Gut microbiota and faecal cytokine profiles of rural Cambodian infants and links to early life nutrition and diarrhoeal episodes.

Matthew J Dalby[1,2], Raymond Kiu[1,2], Iliana R Serghiou[1], Asuka Miyazaki[4], Holly Acford-Palmer[1], Rathavy Tung[6], Shabhonam Caim[1], Sarah Phillips[1], Magdalena Kujawska[1, 3], Mitsuaki Matsui[4], Azusa Iwamoto[5], Bunsreng Taking[7], Sharon E Cox[4,8,9], Lindsay J Hall[1,3,10].

[1]Gut Microbes & Health, Quadram Institute Bioscience, Norwich Research Park, Norwich, NR4 7UQ, UK

[2]Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

[3]Intestinal Microbiome, School of Life Sciences, ZIEL – Institute for Food & Health, Technical University of Munich, Freising, 80333, Germany

[4]School of Tropical Medicine & Global Health, Nagasaki University, Nagasaki, Japan.

[5]Bureau of International Health Cooperation, National Centre for Global Health and Medicine, Tokyo, Japan

[6]National Maternal and Child Health Centre, Ministry of Health, Phnom Penh, Cambodia

[7]Kampong Cham Provincial Health Department, Ministry of Health, Kampong Cham, Cambodia

[8]London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

[9]Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan

[10]Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

Abstract

Currently, there is an underrepresentation of gut microbiota profiles from infants in low-middle income countries, including those residing in rural setting. This study aimed to investigate the gut microbiota composition and faecal cytokine profiles of infants in rural Cambodia. Longitudinal faecal samples collected from 32 infants with an average age of seven months were analysed using 16S rRNA amplicon sequencing to determine composition of the gut microbiota, bacterial strains isolated and subjected to whole genome sequencing and genomic analysis, and concentrations of faecal cytokines were also measured. Initially, and given potential challenges of sample collection and storage in rural settings, we also tested different collection methods, and determined that faecal samples stored in DNA preservation buffer has a significantly greater abundance of *Bacteroides*, when compared to naïve samples. Broader microbiota analysis indicated that *Bifidobacterium* was the most abundant genus in the infant gut with *Bifidobacterium longum* the most abundant species. Gut microbiota composition was associated with breastfeeding, particularly higher *Bifidobacterium longum*. Most infants had detectable pathogenic taxa, indicating exposure to environmental pathogens. We also determined that potentially pathogenic genera were more abundant in infants with recent diarrheal illness. Surprisingly we did not detect any significant antibiotic-associated perturbations in the gut microbiota, and no associations were found between gut microbiota and infant growth. Profiling of isolated strains indicated the ability to digest specific dietary components e.g. human milk oligosaccharides, and the presence of antibiotic resistance genes in potentially pathogenic species, as well as beneficial genera including *Bifidobacterium*. Faecal cytokine analysis revealed consistent differences in Interlukin-1alpha and vascular endothelial growth factor concentrations, suggesting breastfeeding's influence on the infant gut mucosal immune system. This study offers insights into the gut microbiota composition and faecal immune profiles of rural Cambodian infants, emphasising the differences between infants from high-income countries.

58 **Introduction**

59 Research on the gut microbiota has primarily focussed on populations in high income countries (e.g.
60 Europe and North America) (Abdill et al., 2022). Although more recent studies have increased
61 representation of populations in low- and low-middle income countries (LMICs), infant studies lag
62 behind those working on adult cohorts (Porras and Brito 2019). This is important for defining what is
63 a 'healthy' human gut microbiota (Lackey et al., 2019) across individuals in LMICs (Porras and Brito,
64 2019), and because optimal infant development is particularly important for infants from LMICs due
65 to increased risk of (mal)nutrition and infection. While a small proportion of microbiota samples
66 studied globally so far have come from South Asia, including India and Bangladesh (Abdill et al.,
67 2022), research into the gut microbiota of people living outside these countries remains very limited.
68 Moreover, due to the logistics of sample collection and storage in rural areas, research in LMICs has
69 often focused on individuals in urban areas, neglecting key infant populations. Cambodia in
70 particular has a scarcity of microbiota studies and given that this country has such a high prevalence
71 of infants classified as stunted/wasted, and has amongst the highest infant mortality rates in
72 Southeast Asia, with infections accounting for many of these deaths (Cambodia Demographic and
73 Health Survey, 2022), new studies targeting Cambodian infant populations are needed. Previous,
74 research relating to gut-associated microbes of infants living in Cambodia has included studies
75 focusing on particular bacterial pathogens, including antimicrobial resistance in *Enterobacterales* in
76 faecal samples from babies and children (Auguet et al., 2021) and the epidemiology of antibiotic
77 resistance in *Escherichia coli* and *Klebsiella pneumoniae* in children (van Aartsen et al., 2019).
78 However, the composition of the gut microbiota as a whole in infants from Cambodia remains to be
79 investigated.

80 While the nutritional status of infants is improving in Cambodia, children under the age of 5 in rural
81 areas still face challenges with 22% suffering linear growth faltering or 'stunting' (Cambodia
82 Demographic and Health Survey, 2022). This is also true for infants (<1 year) in rural Cambodia who
83 also suffer from similar levels of growth stunting at 19.2% (Miyazaki et al., 2021), with a high
84 proportion also lacking access to appropriate hygiene facilities. This is coupled with a high
85 proportion reportedly experiencing symptoms of diarrhoea, fever or cough and frequent use of
86 unregulated antibiotic treatment to treat these childhood illnesses (Miyazaki et al., 2020).
87 Malnutrition has previously been linked to differences in the gut microbiota of children from several
88 different low income countries (Robertson, 2020; Surono et al., 2021), although research into these
89 associations is lacking for children from Cambodia.

90 The establishment of the gut microbiota during infancy is important aspect of healthy infant
91 development with roles in immune, metabolic, endocrine, and other host developmental pathways
92 (Robertson et al., 2019). Vaginally-delivered, breast fed infants are characterised by their very high
93 abundance of the bacterial genus *Bifidobacterium.* This keystone microbiota member orchestrates
94 wider microbiome structuring and modulates host physiology through the breakdown of complex
95 dietary components such as human milk oligosaccharides (HMOs) in breast milk and plant-based
96 carbohydrates, leading to production of key metabolites that facilitate immune system development
97 and maturation and cognitive responses (Henrick et al., 2021). Previous research indicates that LMIC
98 infant microbiomes are characterised by *Bifidobacterium* species and strains, although these are
99 often different genetic variants when compared to those infants from high income country settings
100 (Lackey et al., 2019; Taft et al., 2022).

101 Infants born into resource-poor environments can have difficulties in establishing a healthy gut
102 microbial ecosystem (Lang & MAL-ED, 2018) due to multiple factors such as birth and breast feeding
103 practices, antibiotic exposure, maternal microbiota, diet, socio-economic status, host genetics,
104 agricultural dependence, urbanisation, access to clean water and hygiene practices (Porras and
105 Brito, 2019; Robertson et al., 2019). Infant microbiomes of those in LMICs are reportedly higher in
106 *Prevotella* species, associated with both positive and negative health outcomes, and may correlate
107 with higher plant based, fibre rich diets (Porras and Brito 2019). The phylum Pseudomonadota
108 (previously Proteobacteria) that includes a variety of pathogenic and opportunistic pathogens, is
109 highly prevalent in microbiomes of those in LMICs (Porras and Brito 2019). These factors may
110 negatively affect the composition of the early infant gut microbiota (Robertson et al., 2019), and
111 may drive these observed microbiome differences across populations (Porras and Brito 2019).

112 In this study we carried out an exploratory analysis of the gut microbiota in faecal samples collected
113 from 32 infants that are part of the Nutrition for Health of Aka-chan (baby in Japanese) and Mamas'
114 cohort (NHAM) based in Cambodia which included background data on infant diet, health, and living
115 conditions. This provides an opportunity to explore the composition of the gut microbiota in rural
116 Cambodian infants that has not previously been investigated. The samples from this cohort are
117 unusual due to the rural location in which these infants live, with poor access to hygiene facilities,
118 high prevalence of pathogen exposure, and frequent antibiotic usage. We analysed the effects of
119 faecal sample storage on the composition of the microbiota of samples collected under suboptimal
120 conditions. Then described the infant faecal microbiota composition and explored its associations
121 with environmental factors such as infant diet and growth. Individual faecal bacteria cultured and
122 isolated and their genomes sequenced to identify the presence of antibiotic resistance and HMO
123 genes. Finally, the concentrations of faecal cytokines were measured to identify any associations

124 with the gut microbiota. We aimed to explore the composition of gut microbiota of these infants and
125 its associations with their environment.

126

127 **Results**

128 The infant faecal samples analysed in this study were sourced from 32 infants representing a subset
129 of infants that are part of NHAM birth cohort based in the rural Kampong Cham province in
130 Cambodia (Miyazaki et al., 2020, 2021) (**Table 1**). Samples were collected from infants from 10
131 villages along the Mekong River in Kampong Cham province. This area is located around 60
132 kilometres from the provincial capital Kampong Cham and has limited transport links. The infants
133 were living in a rural area, with most families being small-scale farmers.

134

135 **Table 1: Infant baseline characteristics.**

Variable	Infants
n	32
Age (days) (mean (SD))	214.9 (37.3)
Sex = Female (%)	15 (46.9)
Breast feeding = No (%)	9 (28.1)
Complementary feeding = No (%)	17 (53.1)
Feeding bottle use = Yes (%)	21 (65.6)
Low birth weight (<2.5kg) (%)	
Normal	29 (90.6)
Unknown	2 (6.2)
Low	1 (3.1)
Growth = Stunted (%)	3 (9.4)
Length-for-age z-score (mean (SD))	-0.9 (0.8)
Underweight z-score (mean (SD))	-0.6 (0.8)
Weight-for-length (mean (SD))	0.0 (1.0)
Hand washing score (mean (SD))	5.2 (1.6)
Improved toilet facility = Yes (%)	22 (68.8)
Household toilet facility = Septic tank (%)	22 (68.8)
Number in household (mean (SD))	5.8 (2.2)
Home village (%)	

K1	1 (3.1)
K2	2 (6.2)
K5	4 (12.5)
K3	10 (31.2)
K4	2 (6.2)
K6	2 (6.2)
P1	1 (3.1)
P3	6 (18.8)
P4	3 (9.4)
P5	1 (3.1)
Antibiotic use (previous 7 days) = No (%)	28 (87.5)
Any recent illness (previous 7 days) = No (%)	11 (34.4)
Diarrheal illness (previous 7 days) = No (%)	26 (81.2)
Fever (previous 7 days) = No (%)	17 (53.1)
Cough (previous 7 days) = No (%)	20 (62.5)

Sample storage conditions influence gut microbiota profiles

Sample collection and storage is challenging in areas without the facilities to immediately freeze faecal samples on collection. To compare the preservation of DNA in samples under different conditions, half of each faecal sample was placed into a DNA preservation solution (DNAShield), with the remaining half initially remaining unaltered. The faecal samples were collected in a rural location and then transported at ambient air temperature with transportation typically taking ~2 days. The untreated half of each sample was then initially frozen at -20°C and then later stored at -80°C. The half of each sample stored in DNAShield were stored refrigerated at 4°C. The samples were collected from 32 infants and three samples from each infant were collected around 2 weeks apart with a total of 96 samples, with half split between each storage condition. Both frozen and DNAShield paired infant gut microbiota samples underwent 16S rRNA amplicon sequencing to analyse their bacterial composition.

Out of the 96 paired samples undergoing 16S rRNA amplicon sequencing, a total of 94 were successfully sequenced when stored in DNAShield, while only 70 were successfully sequenced from Frozen samples (**Supplementary File 1**). There were 69 pairs of samples with successful 16S rRNA amplicon sequencing from both the Frozen and DNAShield portion of the same faecal sample. The resulting bacterial genus composition was compared between these 69 pairs of samples (**Figure 1**).

154 While the total number of bacterial genera detected was similar between frozen and DNAShield
155 samples (**Figure 1A**), microbiota diversity calculated using the Shannon index and Inverse Simpson
156 index was higher in DNAShield samples than paired frozen samples (**Figure 1B-C**). *Bacteroides*,
157 *Escherichia, Parabacteroides, and Lachnoclostridium* were the four bacterial genera with the largest
158 median difference in relative abundance between storage conditions that were higher in DNAShield
159 samples compared to frozen samples (**Figure 1D-G; Supplementary File 1**). *Bacteroides* showed the
160 largest difference in favour of DNAShield storage with a median relative abundance of 0.4% in frozen
161 samples and 8.6% in DNAShield samples (**Figure 1D; Supplementary File 1**). The relative abundance
162 of *Bifidobacterium*, *Streptococcus*, *Enterococcus,* and *Actinomyces* were the genera with the largest
163 difference and significantly higher abundance in favour of frozen samples (**Figure 1H-K;**
164 **Supplementary File 1**). *Bifidobacterium* comprised a median relative abundance of 63% in frozen
165 samples and 46% in DNAShield samples (**Figure 1H; Supplementary File 1**).

166 These results indicate that sample storage conditions selectively altered the abundance of different
167 genera in analysis of bacterial composition with significantly lower abundance of *Bacteroides* and
168 *Parabacteroides*. Further analysis in this study was performed using samples stored in DNAShield
169 due to the larger number of samples with successful sequencing results and the higher bacterial
170 diversity surviving in DNAShield stored samples.

171

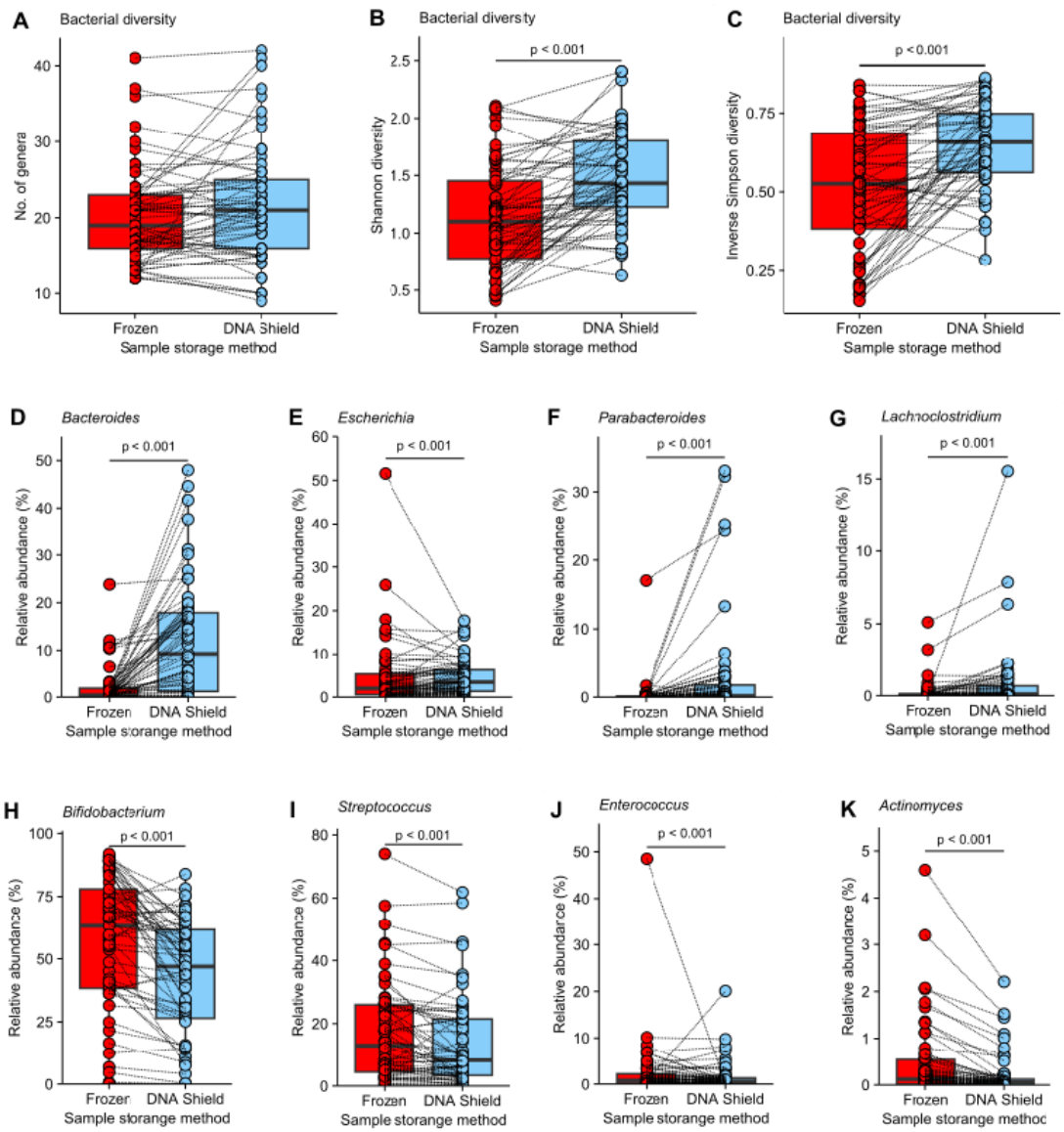


Figure 1. Differences in genus abundance between paired frozen samples and DNAShield stored samples. (A) The number of genera detected in each sample; (B) The Shannon diversity index of each sample; (C) The Inverse Simpson diversity index of each sample; Relative abundance between paired sampled after different storage conditions of (D) *Bacteroides*; (E) *Escherichia*; (F) *Parabacteroides* ; (G) *Lachnoclostridium;* (H) *Bifidobacterium*; (I) *Streptococcus*; (J) *Enterococcus*; (K) *Actinomyces*. Box plots show median and interquartile range with circles showing individual samples. Dotted lines link together paired Frozen and DNAShield samples from the same infant sample. Statistical significance between groups was tested using a Wilcoxon signed rank test with correction for multiple testing amongst bacteria genera using the false discovery rate method.

182 ***Genus bacterial composition of the infant gut microbiota is influenced by breastfeeding and***
183 ***diarrheal illness***

184 The gut microbiota of infants was dominated by *Bifidobacterium,* which comprised a median relative
185 abundance of 48.7% (**Figure 2A**). In addition, other genera including *Streptococcus*, *Bacteroides*,
186 *Escherichia*, and *Lactobacillus* formed the remaining majority of the gut microbiota in these infants
187 (**Figure 2A**).

188 An exploratory analysis was carried out using metadata for the infants included in this study. This
189 included cross-sectional data collected at the start of the study just before collection of the faecal
190 sample at time point 1 and additional information collected at each sample collection visit
191 (**Supplementary File 2**). The exploratory analyses used non-metric multidimensional scaling (NMDS)
192 and sample beta-diversity to cluster the samples by the similarity of their overall microbiota
193 composition at time point 1. Associations between the infant metadata with the beta-diversity of
194 the gut microbiota was analysed using Envfit and significant associations illustrated on the NMDS
195 plot (**Figure 2B**). Only breast-feeding status was significantly associated with microbiota beta-
196 diversity with clear separation in clusters between infants breast fed and those infants not breast
197 fed (**Figure 2B**).

198 To identify associations between the relative abundance of individual genera and infant metadata an
199 analysis using the MaAsLin2 package was carried out that included the samples from all three time
200 points with time point included as a random effect (**Figure 2C; Supplementary File 3**). When
201 background metadata variables were analysed individually, both breastfeeding status and recent
202 diarrhoeal illness within the previous seven days showed significant associations with genus
203 abundance, which was maintained when both variable were included together as a multivariable
204 analysis (**Figure 2C**). Genera with relative abundance that was statistically significantly associated
205 with breast-feeding and diarrheal illness are shown as a heat map with positive and negative
206 symbols indicating the direction of association (**Figure 2C**). Genus positively associated with breast-
207 feeding were *Bifidobacterium* and *Streptococcus* (**Figure 2C; Figure 2D**) whilst those with the
208 strongest negative associations were *Klebsiella* and *Clostridium* (**Figure 2C ; Figure 2E**). Breast-
209 feeding status remained unchanged for individual infants across the three time points. At time point
210 1 the mean age of breast fed infants was 207 days and non-breast fed infants 214 days. Including
211 age as a variable in the analysis did not alter the associations with breastfeeding. Other measures of
212 infant diet, the use of a feeding bottle or complementary feeding, were not associated with genus
213 abundance. MaAsLin2 analysis showed that that recent diarrheal illness within the previous seven
214 days was associated with a strongest lower abundance of *Bacteroides* (**Figure 2F**) and higher relative

abundance of *Clostridium*, *Klebsiella* and *Shigella* (**Figure 2G-H**). *Shigella* was detected in 78% of samples and *Klebsiella* in 82% and *Clostridium* in 65% indicating that these infants have a high prevalence of these potentially pathogenic genera in their gut microbiota (**Supplementary File 4**). Breast feeding was the only variable associated with differences in microbial diversity. The total number of bacterial genus (**Figure 3I**) and the Shannon diversity were lower in breast fed infants than non-breast fed infants (**Figure 3J**).

Only three infants included in the analysis were classified as stunted having a height-for-age z-score of less than 2 and this was not associated with microbiota composition. But, the number of infants was too small for robust testing. Tested as continuous measures, length-for-age z-score , underweight z-score, and weight-for-length z-score were also not associated with microbiota composition. Measures of household hygiene facilities and total number of people in the household were not associated with microbiota composition. Of particular note was the lack of an association between genus composition and recent antibiotic use within the previous 7 days (**Supplementary File 3**).

These results indicate that breast feeding status was the most influential factor shaping the gut microbiota composition in favour of *Bifidobacterium*. Recent diarrheal illness was associated with higher relative abundance of potentially pathogenic genera.

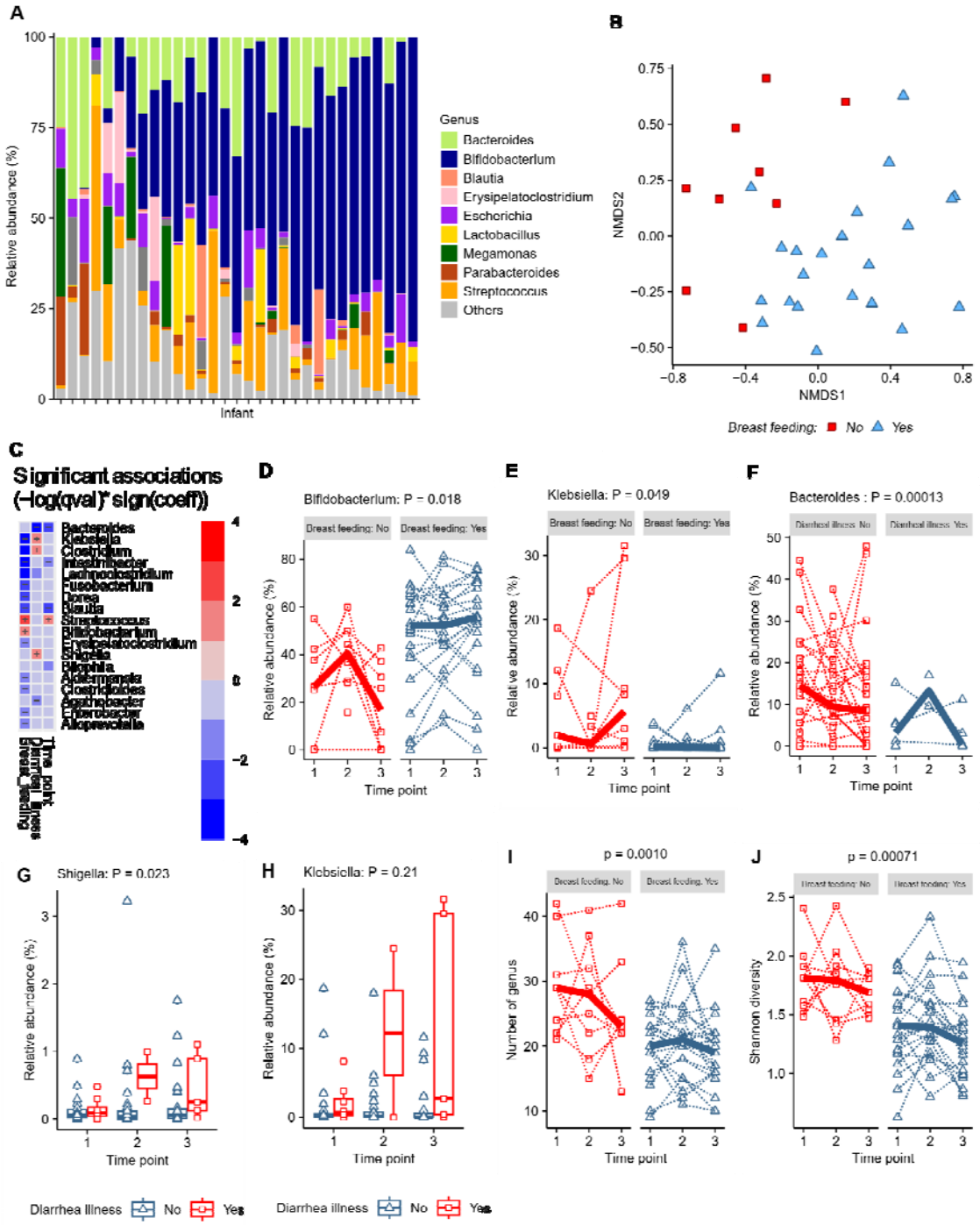


Figure 2. Genus composition and associations of genus composition with breast feeding status and diarrheal illness. (A) The ten most abundant bacterial genera across all samples ordered on the x-axis by *Bifidobacterium* abundance in each infant; (B) Non-metric multidimensional scaling (NMDS) plot representing beta-diversity of infant genus microbiota at time point 1 clustered by similarity and coloured by breastfeeding status with arrows showing top ten bacterial genus diving separation of samples; (C) Heat map showing significant associations determined by MaAsLin2 analysis between

241 genus and breast feeding, diarrheal illness and time point; (D) The relative abundance of the genus
242 *Bifidobacterium* by breastfeeding status; (E) The relative abundance of the genus *Klebsiella* by
243 breastfeeding status; (F) The relative abundance of the genus *Bacteroides* by breastfeeding status;
244 (G) The relative abundance of the genus *Shigella* by diarrheal illness within the previous 7 days; (H)
245 The relative abundance of the genus *Klebsiella* by diarrheal illness within the previous 7 days; (I) The
246 number of genera by breast feeding status; (J) The Shannon diversity index by breast feeding status;
247 Points on plots show individual infant data and dotted lines on spaghetti plots link individual infants
248 across time points with the bold line showing the group median.

249 ***Bacterial species composition of the infant gut microbiota is influenced by breastfeeding status***

250 The same analysis that was carried out on the genus level composition was carried out on the data
251 at the species level with the limitation that species classifications using 16S rRNA amplicon data are
252 less robust than genus (**Supplementary File 5**). The gut microbiota at the species level was
253 dominated by a single *Bifidobacterium* species classified as *Bifidobacterium longum* (**Figure 3A**).
254 Other abundant species detected were members of those genera identified in (**Figure 2A**) with other
255 species of *Bifidobacterium* identified as *B. breve* and *B. bifidum*, while *Escherichia coli*, *Klebsiella*
256 *pneumoniae*, *Bacteroides fragilis* and an unclassified *Streptococcus* comprised the other abundant
257 species present.

258 Repeated exploratory MaAsLin2 analysis of the species level 16S rRNA amplicon sequencing data
259 identified the same association of breast-feeding status on microbiota composition with a stronger
260 effect with species level data (**Figure 3C; Supplementary File 6**). Species with relative abundance
261 that was statistically significantly associated with breast feeding and recent diarrheal illness are
262 shown as a heat map with positive and negative symbols indicating the direction of association
263 (**Figure 2C**). *Bifidobacterium longum* showed the strongest positive association with breast-feeding,
264 with consistently higher median relative abundance in breastfed infants compared to those not
265 breastfed (**Figure 3D**) while *Klebsiella pneumoniae* showed the strongest negative association with
266 breast feeding (**Figure 3E**). The species *Shigella flexneri* (**Figure 3F**) and *Clostridium perfringens*
267 (**Figure 3G**) were higher in samples taken after recent diarrheal illness. Breast feeding was the only
268 variable associated with differences in microbial diversity at the species level of classification. The
269 total number of bacterial species (**Figure 3H**) and the Shannon diversity was lower in breast fed
270 infants than non-breast fed infants (**Figure 3I**). *Shigella flexneri* was detected in 49% of samples and
271 *Clostridium perfringens* detected in 24% of samples indicating that these infants have a high
272 prevalence of these potentially pathogenic genera in their gut microbiota (**Supplementary File 7**).

273 As described at the genus level, classification as stunted with a height-for-age z-score of less that 2
274 was not associated with gut microbiota species composition. Similarly, length-for-age z-score ,
275 underweight z-score, and weight-for-length z-score as continuous measures were also not
276 associated with species composition. Hygiene measures of household toilet facilities and total
277 number of people in the household were again not associated with species composition. Also noted
278 was the lack of an association between the gut microbiota composition and recent antibiotic use
279 (**Supplementary File 6**).

280 These species level results support the finding that breast feeding status was an major influence on
281 the gut microbiota composition and associated with the abundance of *Bifidobacterium longum*.
282 Recent diarrheal illness was associated with higher abundance bacterial species including *Shigella*
283 *flexneri* and *Clostridium perfringens* that are known to be potential causes of diarrheal illness.

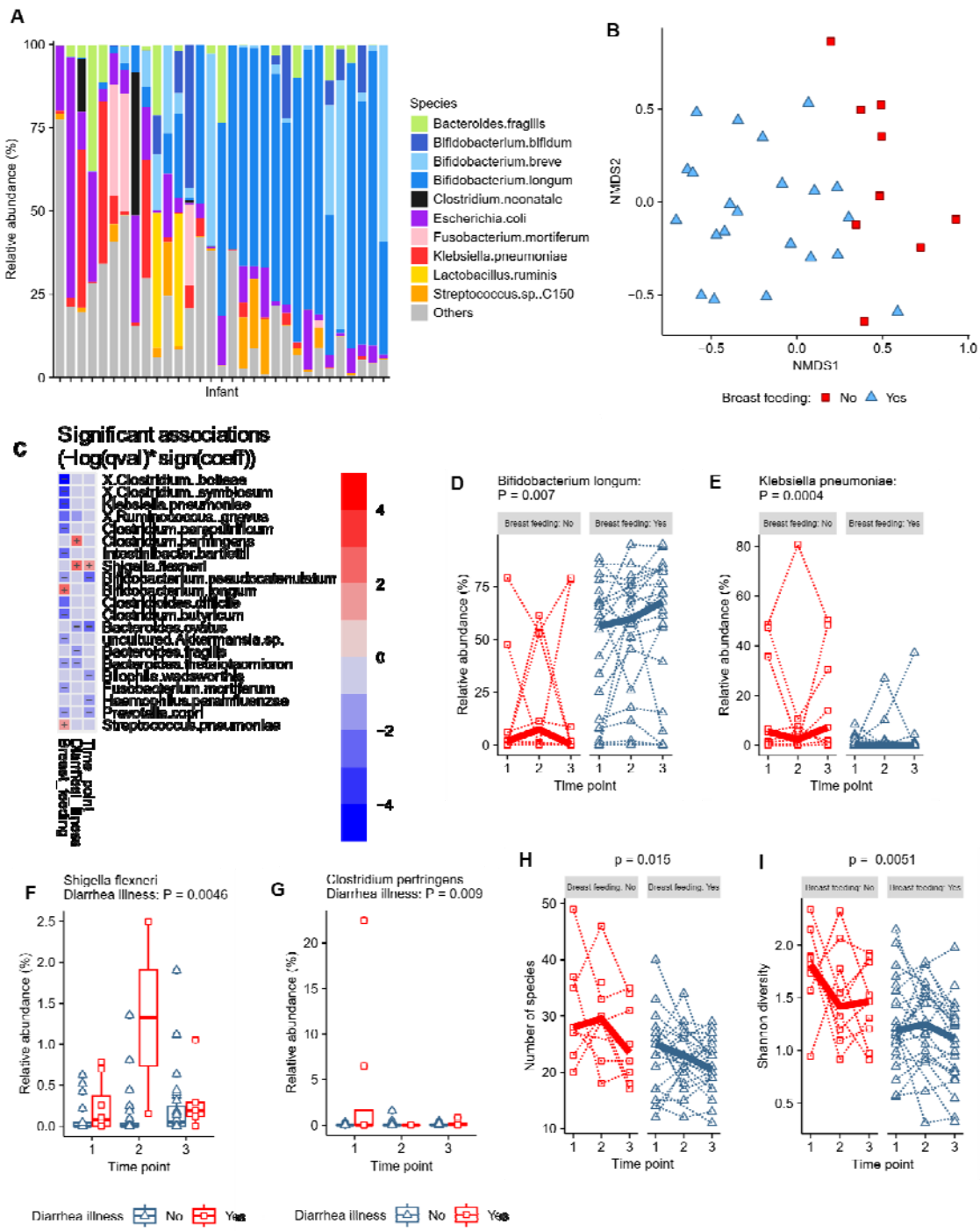


Figure 3. Species composition and association of species composition with breast feeding and diarrheal illness. (A) The top ten most abundant bacterial species across all samples at time point 1 ordered by total *Bifidobacterium* species abundance; (B) Non-metric multidimensional scaling (NMDS) plot representing beta-diversity of infant species microbiota at time point 1 clustered by similarity and coloured by breastfeeding status with arrows showing top five bacterial genus diving

290 separation of samples; (C) Heat map showing significant associations determined by MaAsLin2
291 analysis between species and breast feeding, diarrheal illness and time point; (D) The relative
292 abundance of the *Bifidobacterium longum* by breastfeeding status; (E) The relative abundance of the
293 *Klebsiella pneumoniae* by breastfeeding status; (F) The relative abundance of the *Shigella flexneri* by
294 diarrheal illness within the previous 7 days; (G) The relative abundance of *Clostridium perfringens* by
295 diarrheal illness within the previous 7 days; (H) The number of species by breast feeding status; (I)
296 The Shannon diversity index by breast feeding status, Points on plots show individual infant data and
297 dotted lines on spaghetti plots link individual infants across time points with the bold line showing
298 the group median.

299 ***Infants harbour bacterial strains that encode diverse human milk oligosaccharide and***
300 ***antimicrobial resistance genes***

301 The frozen samples were used to culture and isolate individual colonies for whole genome
302 sequencing to gain further insight into the species of bacteria in the gut microbiota of these infants.
303 Genomic analysis was carried out on the sequenced isolates to identify the species (and strains), and
304 for the presence of virulence traits, antimicrobial resistance genes and HMO gene clusters. From the
305 bacterial isolates genome sequenced 2 were identified as *Bacteroides*, 6 as *Bifidobacterium*, 1 as
306 *Ruminococcus*, 3 as *Enterococcus*, 21 as *Clostridium*, 3 as *Clostridioides*, and 2 as *Paraclostridium*
307 (**Figure 4A**).

308 *Bifidobacterium* isolates comprised three species including *B. pseudocatenulatum*, *B. breve*, and *B.*
309 *longum*. Both *B. breve* and *B. longum* isolates contained complete HMO gene clusters, with none in
310 *B. pseudocatenulatum* (**Figure 4B**). This indicates the ability of the *B. breve* and *B. longum* present to
311 digest the oligosaccharides present in human breast milk with only *B. longum* showed higher relative
312 abundance in breastfed infants compared to non-breast fed (**Figure 4C**). Genomic analysis confirms
313 the presence of *B. longum* species in the gut microbiota of these infant and its ability to metabolise
314 HMOs from breast milk. Comparing the four genome-sequenced bacterial isolates identified as
315 *Bifidobacterium longum* using Average Nucleotide Identity with a 98% cut-off, three were identified
316 as *B. longum* subsp. *longum* and one as *B. longum* subsp. *suis* (**Supplementary File 8; Supplementary**
317 **Figure 1**).

318 Genomic analysis showed the presence of antimicrobial resistance genes across all strains isolated
319 (**Figure 4D**). Of particular note was the presence of the antimicrobial resistance genes *erm*(X), *tet*(O),
320 *tet*(O/23/O), *tet*(Q), and *tet*(W) in the genomes of isolated *Bifidobacterium* species (**Figure 4B**),
321 which provide resistance to erythromycin (*erm*) and tetracycline (*tet*) respectively. The identification
322 of antimicrobial resistance genes in the *Bifidobacterium* strains present may link with the lack of any

323 association between gut microbiota composition and those infants receiving antibiotics, as shown
324 here with NMDS plots showing all samples combined (**Figure 4E**). The relative abundance of
325 *Bifidobacterium* was unchanged in samples from infants receiving antibiotics (**Figure 4F**). Amongst
326 this group of infants the most commonly received antibiotic was amoxycillin, with individual
327 instances of cefixime, and cefadroxile also received (**Supplementary File 2**).

328 Two *Clostridium perfringens* strains were isolated and genome-sequenced. Both contained five toxin
329 producing genes and six colonisation factor genes that relate to the ability to cause infection and
330 disease in the gut (**Figure 4G**). This confirms the presence of *C. perfringens* that was identified using
331 16S rRNA amplicon sequencing and the presence of toxin and colonisation factor genes supports a
332 potential role in its association with cases of infant diarrhoea.

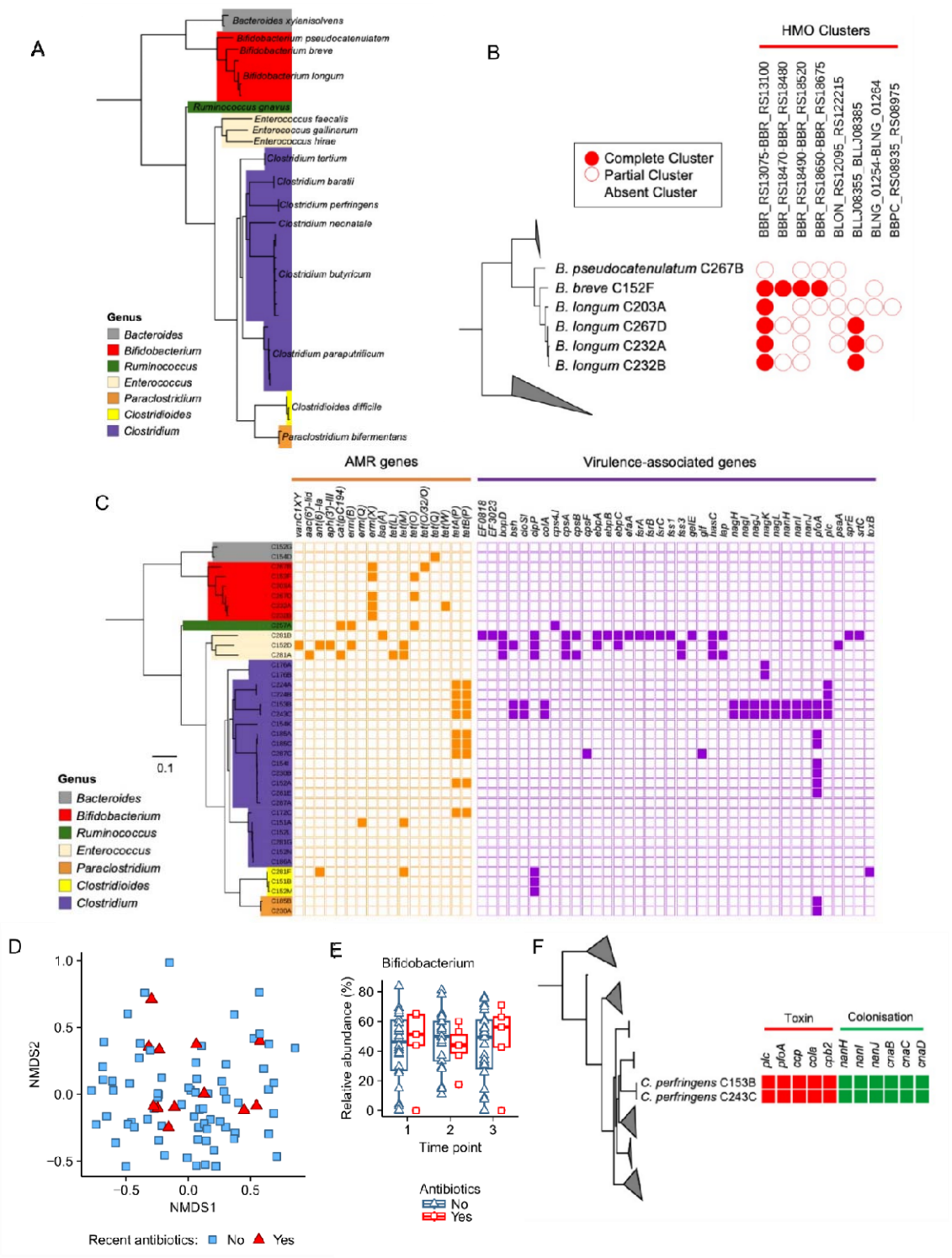


Figure 4. Genomic analysis of isolates for species determination, human milk oligosaccharide (HMO) gene clusters, antimicrobial resistance genes, and virulence genes. (A) Tree showing species and relatedness of whole genome-sequenced isolates; (B) Presence of HMO gene clusters in *Bifidobacterium* isolates; (C) Antimicrobial resistance genes and virulence associated genes identified

338 in the isolate genomes; (D) NMDS plot showing all clustered faecal samples highlighted with recent
339 antibiotic use; (E) *Bifidobacterium* genus abundance by antibiotic use; (F) Presence of toxin genes
340 and colonisation factors in *Clostridium perfringens* isolates.

341 ***Faecal cytokines profiles are associated with breastfeeding status***

342 The early life gut microbiota is known to play a key role in shaping gut and mucosal immune
343 responses. A multi-plex assay including a panel of thirty cytokines was used to determine
344 concentrations in the infant faecal samples (**Supplementary File 9**).

345 A linear mixed model incorporating infant metadata variables as a fixed effect and the infant as a
346 random effect was used as an exploratory analysis of associations between each cytokine and
347 metadata variables in all sample across the three time points (**Supplementary File 10**). The most
348 robust and consistent association was identified between two cytokines and breast feeding (**Figure**
349 **5A-C**). The cytokines Interleukin-1-alpha (IL-1α) (**Figure 5A**) and Vascular endothelial growth factor
350 (VEGF) (**Figure 5B**) were at significantly higher concentration in breast-fed infants. Both cytokines
351 showed substantial variation in concentration across time points within samples from the same
352 infant. IL-1α (**Figure 5A**) and VEGF (**Figure 5B**) were also the two cytokines with the highest
353 concentrations in the faecal samples across all infants.

354 To identify whether differences in major faecal cytokines by breastfeeding status were linked to
355 differences seen in the microbiota composition a correlation analysis was carried out between faecal
356 cytokine concentrations and the relative abundance of the 5 most abundant genus (**Figure 5C**) and
357 species (**Supplementary Figure 2**) in the microbiota. Weakly consistent positive associations were
358 seen between VEGF and *Lactobacillus* (**Figure 5D**) and *Streptococcus* (**Figure 5E**) across the three
359 time points, significant at two time points after correction for multiple testing. There were no
360 consistent correlations between the concentration of other cytokines and relative abundance of
361 individual bacterial genus or species, including *Bifidobacterium*.

362

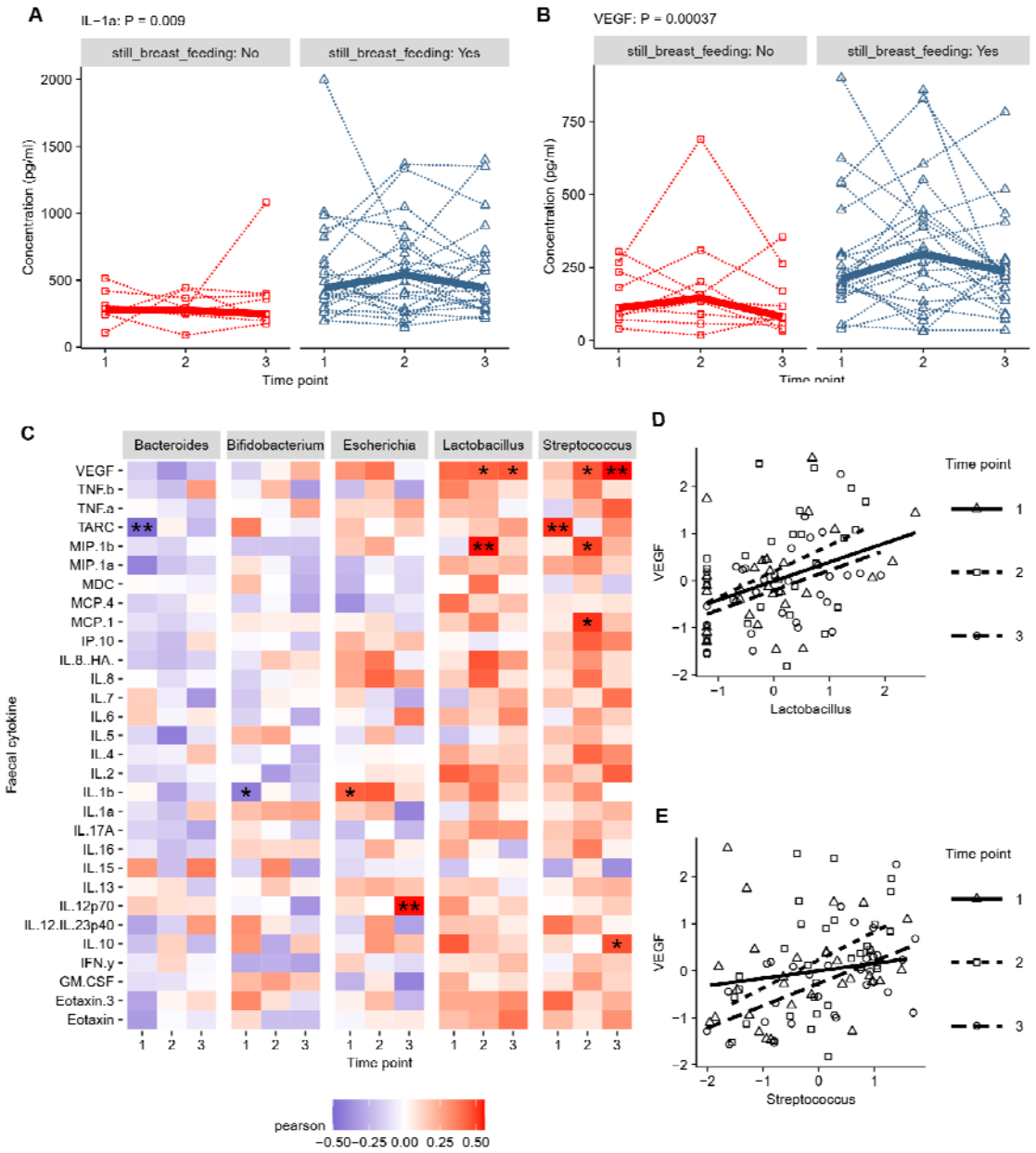


Figure 5. Infant faecal cytokines concentration and associations with breast feeding status. (A) Individual infant IL-1α concentration by breastfeeding status across time points; (B) Individual infant VEGF concentration by breastfeeding status across time points; (C) Pearson correlations between faecal cytokine concentrations and five highest abundance genus with variables in both datasets individually normalised using the bestNormalize package; (D) Scatter plot of VEGF and *Lactobacillus* correlations labelled for each time point; (E) Scatter plot of VEGF vs *Streptococcus* correlations labelled for each time point.

372 **Discussion**

373 This study represents the first time that the infant gut microbiota composition of a cohort of rural
374 Cambodian infants has been analysed in detail. We show that shortage of samples under rural
375 conditions can radically reduce the abundance of certain genus within the samples. The infants
376 showed a gut microbiota is dominated by the genus *Bifidobacterium* and is shaped to a large degree
377 by breastfeeding, which also impacts faecal cytokine profiles.

378 Our study results represent an unusual example of different sample storage conditions as these
379 samples were stored without immediate freezing. Overall the effect of storage conditions on the
380 composition of the microbiota was small but a higher relative abundance of *Bacteroides* in particular
381 was seen in samples stored in DNA preservation solution and a lower abundance of common species
382 including *Bifidobacterium* due to the proportional nature of the microbiota data. *Bacteroides*
383 represents an important genus in the infant gut microbiota, transferred from the mother during
384 delivery and has been found to be higher in vaginally delivered infants (Matharu et al., 2022, p. 1)
385 with some *Bacteroides* species having the ability to consume components of breast milk (Kijner et
386 al., 2022). The relatively high proportion of *Bacteroides* in these infants and their almost
387 disappearance from samples without DNA preservation solution shows the importance of a storage
388 method that preserves this significant component of the infant microbiota.

389 Previous research has investigated the effects of storage conditions on microbiota analysis and
390 reported alterations in *Bacteroides* abundance, although this has focused on storage conditions
391 while frozen or number of freeze thaw cycles (Pérez-Burillo et al., 2021; Poulsen et al., 2021). Room
392 temperature storage of faecal samples for up to 72 hours has been reported to reduce genus
393 including *Bacteroides* in adult faecal samples over time (Roesch et al., 2009). While defrosting frozen
394 samples for three hours before analysis reduced *Bacteroides* abundance by half (Cardona et al.,
395 2012). The duration of cool but unfrozen temperatures that samples in the present study were
396 exposed to appear to have resulted in the particular low of *Bacteroides*. However, the unpreserved
397 samples were also important for culturing and isolating live bacteria for genomic sequencing,
398 highlighting the importance of having samples stored under both conditions.

399 *Bifidobacterium* was present at lower abundance in non-breast fed infants and not breast feeding
400 was associated with higher abundance of potentially pathogenic bacterial genus like *Klebsiella* and
401 *Clostridium*. Although the numbers of infants not breast fed was too low to draw strong conclusions,
402 previous research has indicated that increased *Bifidobacterium* in the preterm infant gut can help
403 reduce the proportions of potential pathogens (Alcon-Giner et al., 2020). This is linked to the ability
404 of *Bifidobacterium* to metabolise HMOs in breast milk and reduce gut pH to make it a more

405 inhospitable environment for overgrowth of opportunistic pathogens. *Bifidobacterium longum* was
406 the dominant species in this cohort of infants and *B. longum* was the only *Bifidobacterium* species
407 with higher abundance in breast fed infants. From the four *B. longum* isolates genome-sequenced,
408 all contained one or two HMO gene clusters indicating that they were potentially capable of
409 metabolising the sugars present in breast milk. Interestingly, three of the *B. longum* isolates were
410 confirmed to be *B. longum* subsp. *longum*, which do not typically dominate the gut microbiota of
411 infants in developed western countries nor other low- or middle income countries with high rates of
412 breast feeding. It was recently reported that *B. longum* subsp. *infantis* was the dominant
413 *Bifidobacterium* in 1 to 2 month old infants from the Gambia and Bangladesh with *B. longum* subsp.
414 *longum* only forming a minor proportion of the microbiota (Taft et al., 2022). The isolation of *B.*
415 *longum* subsp. *longum* from Cambodian samples may be due to sampling at an older age of around 7
416 months, with recent research from Bangladesh indicating that a transition occurs from *B. longum*
417 subsp. *infantis* to *B. longum* subsp. *longum* with increasing infant age (Vatanen et al., 2022). One *B.*
418 *longum* isolate was identified as most closely matching *B. longum* subsp. *suis*, a subspecies originally
419 isolated from pigs (Matteuzzi et al., 1971) and found to be a dominant member of the piglet gut
420 microbiota (Zani et al., 1974). *Bifidobacterium* isolates corresponding to *B. longum* subsp. *suis* were
421 recently identified in Bangladeshi infants (Vatanen et al., 2022), which suggests that this may be an
422 under-recognised member of the infant gut microbiota. From only a single isolate we cannot
423 determine the proportion of *B. longum* subsp. *suis* within Cambodian samples, and this could
424 represent the transient colonisation of this bacterium from close proximity to pigs or that this is a
425 resident of the gut microbiota in this region. The presence of a HMO gene cluster capable of
426 metabolising breast milk sugars indicates that this *B. longum* subsp. *suis* isolate may be capable of
427 metabolising breast milk HMOs. The dominant abundance of the species *B. longum* in breastfed
428 infants underline the central effect of diet on the gut microbiota composition of these infants,
429 overwhelming any other environmental factor in this small cohort.

430 Previous research has identified associations between restricted infant growth and stunting due to
431 malnutrition and differences in the infant gut microbiota (Monira et al., 2011; Surono et al., 2021).
432 We did not find any associations between the gut microbiota and infants growth. This was
433 potentially due to the small cohort of infants involved with only three out of the thirty infants
434 included showing stunted growth and the relatively young age of the infants starting at an average
435 of 7 months of age. Previous research has looked at older children (Monira et al., 2011) or older
436 infants with a higher proportion of the infants suffering from stunting (Surono et al., 2021). This
437 indicates that in this small cohort of infants it is the presence or absence of breast milk having the
438 largest effect, potentially overwhelming the ability to detect other environmental influences.

439 In Cambodia, and many other LMICs, infants can be exposed to pathogens through provision of
440 contaminated drinking water in the first months of life (Poirot et al., 2018). Whilst we show that
441 there are high proportions of beneficial bacteria in most samples, pathogenic taxa including *Shigella*,
442 *Escherichia*, and *Clostridium* genus were also detected in most infants. Although *Shigella flexneri* and
443 *Clostridium perfringens* abundance in the gut was associated with recent diarrhoea these bacterial
444 species were also detectable in a large minority of infant samples indicating environmental exposure
445 to these organisms and potential asymptomatic carriage.

446 Antibiotics can significantly alter the profile of the early gut colonization (Alcon-Giner, et al., 2019)
447 resulting in reduced diversity (Gallego et al. 2019) and susceptibility to other infections (Porras and
448 Brita 2019). Notably, in Cambodia there is widespread, unregulated and inappropriate antibiotic use,
449 and these selection pressures may further facilitate the dissemination of AMR genes (van Aartsen et
450 al., 2019). An estimate of one-third of the Cambodian population is < 18 years old, and this group
451 may be a significant reservoir for the spread of antimicrobial resistant organisms (van Aartsen et al.,
452 2019). Indeed, we observed multiple antibiotic resistance genes in potentially pathogenic species,
453 including those associated with diarrhoeal episodes. We also observed presence of antimicrobial
454 resistance genes in *Bifidobacterium*. Although this may seem unexpected, as *Bifidobacterium* is very
455 rarely reported to be antibiotic resistant, a recent report from the neighbouring country of Vietnam
456 indicates that *Bifidobacterium* isolated from healthy infants harbour antibiotic resistance genes
457 (Chung The et al., 2021). We also found that Cambodian *Bifidobacterium* isolates contained
458 resistance genes to the antibiotics erythromycin and tetracycline. This may link to the fact that this
459 cohort of infants has previously been reported to be exposed to frequent and unregulated use of
460 antibiotics (Miyazaki et al., 2020), thus potentially driving AMR dynamics in beneficial members of
461 the infant gut microbiota. However, although amoxicillin was the most commonly used antibiotic in
462 this cohort, we did not observe any beta-lactam resistance genes in the isolates. Previous studies
463 have indicated that while it appears there is high and unregulated use of antibiotics within the
464 Cambodian population – they may be ineffective due to inappropriate storage thus reducing their
465 effectiveness. However, the widespread presence of resistance genes in other isolates may point
466 towards active selection through exposure to antibiotics. The apparent maintenance of
467 *Bifidobacterium* abundance despite antibiotic treatment may point to the presence of other novel
468 resistance genes in these *Bifidobacterium*, which if identified may have future potential for
469 combining effective probiotic therapy with antibiotic treatment.

470 One genus notable by its absence from this cohort of Cambodian infants is the genus *Prevotella*.
471 Previously found to be a dominant member of the gut microbiota in rural West Africa while rare in
472 Italian children (De Filippo et al., 2010). More recently this high prevalence of *Prevotella* in children

473 in low income countries has been confirmed in The Gambia, Kenya, Mali, and Bangladesh (Pop et al.,
474 2014). Whilst *Prevotella* abundance was lower in children from Bangladesh it was still detectable in
475 children of a similar age to this cohort of Cambodian infants (Pop et al., 2014). In contrast
476 Cambodian infants contained *Bacteroides*, a genus previously associated with the gut microbiota of
477 European and US inhabitants (De Filippo et al., 2010). What dietary differences may account for this
478 difference cannot be determined in this study, however, it does show that further investigation is
479 needed into underrepresented populations. Whether the presence of *Bacteroides* and absence of
480 *Prevotella* persists into adulthood remains unknown as the adult gut microbiota in this region
481 appears to have not yet been investigated.

482 Consistent differences in the faecal concentrations of the cytokines IL-1α and VEGF point to effects
483 of breastfeeding on the infant gut mucosal immune system. VEGF is known to be secreted by the
484 mammary glands and breast milk contains high concentrations of free VEGF that binds to receptors
485 on the surface of cells in the infant intestine (Kobata et al., 2008; Siafakas et al., 1999; Vuorela et al.,
486 2000). Reduced VEGF in preterm infants is suspected to be involved in necrotising enterocolitis
487 which is associated with overgrowth of opportunistic pathogens including *C. perfringens* (Kiu et al.,
488 2023; Sabnis et al., 2015). While the role of VEGF originating from breast milk in term infants
489 remains unexplored this growth factors could exert growth-promoting and protective effects on the
490 infant gut (Kobata et al., 2008). The role of IL-1α in the infant gut has not been explored, however it
491 has also been reported to be present in breast milk (Zanardo et al., 2005) although any role on the
492 infant gut is currently unknown. Although we are unable to demonstrate a breast milk origin from
493 these results, the presence of higher concentrations of these three faecal cytokines in breast fed
494 infants does suggest that they may be derived from breast milk, with a potential to also benefit
495 infant development.

496 While there was little consistent association across time points between the gut microbiota at the
497 genus level and faecal cytokines, a weak positive correlation was noted between *Lactobacillus* and
498 *Streptococcus* and the faecal cytokine VEGF. As *Streptococcus* abundance was higher in breastfed
499 infants this correlation is potentially linked to breastfeeding. As both microbiota composition and
500 faecal cytokines is variable with time it may not be possible to identify links between factors in a
501 limited sample size of infants. Interestingly, despite being the genus most differentiated by breast
502 feeding status there no consistent correlations between *Bifidobacterium* IL-1α or VEGF. Indeed,
503 many species and strains from this genus are known to positively influence the gut and systemic
504 immune system – including *B. longum,* which is proposed to be in part mediated by metabolites
505 produced as a result of HMO degradation e.g. acetate (Henrick et al., 2021; Laursen et al., 2021).
506 However, further studies including larger number of infants would need to be performed to explore

507 microbial-immune interactions in more detail. In addition, the use of the frozen samples for faecal
508 cytokine analysis may be limited as it is unknown how such storage conditions may impact the
509 degradation of the cytokines tested.

510 We show that this cohort of infants from rural Cambodia have a *Bifidobacterium* dominated gut
511 microbiota that is primarily shaped by breastfeeding. This highlights the influence of breast feeding
512 in supporting a beneficial infant microbiota in rural conditions. The identification of an unusual
513 *Bifidobacterium* subspecies *B. longum* subsp. *suis* in these infants that has previously only been
514 identified from infants in Bangladesh shows the importance of looking at infants from
515 underrepresented areas for defining what can constitute a normal infant microbiota. As *Bacteroides*
516 represents a common and important genus in the human gut the effects of sample storage are an
517 important consideration when immediate freezing of samples is not possible. The lack of impact of
518 antibiotic use may indicates either widespread antibiotic resistance or the use of ineffective
519 antibiotics. The finding that faecal cytokines are shaped by breast feeding and could potentially
520 originate from breast milk is a finding with unknown but potentially interesting implications. This
521 study adds to the available evidence exploring the infant gut microbiota communities in a previously
522 neglected geographical area of the world.

523

524 **Materials & Methods**

525 ***Study design & ethics***

526 Samples were collected from a sub-group of infants enrolled in the Nutrition for health Aka-chan
527 (baby in Japanese) and Mamas' (NHAM) cohort based in villages along the Mekong river in rural
528 Kampong Cham province in Cambodia. This registered all infants born between April 2016 and
529 March 2019 with the objective of investigating causes of stunting. Methods for the collection of
530 detailed meta-data for infants included in this sub-study are as previously described (Miyazaki et al.,
531 2021, 2020). Weight and length were measured to the nearest 0.1kg and 1 mm by trained field
532 workers at each home visit using a weighing scale (Seca 877, Seca, Hamburg, Germany) and a length
533 board (Seca 417, Seca).

534 ***Sample collection***

535 Infant faecal samples were analysed from a subset of 32 infants out of 47 enrolled in a detailed 3-
536 month longitudinal follow-up study (Miyazaki et al., 2021, 2020) collected between March and July
537 in 2017. Mothers/primary care-givers and their infants were visited by study field workers every 14
538 days. Faecal samples were collected from 32 infants from 47 infants. Three samples were analysed

539 from each infant for a total of 96 individual samples. The trained field worker gave instructions on
540 how to collect and equipped the caregivers with the material for the child stool collection prior to
541 routine home visits. Once stool sample collection by mothers or caregivers was completed, the
542 sample was kept in relatively cooler place in the house or in cooler bags with ice until the field
543 worker visited to collect them. The majority of the samples were collected by the field worker within
544 10 hours of the time of child defecation reported by mothers (maximum: 47 hours and minimum: 15
545 min among all sample). Each sample collected was divided in two and half stored in DNAShield and
546 stored unaltered at ambient temperatures until delivered to laboratory. The other half of the sample
547 was stored in iceboxes with refrigerant and transported to a -20° freezer on the same day and once
548 delivered was stored at -80°. The first sample was collected after an initial visit where cross-sectional
549 metadata was collected. Subsequent samples were collected at approximately two week intervals
550 along with further individual visit metadata. Underweight z-scores (weight-for-age), weight-for-
551 length z-scores, length-for-age z-score were calculated previously (Miyazaki et al., 2021). A hand
552 washing score was generated as a composite score combining the frequency and consistency of
553 handwashing with soap. Improved toilet facility was defined as having access to a toilet flushing to a
554 septic tank or pit latrine compared to unimproved having only access to a pit latrine or no toilet
555 facility. Household toilet facility was defined as having a toilet within the home flushing to a septic
556 tank.

557 ***DNA extraction of preterm stool samples***

558 FastDNA Spin Kit for Soil (MP) was used to extract DNA from preterm faeces following manufacturer
559 instructions, with extended 3 min bead-beating. DNA concentration and quality were quantified
560 using a Qubit® 2.0 fluorometer (Invitrogen).

561 ***16S rRNA amplicon sequencing of faecal samples***

562 16S rRNA region (V1-V2) primers were used for library construction as described previously (Alcon-
563 Giner et al., 2020). This set of primers allowed the amplification of one 16S rRNA gene sequencing
564 library containing 96 different samples. PCR conditions used were: cycle of 94°C 3 min and 25 cycles
565 of 94°C for 45 s, 55°C for 15 s and 72°C for 30 s. Sequencing of the 16S rRNA gene libraries was
566 performed using Illumina MiSeq platform with 300 bp paired end reads.

567 ***16S rRNA amplicon data analysis***

568 The raw sequence reads obtained were subjected to quality filtering using trim_galore v0.4.3. The
569 resulting filtered sequence reads were then mapped against the 16S rRNA SILVA database
570 (SILVA_132_SSURef_tax_silva) using BLASTn v2.2.25+ with a maximum e-value threshold of 10e-3.

571 This mapping process was performed separately for both pairs of sequences (Quast et al., 2013).
572 After completing the BLASTn alignments, the output files were annotated using the paired-end
573 protocol of MEGAN6, employing the default Lowest Common Ancestor (LCA) parameters (Huson et
574 al., 2016; Huson and Mitra, 2012).

575 Statistical analysis was conducted using R version 4.2.2 (2022-10-31 ucrt) and RStudio
576 (2023.06.0+421). To determine the number of reads required to adequately represent the
577 microbiota in each sample, rarefaction curves were generated using the vegan package in R. For the
578 analysis at the genus level, the sample with the smallest number of reads contained 37,306 classified
579 reads, which was deemed an acceptable minimum. For the species level analysis, the sample with
580 the second lowest number of reads contained 9,317 reads, which was considered an acceptable
581 minimum. The 16S rRNA gene sequence data were then normalized by subsampling using the rarefy
582 function in the phyloseq package, resulting in an even depth of 37,306 reads for the genus data and
583 9,317 reads for the species data. One sample was excluded from the species data analysis as it only
584 contained 3,965 reads.

585 To examine differences in the relative abundance of genera among the 69 infant faecal samples,
586 which had 16S sequence data obtained from both samples stored in DNAShield solution and Frozen,
587 a Wilcoxon rank sum test was performed in R. This test was chosen due to the non-parametric
588 nature of the majority of the genus abundance data. To account for multiple testing, a Benjamini &
589 Hochberg false discovery rate (FDR) method was applied, and a p-value of less than 0.05 was
590 considered statistically significant for adjusted p-values.

591 The overall composition of the 16S rRNA gene sequence data was analysed using NMDS (Non-metric
592 multidimensional scaling) plots. These plots were generated in R Studio using the vegan package
593 version 2.6-4, employing a Bray-Curtis dissimilarity calculation. To determine the statistical
594 significance of differences in microbial community structure based on infant metadata variables,
595 permutational MANOVA was conducted using the Adonis function from the vegan R package version
596 2.6-4.

597 Microbial diversity at both the genus and species taxa levels was calculated using the vegan package
598 version 2.6-4. The number of genera or species was determined for each sample, and the Shannon
599 diversity index was calculated accordingly. Linear mixed-effect models, utilizing the "lmer" function
600 in R, were employed to test the associations between the relative abundance of genera or species
601 and metadata variables. The number of genera or species, as well as the Shannon diversity index,
602 were assessed against each individual metadata variable, with Time Point included as a fixed effect
603 and Patient ID included as a random effect in the model.

604 To test for differential abundance of genera and species in relation to available background
605 metadata variables for the infant samples, the Maaslin2 package (Microbiome Multivariable
606 Associations with Linear Models) (Mallick et al., 2021), was utilized. The Maaslin2 analysis was
607 performed with default parameters, including Benjamini-Hochberg FDR multiple testing correction
608 and a significance threshold of 0.25. No minimum prevalence or abundance cut-offs were applied,
609 and the data input was previously normalized. Each metadata variable was individually tested,
610 incorporating Time Point and Patient ID as random effect grouping variables in a mixed-effect
611 model. Positive associations were identified between Breast Feeding and Diarrheal Illness. These
612 two variables were subsequently tested together as fixed effects, with Time Point and Patient ID
613 included as random effects in the final analysis.

614 ***Anaerobic bacterial culturing and isolation from infant faecal samples***

615 Anaerobic bacterial culturing and isolation was performed on five Cambodian infant faecal samples
616 (C152, C154, C257, C267, and C281) following (Browne, 2018; Browne et al., 2016; Forster et al.,
617 2019). Sample preparation was performed under anaerobic conditions. YCFA media and PBS was
618 pre-reduced in an anaerobic cabinet at 37 ˚C for a minimum of 12hs before culturing. Two samples,
619 weighing 100mg each, were prepared from each faecal sample. The first sample was homogenised in
620 reduced PBS, serially diluted, and plated directly onto the YCFA agar. The second sample was
621 ethanol treated using 70% ethanol under ambient aerobic conditions for 1h to kill vegetative cells;
622 the sample was then washed three times with reduced PBS to remove the ethanol, homogenised in
623 the reduced PBS, serially diluted and plated onto YCFA + Sodium taurocholate agar to stimulate
624 spore germination. All agar plates were placed in the anaerobic cabinet at 37˚C for 48hs. Single
625 colonies were picked and streaked onto pre-reduced YCFA media to isolate and purify the colonies;
626 YCFA media was pre-reduced in the anaerobic cabinet at 37 ˚C for a minimum of 12hs before
627 isolating colonies. Colonies from the non-ethanol treated samples were streaked onto the reduced
628 YCFA agar plates, and colonies from the ethanol treated samples were streaked onto the reduced
629 YCFA + Sodium taurocholate agar plates. Plates were then left to incubate in the anaerobic cabinet
630 at 37˚C for 48-72hs. This was carried out a second time to make sure isolates were pure. Falcon
631 tubes containing reduced YCFA broth was inoculated with a single colony from the YCFA and YCFA +
632 Sodium taurocholate agar plates, and left to incubate in the anaerobic cabinet at 37˚C for 48hs, with
633 bacterial stocks prepared in 1.5ml cryovials with 30% glycerol YCFA broth, and stored at -80˚C.
634 Additionally, two infant faecal samples C203 and C232 were processed for targeted *Bifidobacterium*
635 isolation. Briefly, serially diluted aliquots of faecal homogenates (10^{-1} – 10^{-6} in PBS) were plated onto
636 MRS agar supplemented with 50mg/l mupirocin and 0.5g/l L-cysteine HCL and incubated in an
637 anaerobic cabinet at 37 °C for 48h. Colonies with desired morphology were randomly selected and

638 re-streaked on MRS agar to purity. Pure cultures were prepared in 1.5 ml cryovials with 30% glycerol
639 Reinforced Clostridial Medium (RCM) and stored at –80 °C.

640 ***Whole Genome Sequencing (WGS)***

641 DNA extractions were performed using the MPbio Fast DNA Spin Kit for soil and bead beating (but
642 with an extended 3 min bead beating time). DNA quantification was carried out using the Qubit 2.0
643 fluorometer. DNA from pure bacterial cultures was sequenced at Wellcome Trust Sanger Institute
644 using 96-plex Illumina HiSeq 2500 platform to generate 125 bp paired end reads as described
645 previously (Harris et al., 2010).

646 ***Genomic analysis***

647 Paired-end raw sequence reads (fastq) were quality-filtered (-q 20) via fastp v0.20.0 (Chen et al.,
648 2018) prior to de novo genome assembly using SPAdes v3.14.1 (Bankevich et al., 2012) at default
649 parameters. Next, contigs smaller than 500bp were discarded in each genome assembly. All genome
650 assemblies were subjected to contamination check via CheckM v1.1.3 (Parks et al., 2015),
651 contaminated (contamination >5%) and/or incomplete (completeness <90%) genome assemblies
652 were excluded from further analyses (n=5). All genomes were identified taxonomically using gtdb-tk
653 v1.5.1 (Chaumeil et al., 2020), using ANI cut-off 95% against type strains of each species.

654 These high-quality genome assemblies (n=71) were then dereplicated at ANI 99.9% (cut-off for
655 strain level) using dRep v3.2.2 (Olm et al., 2017) prior to generation of a distance tree using
656 Mashtree v1.2.0 (Katz et al., 2019) on default parameters. Tree was annotated in iTOL v6.5.8
657 (Letunic and Bork, 2019).

658 These genomes were also subjected to sequence search analysis via ABRicate v1.0.1 (Seemann,
659 2023) based on ResFinder v4.0 database (Bortolaia et al., 2020) (--minid=95 and --mincov=90) to
660 determine antimicrobial resistance (AMR) profiles, and VFDB database (Liu et al., 2019) (--minid=90
661 and –mincov=90) to determine virulence profiles. Furthermore, HMO clusters were detected using
662 an in-house HMO sequence database via ARBricate at –minid=70 (Lawson et al., 2020). In addition,
663 *C. perfringens* virulence profiles were predicted via database TOXIper v1.1 using ABRicate at –
664 minid=90 (Kiu, 2023).

665 For *Bifidobacterium longum* subspecies analysis, *B. longum* strains were compared with four
666 subspecies type strains namely subspecies infantis, subspecies longum, subspecies suillum and
667 subspecies suis, to generate a core gene alignment via Panaroo v1.2.8 (Tonkin-Hill et al., 2020),
668 resulting in 1,212 core genes, which core-gene alignment constructed by MAFFT (Katoh and
669 Standley, 2013) were SNP-extracted (via SNP-sites v2.3.3)(Page et al., 2016) and subsequently used

670 to generate a phylogenetic tree via IQ-TREE v2.0.5 (Minh et al., 2020, p. 2) with ultra-fast bootstrap
671 replicates -B 1000 and automatic evolutionary model selection option -m TEST (best fit model was
672 determined to be GTR+F+ASC+G4). Subspecies identification was performed using fastANI v1.3 (Jain
673 et al., 2018) at cut-off 98%.

674 ***Faecal cytokine measurement***

675 Faecal samples were homogenized with PBS using a FastPrep® Bead Beater (4.0m/s, 3min),
676 centrifuged (14,000rpm, 15min) and 25µl of supernatant was used for the assay. Samples were
677 analysed using MULTI-SPOT™ plates, MESO Quickplex SQ120 and discovery workbench software
678 according to the manufacturer's protocol. Pre-coated immunoassays V-PLEX Proinflammatory Panel
679 1 (human) and V-PLEX Cytokine Panel 1 (human) were used to detect a set of 20 different cytokines:
680 IFNγ, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p70, IL-13, TNFα, GMCSF, IL-1α, IL-5, 273 IL-7, IL-12p40, IL-
681 15, IL-16, IL-17A, TNFβ, and VEGF-A.

682 ***Data and Code Availability***

683 Bacterial pure genome assemblies are now deposited at NCBI GenBank under project accession
684 PRJNA958097 while 16S rRNA amplicon sequencing raw reads are deposited in NCBI SRA under
685 project PRJNA958098.

686 ***Statistical analysis***

687 Associations between faecal cytokine concentrations and metadata variables were tested using the
688 MaAsLin2 package (Mallick et al., 2021), a differences in concentration against background metadata
689 variables available for the infant samples. For faecal cytokine analysis MaAsLin2 was run with default
690 parameters including multiple testing correction using Benjamini-Hochberg FDR with a significance
691 threshold of 0.25, except: no minimum prevalence or abundance cut-offs were applied, normalised
692 methods used wans "CLR", and the transformation "LOG". Each metadata variable was tested
693 individually with Time Point and with Patient ID included as a Random Effect grouping variable to
694 run a mixed effect model. Positive associations were identified with Breast Feeding and Diarrheal
695 Illness and these two variables where tested together as Fixed Effects with Time Point and Patient ID
696 included as a Random Effect in the final analysis.

697 Correlations between faecal cytokine concentration and the relative abundance of genus and species
698 was tested using a Pearson correlation with correction for multiple testing using Benjamini &
699 Hochberg FDR correcting between both cytokines and time point. Both cytokine concentration and
700 genus/species abundance were first normalised using the bestNormalize package in R.

Author contributions

Conceptualisation, L.J.H and S.E.C.; Methodology, M.J.D, R.K., I.R.S, A.M., HA-P., R.T., S.C., S.P., M.K., M.M., A.I., and B.T; Software, M.J.D., R.K., S.C., and M.K.; Validation, M.J.D, K.S., S.E.C., and L.J.H.; Formal analysis, M.J.D, R.K., A.M, S.E.C., and L.J.H.; Investigation, M.J.D, R.K., I.R.S, A.M., HA-P., R.T., S.C., S.P., M.K., M.M., A.I., B.T., S.E.C., and L.J.H.; Resources, M.J.D., A.M., R.T., M.M., A.I., B.T., S.E.C., and L.J.H; Data Curation, M.J.D., R.K., A.M., S.E.C.; Writing - Original Draft, M.J.D. and L.J.H.; Writing - Review and Editing, M.J.D, R.K., I.R.S, A.M., HA-P., R.T., S.C., S.P., M.K., M.M., A.I., B.T., S.E.C, and L.J.H; Visualisation, M.J.D and R.K.; Supervision, S.E.C,. and L.J.H.; Project Administration, M.J.D., S.E.C, and L.J.H.; Funding Acquisition, S.E.C, and L.J.H.

Competing interest statement

The authors declare that they have no competing interests.

Acknowledgments

We thank the families for participating in the study.

Funding information

L.J.H. is supported by Wellcome Trust Investigator Awards 100974/C/13/Z and 220876/Z/20/Z; the Biotechnology and Biological Sciences Research Council (BBSRC), Institute Strategic Programme Gut Microbes and Health BB/R012490/1, and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356.

References

Abdill, R.J., Adamowicz, E.M., Blekhman, R., 2022. Public human microbiome data are dominated by highly developed countries. PLoS Biol 20, e3001536. https://doi.org/10.1371/journal.pbio.3001536

Alcon-Giner, C., Dalby, M.J., Caim, S., Ketskemety, J., Shaw, A., Sim, K., Lawson, M.A.E., Kiu, R., Leclaire, C., Chalklen, L., Kujawska, M., Mitra, S., Fardus-Reid, F., Belteki, G., McColl, K.,

730 Swann, J.R., Kroll, J.S., Clarke, P., Hall, L.J., 2020. Microbiota Supplementation with
731 Bifidobacterium and Lactobacillus Modifies the Preterm Infant Gut Microbiota and
732 Metabolome: An Observational Study. Cell Rep Med 1, 100077.
733 https://doi.org/10.1016/j.xcrm.2020.100077
734 Auguet, O.T., Niehus, R., Gweon, H.S., Berkley, J.A., Waichungo, J., Njim, T., Edgeworth, J.D., Batra,
735 R., Chau, K., Swann, J., Walker, S.A., Peto, T.E.A., Crook, D.W., Lamble, S., Turner, P., Cooper,
736 B.S., Stoesser, N., 2021. Population-level faecal metagenomic profiling as a tool to predict
737 antimicrobial resistance in Enterobacterales isolates causing invasive infections: An
738 exploratory study across Cambodia, Kenya, and the UK. EClinicalMedicine 36, 100910.
739 https://doi.org/10.1016/j.eclinm.2021.100910
740 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko,
741 S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev,
742 M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to
743 single-cell sequencing. J Comput Biol 19, 455–477. https://doi.org/10.1089/cmb.2012.0021
744 Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R.L.,
745 Rebelo, A.R., Florensa, A.F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G.,
746 Bender, J.K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B.B., Malhotra-Kumar, S., Westh, H.,
747 Pinholt, M., Anjum, M.F., Duggett, N.A., Kempf, I., Nykäsenoja, S., Olkkola, S., Wieczorek, K.,
748 Amaro, A., Clemente, L., Mossong, J., Losch, S., Ragimbeau, C., Lund, O., Aarestrup, F.M.,
749 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. Journal of Antimicrobial
750 Chemotherapy 75, 3491–3500. https://doi.org/10.1093/jac/dkaa345
751 Browne, H.P., 2018. Development of anaerobic culturing methods to study the evolution and
752 transmission of the human intestinal microbiota (doctoral). London School of Hygiene &
753 Tropical Medicine. https://doi.org/10.17037/PUBS.04647551
754 Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., Goulding, D., Lawley,
755 T.D., 2016. Culturing of "unculturable" human microbiota reveals novel taxa and extensive
756 sporulation. Nature 533, 543–546. https://doi.org/10.1038/nature17645
757 Cambodia Demographic and Health Survey, 2022. Cambodia Demographic and Health Survey [WWW
758 Document]. URL https://dhsprogram.com/publications/publication-PR136-Preliminary-
759 Reports-Key-Indicators-Reports.cfm (accessed 1.19.23).
760 Cardona, S., Eck, A., Cassellas, M., Gallart, M., Alastrue, C., Dore, J., Azpiroz, F., Roca, J., Guarner, F.,
761 Manichanh, C., 2012. Storage conditions of intestinal microbiota matter in metagenomic
762 analysis. BMC Microbiology 12, 158. https://doi.org/10.1186/1471-2180-12-158

763 Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2020. GTDB-Tk: a toolkit to classify
764 genomes with the Genome Taxonomy Database. Bioinformatics 36, 1925–1927.
765 https://doi.org/10.1093/bioinformatics/btz848
766 Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
767 Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560
768 Chung The, H., Nguyen Ngoc Minh, C., Tran Thi Hong, C., Nguyen Thi Nguyen, T., Pike, L.J., Zellmer,
769 C., Pham Duc, T., Tran, T.-A., Ha Thanh, T., Van, M.P., Thwaites, G.E., Rabaa, M.A., Hall, L.J.,
770 Baker, S., 2021. Exploring the Genomic Diversity and Antimicrobial Susceptibility of
771 Bifidobacterium pseudocatenulatum in a Vietnamese Population. Microbiology Spectrum 9,
772 e00526-21. https://doi.org/10.1128/Spectrum.00526-21
773 De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S.,
774 Pieraccini, G., Lionetti, P., 2010. Impact of diet in shaping gut microbiota revealed by a
775 comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. U.S.A. 107,
776 14691–14696. https://doi.org/10.1073/pnas.1005963107
777 Forster, S.C., Kumar, N., Anonye, B.O., Almeida, A., Viciani, E., Stares, M.D., Dunn, M., Mkandawire,
778 T.T., Zhu, A., Shao, Y., Pike, L.J., Louie, T., Browne, H.P., Mitchell, A.L., Neville, B.A., Finn, R.D.,
779 Lawley, T.D., 2019. A human gut bacterial genome and culture collection for improved
780 metagenomic analyses. Nat Biotechnol 37, 186–192. https://doi.org/10.1038/s41587-018-
781 0009-7
782 Harris, S.R., Feil, E.J., Holden, M.T.G., Quail, M.A., Nickerson, E.K., Chantratita, N., Gardete, S.,
783 Tavares, A., Day, N., Lindsay, J.A., Edgeworth, J.D., de Lencastre, H., Parkhill, J., Peacock, S.J.,
784 Bentley, S.D., 2010. Evolution of MRSA during hospital transmission and intercontinental
785 spread. Science 327, 469–474. https://doi.org/10.1126/science.1182395
786 Henrick, B.M., Rodriguez, L., Lakshmikanth, T., Pou, C., Henckel, E., Arzoomand, A., Olin, A., Wang, J.,
787 Mikes, J., Tan, Z., Chen, Y., Ehrlich, A.M., Bernhardsson, A.K., Mugabo, C.H., Ambrosiani, Y.,
788 Gustafsson, A., Chew, S., Brown, H.K., Prambs, J., Bohlin, K., Mitchell, R.D., Underwood,
789 M.A., Smilowitz, J.T., German, J.B., Frese, S.A., Brodin, P., 2021. Bifidobacteria-mediated
790 immune system imprinting early in life. Cell 184, 3884-3898.e11.
791 https://doi.org/10.1016/j.cell.2021.05.030
792 Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R.,
793 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale
794 Microbiome Sequencing Data. PLOS Computational Biology 12, e1004957.
795 https://doi.org/10.1371/journal.pcbi.1004957

796 Huson, D.H., Mitra, S., 2012. Introduction to the Analysis of Environmental Sequences:
797 Metagenomics with MEGAN, in: Anisimova, M. (Ed.), Evolutionary Genomics: Statistical and
798 Computational Methods, Volume 2, Methods in Molecular Biology. Humana Press, Totowa,
799 NJ, pp. 415–429. https://doi.org/10.1007/978-1-61779-585-5_17
800 Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI
801 analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9, 5114.
802 https://doi.org/10.1038/s41467-018-07641-9
803 Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7:
804 Improvements in Performance and Usability. Molecular Biology and Evolution 30, 772–780.
805 https://doi.org/10.1093/molbev/mst010
806 Katz, L.S., Griswold, T., Morrison, S.S., Caravas, J.A., Zhang, S., den Bakker, H.C., Deng, X., Carleton,
807 H.A., 2019. Mashtree: a rapid comparison of whole genome sequence files. J Open Source
808 Softw 4. https://doi.org/10.21105/joss.01762
809 Kijner, S., Cher, A., Yassour, M., 2022. The Infant Gut Commensal Bacteroides dorei Presents a
810 Generalized Transcriptional Response to Various Human Milk Oligosaccharides. Frontiers in
811 Cellular and Infection Microbiology 12.
812 Kiu, R., 2023. TOXIper: Toxinotype assignment of Clostridium perfringens via ABRicate.
813 Kiu, R., Shaw, A.G., Sim, K., Acuna-Gonzalez, A., Price, C.A., Bedwell, H., Dreger, S.A., Fowler, W.J.,
814 Cornwell, E., Pickard, D., Belteki, G., Malsom, J., Phillips, S., Young, G.R., Schofield, Z., Alcon-
815 Giner, C., Berrington, J.E., Stewart, C.J., Dougan, G., Clarke, P., Douce, G., Robinson, S.D.,
816 Kroll, J.S., Hall, L.J., 2023. Particular genomic and virulence traits associated with preterm
817 infant-derived toxigenic Clostridium perfringens strains. Nat Microbiol 8, 1160–1175.
818 https://doi.org/10.1038/s41564-023-01385-z
819 Kobata, R., Tsukahara, H., Ohshima, Y., Ohta, N., Tokuriki, S., Tamura, S., Mayumi, M., 2008. High
820 levels of growth factors in human breast milk. Early Human Development 84, 67–69.
821 https://doi.org/10.1016/j.earlhumdev.2007.07.005
822 Lackey, K.A., Williams, J.E., Meehan, C.L., Zachek, J.A., Benda, E.D., Price, W.J., Foster, J.A., Sellen,
823 D.W., Kamau-Mbuthia, E.W., Kamundia, E.W., Mbugua, S., Moore, S.E., Prentice, A.M., K.,
824 D.G., Kvist, L.J., Otoo, G.E., García-Carral, C., Jiménez, E., Ruiz, L., Rodríguez, J.M., Pareja,
825 R.G., Bode, L., McGuire, M.A., McGuire, M.K., 2019. What's Normal? Microbiomes in Human
826 Milk and Infant Feces Are Related to Each Other but Vary Geographically: The INSPIRE Study.
827 Front. Nutr. 6, 45. https://doi.org/10.3389/fnut.2019.00045
828 Laursen, M.F., Sakanaka, M., von Burg, N., Mörbe, U., Andersen, D., Moll, J.M., Pekmez, C.T.,
829 Rivollier, A., Michaelsen, K.F., Mølgaard, C., Lind, M.V., Dragsted, L.O., Katayama, T.,

830 Frandsen, H.L., Vinggaard, A.M., Bahl, M.I., Brix, S., Agace, W., Licht, T.R., Roager, H.M.,
831 2021. Bifidobacterium species associated with breastfeeding produce aromatic lactic acids in
832 the infant gut. Nat Microbiol 6, 1367–1382. https://doi.org/10.1038/s41564-021-00970-4
833 Lawson, M.A.E., O'Neill, I.J., Kujawska, M., Gowrinadh Javvadi, S., Wijeyesekera, A., Flegg, Z.,
834 Chalklen, L., Hall, L.J., 2020. Breast milk-derived human milk oligosaccharides promote
835 Bifidobacterium interactions within a single ecosystem. ISME J 14, 635–648.
836 https://doi.org/10.1038/s41396-019-0553-2
837 Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments.
838 Nucleic Acids Research 47, W256–W259. https://doi.org/10.1093/nar/gkz239
839 Liu, B., Zheng, D., Jin, Q., Chen, L., Yang, J., 2019. VFDB 2019: a comparative pathogenomic platform
840 with an interactive web interface. Nucleic Acids Research 47, D687–D692.
841 https://doi.org/10.1093/nar/gky1080
842 Mallick, H., Rahnavard, A., McIver, L.J., Ma, S., Zhang, Y., Nguyen, L.H., Tickle, T.L., Weingart, G., Ren,
843 B., Schwager, E.H., Chatterjee, S., Thompson, K.N., Wilkinson, J.E., Subramanian, A., Lu, Y.,
844 Waldron, L., Paulson, J.N., Franzosa, E.A., Bravo, H.C., Huttenhower, C., 2021. Multivariable
845 association discovery in population-scale meta-omics studies. PLOS Computational Biology
846 17, e1009442. https://doi.org/10.1371/journal.pcbi.1009442
847 Matharu, D., Ponsero, A.J., Dikareva, E., Korpela, K., Kolho, K.-L., de Vos, W.M., Salonen, A., 2022.
848 Bacteroides abundance drives birth mode dependent infant gut microbiota developmental
849 trajectories. Frontiers in Microbiology 13.
850 Matteuzzi, D., Crociani, F., Zani, O., Trovatelli, L.D., 1971. Bifidobacterium suis n. sp.: A new species
851 of the genus Bifidobacterium isolated from pig faces. Zeitschrift für allgemeine Mikrobiologie
852 11, 387–395. https://doi.org/10.1002/jobm.19710110504
853 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear,
854 R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
855 Genomic Era. Molecular Biology and Evolution 37, 1530–1534.
856 https://doi.org/10.1093/molbev/msaa015
857 Miyazaki, A., Matsui, M., Tung, R., Taing, B., White, L.V., Iwamoto, A., Cox, S.E., 2021. Determinants
858 of growth measurements in rural Cambodian infants: a cross-sectional study. Int Health 13,
859 49–56. https://doi.org/10.1093/inthealth/ihaa018
860 Miyazaki, A., Tung, R., Taing, B., Matsui, M., Iwamoto, A., Cox, S.E., 2020. Frequent unregulated use
861 of antibiotics in rural Cambodian infants. Transactions of The Royal Society of Tropical
862 Medicine and Hygiene 114, 401–407. https://doi.org/10.1093/trstmh/traa020

863 Monira, S., Nakamura, S., Gotoh, K., Izutsu, K., Watanabe, H., Alam, N., Endtz, H., Cravioto, A., Ali,
864 Sk., Nakaya, T., Horii, T., Iida, T., Alam, M., 2011. Gut Microbiota of Healthy and
865 Malnourished Children in Bangladesh. Frontiers in Microbiology 2.
866 Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic
867 comparisons that enables improved genome recovery from metagenomes through de-
868 replication. ISME J 11, 2864–2868. https://doi.org/10.1038/ismej.2017.126
869 Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites:
870 rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2, e000056.
871 https://doi.org/10.1099/mgen.0.000056
872 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the
873 quality of microbial genomes recovered from isolates, single cells, and metagenomes.
874 Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114
875 Pérez-Burillo, S., Hinojosa-Nogueira, D., Navajas-Porras, B., Blasco, T., Balzerani, F., Lerma-Aguilera,
876 A., León, D., Pastoriza, S., Apaolaza, I., Planes, F.J., Francino, M.P., Rufián-Henares, J.Á., 2021.
877 Effect of Freezing on Gut Microbiota Composition and Functionality for In Vitro
878 Fermentation Experiments. Nutrients 13, 2207. https://doi.org/10.3390/nu13072207
879 Pop, M., Walker, A.W., Paulson, J., Lindsay, B., Antonio, M., Hossain, M.A., Oundo, J., Tamboura, B.,
880 Mai, V., Astrovskaya, I., Bravo, H.C., Rance, R., Stares, M., Levine, M.M., Panchalingam, S.,
881 Kotloff, K., Ikumapayi, U.N., Ebruke, C., Adeyemi, M., Ahmed, D., Ahmed, F., Alam, M.T.,
882 Amin, R., Siddiqui, S., Ochieng, J.B., Ouma, E., Juma, J., Mailu, E., Omore, R., Morris, J.G.,
883 Breiman, R.F., Saha, D., Parkhill, J., Nataro, J.P., Stine, O.C., 2014. Diarrhea in young children
884 from low-income countries leads to large-scale alterations in intestinal microbiota
885 composition. Genome Biology 15, R76. https://doi.org/10.1186/gb-2014-15-6-r76
886 Porras, A.M., Brito, I.L., 2019. The internationalization of human microbiome research. Current
887 Opinion in Microbiology 50, 50–55. https://doi.org/10.1016/j.mib.2019.09.012
888 Poulsen, C.S., Kaas, R.S., Aarestrup, F.M., Pamp, S.J., 2021. Standard Sample Storage Conditions Have
889 an Impact on Inferred Microbiome Composition and Antimicrobial Resistance Patterns.
890 Microbiology Spectrum 9, e01387-21. https://doi.org/10.1128/Spectrum.01387-21
891 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013.
892 The SILVA ribosomal RNA gene database project: improved data processing and web-based
893 tools. Nucleic Acids Research 41, D590–D596. https://doi.org/10.1093/nar/gks1219
894 Robertson, R.C., 2020. The Gut Microbiome in Child Malnutrition, in: Michaelsen, K.F., Neufeld, L.M.,
895 Prentice, A.M. (Eds.), Nestlé Nutrition Institute Workshop Series. S. Karger AG, pp. 133–144.
896 https://doi.org/10.1159/000503352

897 Robertson, R.C., Manges, A.R., Finlay, B.B., Prendergast, A.J., 2019. The Human Microbiome and
898 Child Growth – First 1000 Days and Beyond. Trends in Microbiology 27, 131–147.
899 https://doi.org/10.1016/j.tim.2018.09.008
900 Roesch, L.F.W., Casella, G., Simell, O., Krischer, J., Wasserfall, C.H., Schatz, D., Atkinson, M.A., Neu, J.,
901 Triplett, E.W., 2009. Influence of Fecal Sample Storage on Bacterial Community Diversity.
902 Open Microbiol J 3, 40–46. https://doi.org/10.2174/1874285800903010040
903 Sabnis, A., Carrasco, R., Liu, S.X.L., Yan, X., Managlia, E., Chou, P.M., Tan, X.-D., De Plaen, I.G., 2015.
904 Intestinal Vascular Endothelial Growth Factor Is Decreased in Necrotizing Enterocolitis.
905 Neonatology 107, 191–198. https://doi.org/10.1159/000368879
906 Seemann, T., 2023. ABRicate: Mass screening of contigs for antimicrobial and virulence genes.
907 Siafakas, C.G., Anatolitou, F., Fusunyan, R.D., Walker, W.A., Sanderson, I.R., 1999. Vascular
908 endothelial growth factor (VEGF) is present in human breast milk and its receptor is present
909 on intestinal epithelial cells. Pediatr Res 45, 652–657. https://doi.org/10.1203/00006450-
910 199905010-00007
911 Surono, I.S., Widiyanti, D., Kusumo, P.D., Venema, K., 2021. Gut microbiota profile of Indonesian
912 stunted children and children with normal nutritional status. PLOS ONE 16, e0245399.
913 https://doi.org/10.1371/journal.pone.0245399
914 Taft, D.H., Lewis, Z.T., Nguyen, N., Ho, S., Masarweh, C., Dunne-Castagna, V., Tancredi, D.J., Huda,
915 M.N., Stephensen, C.B., Hinde, K., von Mutius, E., Kirjavainen, P.V., Dalphin, J.-C., Lauener,
916 R., Riedler, J., Smilowitz, J.T., German, J.B., Morrow, A.L., Mills, D.A., 2022. Bifidobacterium
917 Species Colonization in Infancy: A Global Cross-Sectional Comparison by Population History
918 of Breastfeeding. Nutrients 14, 1423. https://doi.org/10.3390/nu14071423
919 Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A., Gladstone, R.A., Lo, S.,
920 Beaudoin, C., Floto, R.A., Frost, S.D.W., Corander, J., Bentley, S.D., Parkhill, J., 2020.
921 Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biology 21,
922 180. https://doi.org/10.1186/s13059-020-02090-4
923 van Aartsen, J.J., Moore, C.E., Parry, C.M., Turner, P., Phot, N., Mao, S., Suy, K., Davies, T., Giess, A.,
924 Sheppard, A.E., Peto, T.E.A., Day, N.P.J., Crook, D.W., Walker, A.S., Stoesser, N., 2019.
925 Epidemiology of paediatric gastrointestinal colonisation by extended spectrum
926 cephalosporin-resistant Escherichia coli and Klebsiella pneumoniae isolates in north-west
927 Cambodia. BMC Microbiol 19, 59. https://doi.org/10.1186/s12866-019-1431-9
928 Vatanen, T., Ang, Q.Y., Siegwald, L., Sarker, S.A., Le Roy, C.I., Duboux, S., Delannoy-Bruno, O., Ngom-
929 Bru, C., Boulangé, C.L., Stražar, M., Avila-Pacheco, J., Deik, A., Pierce, K., Bullock, K., Dennis,
930 C., Sultana, S., Sayed, S., Rahman, M., Ahmed, T., Modesto, M., Mattarelli, P., Clish, C.B.,

931 Vlamakis, H., Plichta, D.R., Sakwinska, O., Xavier, R.J., 2022. A distinct clade of
932 Bifidobacterium longum in the gut of Bangladeshi children thrives during weaning. Cell 185,
933 4280-4297.e12. https://doi.org/10.1016/j.cell.2022.10.011
934 Vuorela, P., Andersson, S., Carpén, O., Ylikorkala, O., Halmesmäki, E., 2000. Unbound vascular
935 endothelial growth factor and its receptors in breast, human milk, and newborn intestine.
936 The American Journal of Clinical Nutrition 72, 1196–1201.
937 https://doi.org/10.1093/ajcn/72.5.1196
938 Zanardo, V., Nicolussi, S., Cavallin, S., Trevisanuto, D., Barbato, A., Faggian, D., Favaro, F., Plebani, M.,
939 2005. Effect of Maternal Smoking on Breast Milk Interleukin-1α, β-Endorphin, and Leptin
940 Concentrations. Environ Health Perspect 113, 1410–1413. https://doi.org/10.1289/ehp.7702
941 Zani, G., Biavati, B., Crociani, F., Matteuzzi, D., 1974. Bifidobacteria from the Faeces of Piglets.
942 Journal of Applied Bacteriology 37, 537–547. https://doi.org/10.1111/j.1365-
943 2672.1974.tb00479.x
944

945

946

947

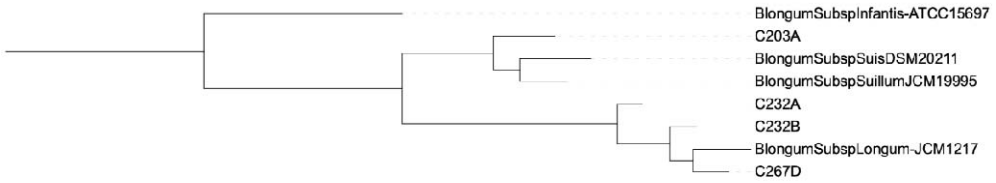
948 **Supplementary Files**

949 Supplementary file 1. Genus differences between DNAShield and Frozen storage conditions.
950 Supplementary File 2. Metadata and Genus Sequence Data
951 Supplementary file 3. Genus data Maaslin2 analysis results.
952 Supplementary file 4. Genus prevalence in samples.
953 Supplementary file 5. Combined sample metadata and species sequencing data for all samples.
954 Supplementary file 6. Species data Maaslin2 analysis results.
955 Supplementary file 7. Species prevalence in samples.
956 Supplementary File 8. Pairwise Average Nucleotide Identity (ANI)
957 Supplementary file 9. Metadata and Faecal Cytokine Concentrations.
958 Supplementary file 10. MaAsLin2 Cytokine Results
959

960

961 **Supplementary Figures:**

962

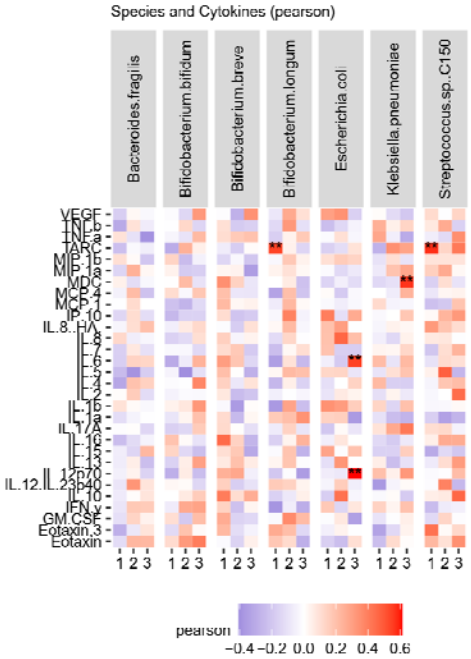


963

964 **Supplementary Figure 1. T**he subspecies of *Bifidobacterium longum* isolates determined using
965 average nucleotide identity with a 98% cut-off.

966

967



968

969 **Supplementary Figure 2.** Correlations between cytokines and Genus or Species relative abundance.

970

971