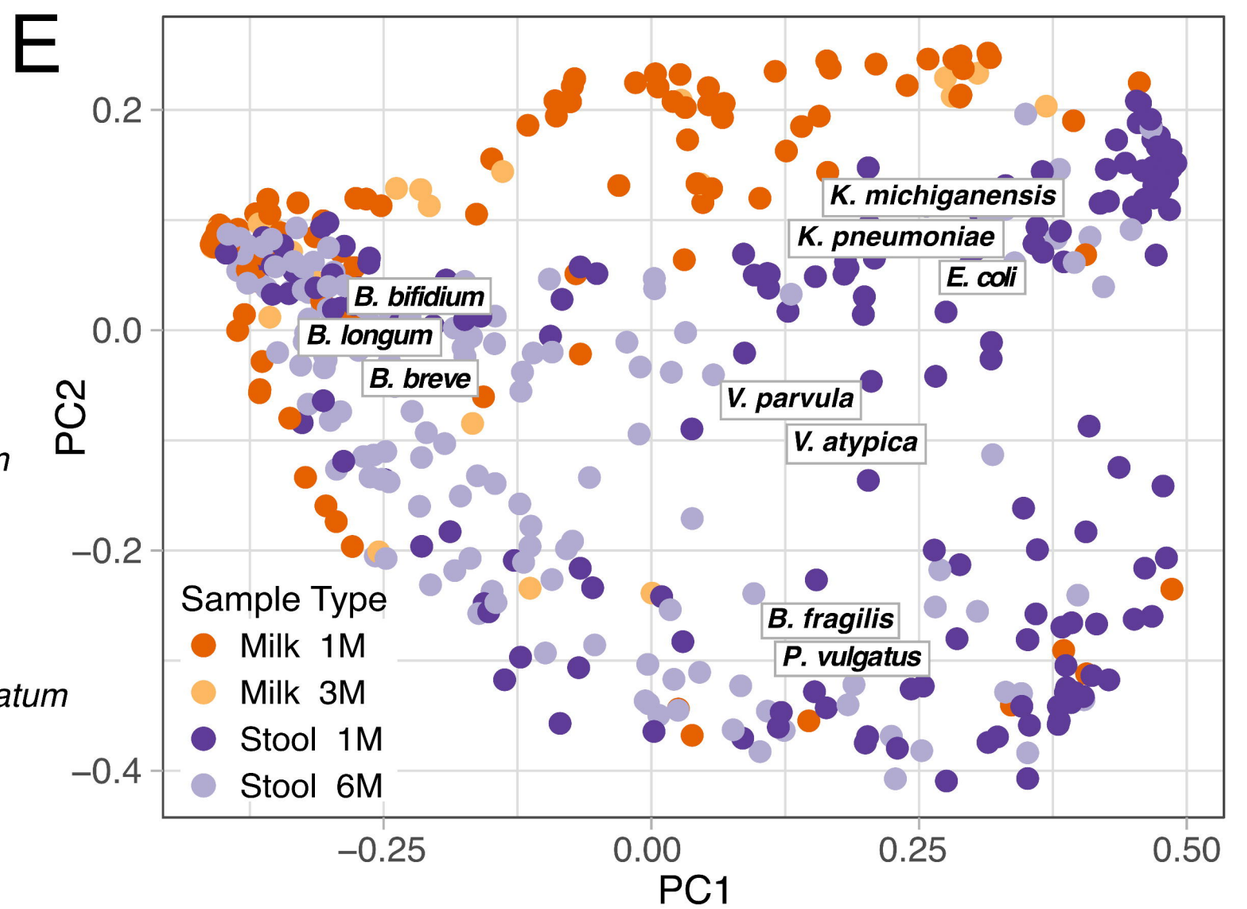
72. Wickham, H. & Wickham, M. H. The ggplot package. *Google Scholar*.











A

B

C

D

E

F

G

A *B. longum* (SGB17248)

Infant 174

Mother-Infant Pair 1

Infant 110

Infant 130

Infant 138

Infant 55

Infant 39

Infant 28

Infant 180

Infant 41

Tree scale: 0.1

B *K. pneumoniae* (SGB10115)

Mother-Infant Pair 173

Infant 169

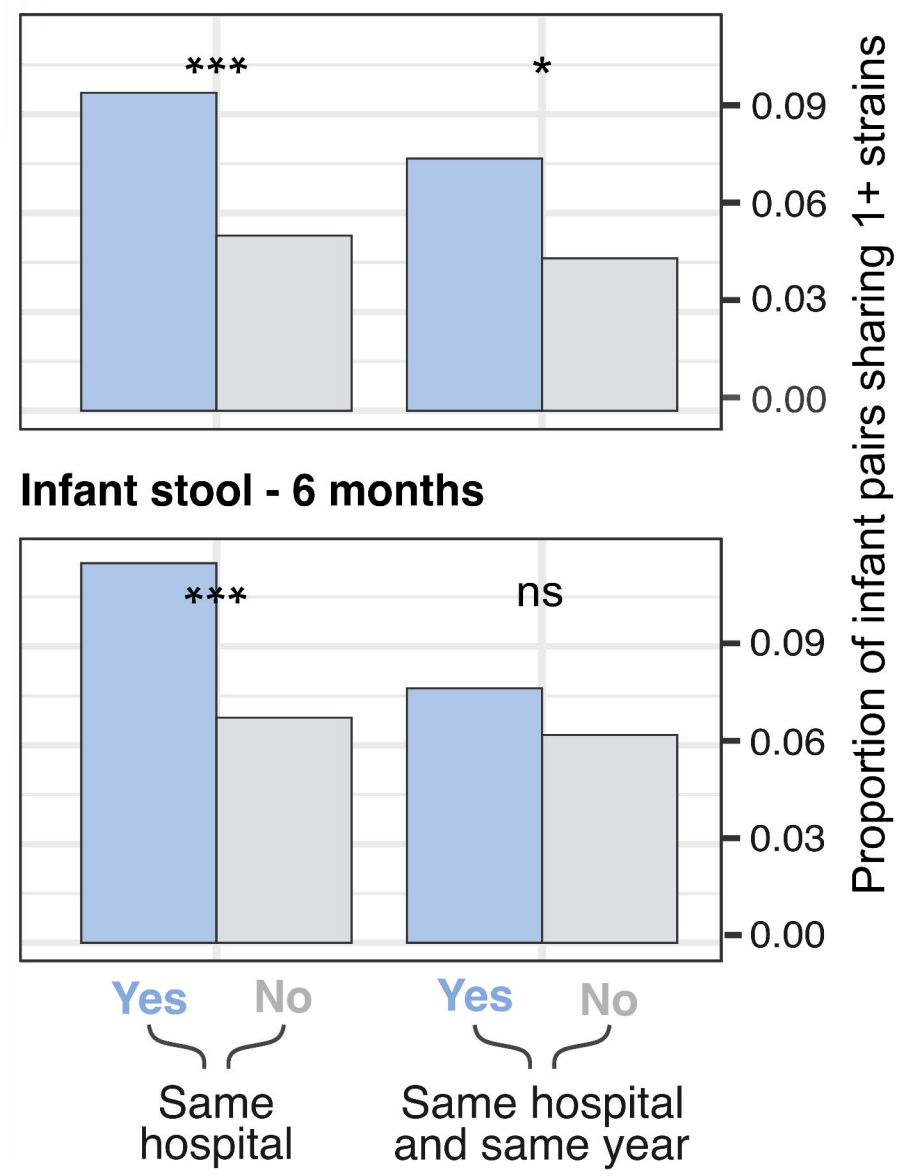
Sample Type
Milk 1M
Stool 1M
Stool 6M

Tree scale: 0.01

C Unrelated Infants

Infant stool - 1 month

Infant stool - 6 months



A

Infant Stool

Rel. abundance

10^{1} 10^{-2} 10^{-3} 0

Pathway Class

Sample Type

Predominance group

peptidoglycan maturation
glycolysis IV
pentose phosphate pathway I
L-valine biosynthesis
pyruvate fermentation to isobutanol
L-isoleucine biosynthesis I
superpath. branched chain amino acid biosynthesis
L-isoleucine biosynthesis III
sucrose biosynthesis II
glycogen degradation II
guanosine ribonucleotides de novo biosynthesis
peptidoglycan biosynthesis III
UMP biosynthesis I
L-lysine biosynthesis III
L-lysine biosynthesis VI
UDP-N-acetylmuramoyl-pentapeptide biosynthesis II
UDP-N-acetylmuramoyl-pentapeptide biosynthesis I
peptidoglycan biosynthesis I
S-adenosyl-L-methionine salvage I
5–aminoimidazole ribonucleotide biosynthesis II
superpath. 5-aminoimidazole ribonucleotide biosynth.
L-arginine biosynthesis I
superpath. L-isoleucine biosynthesis I
superpath. L-threonine biosynthesis
L-histidine biosynthesis
folate transformations II
chorismate biosynthesis I
chorismate biosynthesis from 3-dehydroquinate
superpath. aromatic amino acid biosynthesis
dTDP-β-L-rhamnose biosynthesis

B

Milk

Rel. abundance

10^{1} 10^{-3} 10^{-4} 10^{-5} 0

Sample Type

Predominance group

L-valine biosynthesis
pyruvate fermentation to isobutanol
sucrose biosynthesis II
glycogen degradation II
guanosine ribonucleotides de novo biosynth.
UDP-N-acetyl-D-glucosamine biosynthesis I
5-aminoimidazole ribonucleotide biosynthesis II
superpath. 5-aminoimidazole ribonucleotide biosynth.
5-aminoimidazole ribonucleotide biosynthesis I
L-isoleucine biosynthesis I
superpath. branched chain amino acid biosynth.
L-isoleucine biosynthesis III
UMP biosynthesis I
superpath. of L-threonine biosynthesis
folate transformations II
S-adenosyl-L-methionine salvage I
peptidoglycan maturation
gondoate biosynthesis
adenosine deoxyribonucleotides de novo biosynth. II
guanosine deoxyribonucleotides de novo biosynth. II

Sample Type

Stool 1M
Stool 6M
Milk 1M
Milk 3M

Predominance group

B. longum
B. breve
B. bifidum
Other

Class pathway

Cell structure
Cell metabolism
Amino acid biosynthesis
Nucleotide biosynthesis
Other

C

Essential aa biosynthesis

Rel. abundance

2×10^{-5} 3×10^{-5} 4×10^{-5} 5×10^{-5}

**** **

Excl. BF
Excl. BF
Non excl. BF

1M 3M 1M 6M 6M

Milk Stool

D

Milk-stool shared pathways across pairs

Mother-infant pair

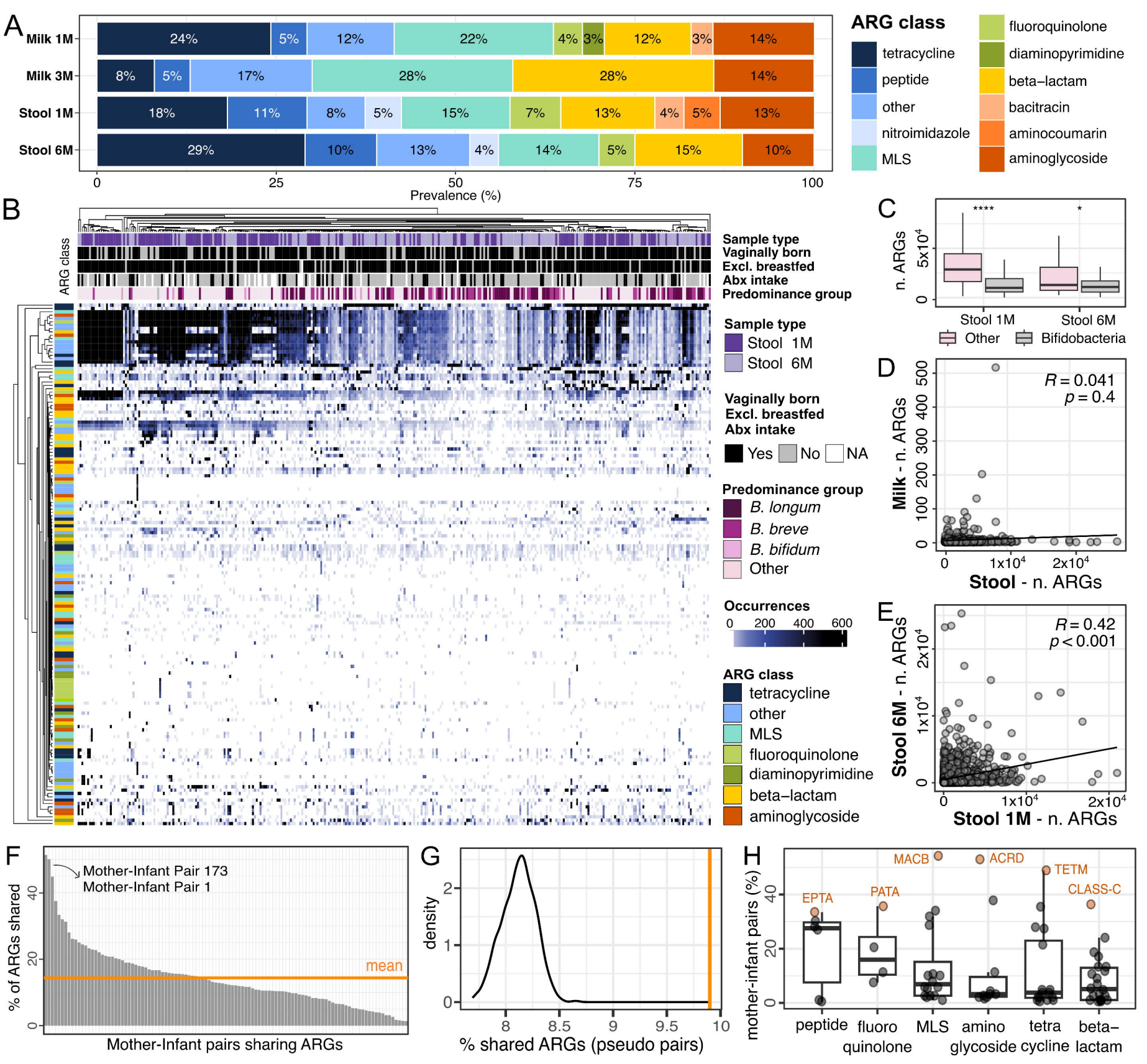
Pair 172
Pair 7
Pair 40
Pair 36
Pair 41
Pair 13
Pair 22
Pair 71
Pair 97
Pair 154
Pair 149
Pair 144
Pair 95
Pair 3
Pair 9
Pair 135
Pair 161
Pair 157
Pair 1
Pair 173

Spearman's correlation rho

−0.25
0.00
0.25
0.50

−20 −15 −10 −5 0

log10(adjusted p-value)


A
Milk 1M
Milk 3M
Stool 1M
Stool 6M
24% 5% 12% 22% 4% 3% 12% 3% 14%
8% 5% 17% 28% 28% 14%
18% 11% 8% 5% 15% 7% 13% 4% 5% 13%
29% 10% 13% 4% 14% 5% 15% 10%
0 25 50 75 100
Prevalence (%)
ARG class
tetracycline
peptide
other
nitroimidazole
MLS
fluoroquinolone
diaminopyrimidine
beta–lactam
bacitracin
aminocoumarin
aminoglycoside
B
ARG class
Sample type
Vaginally born
Excl. breastfed
Abx intake
Predominance group
Sample type
Stool 1M
Stool 6M
Vaginally born
Excl. breastfed
Abx intake
Yes No NA
Predominance group
B. longum
B. breve
B. bifidum
Other
Occurrences
0 200 400 600
ARG class
tetracycline
other
MLS
fluoroquinolone
diaminopyrimidine
beta–lactam
aminoglycoside
C
n. ARGs
0 5x10^4
****
*
Stool 1M
Stool 6M
Other
Bifidobacteria
D
R = 0.041
p = 0.4
Milk - n. ARGs
0 100 200 300 400 500
0 1x10^4 2x10^4
Stool - n. ARGs
E
R = 0.42
p < 0.001
Stool 6M - n. ARGs
0 1x10^4 2x10^4
0 1x10^4 2x10^4
Stool 1M - n. ARGs
F
Mother-Infant Pair 173
Mother-Infant Pair 1
mean
% of ARGs shared
0 20 40
Mother-Infant pairs sharing ARGs
G
density
0 1 2
8 8.5 9 9.5 10
% shared ARGs (pseudo pairs)
H
mother-infant pairs (%)
0 20 40
EPTA
PATA
MACB
ACRD
TETM
CLASS-C
peptide
fluoro quinolone
MLS
amino glycoside
tetra cycline
beta– lactam